History

Joint Genotyping Task Force

Charles Prober Dean
Russ Altman Genetics
Pat Brown Biochem.
Mike Grecius Neur.
Carlos Bustamente Gen.
Ralph Horwitz Psych
Anne James Legal Counsel
Stuart Kim Dev. Bio.
Phil Lavori HRP
Kelly Ormond Genetics
Mike Snyder Genetics
Keyan Salari Med. School

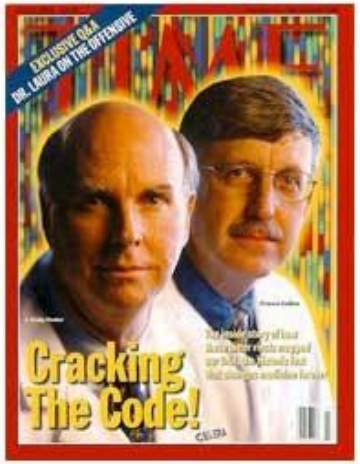
Hank Greely Law School
Clarence Braddock Med School
Gil Chu Biochem
Sean David Med. School
Harry Greenberg Dean
Louanne Hudgins Epidemiology
Jesse Karmazin Med. School
Mark Krasnow Biochem
David Magnus Cen. BME
Alan Schatzberg Psych.
Atul Butte BMI
Mildred Cho Pediatrics

Personal Genotyping

- Voluntary. You can use a public genome file instead of your own.
- Confidential – instructors will not know who opted to be genotyped.
- Private – You will not be asked to reveal your own private DNA information.
- Counseling - genetic counseling via 23andMe and medical/psychological counseling via Dr. Alan Schatzberg (Psychology, Stanford).

