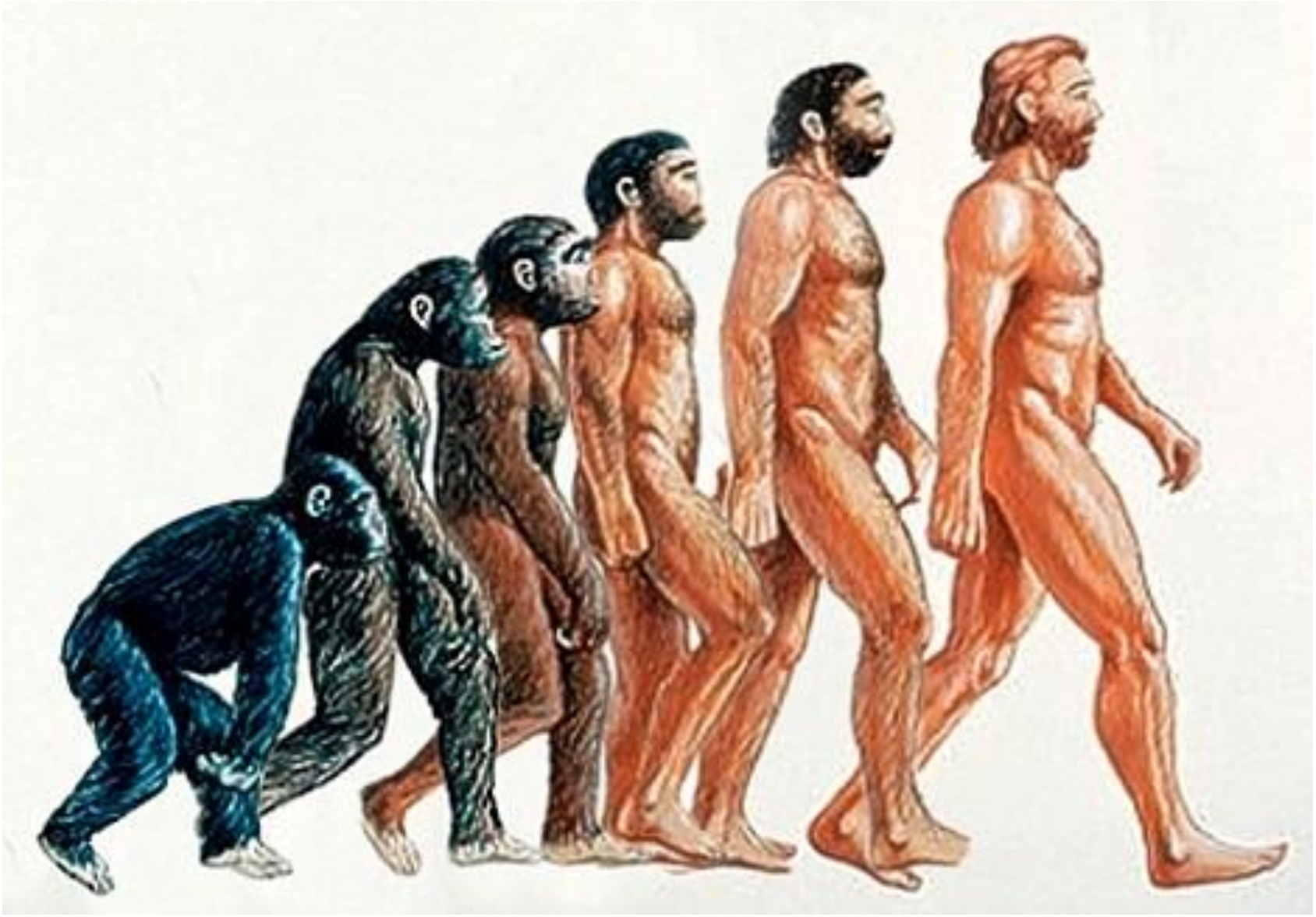
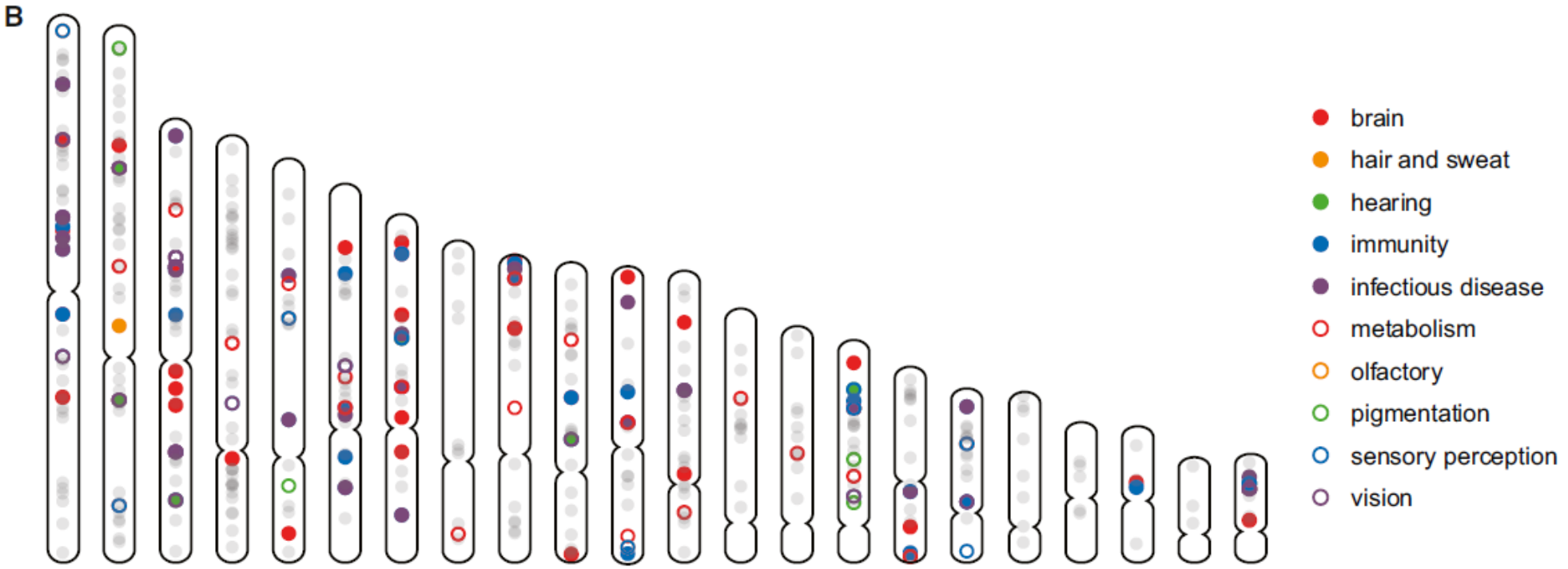
San Jose Mercury News
Lisa Krieger
lkriger@mercurynews.com
650-793-0720

