Human Population Genetics & Gene-Environment Interactions

J. Claiborne Stephens, Ph.D.

Chief Scientific Officer

Motif BioSciences Inc

May 2007

www.motifbio.com

Pre-WGA Genetic Analysis

- Emphasis on pedigrees
  - Linkage (transmission in families)
- Knowledge of molecular biology
  - Candidate gene approach
- New emphasis on “susceptibility”
  - Common Disease/Common Variant hypothesis (CD/CV)

Premise of CD/CV

Majority of health-related gene variants are common

↓

If common, then old

↓

If old, then shared

Flow Chart of Genetic Association Analysis

Study Design

↓

Population

↓

Phenotyping

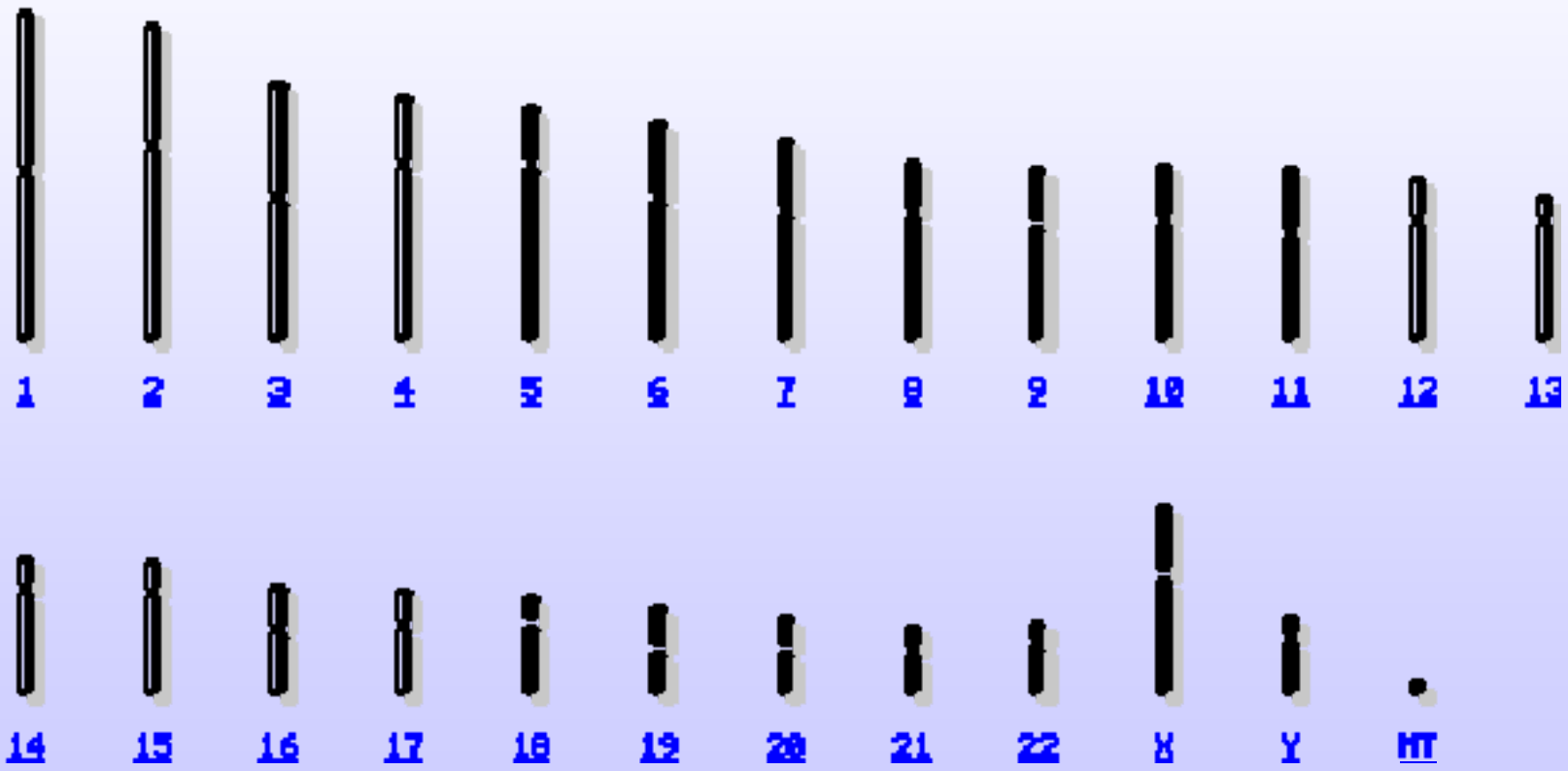
↓

Genotyping

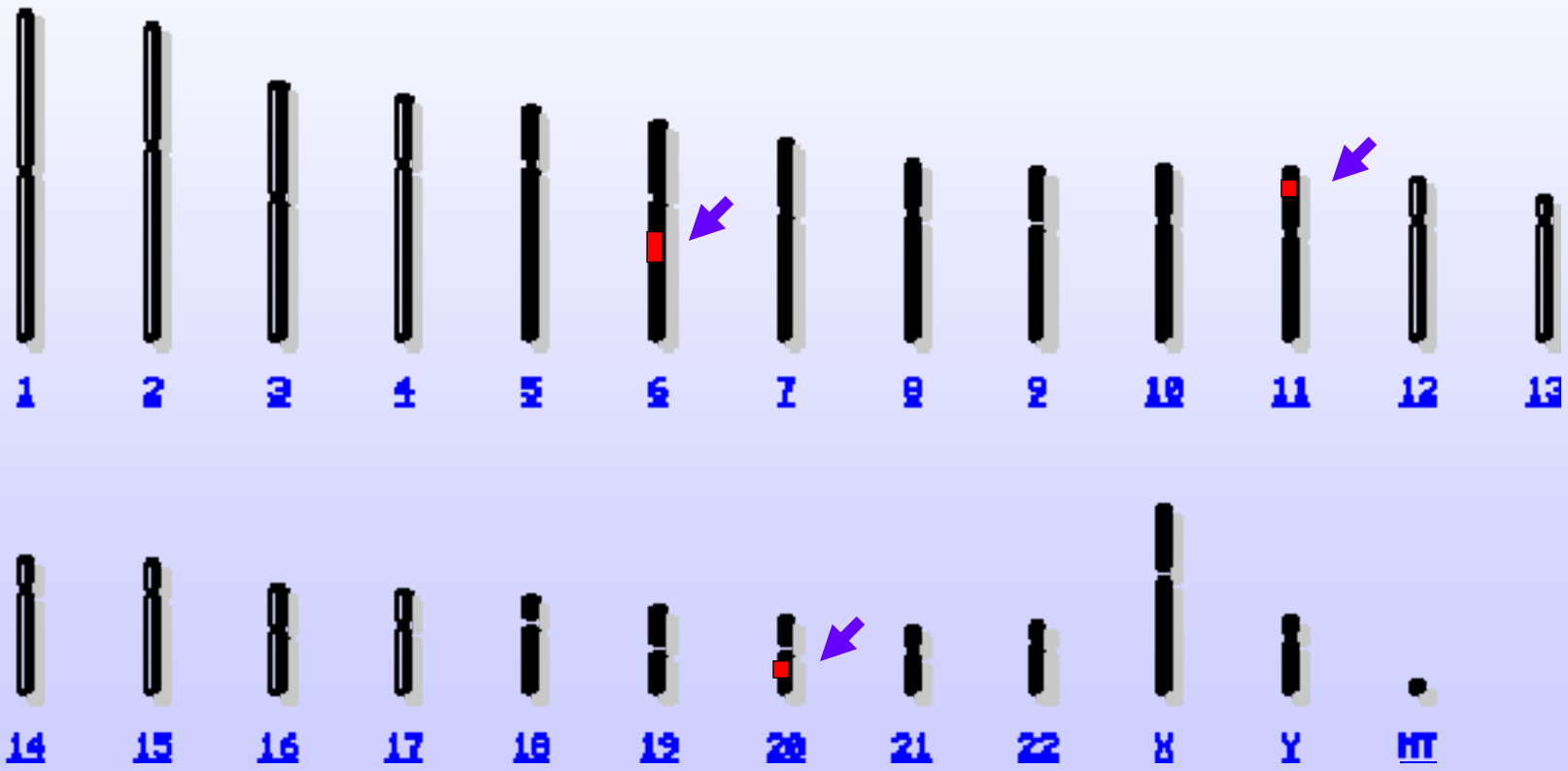
↓

Statistical Analysis

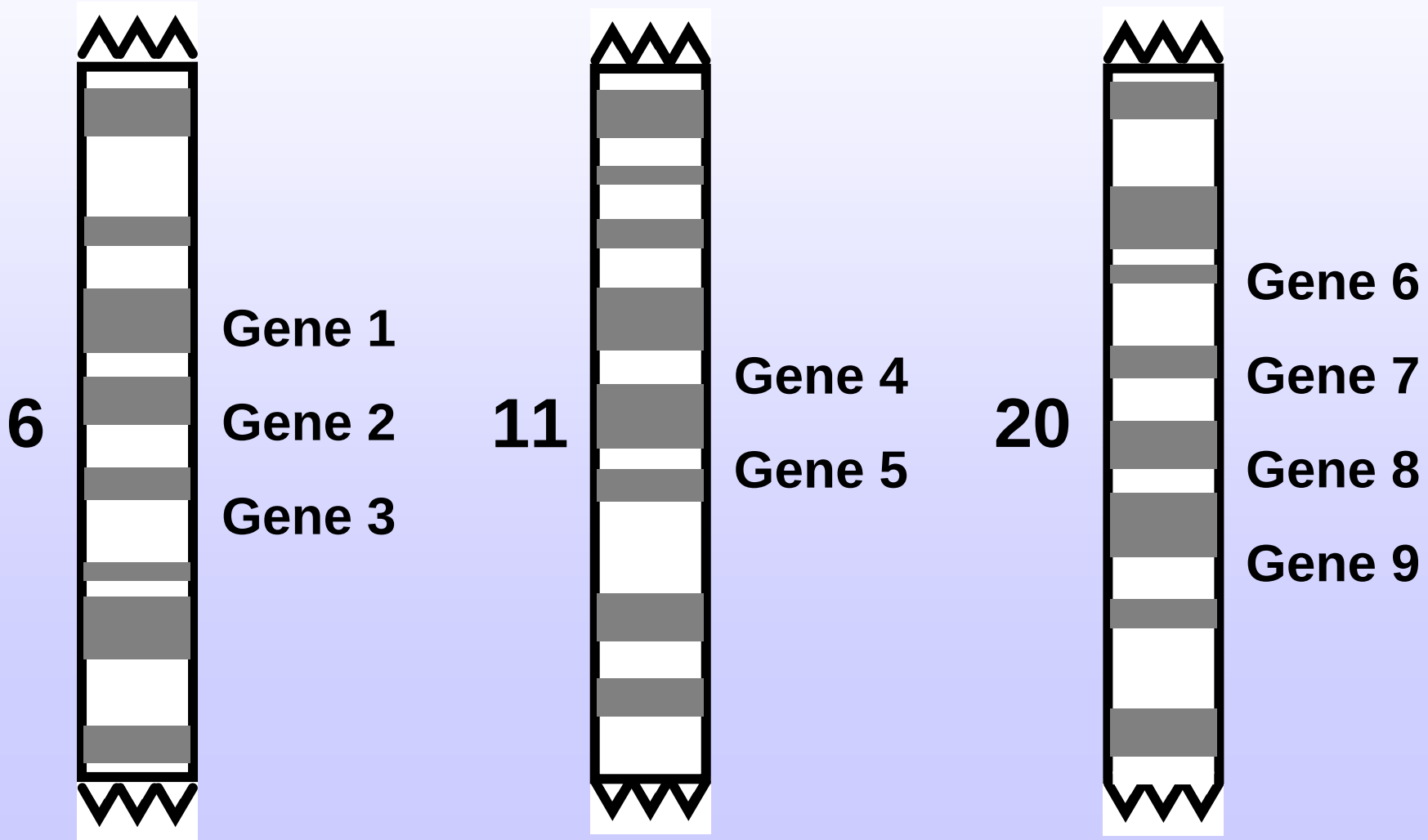