Human Genetic Diversity

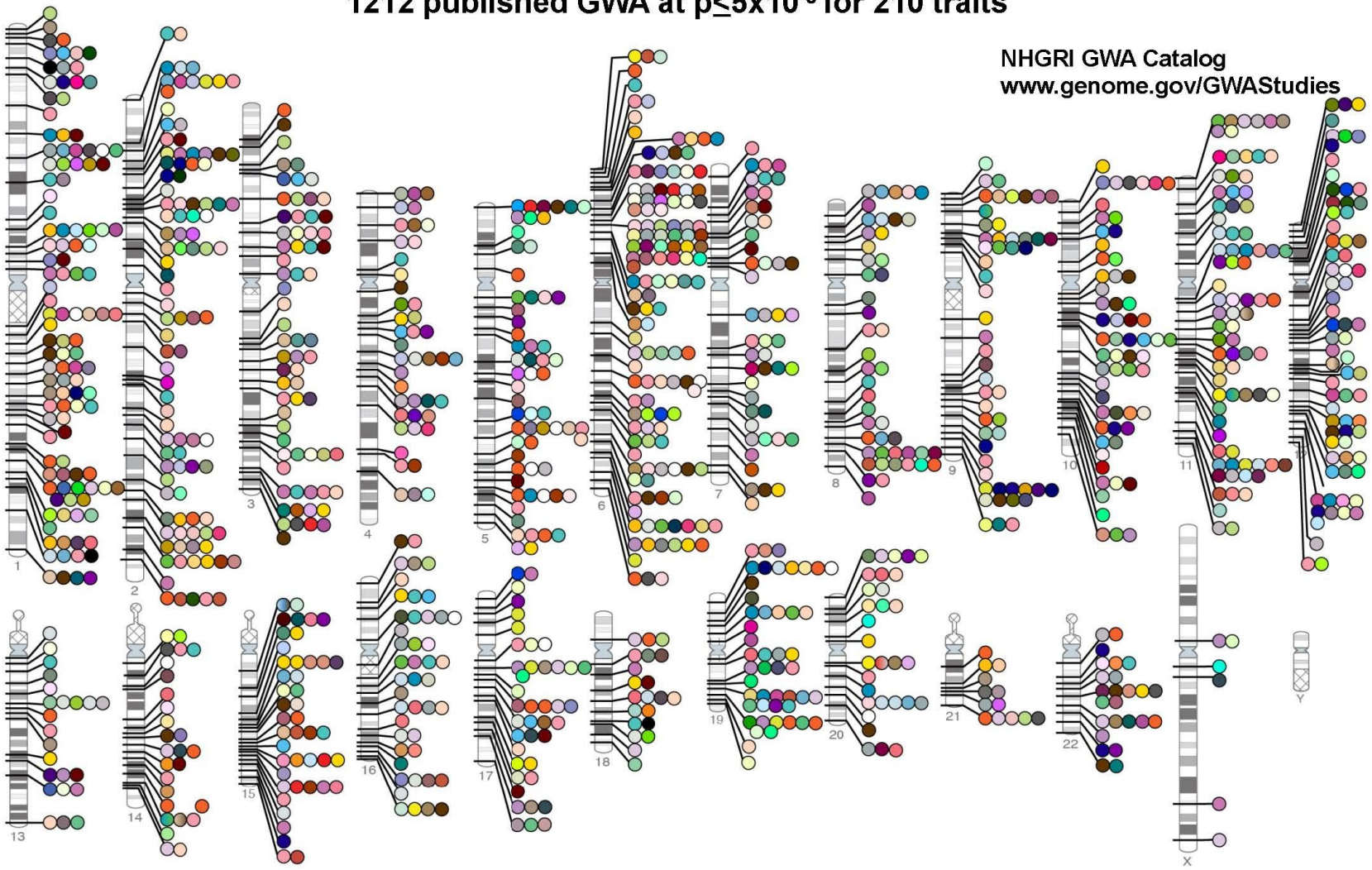
2000



0.1% x 3.3 billion
= 3,300,000 bp of differences

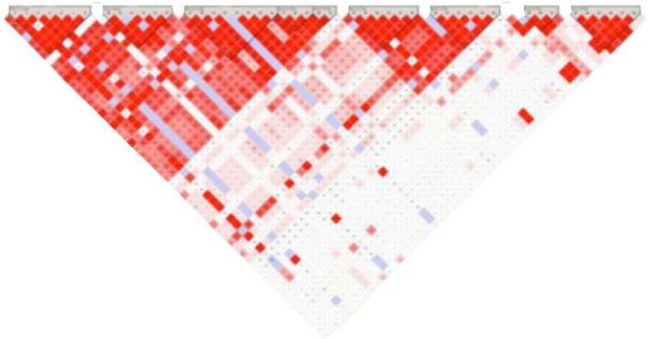
"I believe one of the great truths to emerge from this triumphant expedition inside the human genome is that in genetic terms, all human beings, regardless of race, are more than 99.9 percent the same."
President Bill Clinton, June 26, 2000, The White House East Room

Published Genome-Wide Associations through 12/2010, 1212 published GWA at $p \leq 5 \times 10^{-8}$ for 210 traits