Human Positive Selection



Human Positive Selection Loci



Positive selection regions

Mostly changes in expression. Only 35 affect protein coding

Examples:

- SCL24A5 lighter skin
- MATP lighter skin
- LCT milk drinking
- EDAR hair thickness
- ARHGEF3 bone mineral density
- BTLA rheumatoid arthritis
- ITPR3 Type 1 diabetes
- TLR5 interferon gamma secretion

Genotation

Run selection exercise.

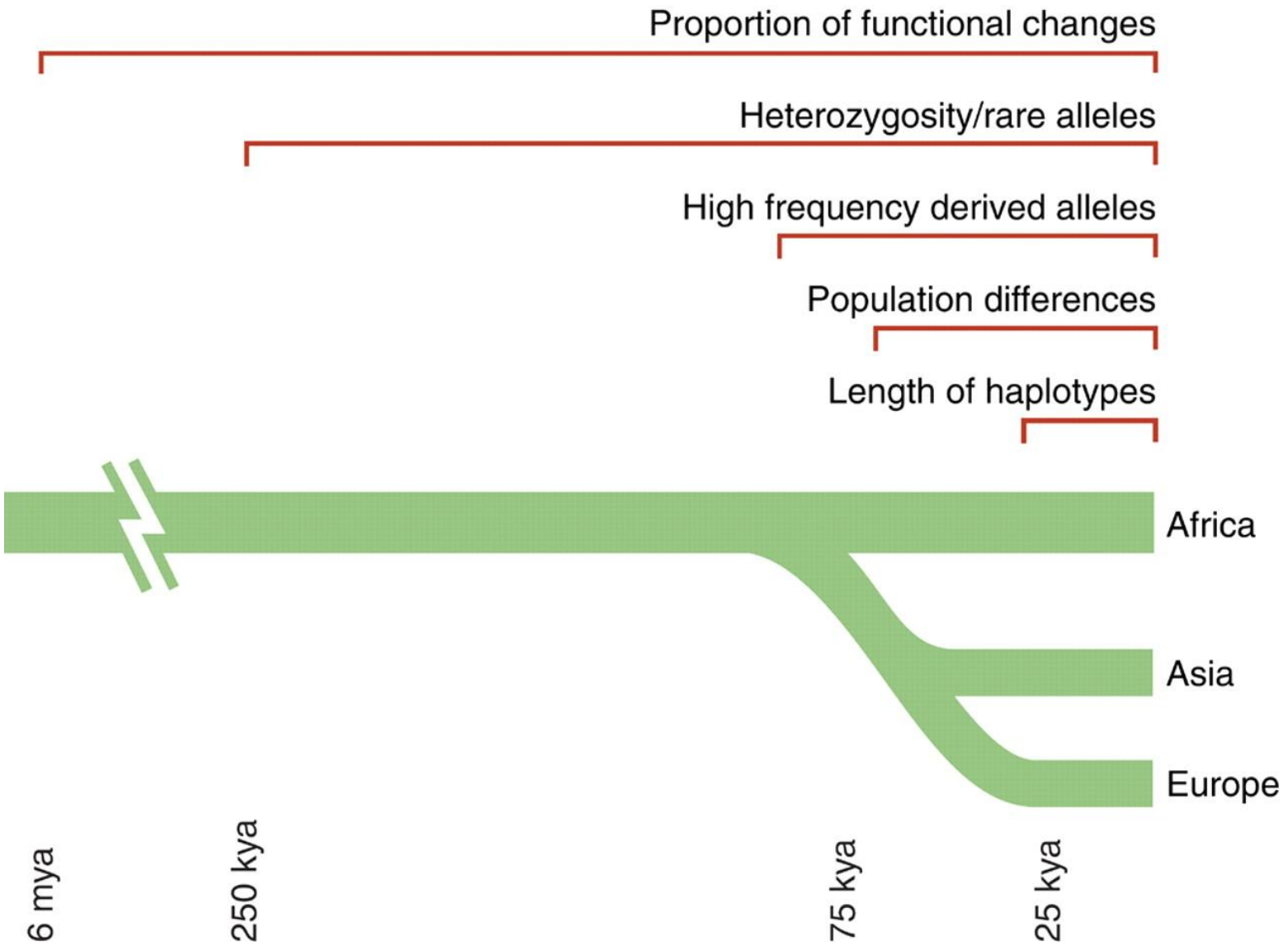
Ways to detect genes under positive selection

1. Proportion of functional changes

- Positive selection may favor many alleles, not just one

- This can be detected by a large number of coding changes relative to neutral changes in the gene.