Finding “Disease” Genes



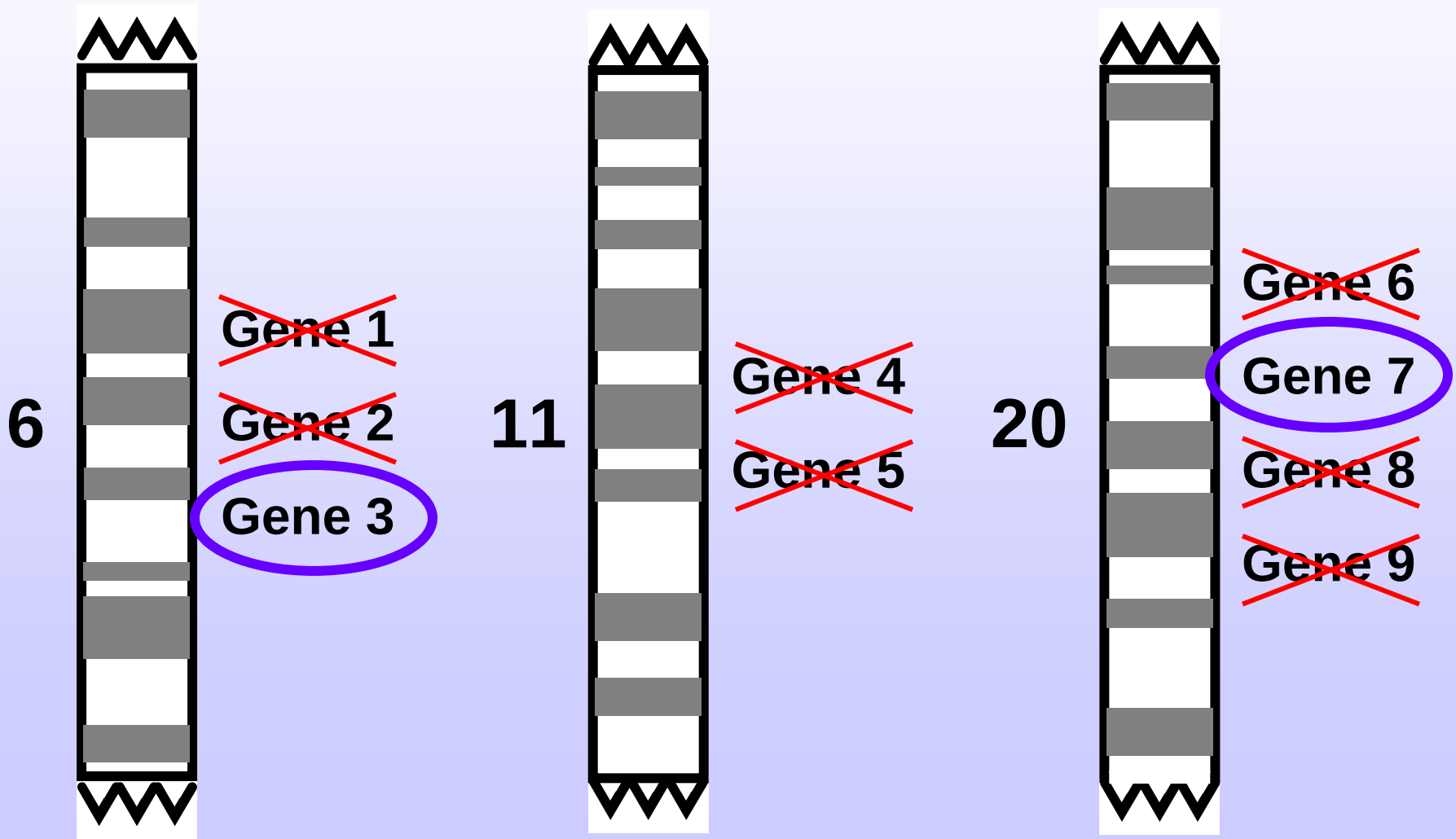
Finding “Disease” Genes



Chromosome region → Specific Genes

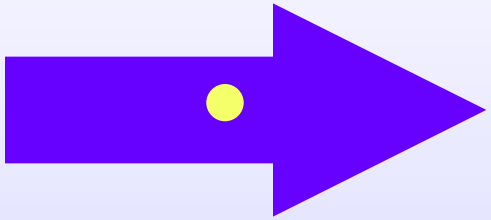


Chromosome region → Specific Genes

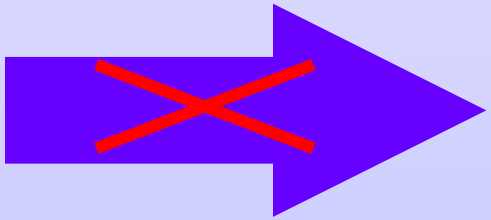


Gene → Specific Mutation(s)

Gene 3

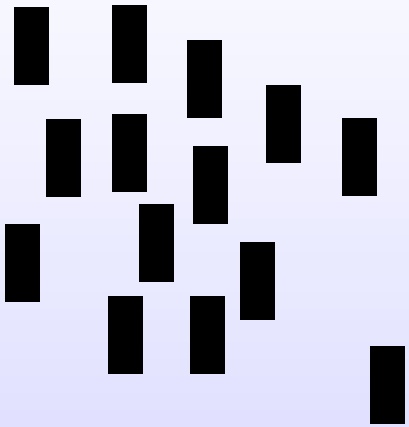


Gene 7



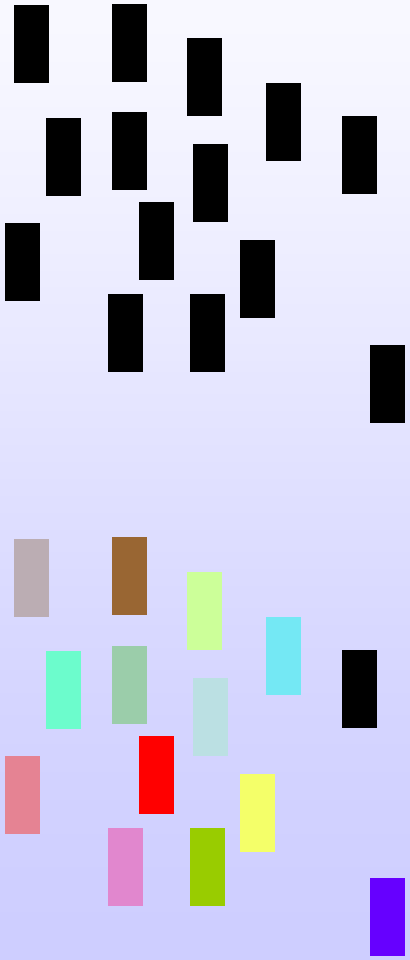
Now what?

