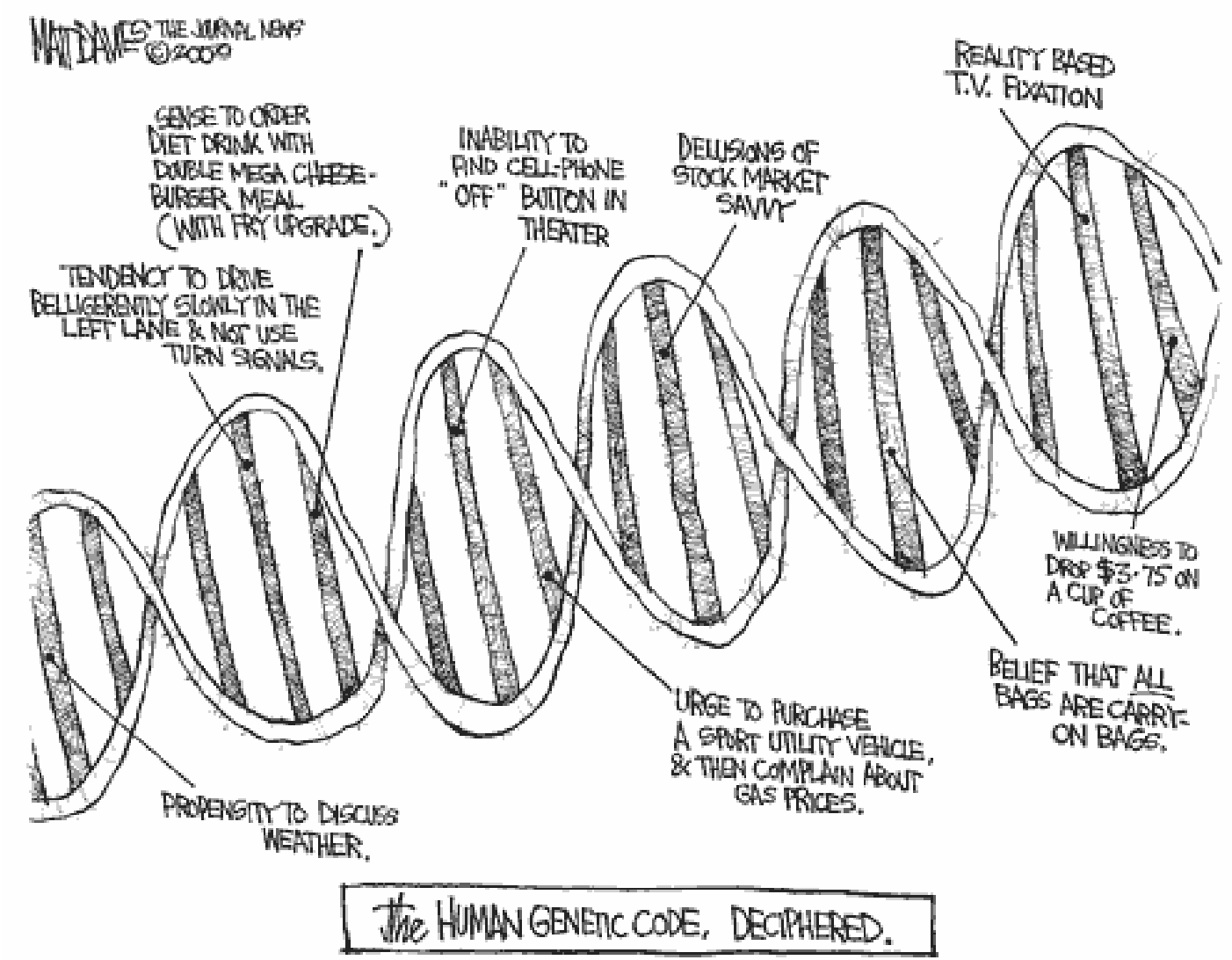
Epidemiology 217
Molecular and Genetic Epidemiology I



Outline

- **Course structure**
- Overview of genetics
- Overview of genetic epidemiology and the rest of the course

Course Goals

- Develop a framework for interpreting and incorporating genetic information in your research

- **Learn:**
  - Common genetic measures.
  - Some population genetics.
  - Approaches to search for disease-causing genes:
    - **Association** (key aspect of course)
    - Linkage
    - Admixture

Course Details

11 Tuesdays from 01/08/2013 – 03/19/2013, 1:10-3:00 pm, CB6702 (China Basin)

Course Directors:

Thomas Hoffmann, HoffmannT@humgen.ucsf.edu
John Witte, WitteJ@humgen.ucsf.edu

Lecturers:

Joe Wiemels, joe.wiemels@ucsf.edu
Eric Jorgenson, eric.jorgenson@kp.org
Neil Risch, rischn@humgen.ucsf.edu

Teaching Assistant:

Kirsten Kangelaris, kkangelaris@ucsf.edu

website:

https://courses.ucsf.edu/course/view.php?id=901

(Lectures, homework assignments, and answers)

Assignments

- **Problem sets (50%)**

Due at noon on Mondays to Kirsten Kangelaris, kkangelaris@ucsf.edu

- **Reading / class participation (20%):**

Genetic Epidemiology: Methods and Applications, by Melissa A. Austin. CABI, 2013. ISBN: 978-1780641812.

Books may be purchased either through the publisher or a variety of commercial venues (e.g., Amazon.com), where you can also read the beginning of the first chapter.

Students may be called upon during class to answer questions about the assigned chapters.

- **Final project (design study)**
  - 30% of grade (due Friday, 3/15 at Noon)
  - Present to class

Syllabus - https://courses.ucsf.edu/course/view.php?id=901

Date	Topic / Content	Lecturer	Required Reading (pre-lecture)	Assignment Due (Monday @ noon)
01/07/2014	**Introduction: The Big Picture** The process of genetic epidemiology; general approaches to assess the genetic basis of disease.	T Hoffmann	Austin Chapter 1	
01/14/2014	**Population Genetics, Modeling Genetic Inheritance** Basics of population genetics; Hardy-Weinberg Equilibrium; aggregation; heritability.	T Hoffmann	Austin Chapter 2	Assignment 1 (Due 1/13 at Noon)
01/21/2014	**Mendel's Laws and Molecular Genetics** Mendel's laws (segregation, assortment); molecular measures; genotyping; arrays; sequencing.	J Wiemels	Austin Chapter 3	Assignment 2 (Due 1/20 at Noon)
01/28/2014	**Family-based Studies** Linkage analysis, Family-based association studies, TDT, FBAT	T Hoffmann	Austin Chapter 4	Assignment 3 (Due 1/27 at Noon)
02/04/2014	**Genetic Association Studies** General principles; candidate gene studies.	J Witte	Austin Chapter 5	Assignment 4 (Due 2/03 at Noon)
02/11/2014	**Genome-wide Association Studies (GWAS)** Agnostic searches across genome for associated SNPs; multiple testing; Imputation; Consortia.	J Witte	Austin Chapter 6	Assignment 5 (Due 2/10 at Noon)
02/18/2014	**Gene-Environment Interactions and Epistasis** Interactions; pharmocogenetics; less common & rare variants	J Witte	Austin Chapter 7	Assignment 6 (Due 2/17 at Noon)
02/25/2014	**Non-mendelian genetics** Mitochondrial (MT), De novo, Parent-of-origin, Methylation	E Jorgenson	Austin Chapter 8	Assignment 7 (Due 2/24 at Noon)
03/04/2014	**Next Generation Sequencing** Full genome and exome sequencing studies. Software for genetic analysis.	E Jorgenson	Austin Chapter 9	Assignment 8 (Due 3/03 at Noon)
03/11/2014	**Admixture Analysis** Population substructure, admixture mapping.	N Risch	Winkler et al., Admixture mapping comes of age, Ann Rev Genom Human Genet 2010 11:65-89, read through the "Admixed populations" section.	
03/18/2014	**Putting it all Together: Incorporating Molecular and Genetic Measures into Your Research** Final Project presentations	T Hoffmann, J Witte	Austin chapters 10 and 11	Final Project Due **Friday 3/14 @ noon**

Molecular & Genetic Epidemiology

Distinction

- Molecular: molecular, cellular, and other biologic measurements, on disease [e.g., biomarkers - selenium in toe nails, proteins, hormones]
- Genetic: role of inherited factors in disease (encompassed within molecular)

Genetic Epidemiology

Focus of course

- Genetic epidemiology
  - Initially studied single gene disorders
  - Now more complex genetic disorders and environment
- Many designs same as epidemiology (e.g., case-control)
- Some specialized analysis methods.
- Population genetics increasingly important

Aims

- Detect genetic causes of disease
- Understand biological process
- Prevention strategies, lifestyle intervention
- Improved therapeutic strategies, personalized medicine

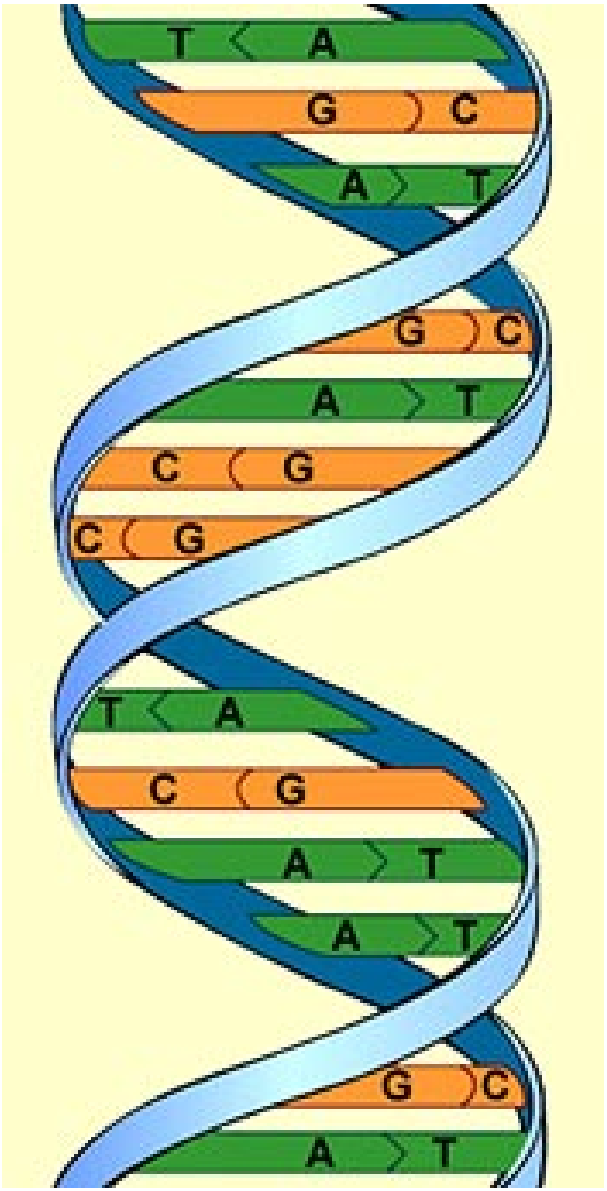
	Genetics	Epidemiology
Origins	Biological science of inheritance Eperimental/breeding	Methodological sciences / public health Oservational/analytic studies
Focus	Mechanisms of inheritance, often on rare diseases	Etiologic factors, primarily on common diseases
Tools	Family studies Laboratory methods Statistical models Population models	Observational studies Case-control studies Cohort designs Clinical trials
Goals	Understand mechanisms of inheritance	Understand distribution etiology, and progression of diseases
Overlap	Diseases of interest Population studies Ultimate goal of disease prevention	

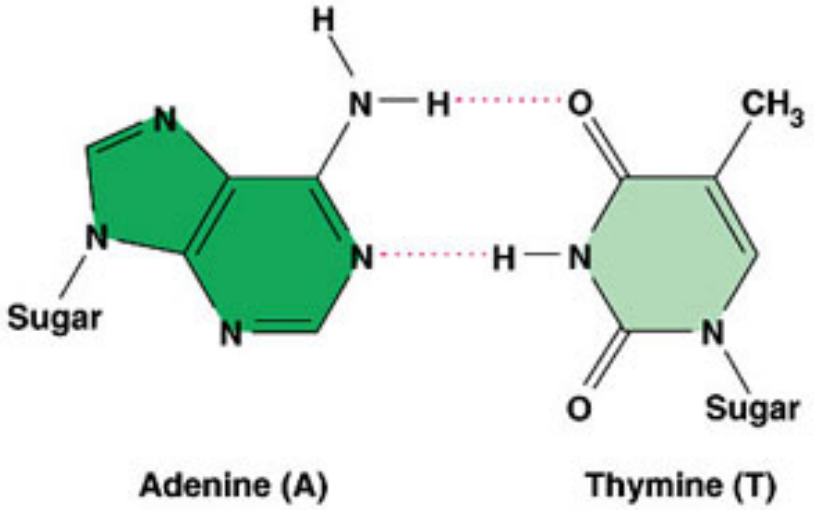
Austin, Ch 1, Table 1.1. Comparison of genetics and epidemiology.