Fig. 1. Time scales for the signatures of selection.



P C Sabeti et al. Science 2006;312:1614-1620

FOXP2: a transcription factor involved in human speech

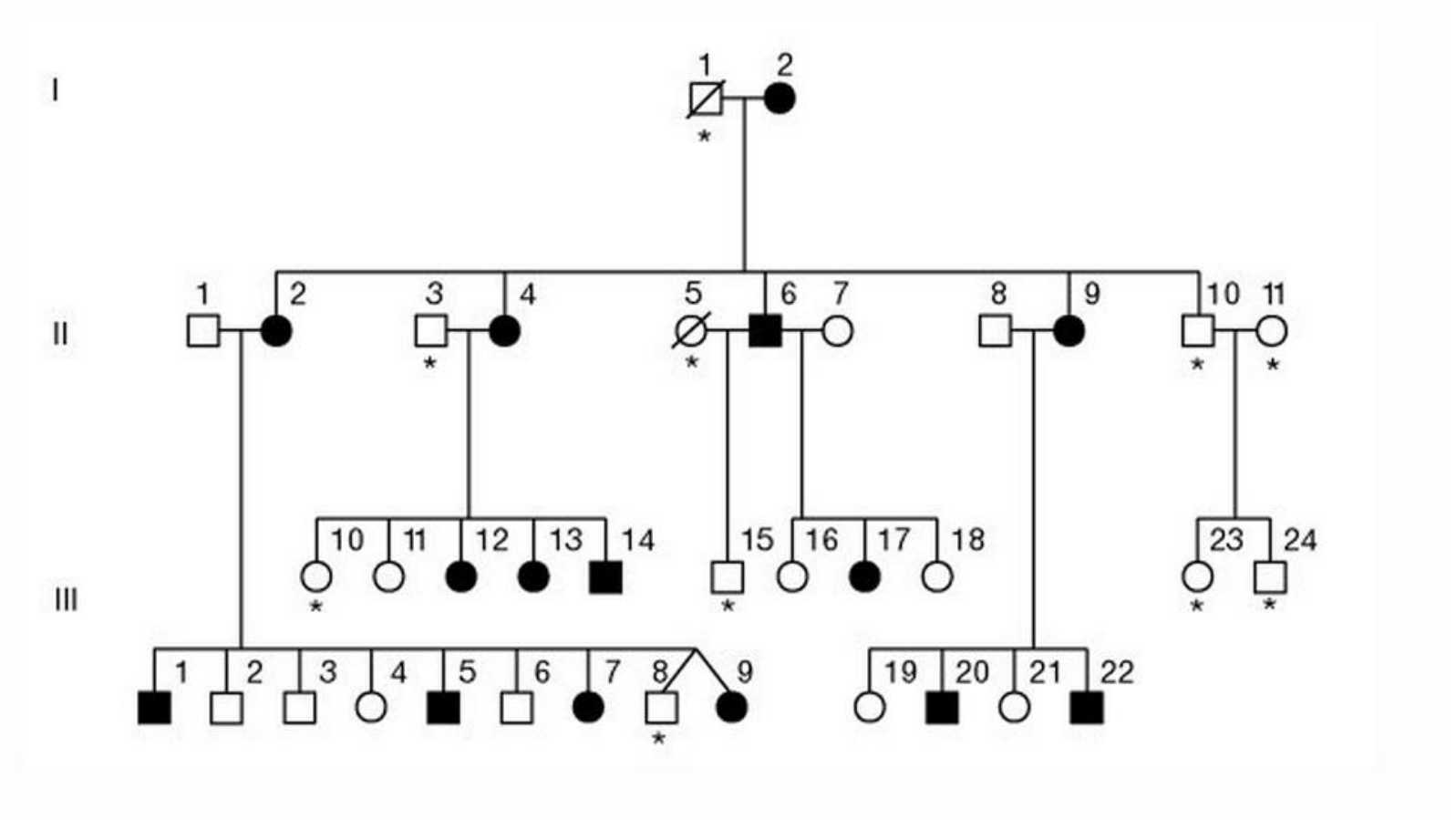
Adaptively evolving along human lineage.

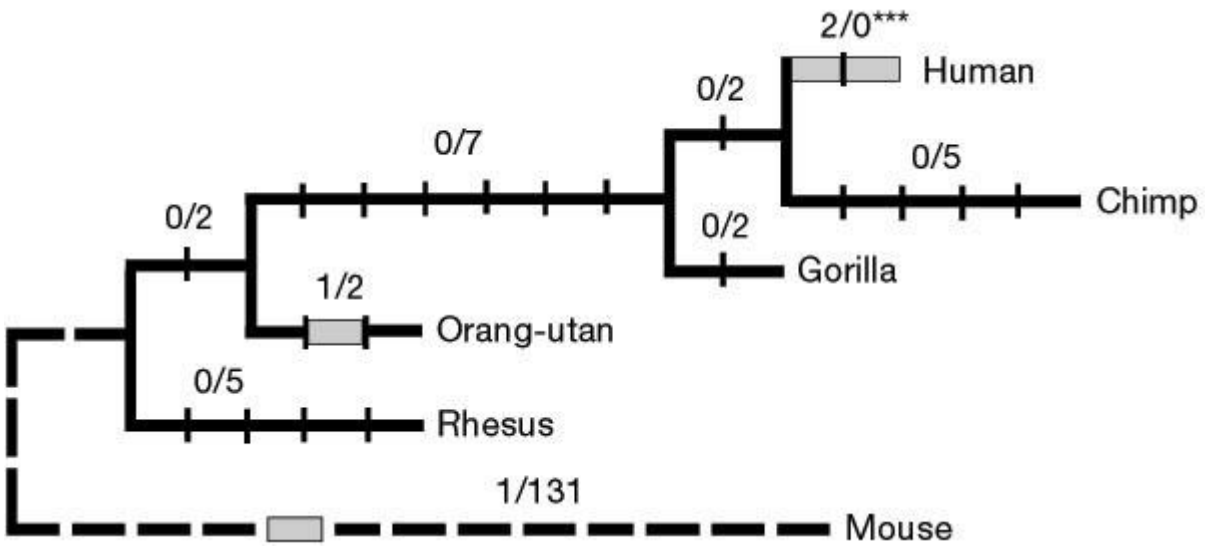
Highly conserved, but has recent human-specific changes