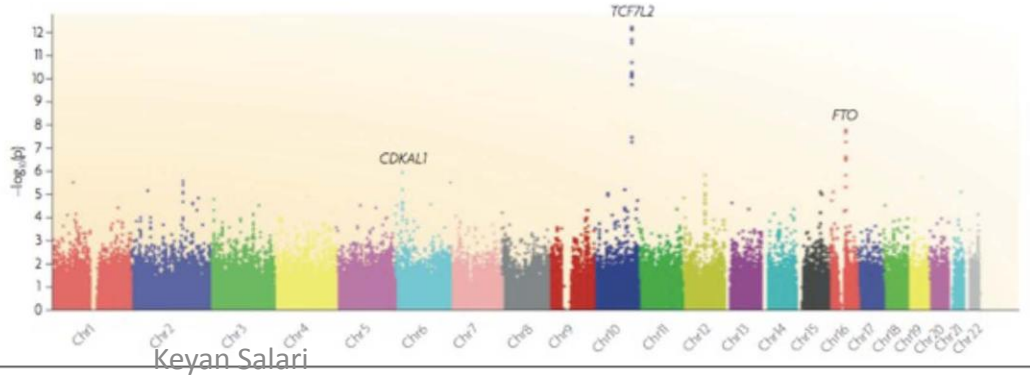


1. Natural variation in the human genome

2. Genetic Association & Linkage Disequilibrium



3. Genome-wide association studies



Steve Quake

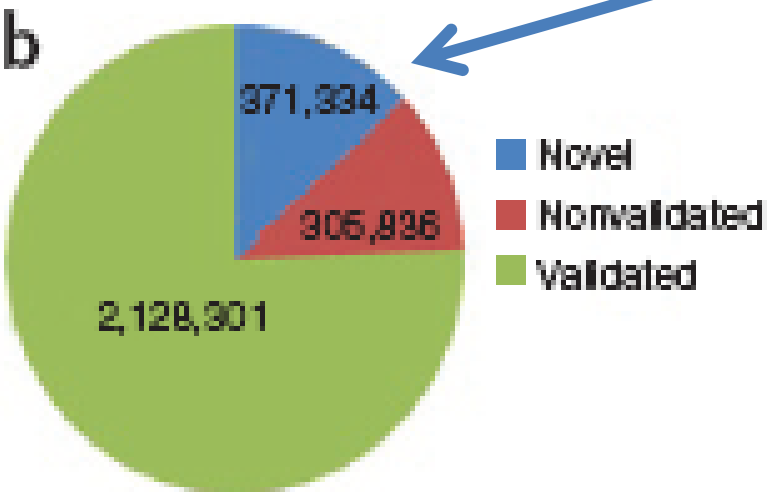
nature biotechnology

LETTERS

Single-molecule sequencing of an individual human genome

Dmitry Pushkarev[1,2], Norma F Neff[1,2] & Stephen R Quake[1]

Nature Biotech 27, 847, 2009



- 2.8 Million SNPs
  - 371 Thousand SNPs (13%) are novel

Genetic variation for a simple trait

Chr12: ALDH2 - SNP rs671

```
...GGGCTGCAGGCATACACTGAAGTGAAAACTGTGAGTGTG
...GGGCTGCAGGCATACACTGAAGTGAAAACTGTGAGTGTG
... G  L  Q  A  Y  T  E  V  K  T  V  S  V
```

Genotype: G/G

Protein: functional

Phenotype: none

Chr12: ALDH2 - SNP rs671

```
... GGGCTGCAGGCATACACTGAAGTGAAAACTGTGAGTGTG
... GGGCTGCAGGCATACACTAAAGTGAAAACTGTGAGTGTG
... G  L  Q  A  Y  T  E/K V  K  T  V  S  V
```