It all starts with Genetic Variation



Measured across 3 billion bases of the human genome, all people are about 99.9% identical

It all starts with Genetic Variation



Measured across 3 billion bases of the human genome, all people are about 99.9% identical

But

The remaining 0.1% is still ~ 3 million bases!

DNA Variation

```
ACGGCTACCTTAACGTAACTGCACACTTCGTGCGTCCGGACGA
ACGGCTACCTTGACGTAACTGGACACATCGTGCGTCCGGATGA
ACGGCTATCTTGACATAACTGGACACTTCGTGCGTCCGGATGA
ACGGCTACCTTGACGTAACTGGACACTTCGTGCATCCGGACGA
ACGGCTACCTTGACGTAACTGGACACTTCGTGCATCCGGACGA
ACGGCTATCTTGACGTAACTGGACACTTCGTGCGTCCGGATGA
ACGGCTATCTTGACATAACTGGACACTTCGTGCGTCCGGATGA
ACGGCTACCTTGACGTAACTGGACACTTCGTGCGTCCGGATGA
ACGGCTATCTTGACGTAACTGGACACTTCGTGCGTCCGGATGA
ACGGCTACCTTAACGTAACTGCACACTTCGTGCGTCCGGACGA
ACGGCTATCTTGACATAACTGGACACTTCGTGCGTCCGGATGA
```

Generally, only a few positions per thousand have common variation

DNA Variation

```
.......C...A..G......C....T......G......C..
.......C...G..G......G....A......G......T..
.......T...G..A......G....T......G......T..
.......C...G..G......G....T......A......C..
.......C...G..G......G....T......A......C..
.......T...G..G......G....T......G......T..
.......T...G..A......G....T......G......T..
.......C...G..G......G....T......G......T..
.......T...G..G......G....T......G......T..
.......C...A..G......C....T......G......C..
.......T...G..A......G....T......G......T..
```

Focus on those positions that have common variation

DNA Variation

```
CAGCTGC
CGGGAGT
TGAGTGT
CGGGTAC
CGGGTAC
TGGGTGT
TGAGTGT
CGGGTGT
TGGGTGT
CAGCTGC
TGAGTGT
```

Ignore constant positions and identify “haplotypes”

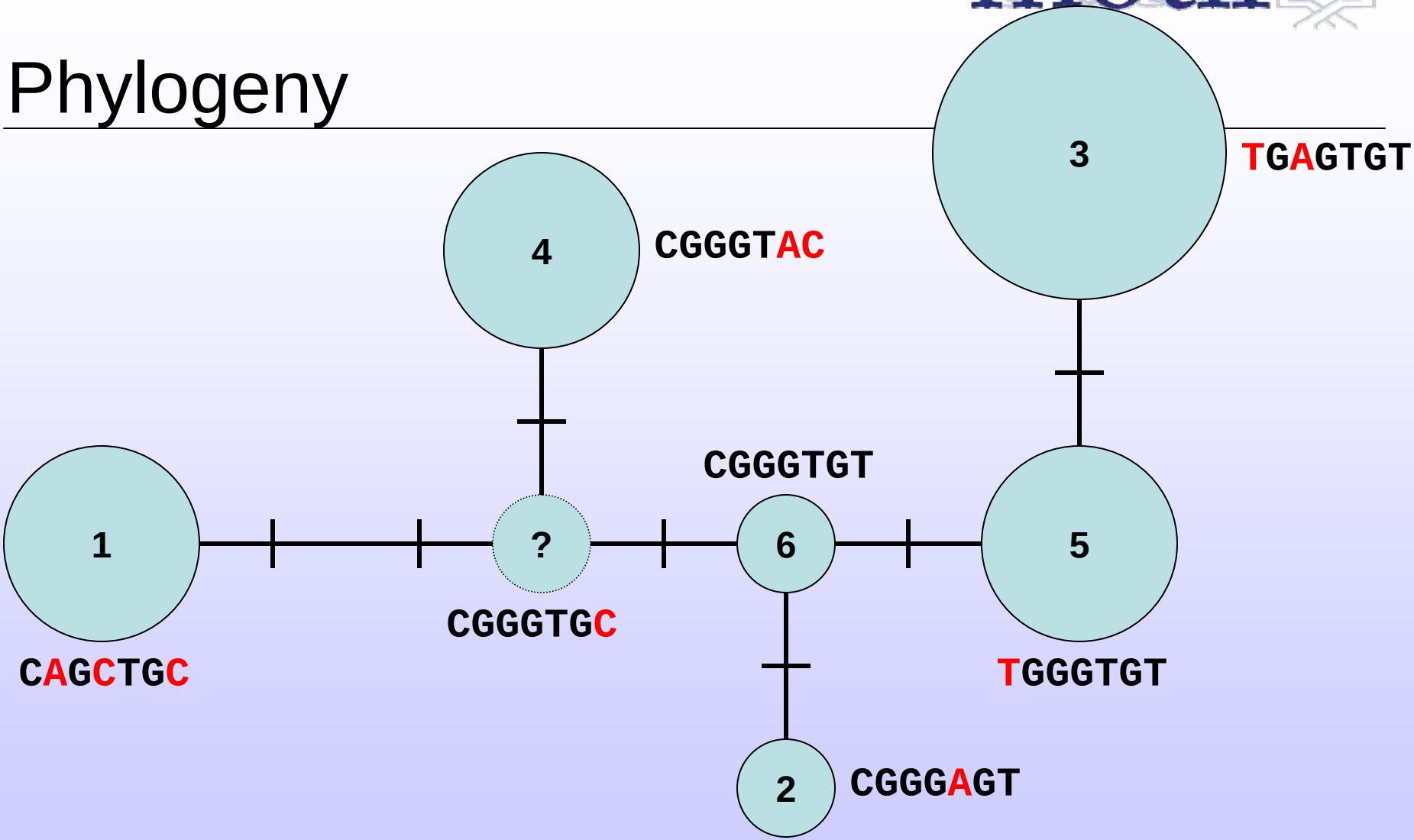
DNA Variation

```
1 CAGCTGC 2
2 CGGGAGT 1
3 TGAGTGT 3
4 CGGGTAC 2
5 TGGGTGT 2
6 CGGGTGT 1
```

Six distinct haplotypes observed, one to three times each.
Based on seven variable positions, or "SNPs".