Outline

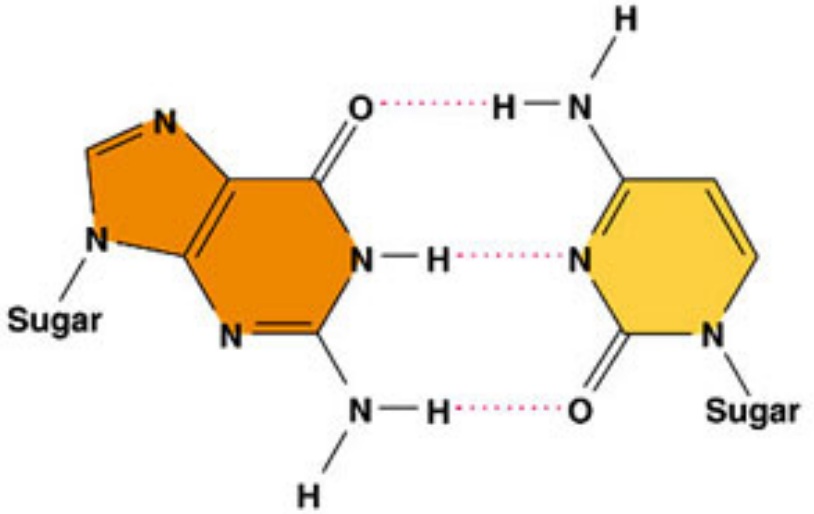
- Course structure
- **Overview of genetics**
- Overview of genetic epidemiology and the rest of the course

DNA



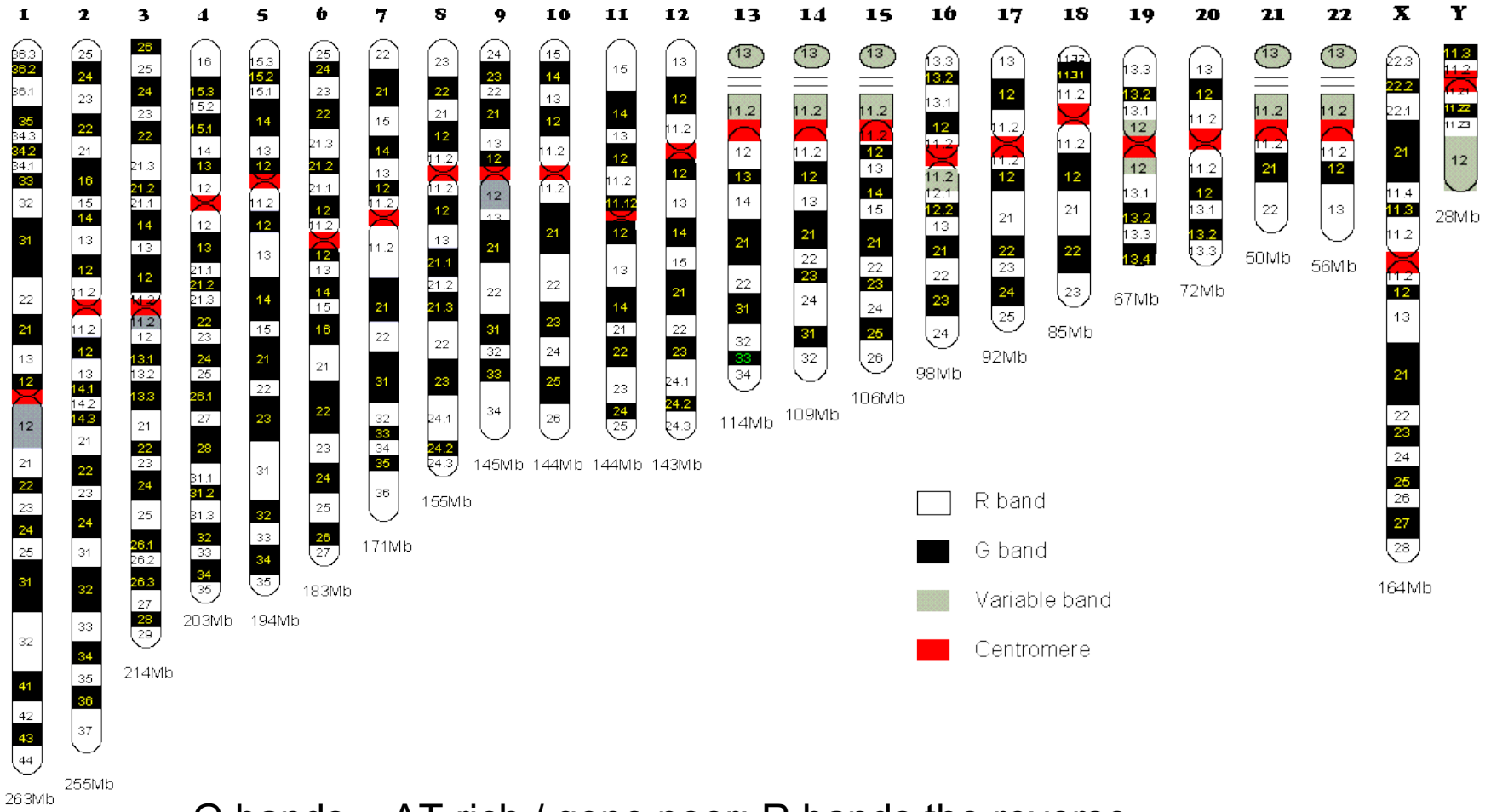


Adenine (A) Thymine (T)



Guanine (G) Cytosine (C)

Human Chromosomes



- G bands – AT rich / gene poor; R bands the reverse
- Look for broad structural abnormalities from these stains

Human Chromosome 21

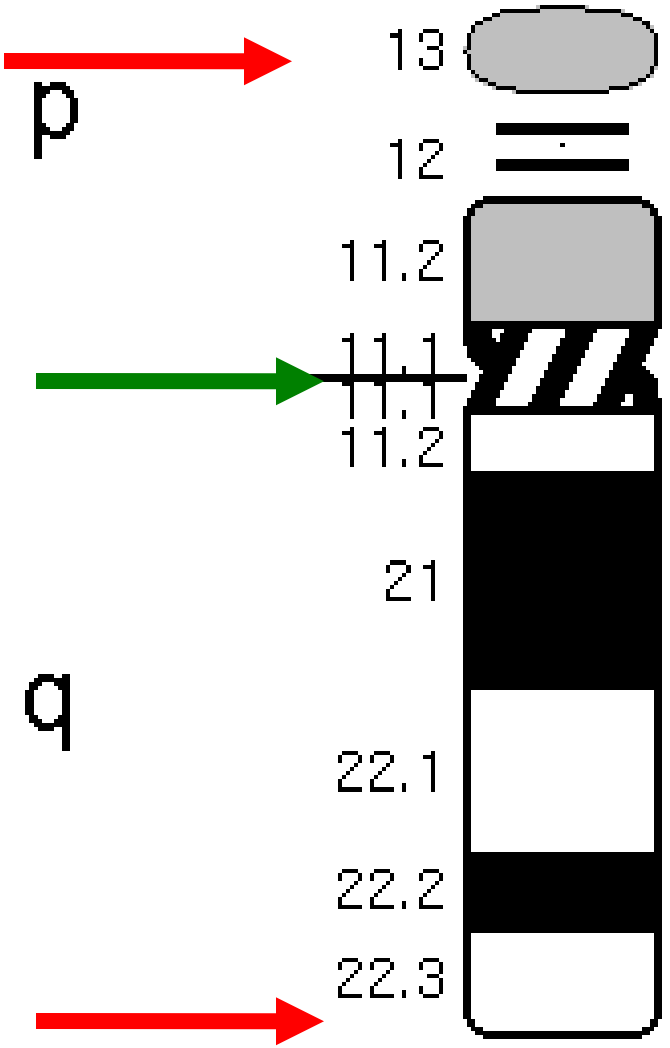
Telomeres

Centromere

p: petit arm

q: queue (tail) or long arm

21q22.1 is pronounced twenty-one q two two point one



Chromosome Bands

- Stain chromosomes so they can be seen by microscope
  - e.g., Giesma stain (G-banding).
- Appear as alternating bands
  - e.g., dark/G-band and light band.
  - Specific to phosphate groups of DNA.
  - Attaches to DNA regions with high adenine-thymine (A-T) bonding.

- With low resolution, few bands seen:
  - … p2, p1 centromere q1, q2, … (count out from centromere).
- With higher resolution sub bands seen:
  - … p12, p11 centromere q11, q12 …

Variation in Genome

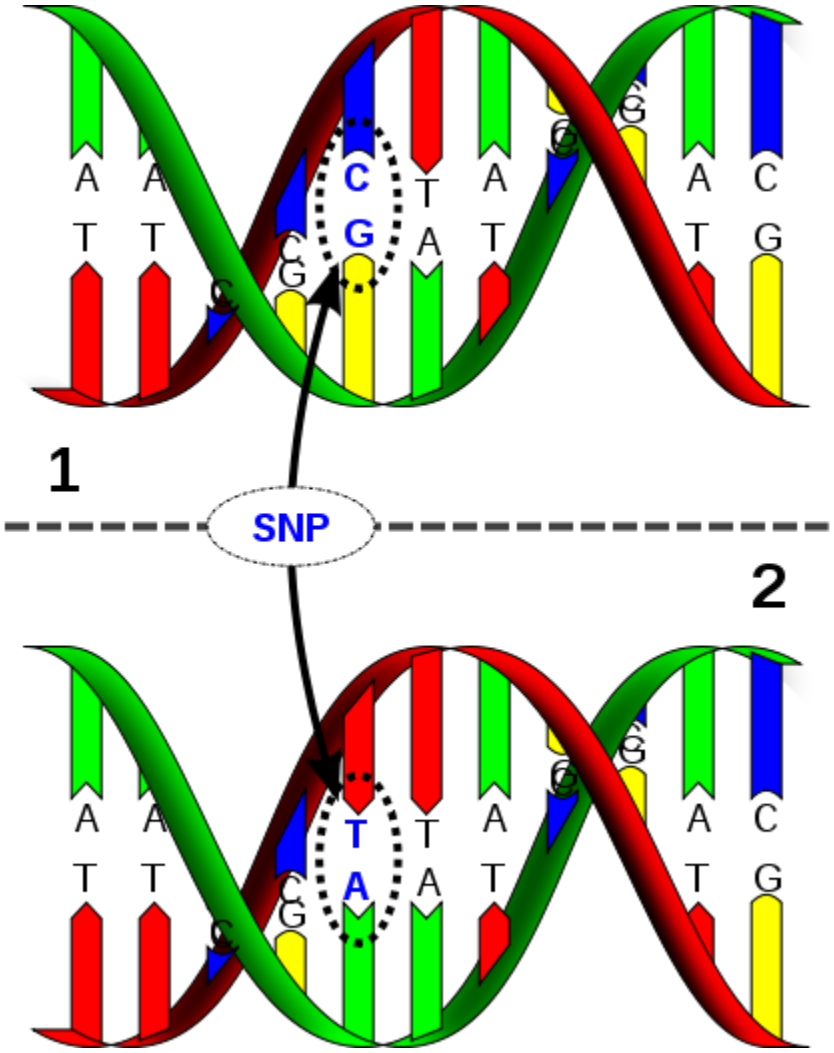
- Mutation
  - When event first occurs in an individual:
    - genetic change due to internal events (e.g., copy errors during cell division) or external agents (e.g., radiation, mutagens).
  - Can end with one generation, or be passed on (germline mutations)

- Polymorphism
  - Means “many forms”
  - Minor allele frequency (MAF) > 1%
  - Generated by old mutations (random, or fitness benefit).

Single Nucleotide Polymorphism (SNPs)

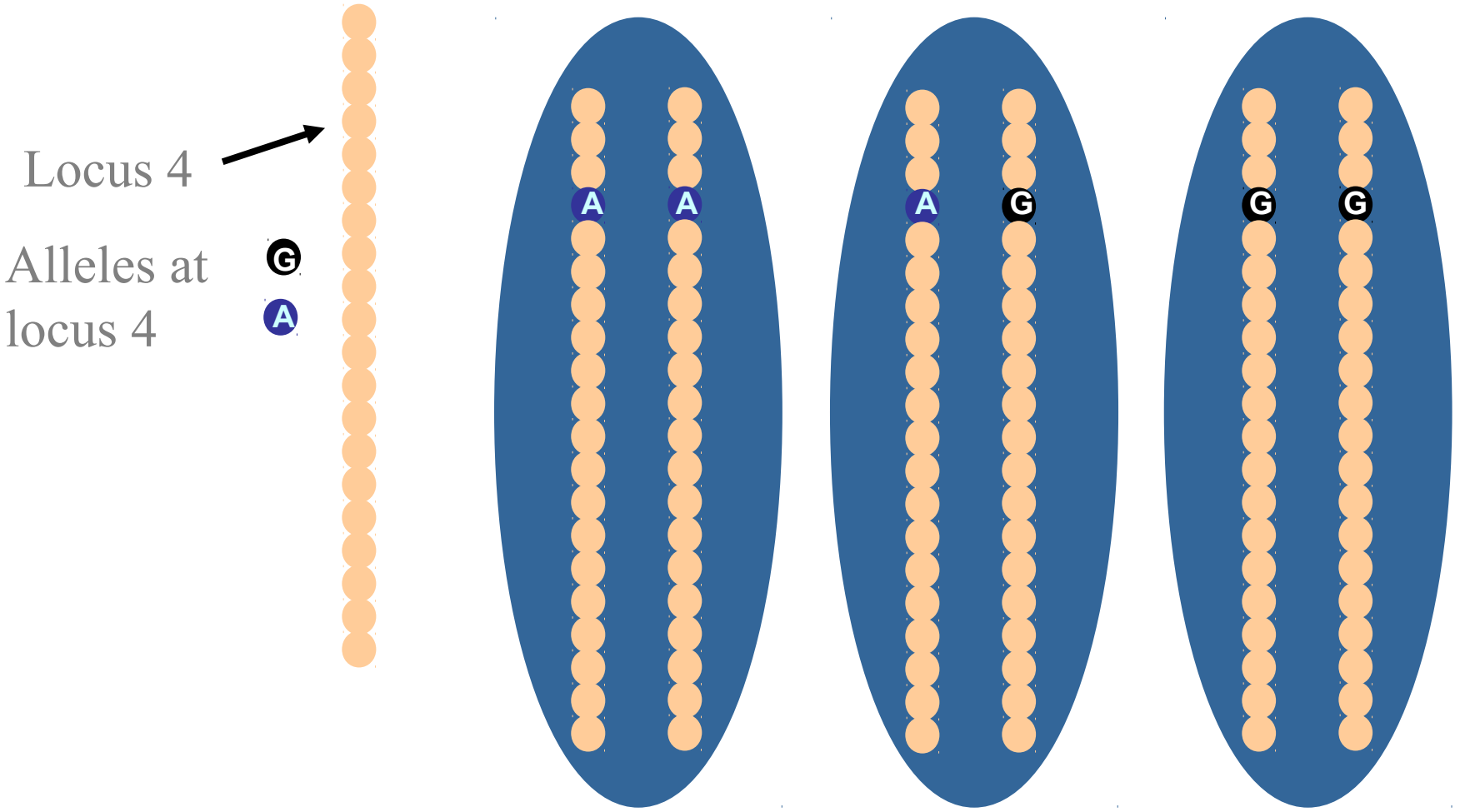
- Change a single DNA letter
- Most frequent genetic variant