Genotype: A/G

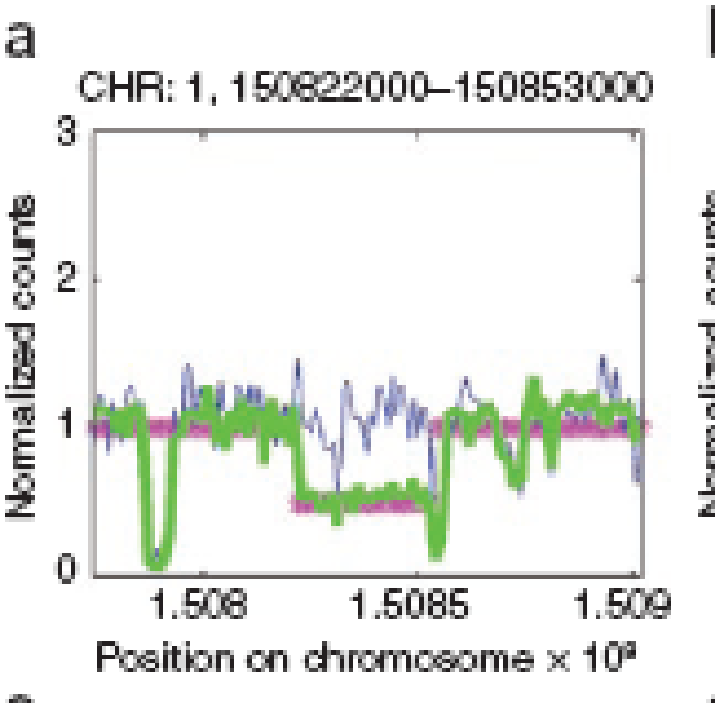
Protein: 1/2 functional

Phenotype: alcohol flush reaction

G allele functional
A allele missense (null)

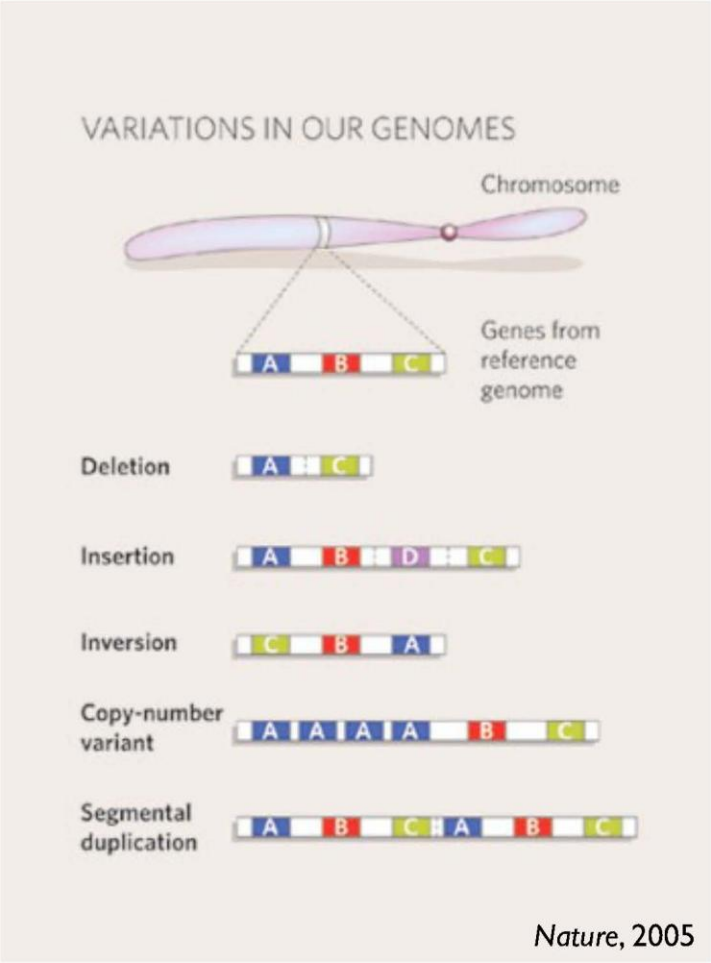
CEU 100% G
YRI 100% G
CHB/JPT 76-84% G

Copy number variation



- 752 copy number variations
- 16 Mb total

Human Genetic Variation



Nature, 2005

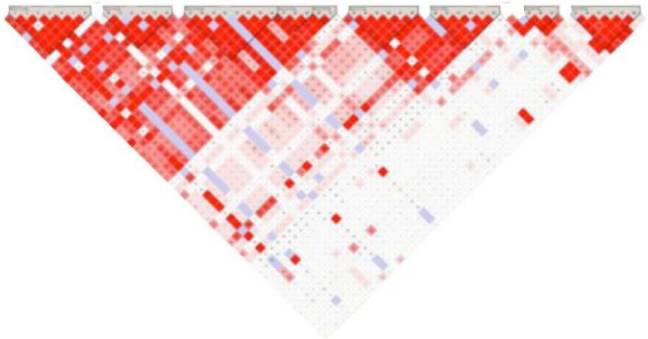
Structural variation

- 12% of our genome
- thousands of genes, disease loci, functional elements
- likely role in phenotypic variation and human disease