Mutations of FOXP2 cause a severe speech and language disorder

FOXP2 mutations affect speech and language

Affected members have deficits in language processing (such as the ability to break up words into their constituent phonemes) and grammatical skills (including production and comprehension of word inflections and syntactical structure).





Foxp2 gene
Has high functional differences.
Bars represent nucleotide changes. Grey bars indicate amino-acid changes

Cell

Volume 137, Issue 5, 29 May 2009, Pages 961–971

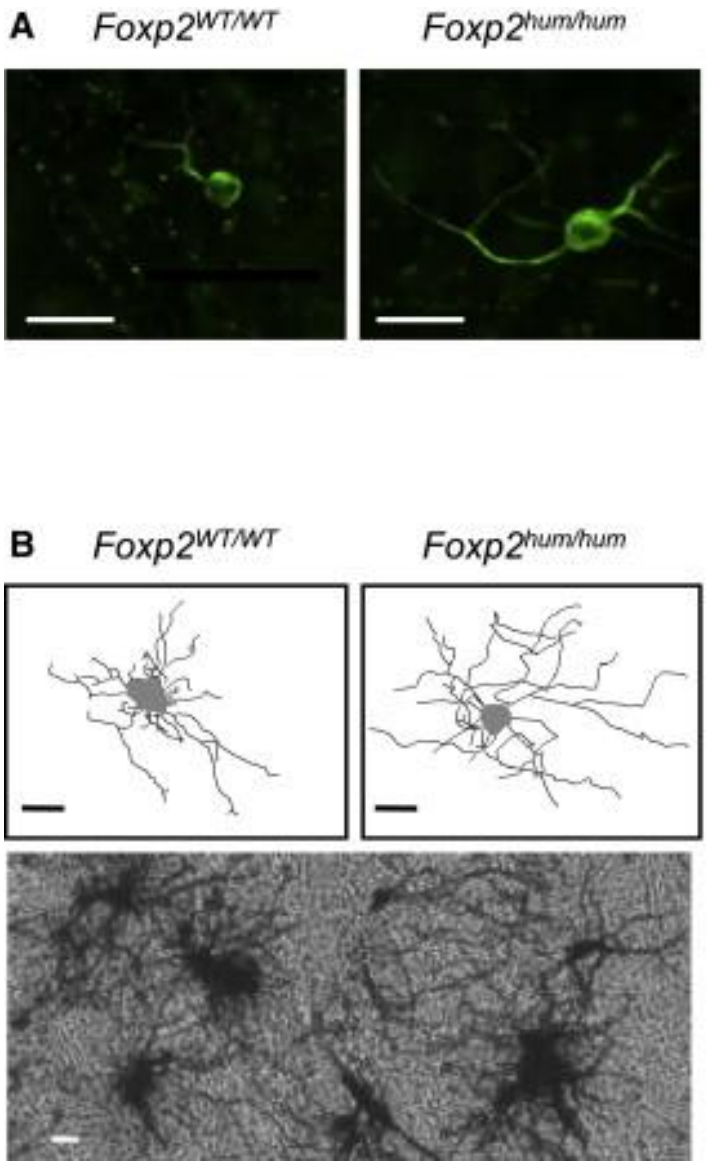
Article

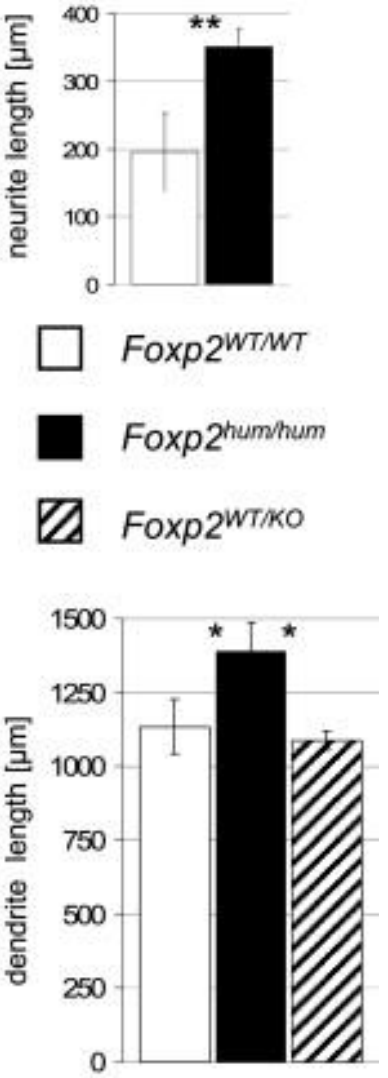
A Humanized Version of Foxp2 Affects Cortico-Basal Ganglia Circuits in Mice