- 1 per 300 base pairs
  - Common (MAF>5%)
  - Less common (1-5%)
  - Rare ‘variants’ (<1%)
    - “SNV”



David Hall

Genotypes



Locus: chromosomal location that's polymorphic.
Alleles: different variants @ locus

- Each somatic cell is diploid (two copies of each autosome)
- Thus 3 genotypes at locus 4 (use only one strand, often forward): AA, AG, GG

Types of Variants in Genes

- Noncoding
- Coding
  - Synonymous = no change in amino acid
  - Nonsynonymous/nonsense = change to stop codon
  - Nonsynonymous/missense = change amino acid

e.g., **MTHFR C677T SNP**

Normal ('wild-type') allele

Gene sequence .....GCG GGA G**C**C GAT.....

Protein Sequence .....Ala Gly **Ala** Asp.....

Variant allele

Gene Sequence .....GCG GGA G**T**C GAT.....

Protein Sequence .....Ala Gly **Val** Asp .....

Human Genome Statistics

- 3,283,984,159 basepairs

- 20,442 known protein coding genes
- 649,964 exons
- Short variants (SNPs, indels, somatic mutations): 41,113,446

- Mutation rate ≈ 10^{-8} per bp per generation

- <u>In each person:</u>
  - 65 new mutations expected
  - 1 variant per 1,331 basepairs
  - 2,444,055 variants

- Most variants are old

http://www.ensembl.org/Homo_sapiens

Course Goals – Please Fill out

Name: Tom Hoffmann

Department/Program: Epidemiology and Biostatistics, and Institute for Human Genetics

Background in Genetics: Undergraduate classical genetics course (Snustad & Simmons), many statistical genetics courses

Background in Statistics: PhD in Biostatistics

Research Interests: Methods for and analysis of the underlying genetic architecture of many diseases, esp. age-related hearing impairment.

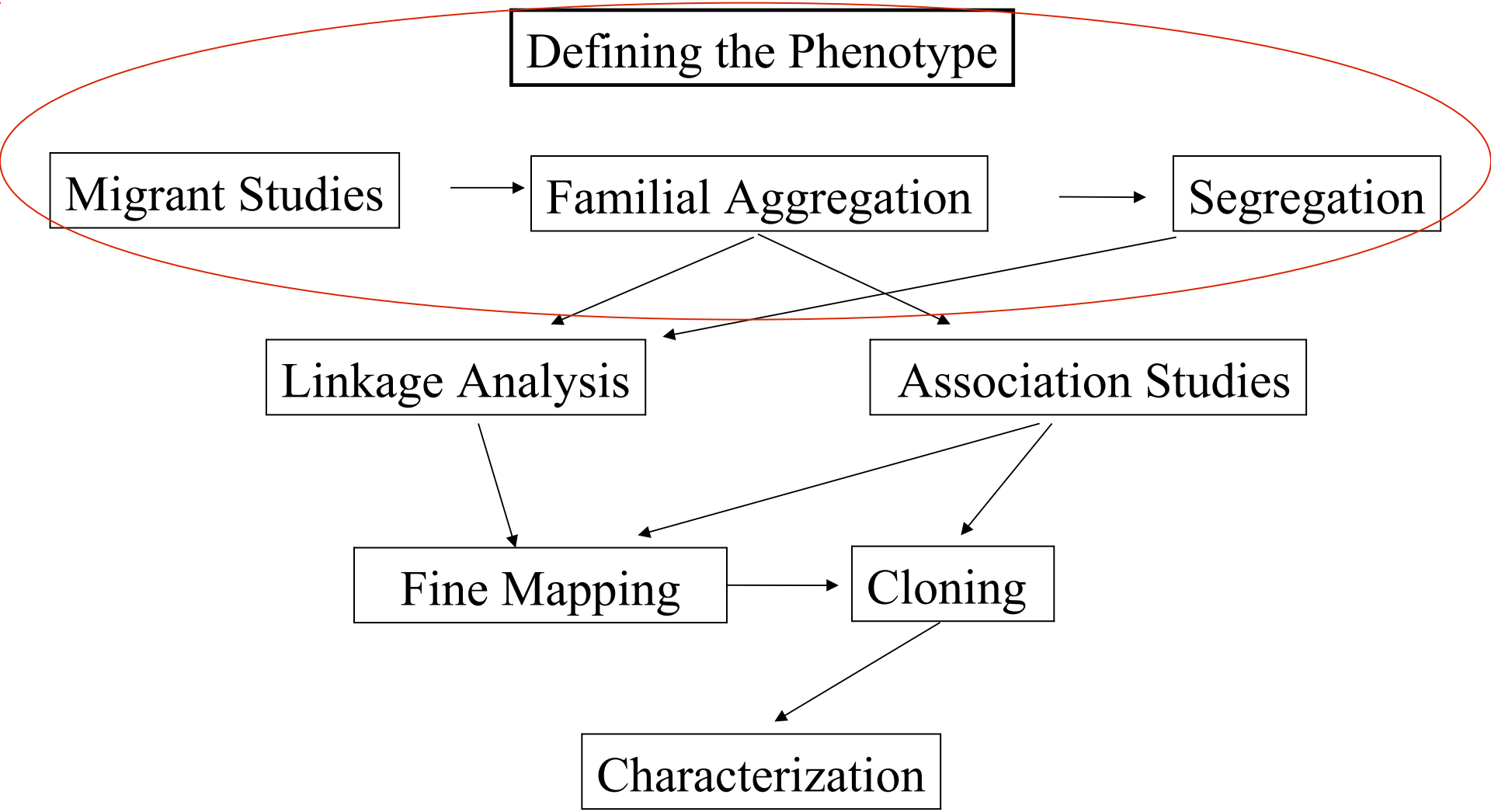
What do you want to get out of this course?

Be able to use genetic information in my research.

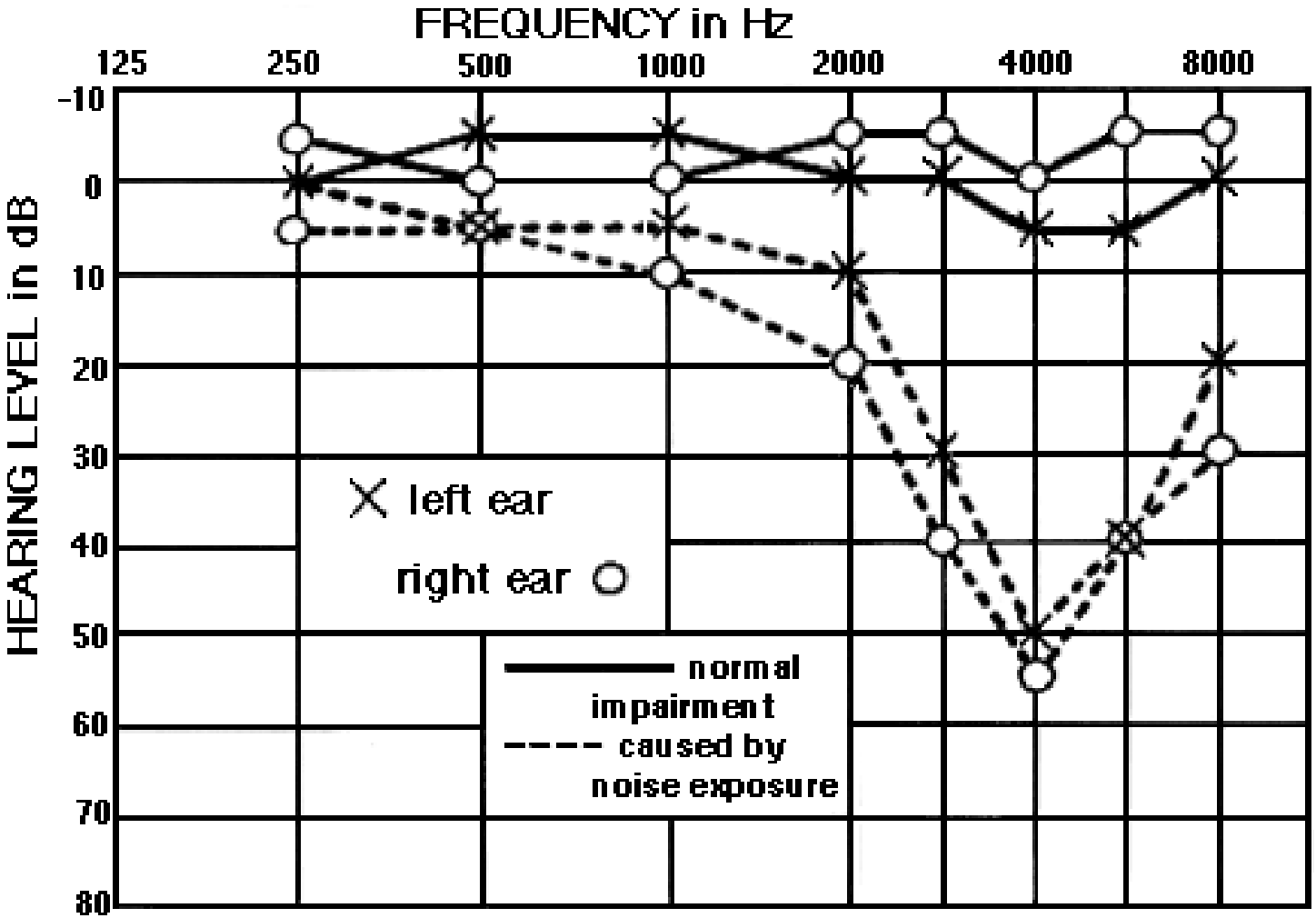
Outline

- Course structure
- Overview of genetics
- **Overview of genetic epidemiology and the rest of the course**

Process of Genetic Epidemiology



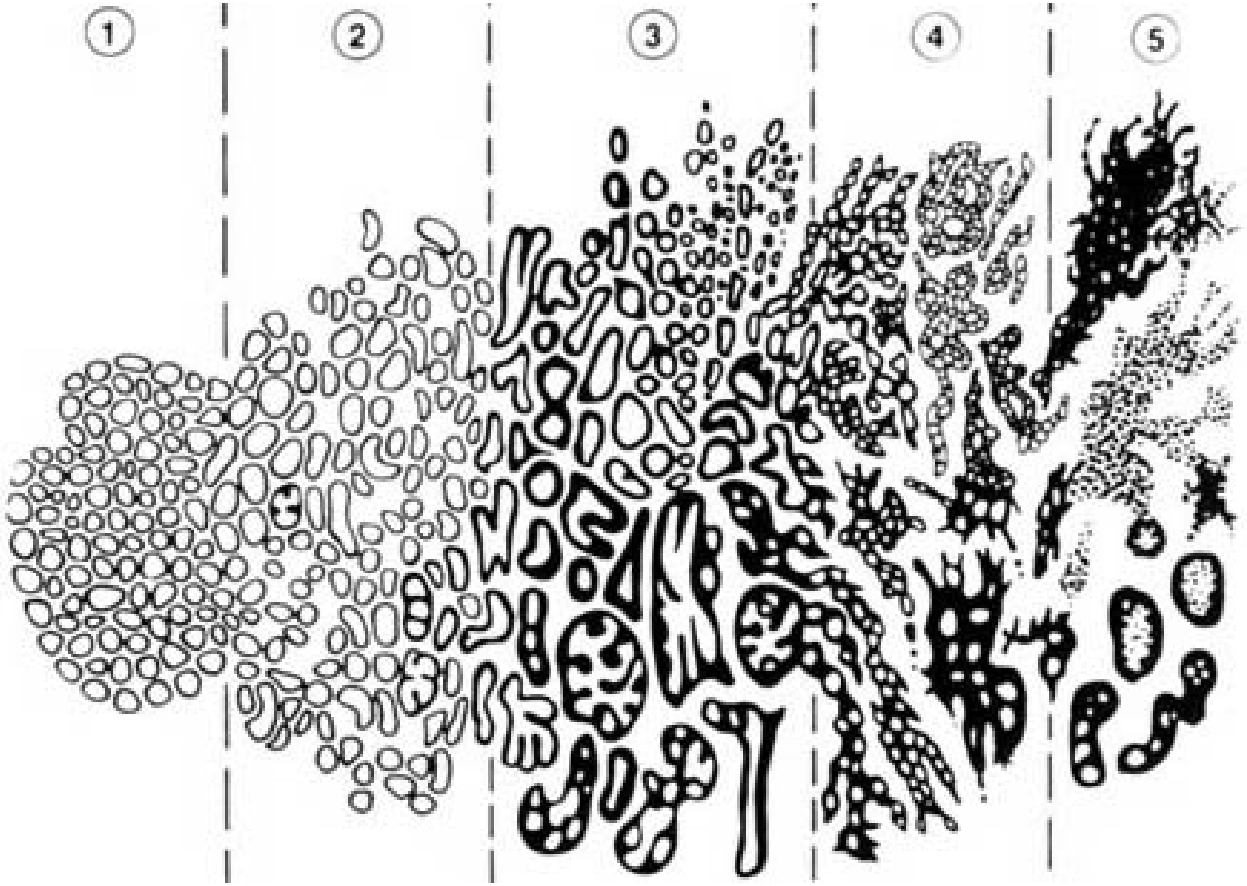
First: Define the Phenotype!