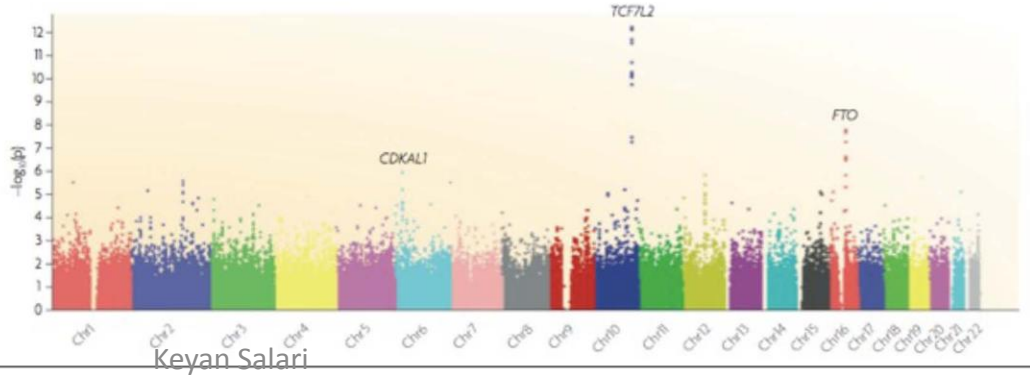
Redon *et al. Nature*, 2006

1. Natural variation in the human genome

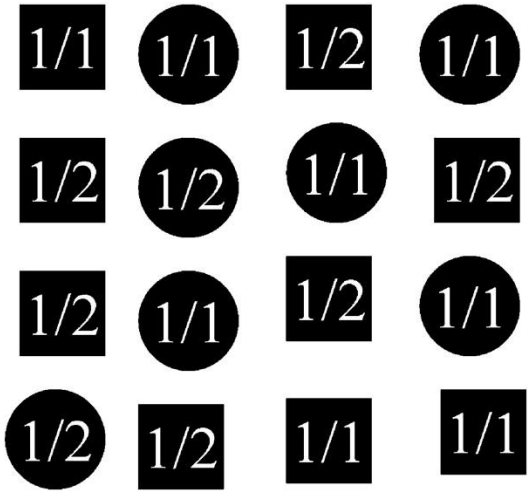
2. Genetic Association & Linkage Disequilibrium