International HapMap Project

Phylogeny



The relationships among the haplotypes, in terms of mutational pattern, can often be inferred.

	D+	D-
M+	500	500
M-	500	500

OR = 1.0

P = 1.0

	D+	D-
M+	400	350
M-	600	650

OR = 1.24

P = 0.02

	D+	D-
M+	500	350
M-	500	650

OR = 1.86

$P < 1.2 \times 10^{-11}$

Recent Results from T2DM WGA

	deCODE		French	
	OR	p-value*	OR	p-value*
TCF7L2	1.38	**9.74**	2.77	**33.82**
CDKAL1	1.20	**8.11**	**-**	**-**
SLC30A8	1.15	**5.48**	1.53	**7.21**
HHEX/IDE	1.19	**3.24**	1.44	**5.52**

French = Sladek et al. 2007
deCODE = Grant et al. 2006; Steinthorsdottir 2007 (Iceland only)

*** -log10 of p-value**

T2DM WGA (cont'd)

	DGI		FUSION		UK		All	
Gene or Genomic Region	OR	p-value*	OR	p-value*	OR	p-value*	OR	p-value*
TCF7L2	1.38	**30.64**	1.34	**7.89**	1.37	**12.17**	1.37	**48.00**
IGF2BP2	1.17	**8.77**	1.18	3.68	1.11	3.80	1.14	**15.05**
CDKN2A/B	1.20	**7.27**	1.20	2.66	1.19	**6.31**	1.20	**14.11**
FTO	1.03	0.60	1.11	1.80	1.23	**13.14**	1.17	**11.89**
CDKAL1	1.08	2.62	1.12	2.02	1.16	**7.89**	1.12	**10.39**
KCNJ11	1.15	**7.00**	1.11	**7.89**	1.15	2.89	1.14	**10.17**
HHEX/IDE	1.14	3.77	1.10	1.59	1.13	**5.34**	1.13	**9.24**
SLC30A8	1.07	1.33	1.18	4.15	1.12	4.15	1.12	**7.28**
"Chr 11"	1.16	0.92	1.48	**7.24**	1.13	1.17	1.25	**6.37**
PPARG	1.09	2.72	1.20	2.85	1.23	2.89	1.14	**5.77**

Adapted from Table 1 in Scott et al. and the following studies:

DGI = Diabetes Genetics Initiative 2007

FUSION = Scott et al. 2007

UK = Zeggini et al. 2007

*** -log10 of p-value**

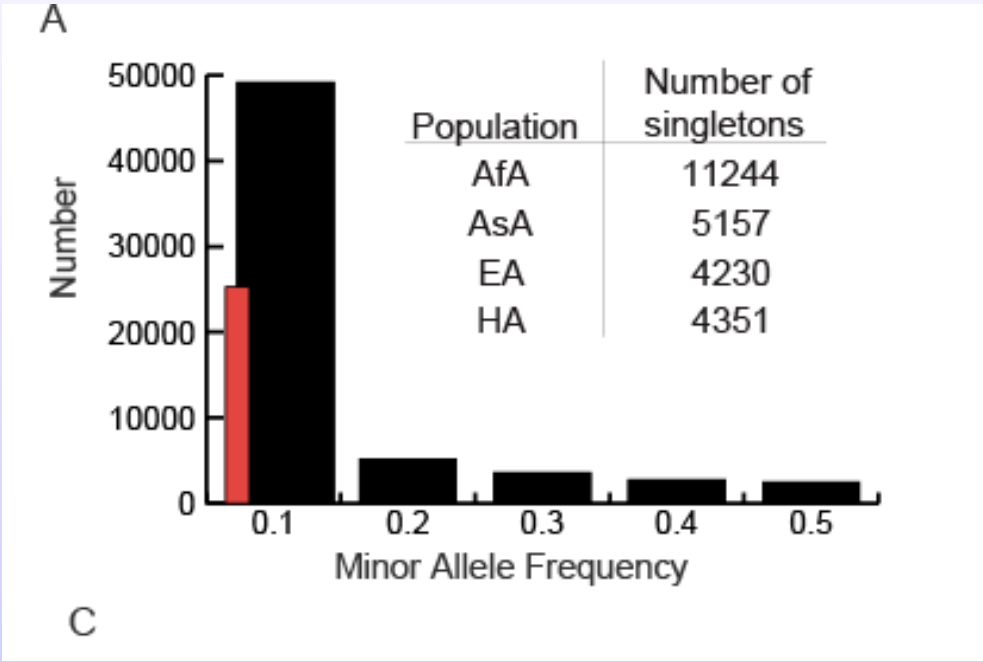
Human Gene Resequencing Projects

Project Name	# chromosomes	Eur	Afr	Asi	His	# genes resequenced
Seattle SNPs	94	46	48	-	-	100
Environ. Genome	180	-	-	-	-	213
Applera	78	40	38	-	-	11,624
ENCODE	128	32	32*	32*	-	10 x 500kb regions
Genaissance	152	40	40	38	34	3,873

*samples were ascertained from native populations, not U.S. populations

(Guthery et al., AJHG, in press)