You'll be doing this in your first homework some...

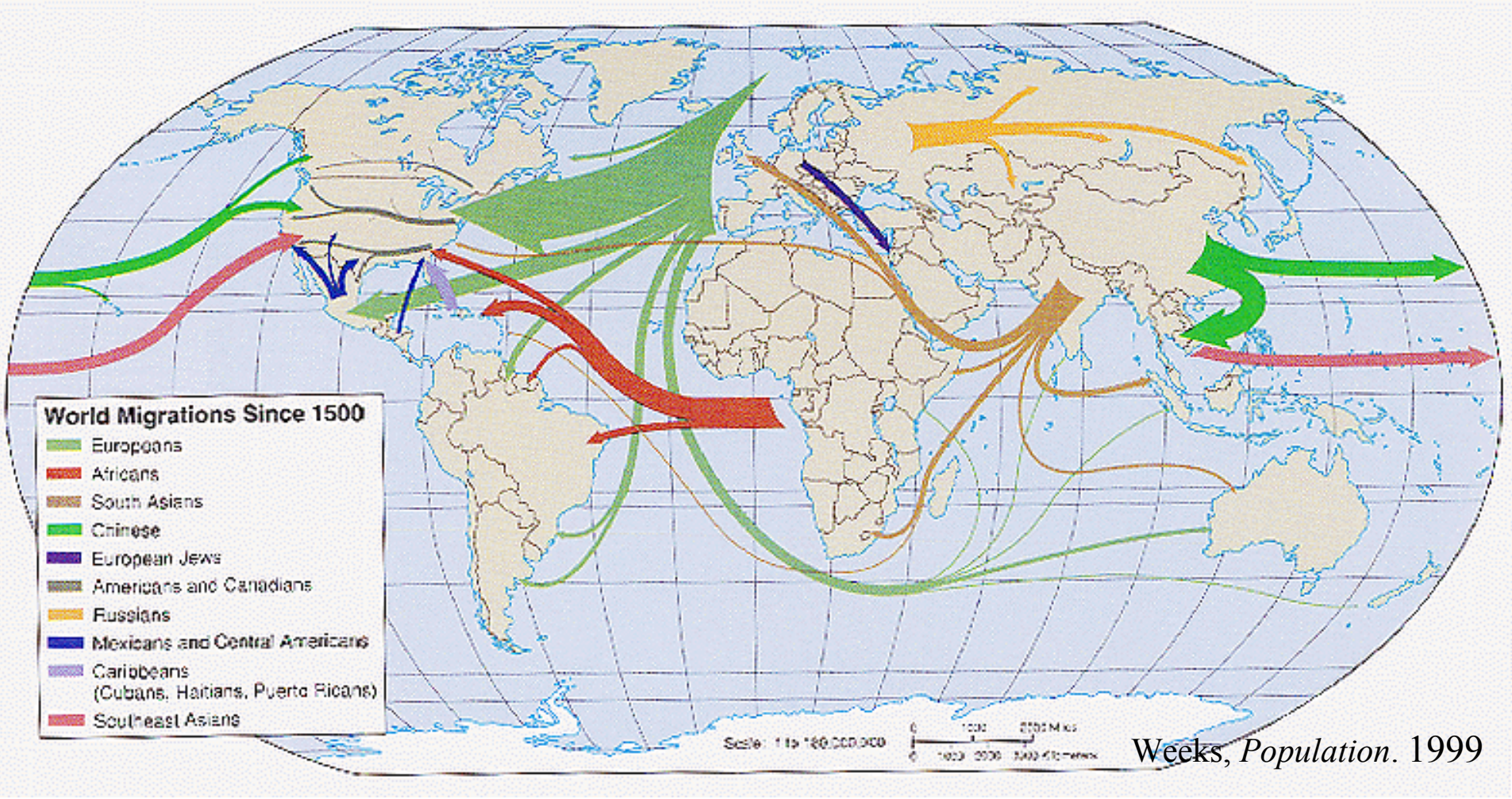
http://www.sfu.ca/sonic-studio/handbook/Audiogram.html

First: Define the Phenotype!



Gleason DF. In Urologic Pathology: The Prostate. 1977; 171-198.

Migrant Studies



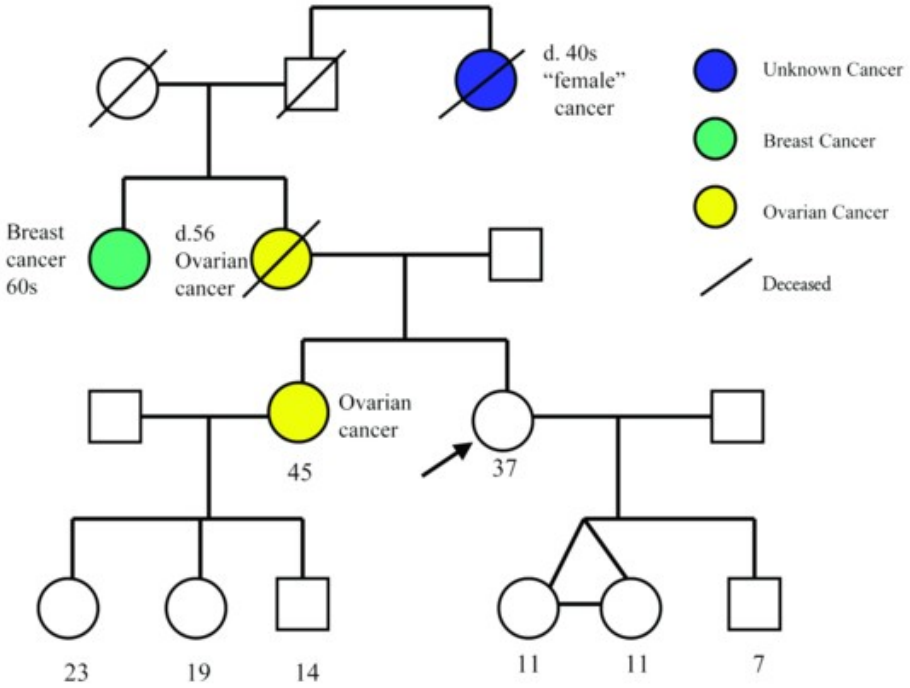
Weeks, *Population*. 1999

Studies based on the assumption that migrants carry a risk that reflects their country of origin, rather than the host country.

"Migrant Studies". Encycolpedia of Biostatistics, John Wiley and Sons 2005.

Familial Aggregation

- Does the phenotype tend to run in families?



Covered more in Austin Ch 2 & Lecture 2: Population Genetics, Modeling Genetic Inheritance

Analysis of Twin Studies

- Compare the disease concordance rates of MZ (identical) and DZ (fraternal) twins.

Twin 1

Twin 2	Disease	Yes	No
	Yes	A	B
	No	C	D

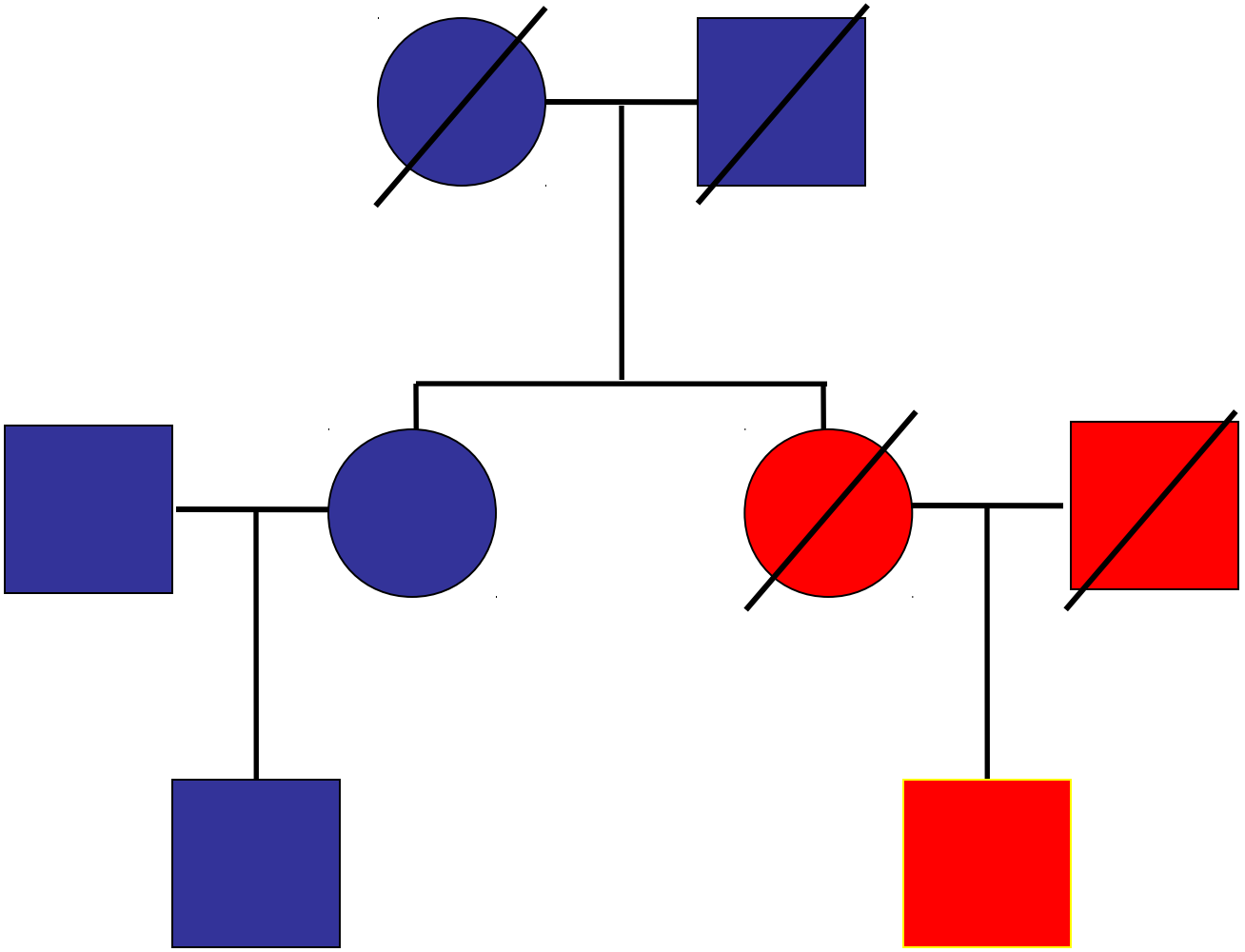
Concordance = 2A/(2A+B+C)

Then one can estimate heritability of a phenotype.