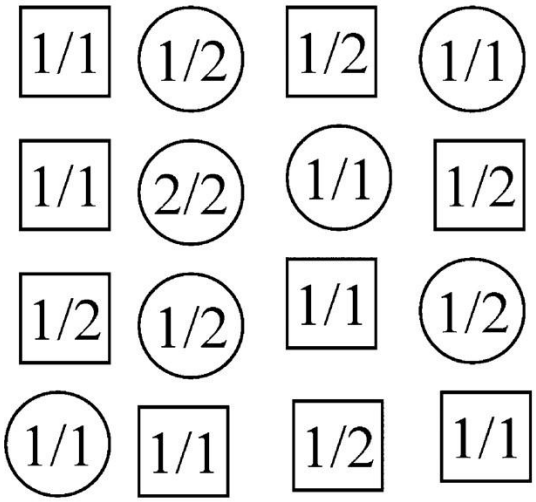


3. Genome-wide association studies



Genetic Association





Cases with Alcoholism

ALDH2: 1/1: 283 (83%)
1/2: 57 (17%)
2/2: 0

ALDH2*2 frequency: 0.08

Control sample

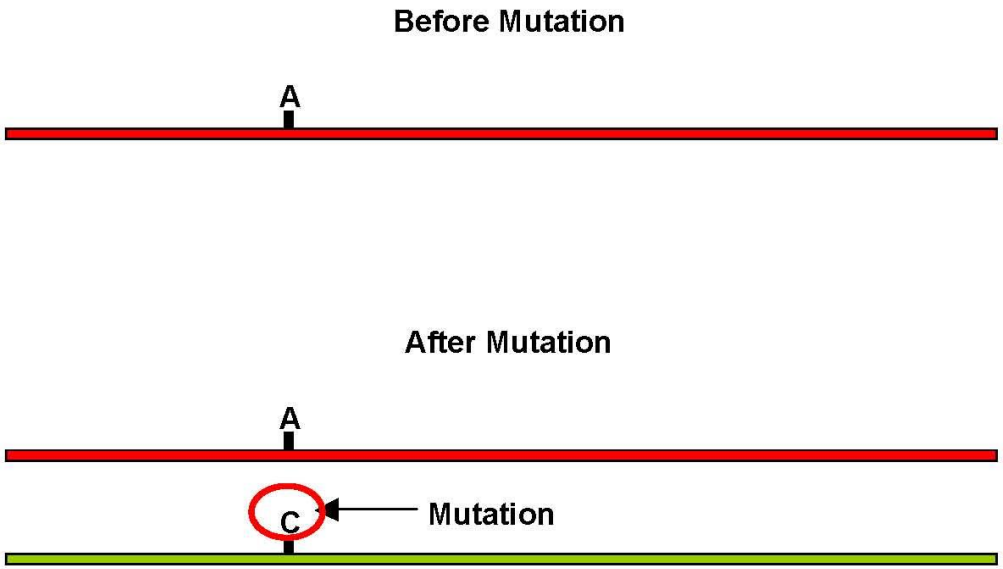
ALDH2: 1/1: 304 (56%)
1/2: 218 (40%)
2/2: 23 (4%)

ALDH2*2 frequency: 0.24

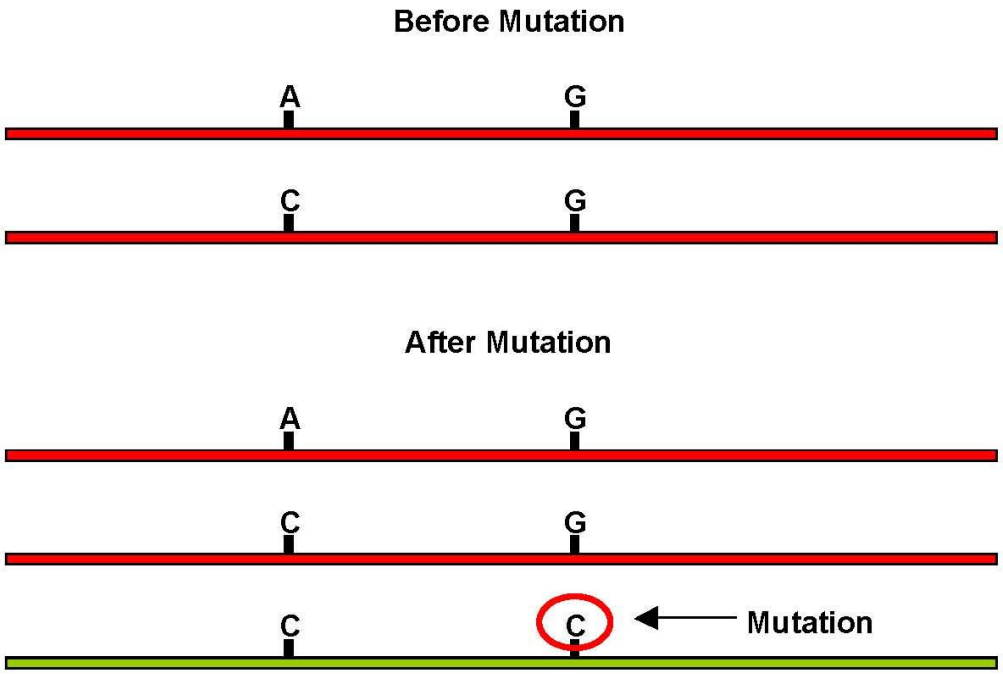
Today …

- We'll consider properties of pairs of alleles
- Haplotype frequencies
- Linkage equilibrium
- Linkage disequilibrium