Wolfgang Enard[1], Sabine Gehre[1], Kurt Hammerschmidt[2], Sabine M. Hölter[3], Torsten Blass[1],

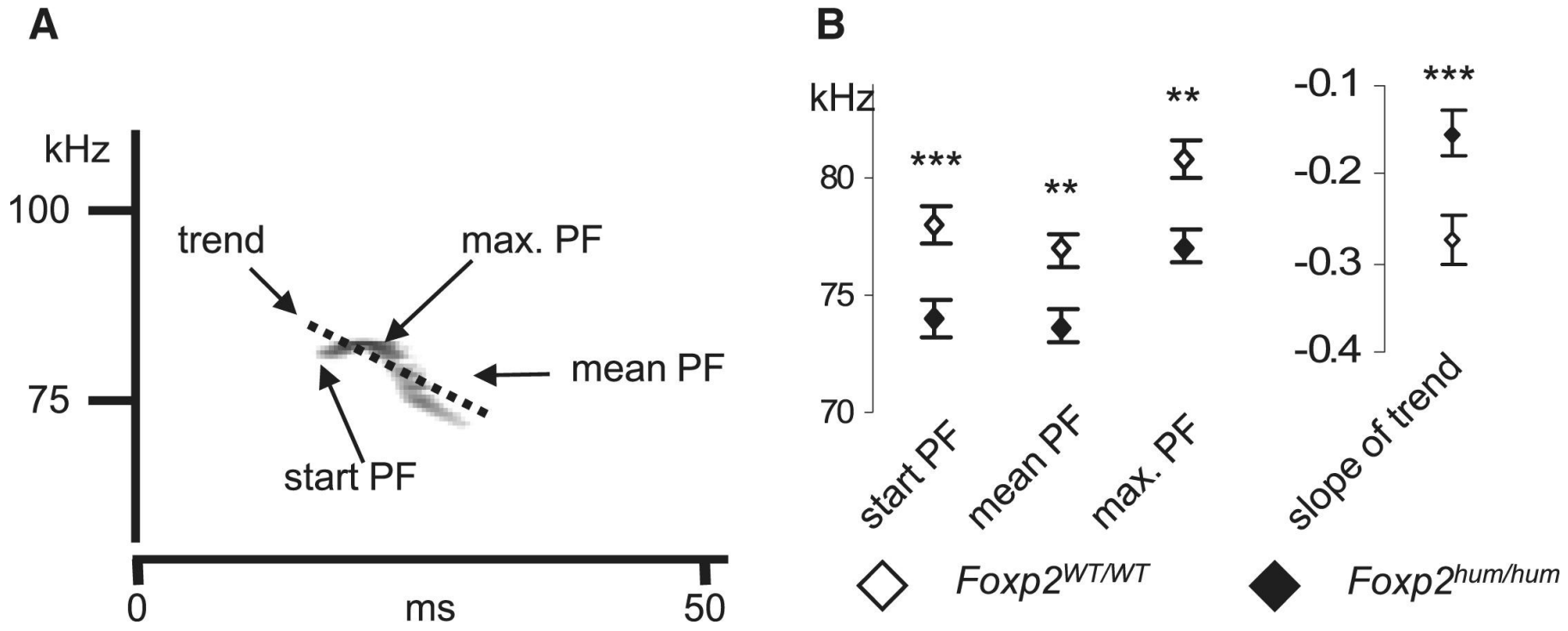
Replaced mouse FOXP2 with human FOXP2
Found gain-of-function changes in brain and behavior