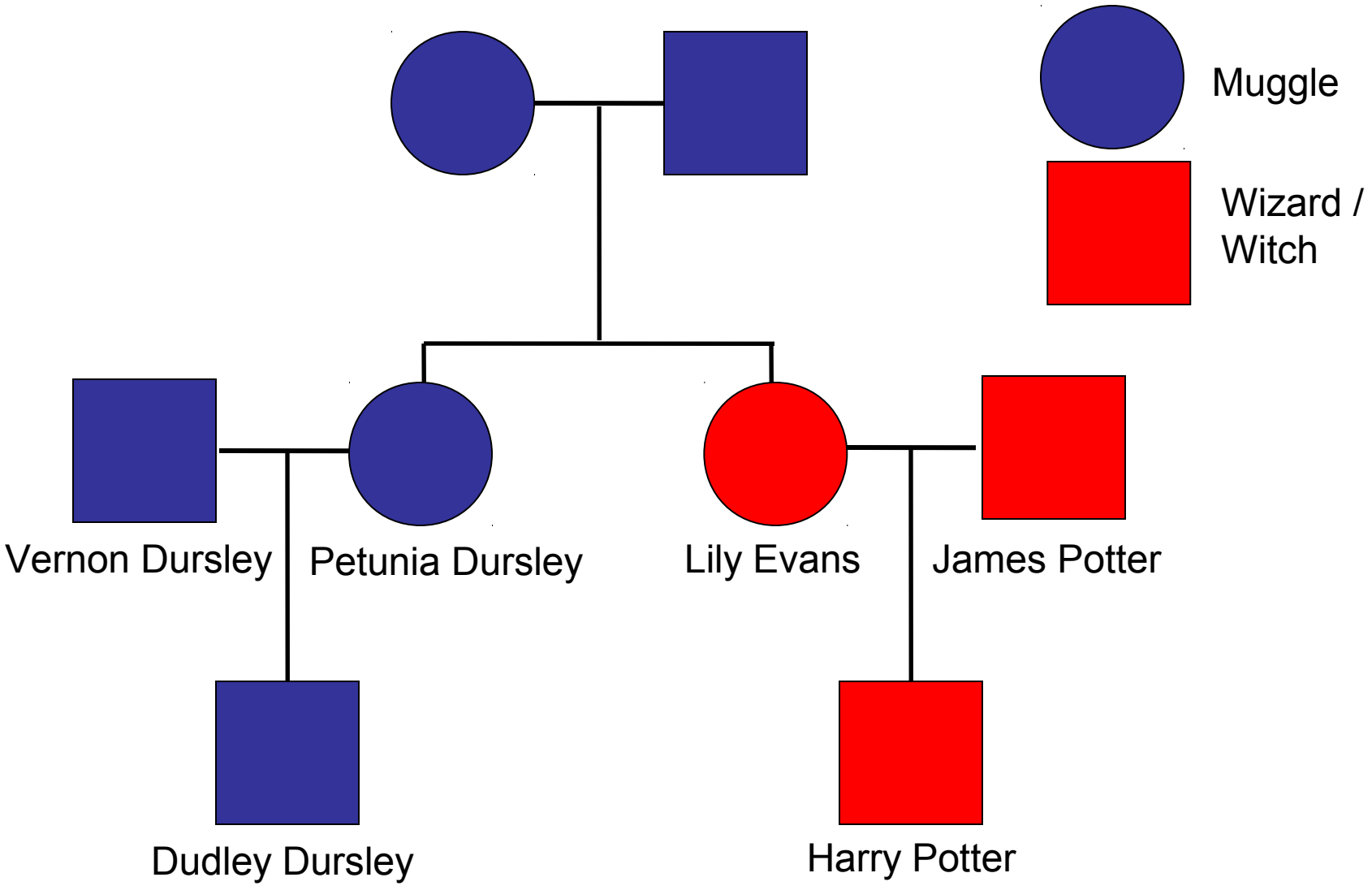
Models of Genetic Susceptibility

- Study families.
- Estimate ‘mode of inheritance’ & what type of genetic variant might be causal.
- Determine whether the disease appears to follow particular patterns across generations.
- Estimate whether variants are rare or common, etc.

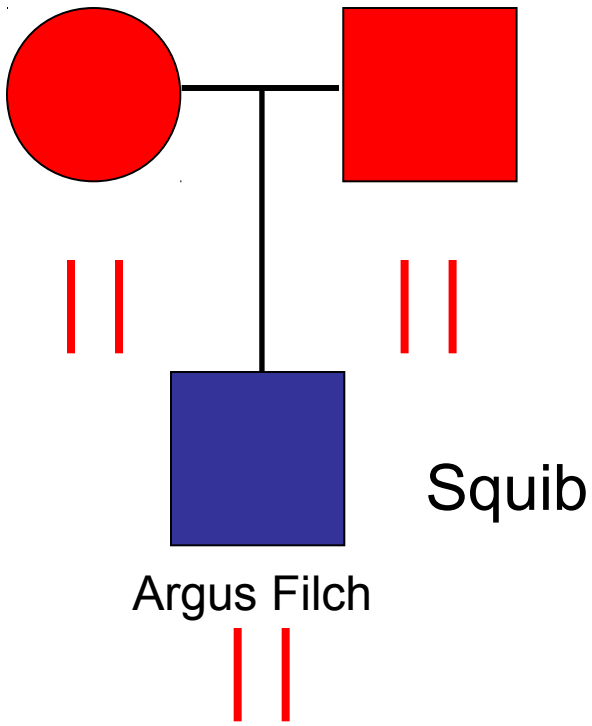
Segregation



Segregation: Harry Potter's Pedigree



Filch?

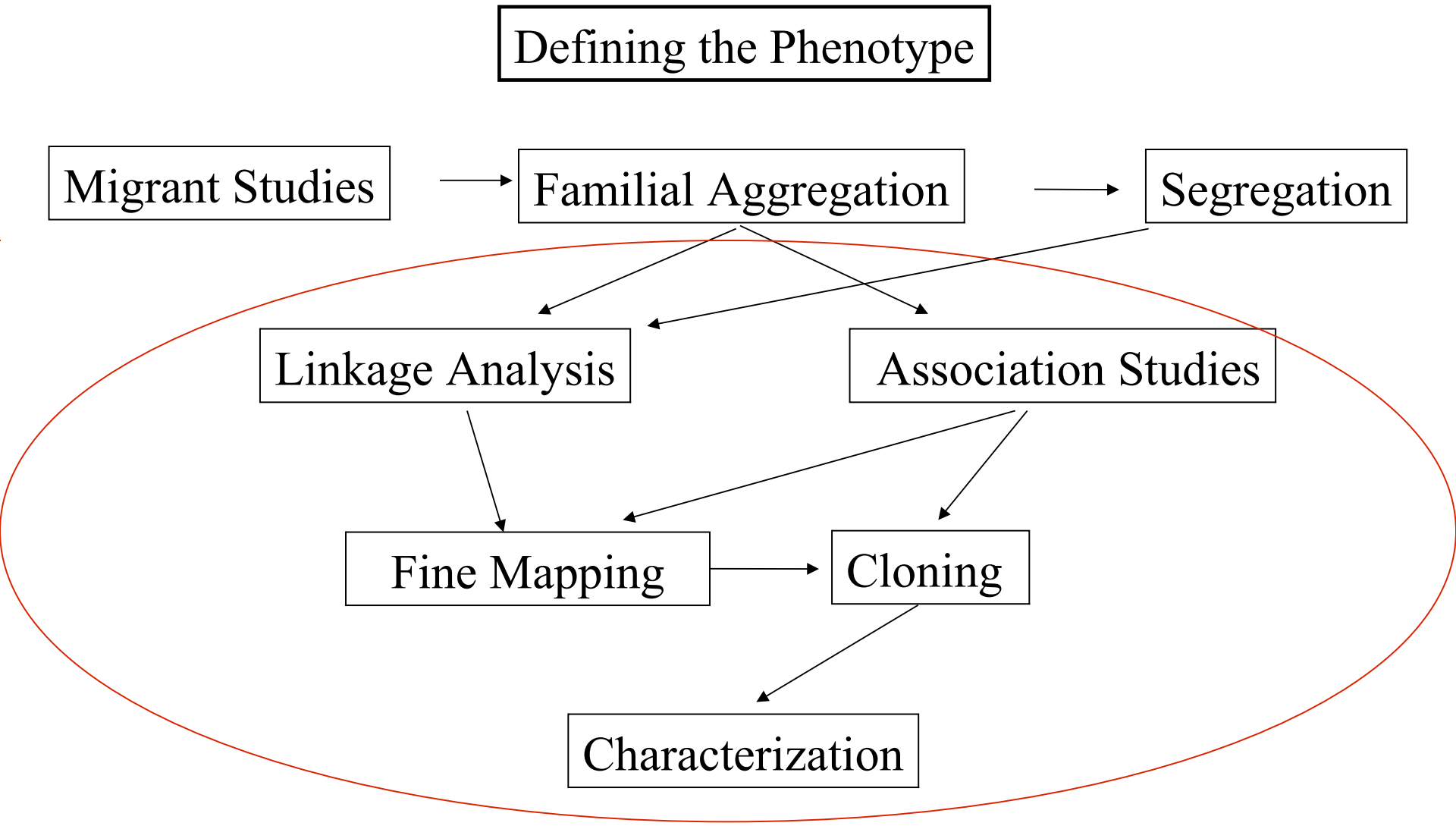


Segregation Analysis

- What is the best model of inheritance for observed families?
    - Dominant
    - Recessive
    - Additive
  - Disease allele frequency?
  - Magnitude of risk?

- Fit formal genetic models to data on disease phenotypes of family members.
- The parameters of the model are generally fitted finding the values that maximize the probability (likelihood) of the observed data.
- This information is useful in parametric linkage analysis, which assumes a defined model of inheritance.

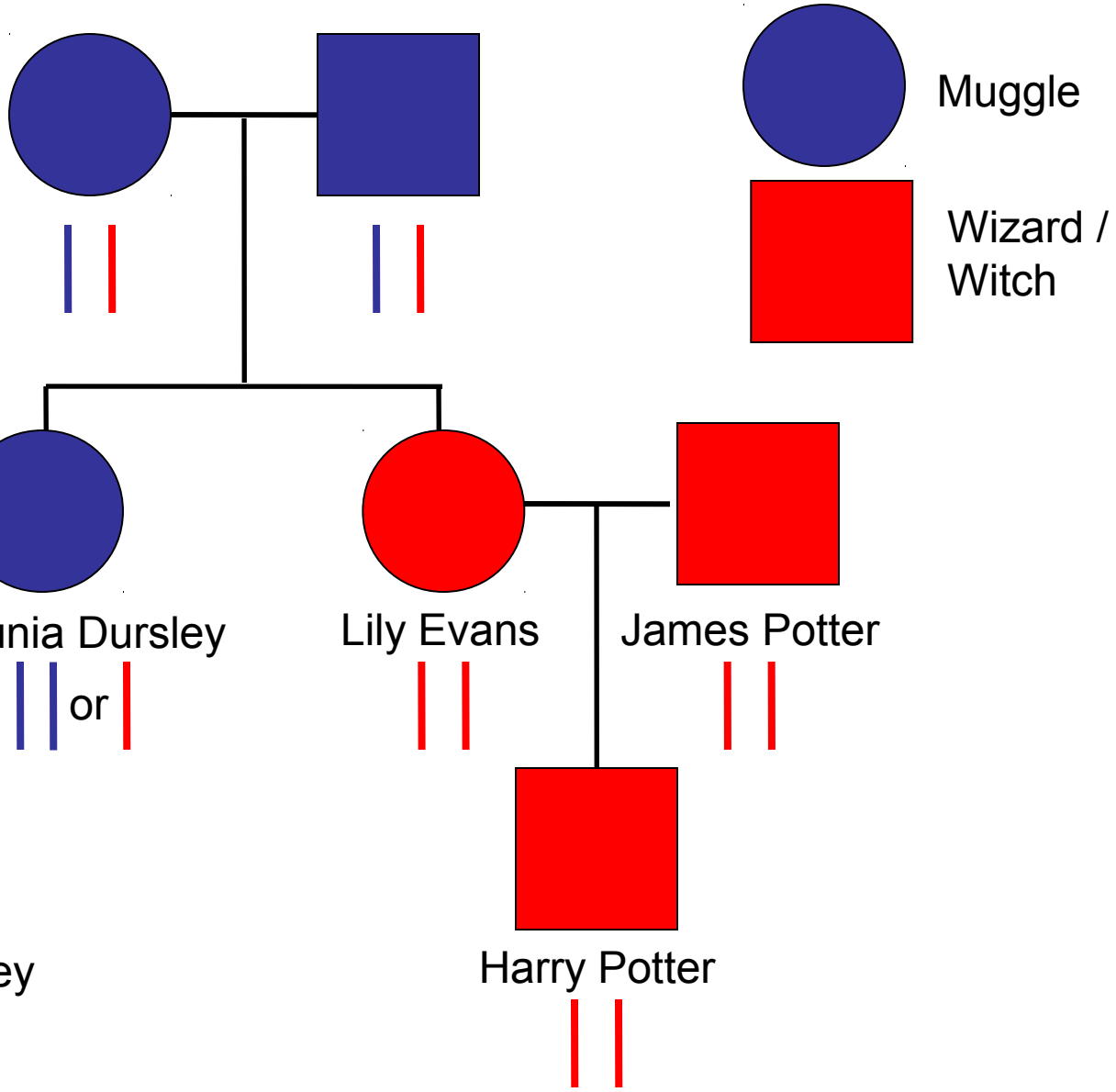
Process of Genetic Epidemiology



Linkage: Harry Potter's Pedigree

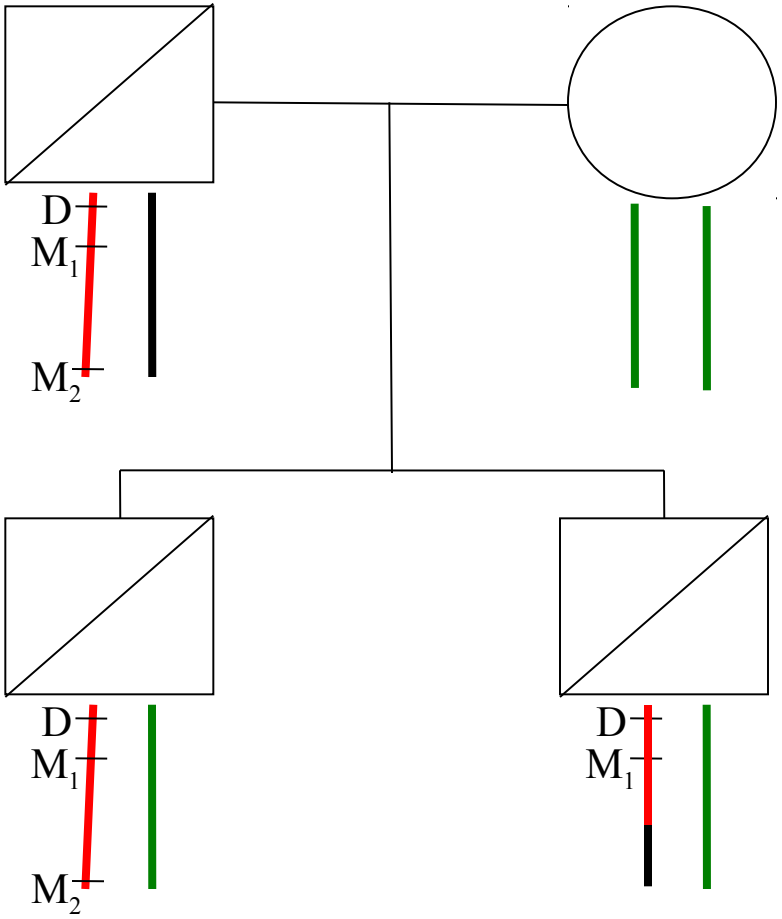
Measure co-segregation in pedigree

Based on detection of recombination events (meiosis)