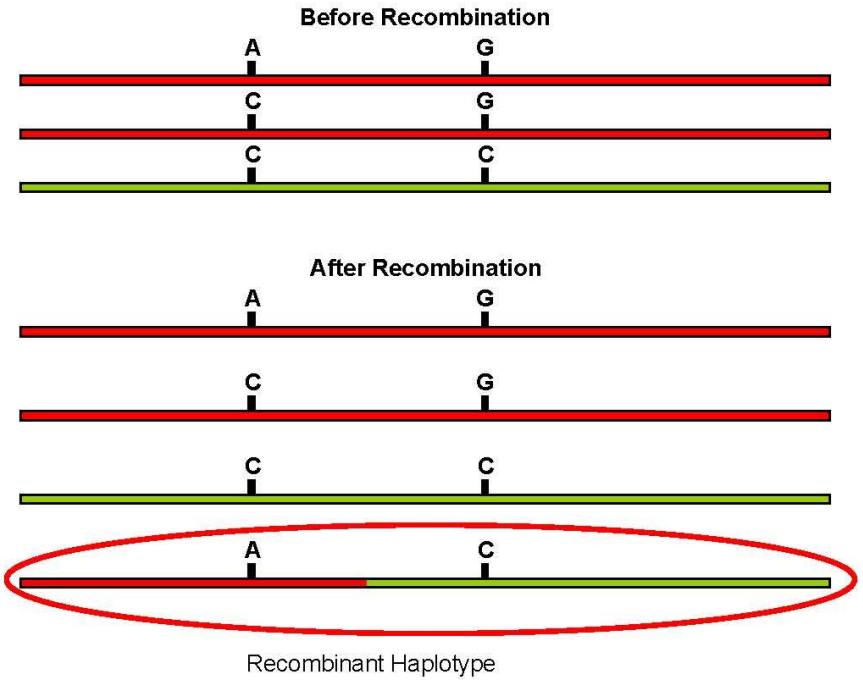
Alleles that exist today arose through ancient mutation events…



One allele arose first, and then the other…

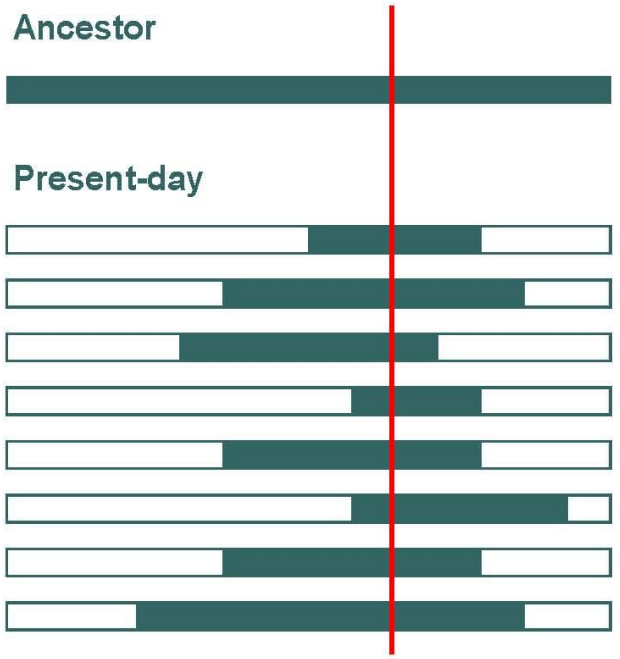


Recombination generates new arrangements for ancestral alleles



Linkage Disequilibrium

- Chromosomes are mosaics
- Extent and conservation of mosaic pieces depends on
  - Recombination rate
  - Mutation rate
  - Population size
  - Natural selection



- **Combinations of alleles at very close markers is called a haplotype**

Why is linkage disequilibrium important for gene mapping?