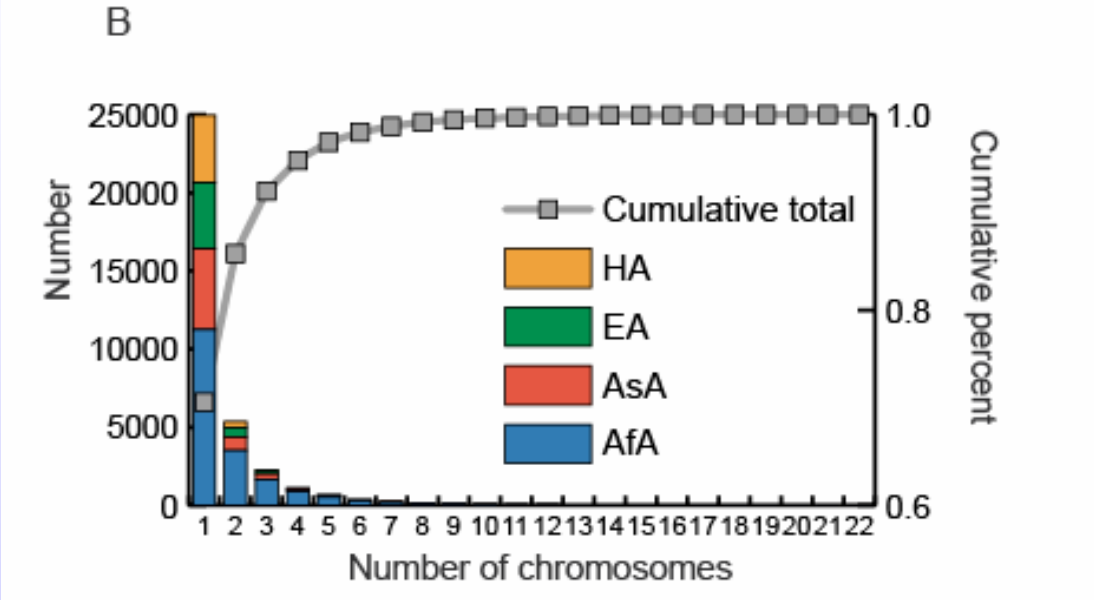
SNP Frequency Distribution in GRP



Population	Number of singletons
AfA	11244
AsA	5157
EA	4230
HA	4351

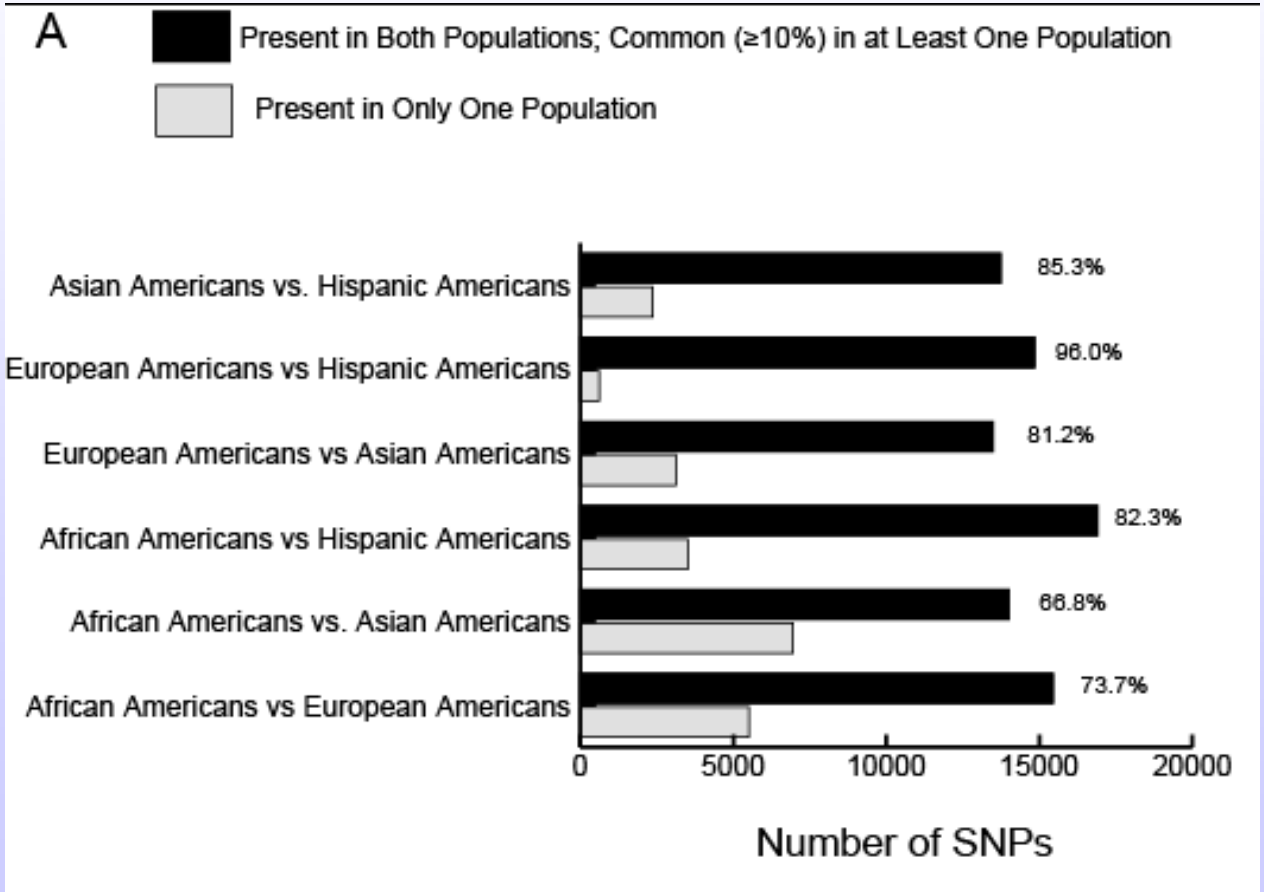
(Guthery et al., AJHG, in press)