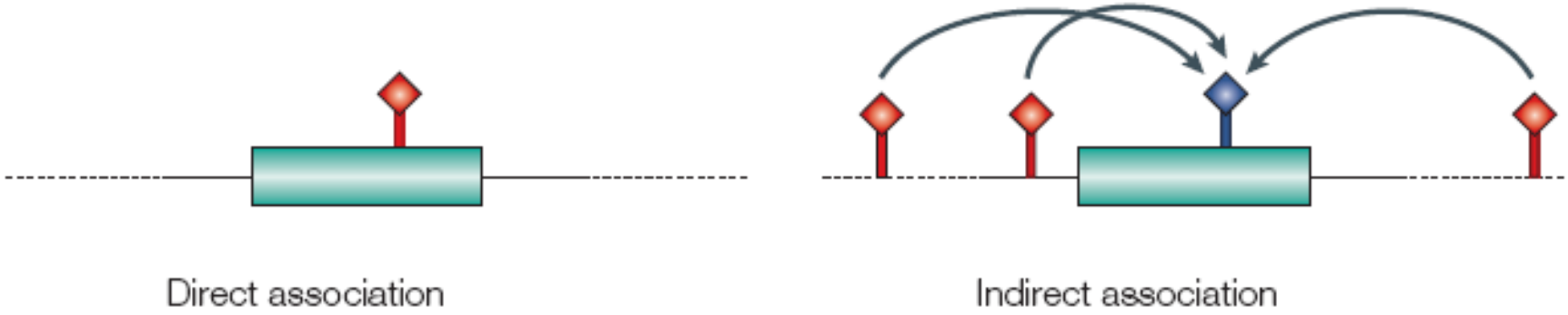


Covered more in Austin Ch 4 & Lecture 4: Family-based studies

Affected sib-pair Linkage



Linkage Disequilibrium

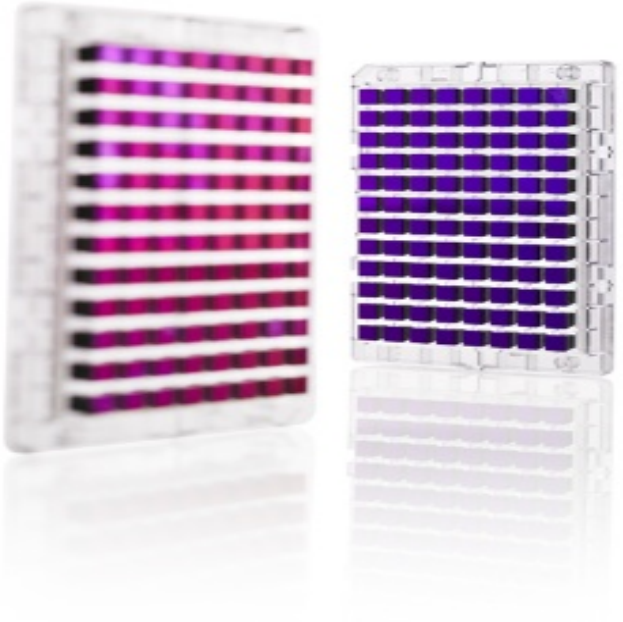


important fundamental concept *will come up repeatedly and be explained further in course*

In population genetics, linkage disequilibrium is the non-random association of alleles at two or more loci, that descend from single, ancestral chromosomes.
http://en.wikipedia.org/wiki/Linkage_disequilibrium

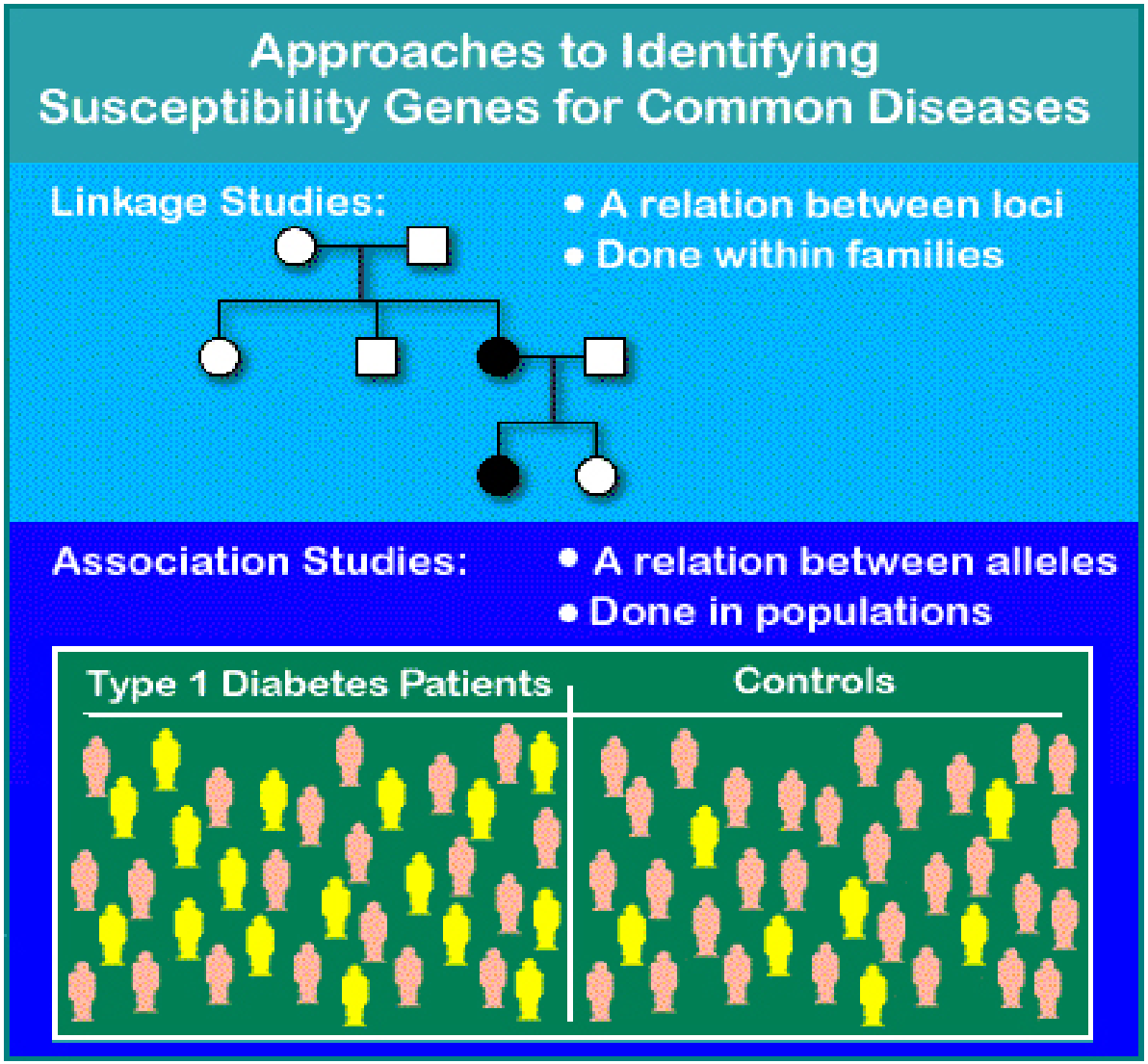
Hirschhorn & Daly, Nat Rev Genet 2005

Genotyping technologies



Covered more in Austin Ch 3 & Lecture 3: Mendel's laws and Molecular Genetics

Association Studies



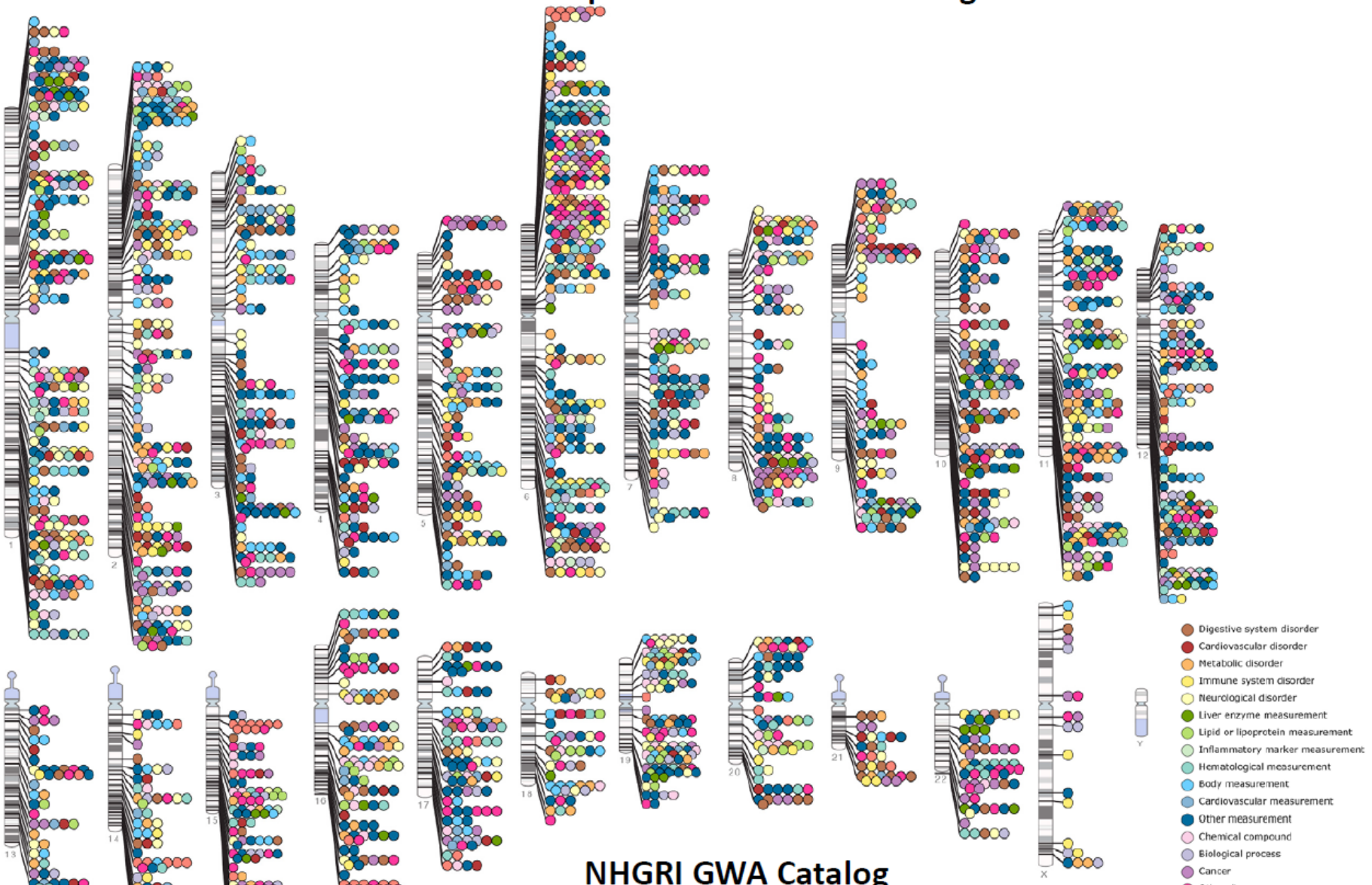
Covered more in Austin Ch 5 & Lecture 5: Genetic Association Studies

Genome-wide Association Studies

Covered more in Austin Ch 5-7 & Lecture 5-7, Genome-wide Association Studies (GWAS)+