Foxp2^{hum} Increases the Length of Dendritic Trees

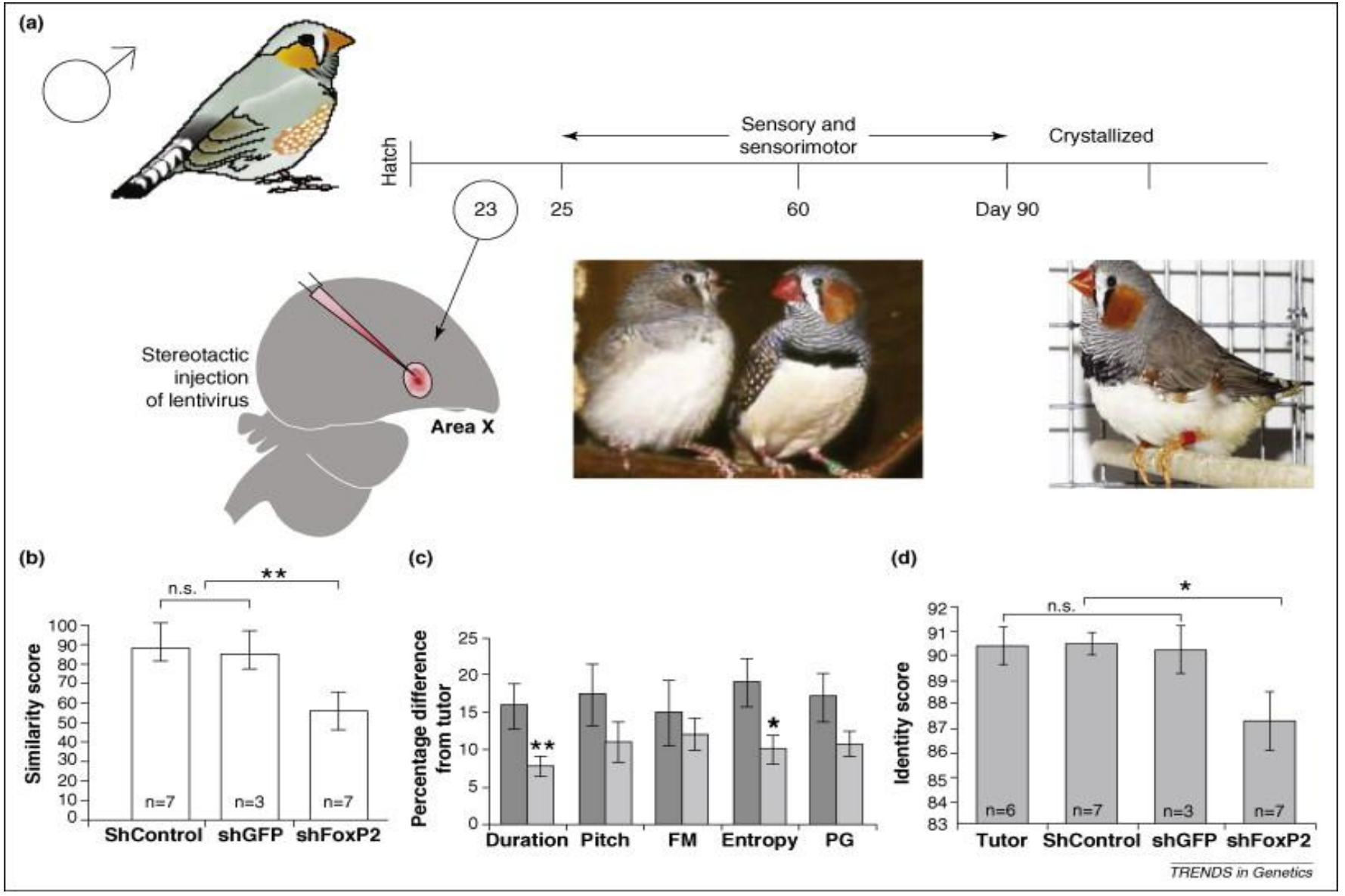




Foxp2^{hum} Affects the Structure of Pup Isolation Calls



FOXP2 controls birdsong



Fischer and Scharff Trends Genet. 2009

Novel mutations in human FOXP2

Exome Variant Server

http://evs.gs.washington.edu/EVS/

Database of sequence variants appearing in cDNAs

11 people have protein changes in FOXP2

Gene Name: FOXP2 (Gene ID: 93986) (+)

Chromosome 7: 114055052 - 114333827

Population: EuropeanAmerican, AfricanAmerican

GWAS Catalog: FOXP2

NHGRI Catalog of Published

C: FOXP2

: FOXP2

Variant Pos	rs ID	Alleles	AA Allele #	All Allele #	Genes	GVS Function	Amino Acid	Protein Pos.
7:114066586	unknown	T/C	T=0/C=4406	T=2/C=13002	FOXP2	missense	ILE,THR	7/716
7:114066586	unknown	T/C	T=0/C=4406	T=2/C=13002	FOXP2	missense	ILE,THR	7/458
7:114066586	unknown	T/C	T=0/C=4406	T=2/C=13002	FOXP2	missense	ILE,THR	7/715
7:114066616	unknown	T/A	T=2/A=4404	T=8/A=12998	FOXP2	missense	LEU,GLN	17/733
7:114066616	unknown	T/A	T=2/A=4404	T=8/A=12998	FOXP2	missense	LEU,GLN	17/433
7:114066616	unknown	T/A	T=2/A=4404	T=8/A=12998	FOXP2	missense	LEU,GLN	17/741
7:114066616	unknown	T/A	T=2/A=4404	T=8/A=12998	FOXP2	missense	LEU,GLN	17/716
7:114066616	unknown	T/A	T=2/A=4404	T=8/A=12998	FOXP2	missense	LEU,GLN	17/458
7:114066616	unknown	T/A	T=2/A=4404	T=8/A=12998	FOXP2	missense	LEU,GLN	17/715
7:114066700	unknown	C/G	C=0/G=4406	C=1/G=13005	FOXP2	missense	THR,SER	45/733
7:114066700	unknown	C/G	C=0/G=4406	C=1/G=13005	FOXP2	missense	THR,SER	45/433
7:114066700	unknown	C/G	C=0/G=4406	C=1/G=13005	FOXP2	missense	THR,SER	45/741
7:114066700	unknown	C/G	C=0/G=4406	C=1/G=13005	FOXP2	missense	THR,SER	45/716
7:114066700	unknown	C/G	C=0/G=4406	C=1/G=13005	FOXP2	missense	THR,SER	45/458
7:114066700	unknown	C/G	C=0/G=4406	C=1/G=13005	FOXP2	missense	THR,SER	45/715
7:114066775	unknown	A/G	A=2/G=4404	A=2/G=12996	FOXP2	intron	none	NA

Ways to detect genes under positive selection
2. Lower Genetic Diversity

Heterozygosity (lower)/rare alleles (higher)

Selective sweep reduces genetic diversity at linked alleles.