“Private” SNP Distribution



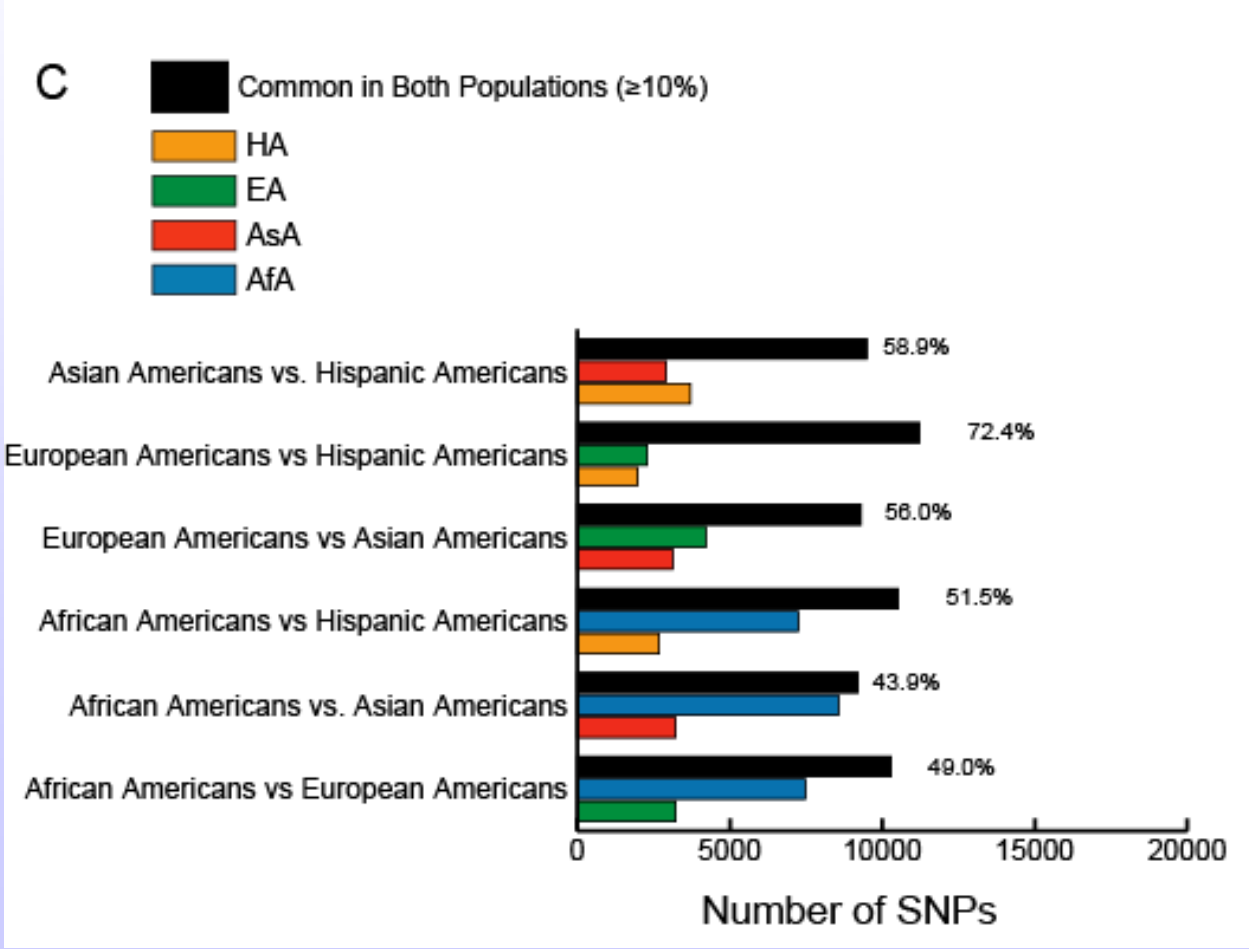
(Guthery et al., AJHG, in press)

Common SNP, Shared Presence



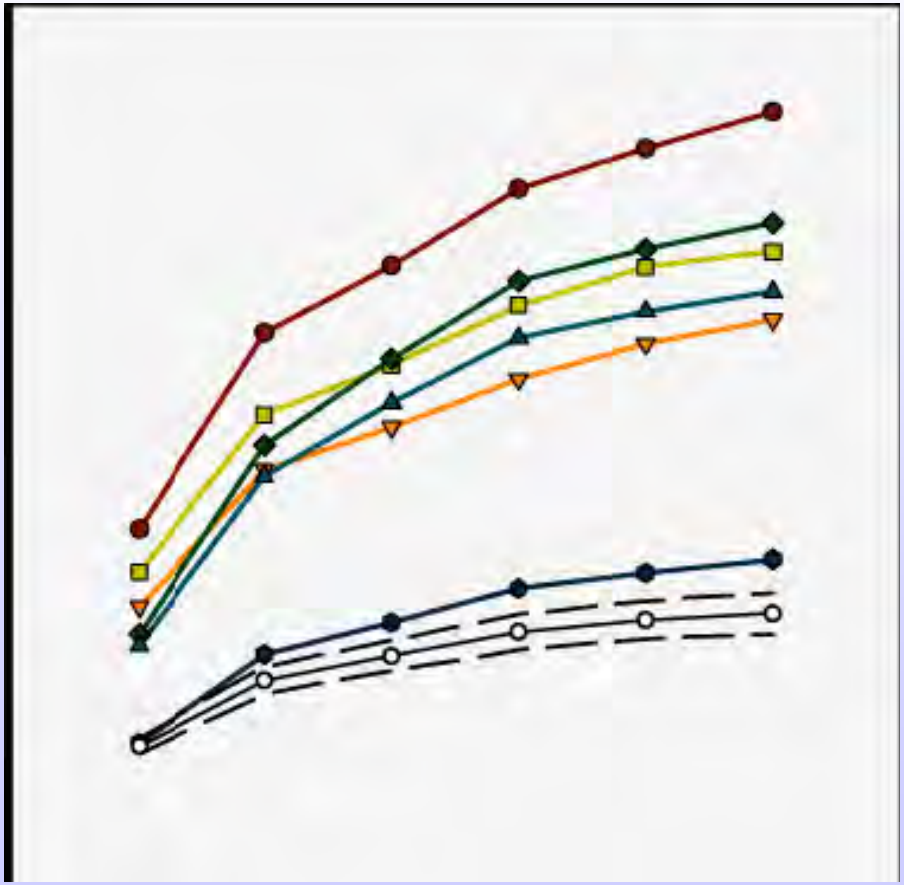
(Guthery et al., AJHG, in press)