Published Genome-Wide Associations through 07/2012

Published GWA at $p \leq 5X10^{-8}$ for 18 trait categories



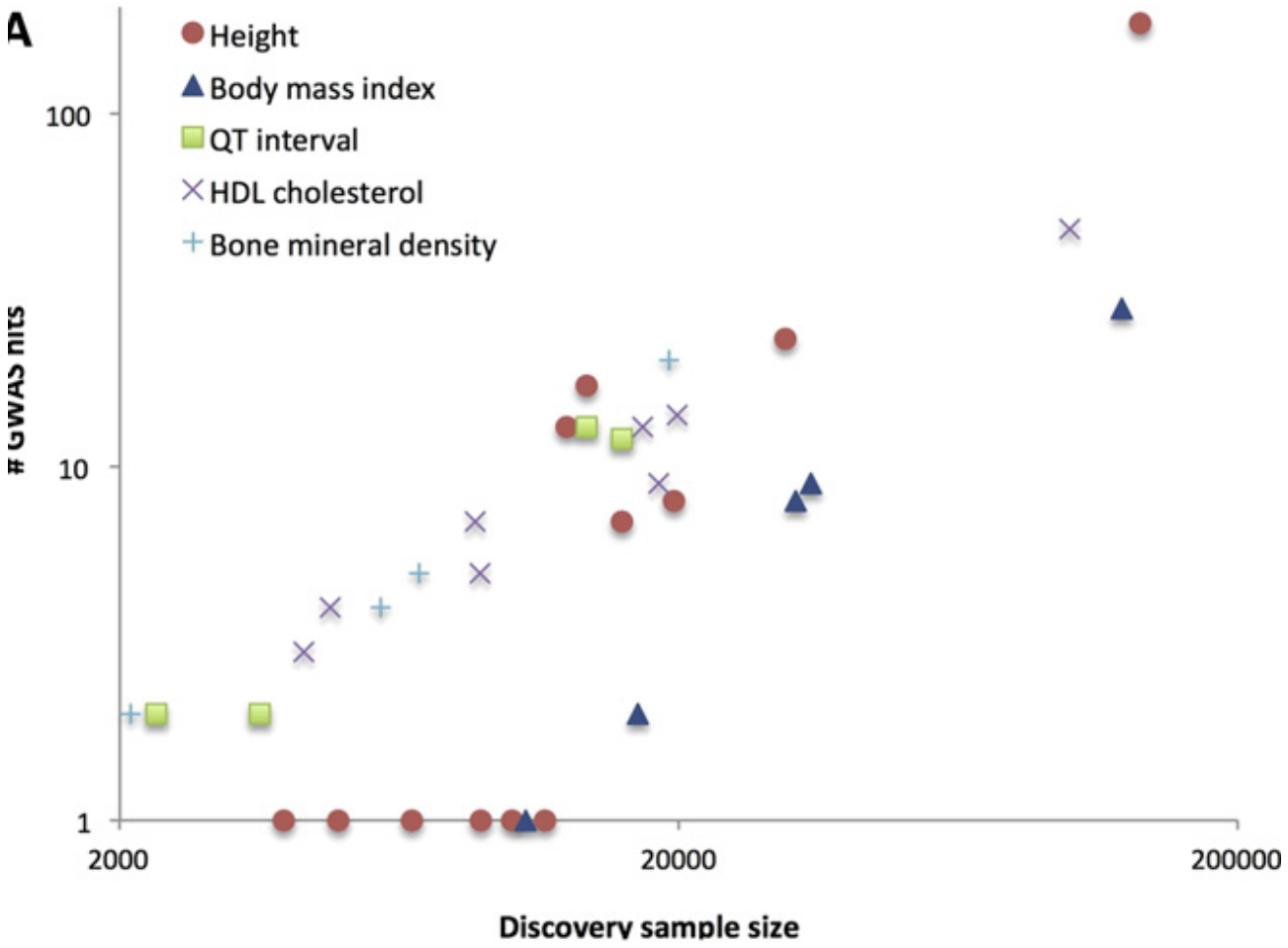
NHGRI GWA Catalog
www.genome.gov/GWAStudies
www.ebi.ac.uk/fgpt/gwas/

National Human Genome Research Institute

EMBL-EBI

Size matters

- Why?
- Consortiums



Visscher, AJHG 2012

Admixture Mapping

- Potentially powerful approach to searching for disease-causing genes
- Requires:
  1. Two populations with naturally occurring phenotypic and genetic differences.
  2. Recent gene flow between the populations (e.g., within 10 generations).
- Markers in the vicinity of the trait locus will also show excess ancestry from the population with the higher allele frequency

Covered more in Lecture 10: Admixture Analysis

Admixture Mapping

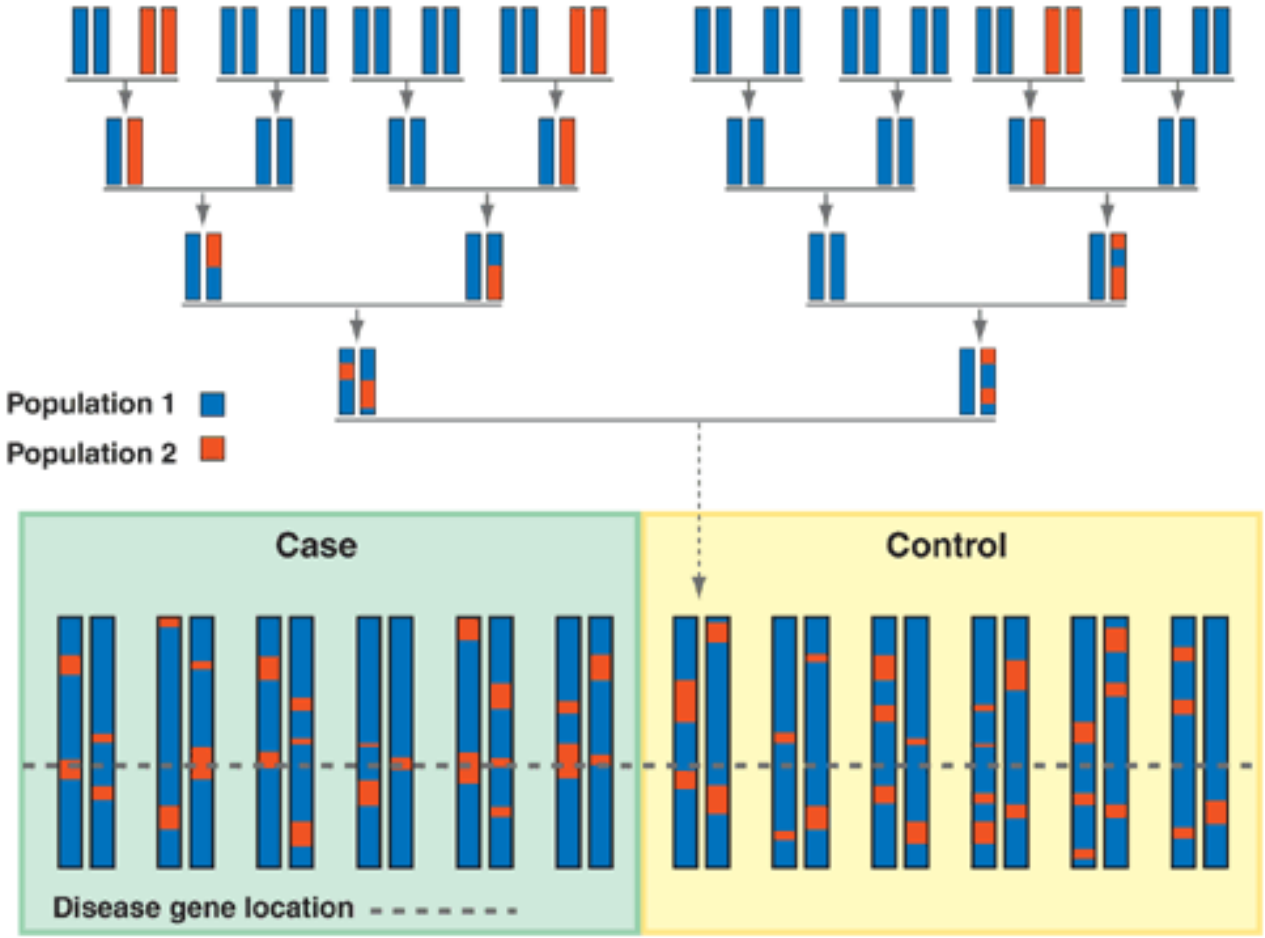


Figure 1 Schematic of one chromosome pair from each of several individuals in an admixed population. A group of cases (for a given disease) and a group of controls are separately presented at the bottom left and the bottom right, respectively. For one of the control individuals (arrow), a schematic presentation of all its ancestors in the last four generations is shown in the upper part of the figure. Admixture mapping can be ideally applied if population 1 (blue) and population 2 (red) carry a different allele at the disease locus (dashed line). Whole-genome scanning under the admixture mapping strategy consists of scanning the genome and identifying the regions with an excess of 'red' ancestry in the cases versus the controls, assuming that the 'red' population carries the predisposition allele. The size of the blocks from different ancestors will depend on the number of generations since the populations were mixed.

Nature Genetics 37, 118 - 119 (2005)

Summary of Main Mapping Approaches

	Linkage Analysis	Admixture	Association Study
Power*	Low	Moderate/High	High
# SNPs required for scan	Low	Low	High
Sensitivity to genetic heterogeneity	Low	Moderate	High
Mapping resolution	Poor	Intermediate	Good

Nature Genetics 37, 118 - 119 (2005)

Cloning a Gene

- Showing that it is clearly causal for disease.
- Generally requires experiments beyond those undertaken by a genetic epidemiologist.

Re-Sequencing Genomes (Ozzy Osbourne?)

cientists to map Ozzy Osbourne's enetic code to find out how he survived o much substance abuse

NICK KLOPSIS
ILY NEWS WRITER

nday, June 14th 2010, 1:39 PM

Sayles/AP

zzy Osbourne has had many issues relating to drug and alcohol abuse, so scientists are apping out his genetic code to find out how his body can take it.

u can't kill rock and roll, but it's not ually this hard to kill a rocker.

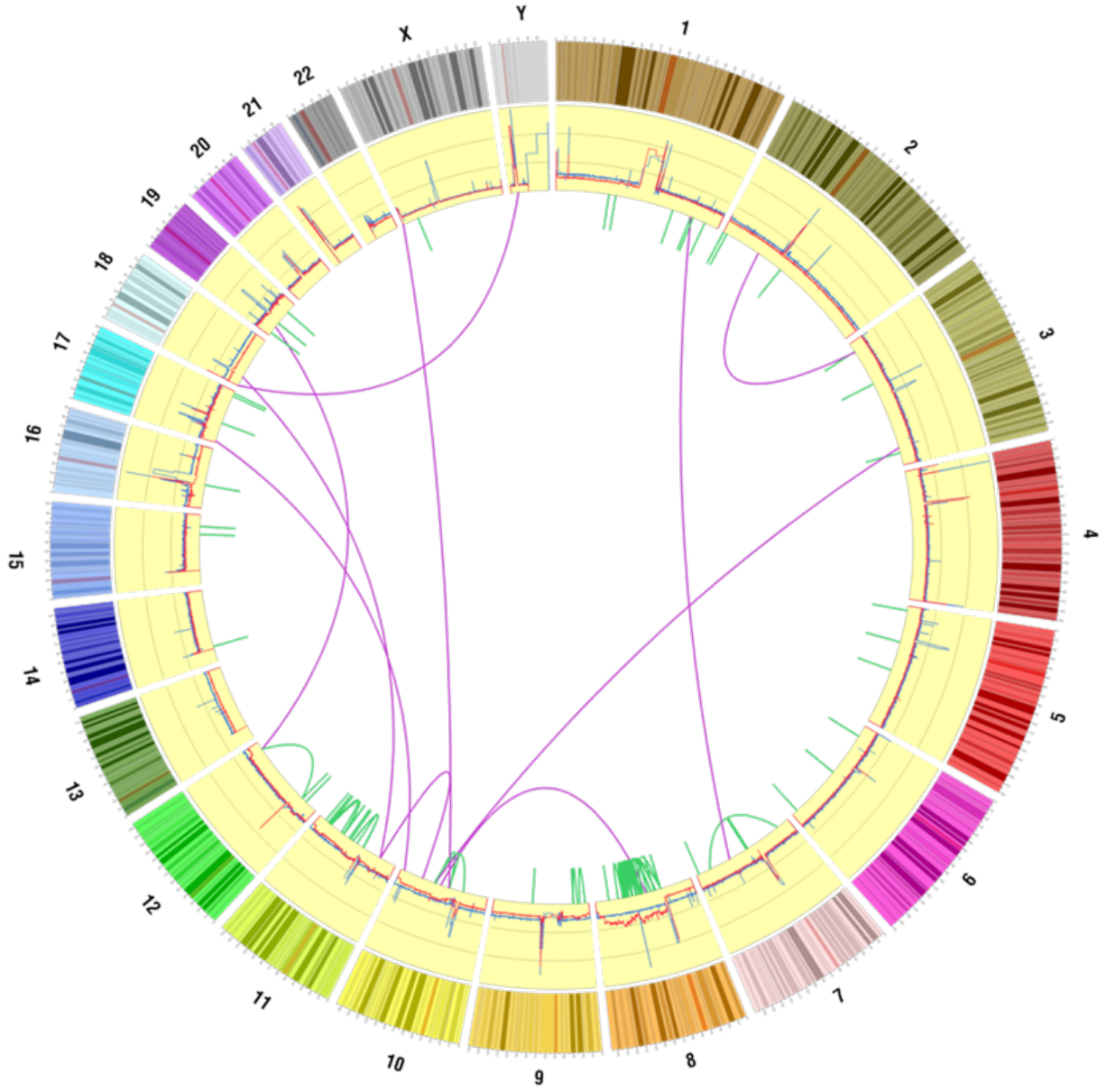
RELATED NEWS

"Sequencing and analysing individuals with extreme medical histories provides the greatest potential scientific value.“

Nathan Pearson, Director of Research Knome

Covered more in Lecture 9: Next-generation sequencing

Circos Plot: Tumor – Normal



Remi Kazma

Characterization

- Once genes are identified, molecular methods are used to determine the structure of the gene, identification of regulatory elements, etc.