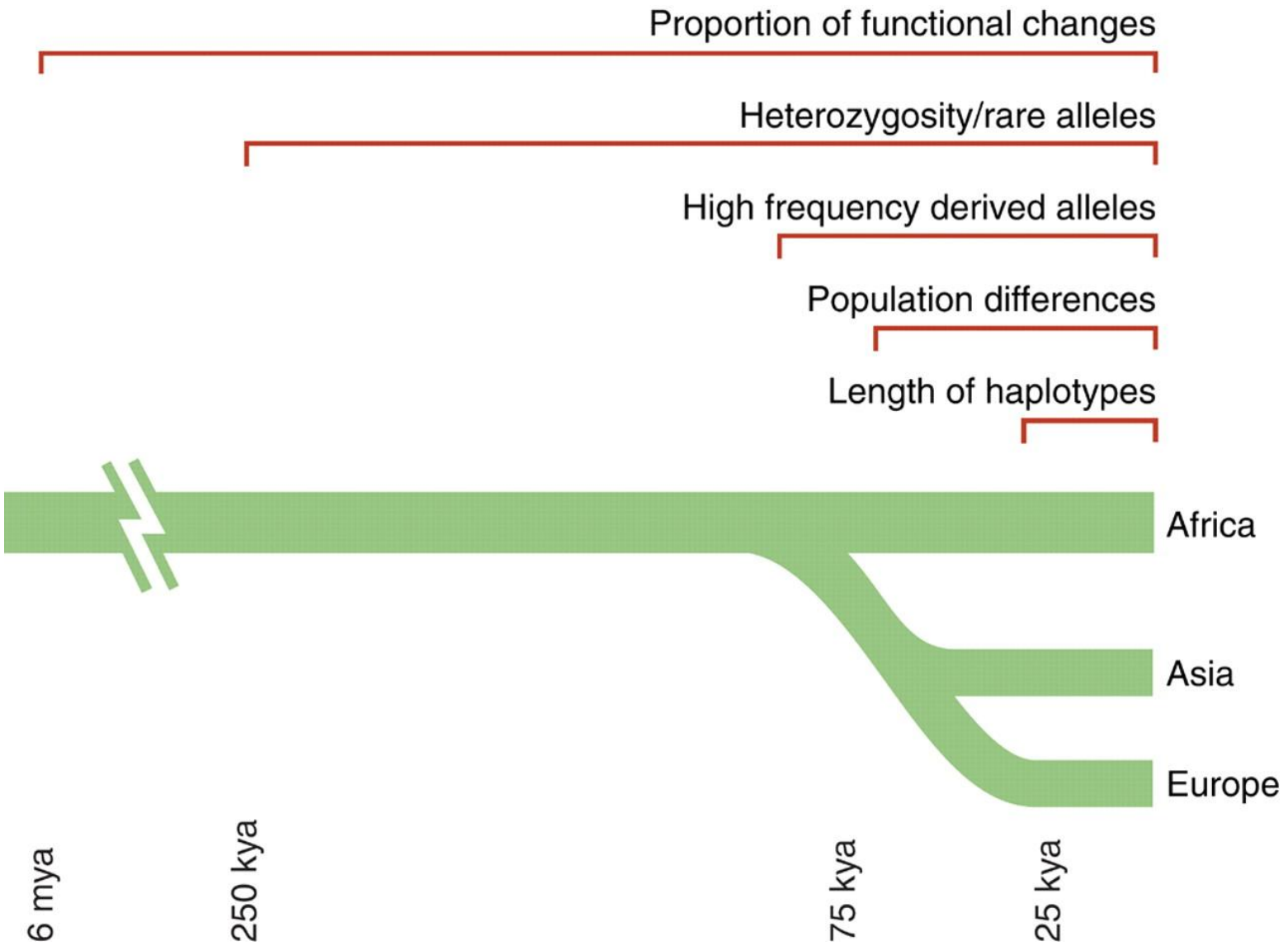
If rare alleles are linked to the selected mutation, they increase in frequency by hitchhiking. So their frequency rises.

Ways to detect genes under positive selection
3. High Frequency Derived Alleles

Most new alleles are at low frequency

One way for a derived allele to become high frequency is to by linked to an allele undergoing positive selection

Fig. 1. Time scales for the signatures of selection.



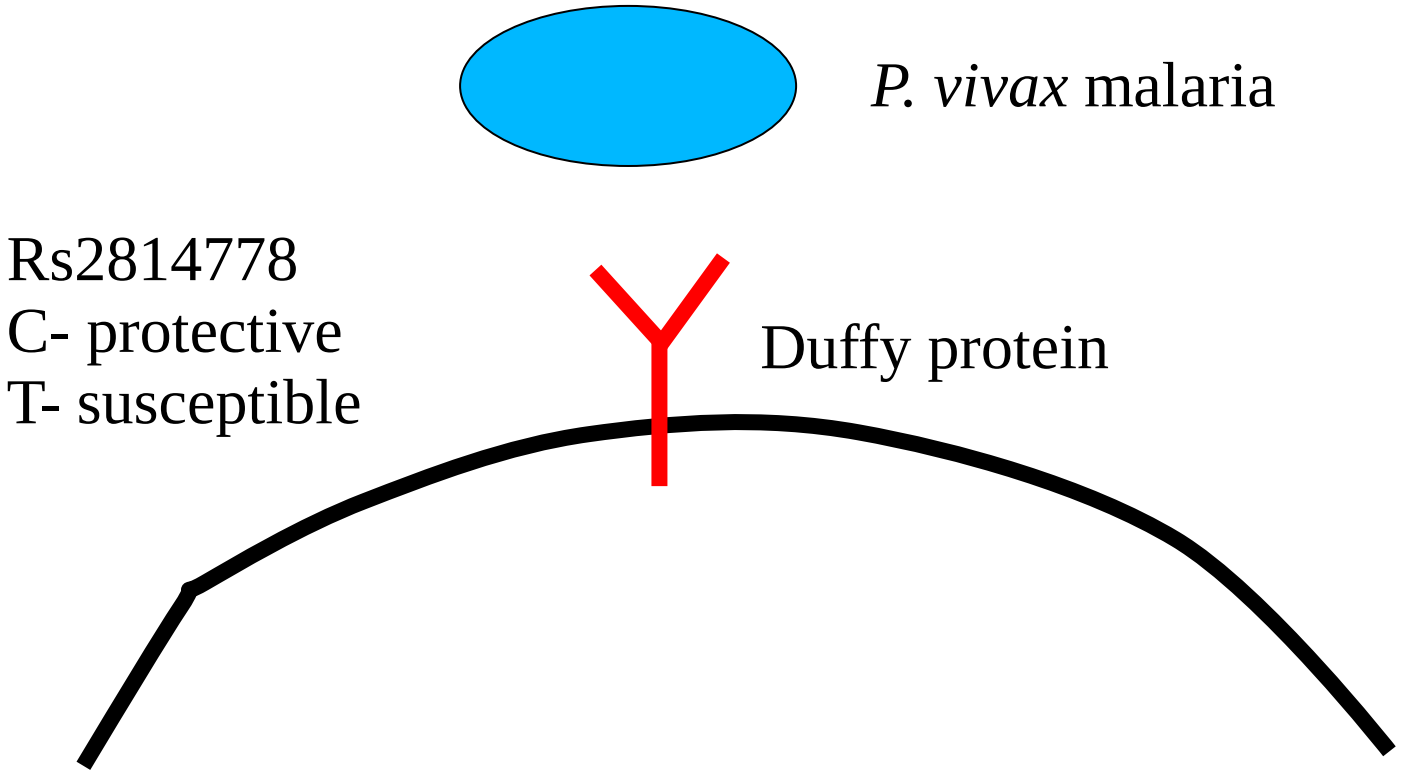
P C Sabeti et al. Science 2006;312:1614-1620