Common SNP, Shared Commonness



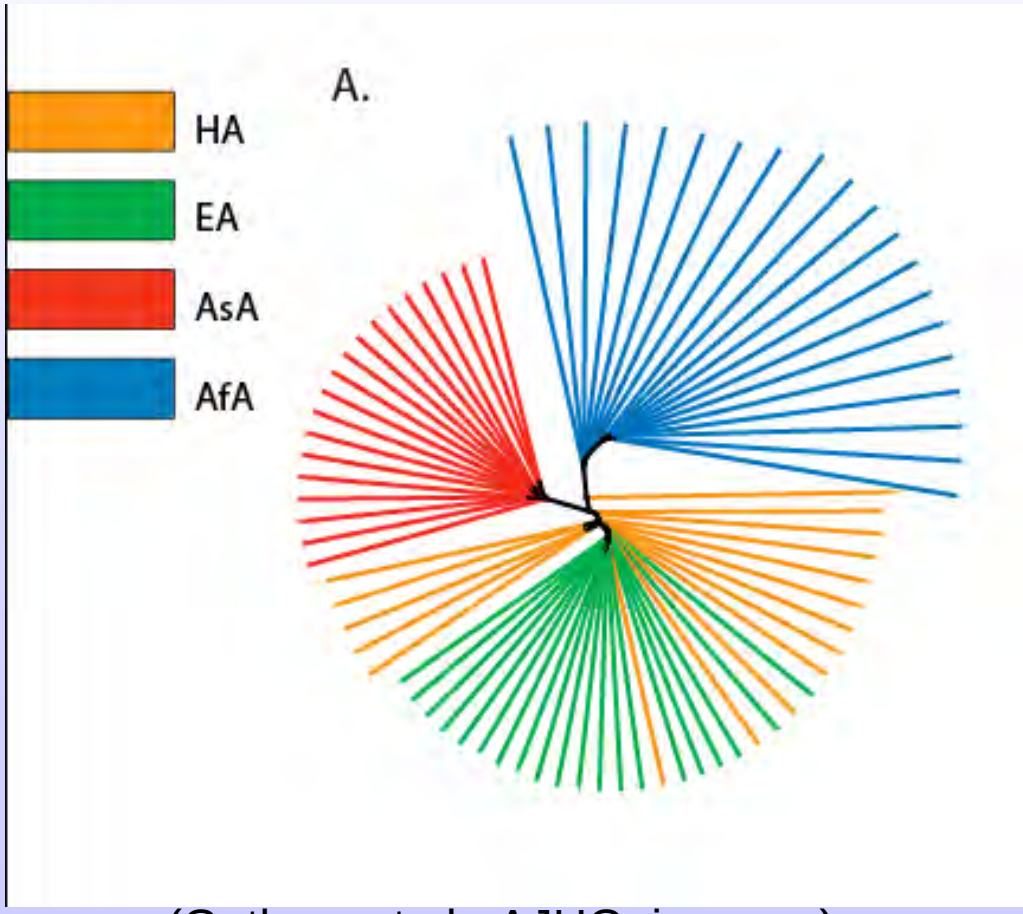
(Guthery et al., AJHG, in press)

PW Population Freq. Differences Increase



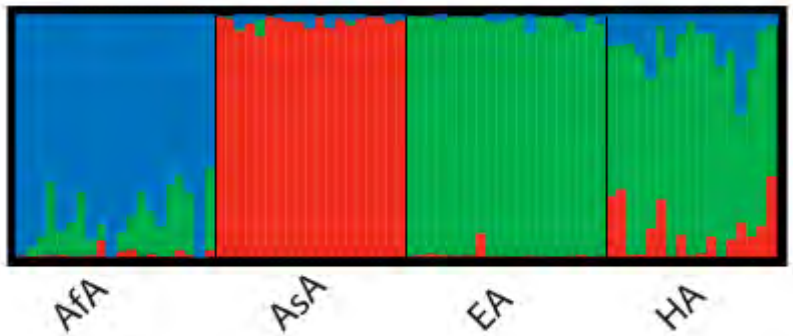
(Guthery et al., AJHG, in press)

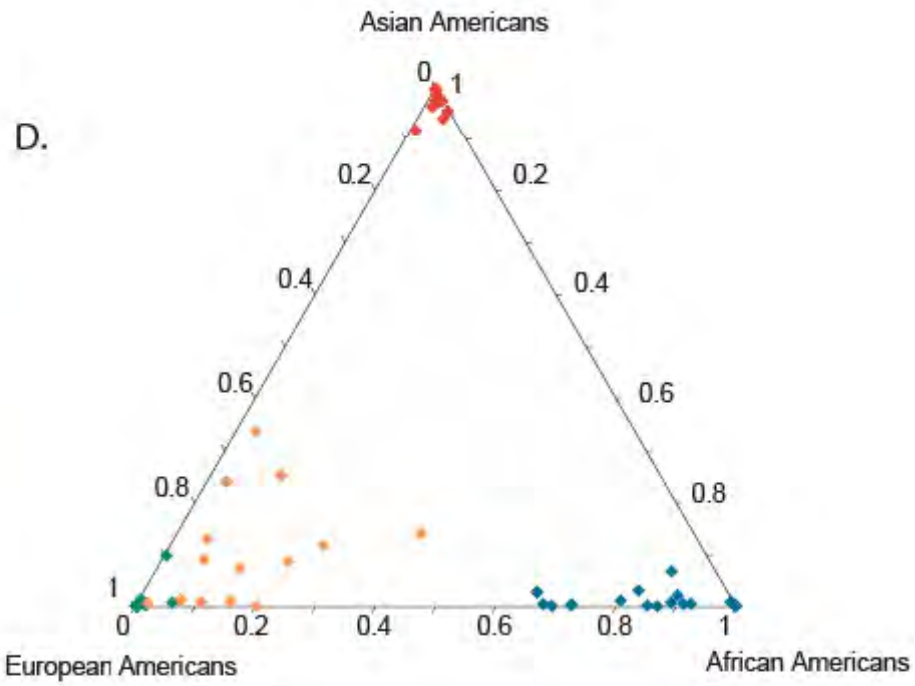
UPGMA Tree of Individuals



(Guthery et al., AJHG, in press)

STRUCTURE Results for GRP Sample





(Guthery et al., AJHG, in press)

The Power of Population Genetics

Advance Online Publication | doi:10.1038/nature05568 | Published online 11 February 2007

NEWS & VIEWS

HUMAN GENETICS

Variants in common diseases

Nelson B. Freimer and Chiara Sabatti

Most common diseases arise from interaction between multiple genetic variations and factors such as diet. Studies of such diseases that exploit the rich data on variation in the human genome are just beginning.