- Use epidemiologic studies to distinguish public health implications:
  1. Determine frequencies of causal alleles; and
  2. Characterize their effects—and interacting environmental factors—on disease rates.

Genetic Testing?

"All the News That's Fit to Print"

The New York Times

Late Edition

VOL. CLVII . . No. 54,393 © 2008 The New York Times NEW YORK, THURSDAY, JANUARY 17, 2008 $1.25

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Cumulative Association of Five Genetic Variants with Prostate Cancer

S. Lilly Zheng, M.D., Jielin Sun, Ph.D., Fredrik Wiklund, Ph.D., Shelly Smith, M.S., Pär Stattin, M.D., Ph.D., Ge Li, M.D., Hans-Olov Adami, M.D., Ph.D., Fang-Chi Hsu, Ph.D., Yi Zhu, B.S., Katarina Bälter, Ph.D., A. Karim Kader, M.D., Ph.D., Aubrey R. Turner, M.S., Wennuan Liu, Ph.D., Eugene R. Bleecker, M.D., Deborah A. Meyers, Ph.D., David Duggan, Ph.D., John D. Carpten, Ph.D., Bao-Li Chang, Ph.D., William B. Isaacs, Ph.D., Jianfeng Xu, M.D., D.P.H., and Henrik Grönberg, M.D., Ph.D.

$300 to Learn Risk of Cancer Of the Prostate

A New Test Is Raising Treatment Questions

FED'S CHAIRMAN IS SAID TO BACK AID TO ECONOMY

TALKS WITH LAWMAKERS

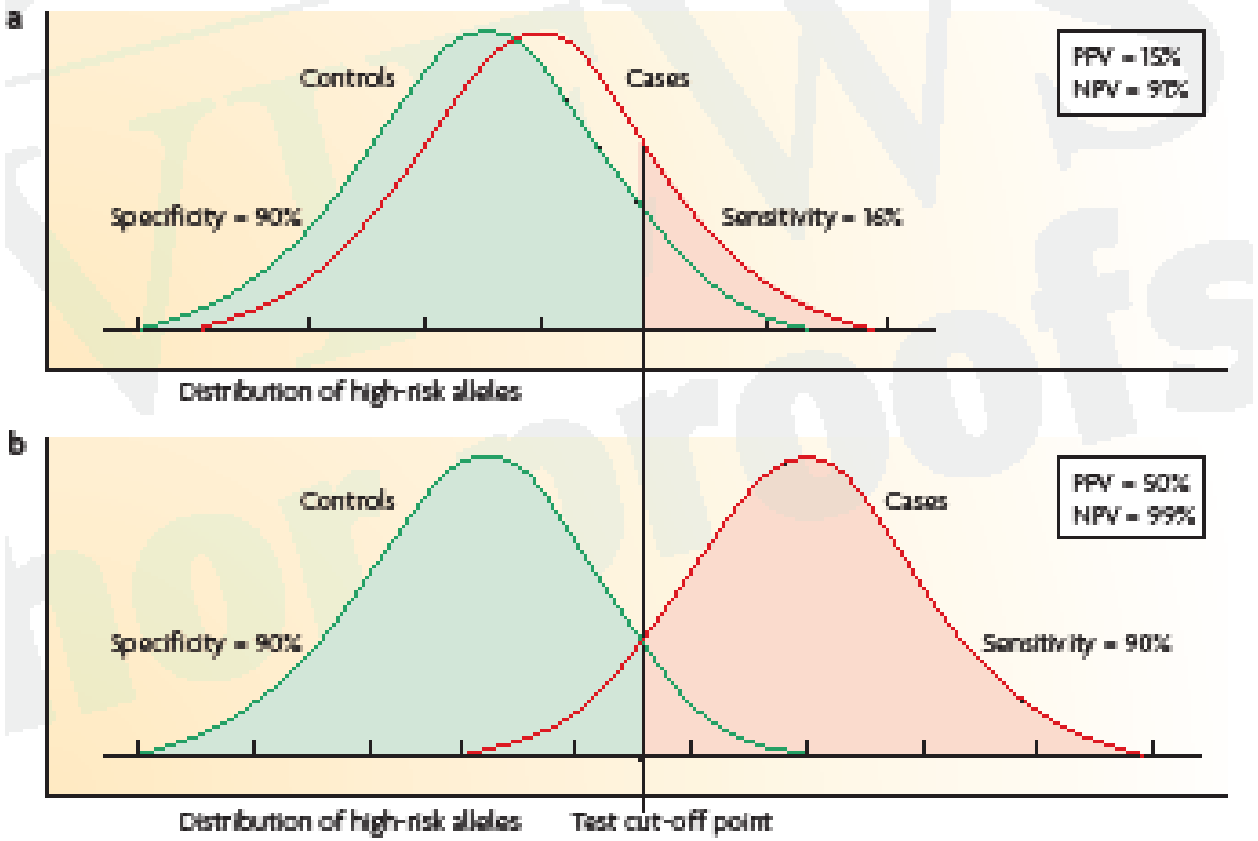
Crucial Support Seen on Short-Term Package for Stimulus

About Giuliani

in South Carolina, Lesson Learned

Tactics

Large RR ≠ Good Prediction



Witte, Nat Rev Genet, 2009

Genetic Testing Based on GWAS?

- Multiple companies marketing direct to consumer genetic ‘test’ kits.
- Send in spit.
- Array technology (Illumina / Affymetrix).
- Many results based on GWAS.

- Companies:
  1. 23andMe
  2. deCODEme
  3. Navigenics

Order

Register

Spit

Send

Discover

For One Price:

Ancestry

Reveal the magic in your DNA, and understand your ancestral history using:

- Relative Finder
- Global Origins
- Ancestral Lineages

Health

Understand how your genes impact your health.

- Carrier Status
- Disease Risk
- Drug Response

Start exploring your DNA...

Add to Cart

23andMe

welcome | health | ancestry | how it works | buy | help

Please read the following statement below and click the "I understand" button to enter 23andMe.com.

Welcome to 23andMe.

At this time, we have suspended our health-related genetic tests to comply with the U.S. Food and Drug Administration's directive to discontinue new consumer access during our regulatory review process.

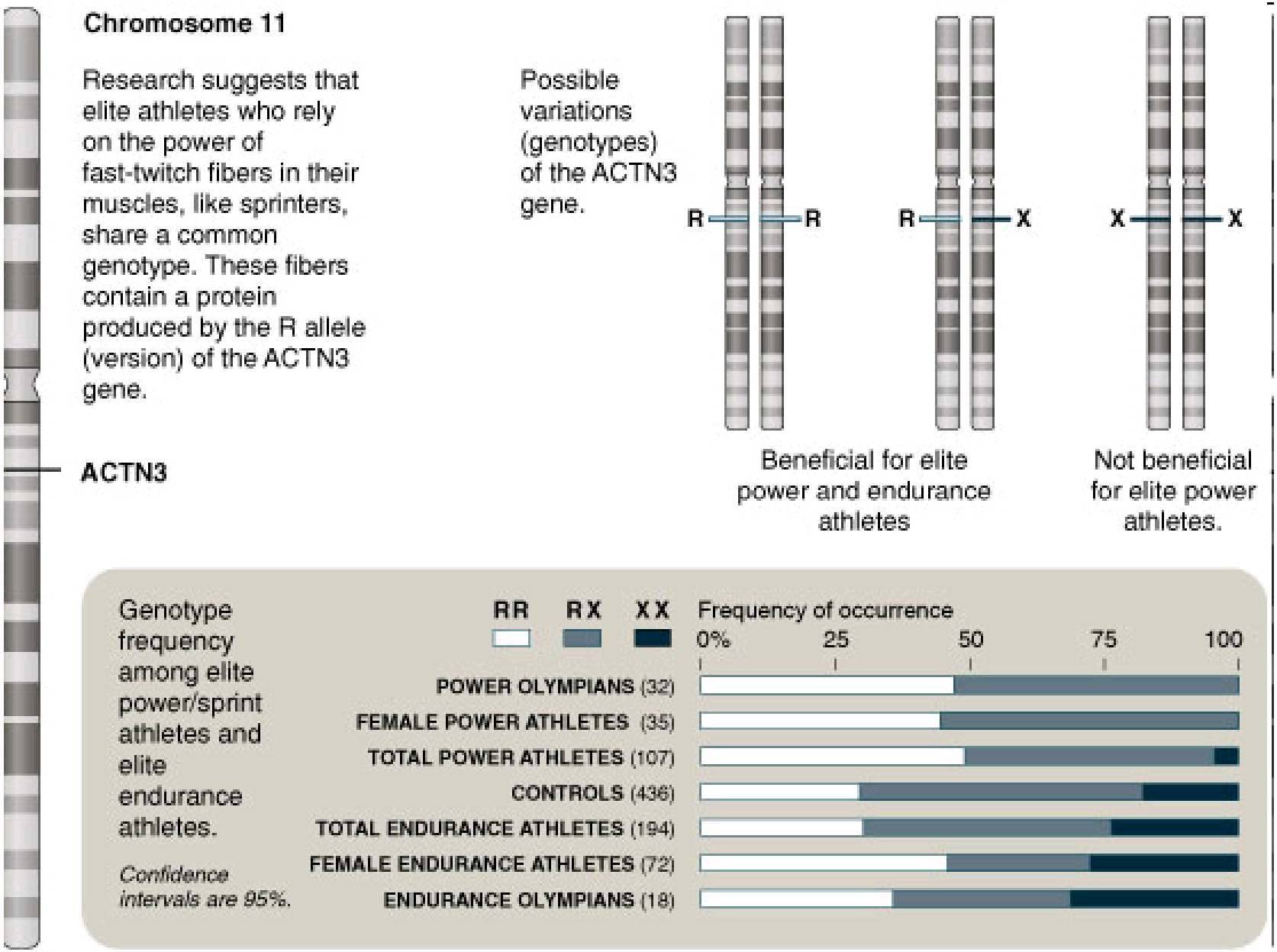
We are continuing to provide you with both ancestry-related genetic tests and raw genetic data, without 23andMe's interpretation.

If you are an existing customer please click the button below and then go to the health page for additional information, including information about refunds.

We remain firmly committed to fulfilling our long-term mission to help people everywhere have access to their own genetic data and have the ability to use that information to improve their lives.

Upon entering the site, please confirm you understand the new changes in our services.

I understand that 23andMe only sells ancestry reports and raw genetic data at this time. I understand 23andMe will not provide health-related reports. However, 23andMe may provide health-related results in the future, dependent upon FDA marketing authorization.



Sources: Stephen M. Roth, Ph.D., University of Maryland; American Journal of Human Genetics

NY Times, 11/30/08

ATLAS SPORTS GENETICS

ATHLETIC TALENT LABORATORY ANALYSIS SYSTEM

Finding any great Olympic champion normally takes years to determine.

What if we knew a part of the answer when we were born?

See How...

The New York Times - Born to Run? Little Ones Get Test for Sports Gene

GIST™ Sports Profile

Genetic Testing for Speed/Power and Endurance Events

ATLAS SPORTS GENETICS

ATHLETIC TALENT LABORATORY ANALYSIS SYSTEM

larger image

ATLAS First

Recommended for ages 1 and up.
Description: Our *Atlas First* product is geared specifically at the youngest of athletes. Doing any type of performance based sport talent identification testing is very difficult below age 6 due to developmental levels of motor skills, strength and eye-hand coordination. *Atlas First* looks at only genetic markers, specifically the presence of ACTN3. Studies have found that individuals having the variant in both copies of their ACTN3 gene may have a natural predisposition to endurance events, one copy of their ACTN3 gene may be equally suited to for both endurance and sports/power event, neither copy of their ACTN3 gene may have a natural predisposition to sprint/power events. Knowing this information may be helpful, not in eliminating choices for sport activities but adding exposure to a host of team or individual sport events that may come easier to a young athlete.

The test is one of tool of many that can help children realize their athletic potential.

Other Products available through Atlas First

- Height monitoring charts
- Weight monitoring charts
- BMI charts
- Height Prediction Calculator (not genetic)

$149.00

Actress Vanessa Williams Explains How DNA Powers Her Family Tree

The following appeared in the Los Angeles Times*, April 27, 2013. Written by Jessica P. Ogilvie.*

The world within Vanessa Williams
5 QUESTIONS

Most of us are curious about our family lineage. For Vanessa Williams, who recently took part in the show "Who Do You Think You Are" and explored her family's history, the task was both surprising and informative. Here, she talks about what she learned and how she plans to use that information.

How did you become interested in finding out about your lineage?

Genetic Taste Testing & Earwax Project

- Taste testing
  - Strips coated with Phenylthiocarbamide (PTC, or phenylthiourea).
  - Bitter or tasteless, depending on variants in the taste receptor TAS2R.
- Earwax
  - Wet or dry?
- What do you think your phenotype is?


Genomic Ads are available NOW!
Ads are available at this genomic spot
www.biocomicals.com