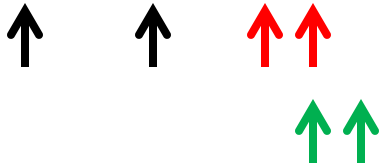
Association Studies and Linkage Disequilibrium

- If all polymorphisms were independent at the population level, association studies would have to examine every one of them…

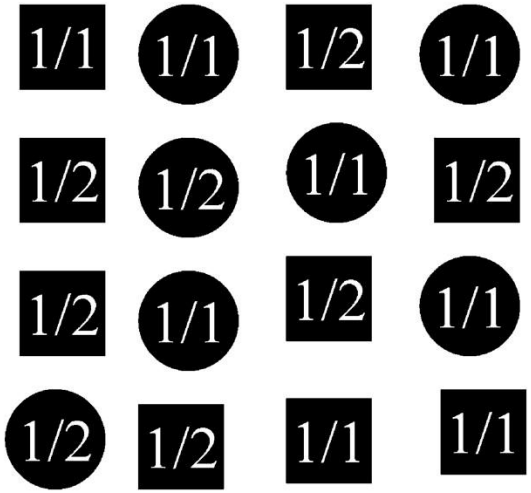
- Linkage disequilibrium makes tightly linked variants strongly correlated producing cost savings for association studies

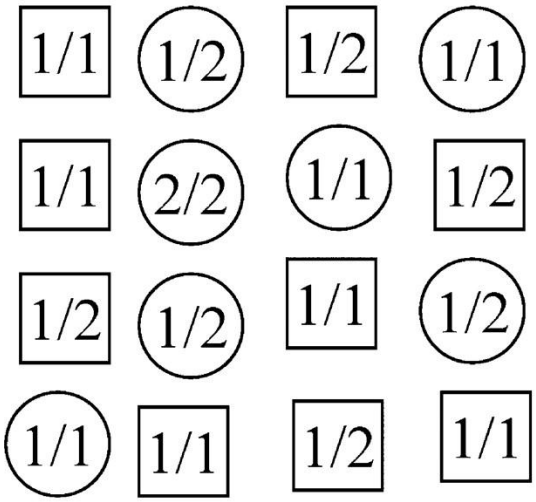
Tag SNPS – Genotype at one polymorphism informs about the genotype of a nearby, linked polymorphism

G.C.C.C.C.T.A.G	reference
A.T.T.A.T.C.G.A	alternate SNP
G.C.C.C.C.T.A.G	114 instances
A.C.C.C.C.T.A.G	2 instances
G.**T**.C.**A**.C.T.A.G	7 instances
G.**T**.C.**A**.C.**C.G.A**	9 instances
G.C.**T**.C.C.T.A.G	2 instances
G.C.**T**.C.C.**C.G**.G	2 instances



Genetic Association





Cases with Alcoholism

ALDH2: 1/1: 283 (83%)
1/2: 57 (17%)
2/2: 0

ALDH2*2 frequency: 0.08

Control sample

ALDH2: 1/1: 304 (56%)
1/2: 218 (40%)
2/2: 23 (4%)

ALDH2*2 frequency: 0.24

Linkage Disequilibrium Enables Genetic Association Studies

- In contrast to linkage studies, association studies can identify variants with relatively small individual contributions to disease risk
- However, they require detailed measurement of genetic variation and there are >10,000,000 catalogued genetic variants
- Until recently, studies limited to candidate genes or regions
  - A hit-and-miss approach…
- Because assay costs are decreasing and a modest number of variants can represent all others, genome-wide association studies are now possible.

The Allelic Architecture of Disease

What is it and how do we discover it?