Look up Duffy red cell antigen Rs2814778

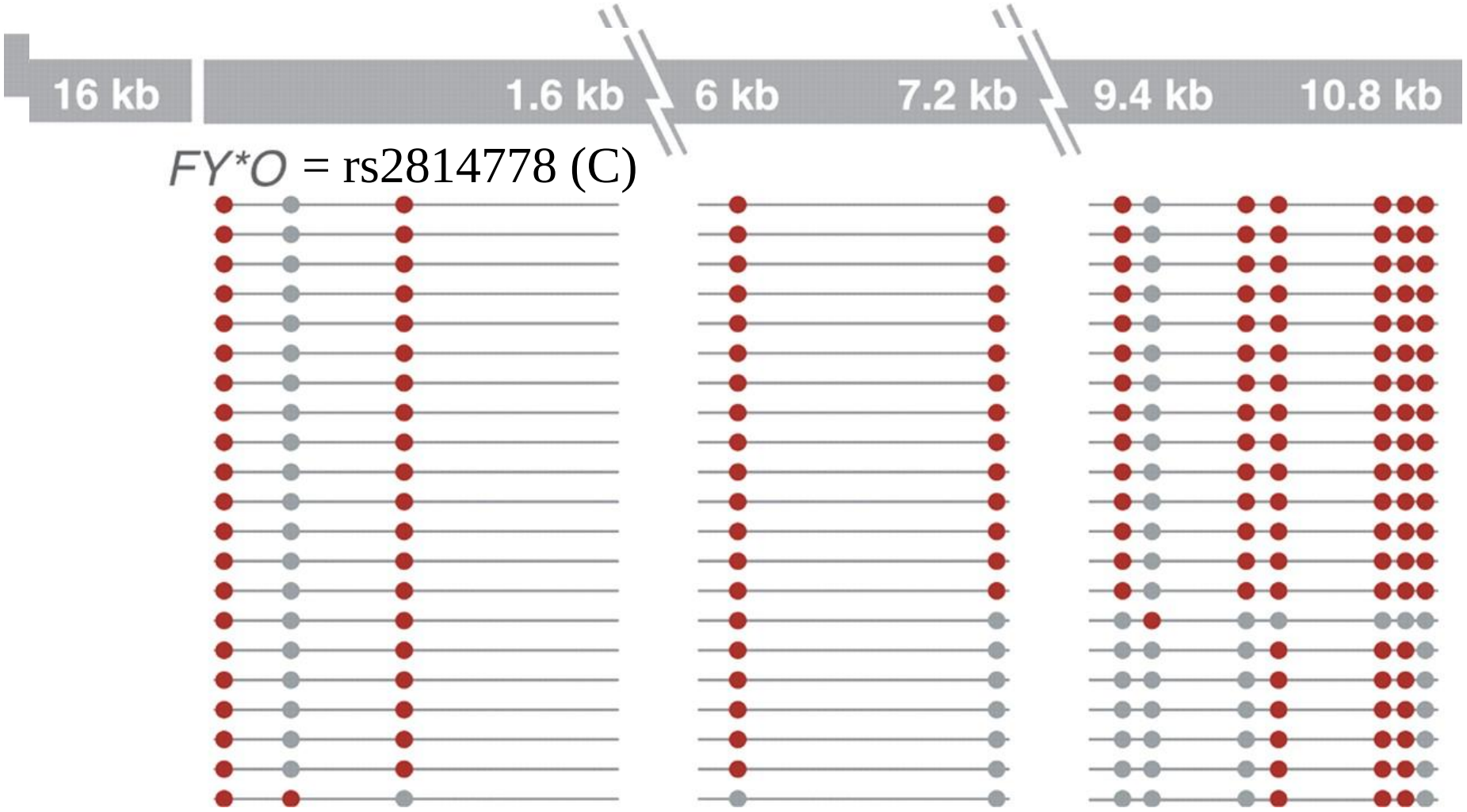
C- protective from malaria
T- susceptible to malaria

Stuart = TT

Duffy red cell antigen (FY)



Excess of high-frequency derived alleles at the Duffy red cell antigen (FY) gene
Resistance to malaria.



Red – derived. Gray – ancestral (Chimp)

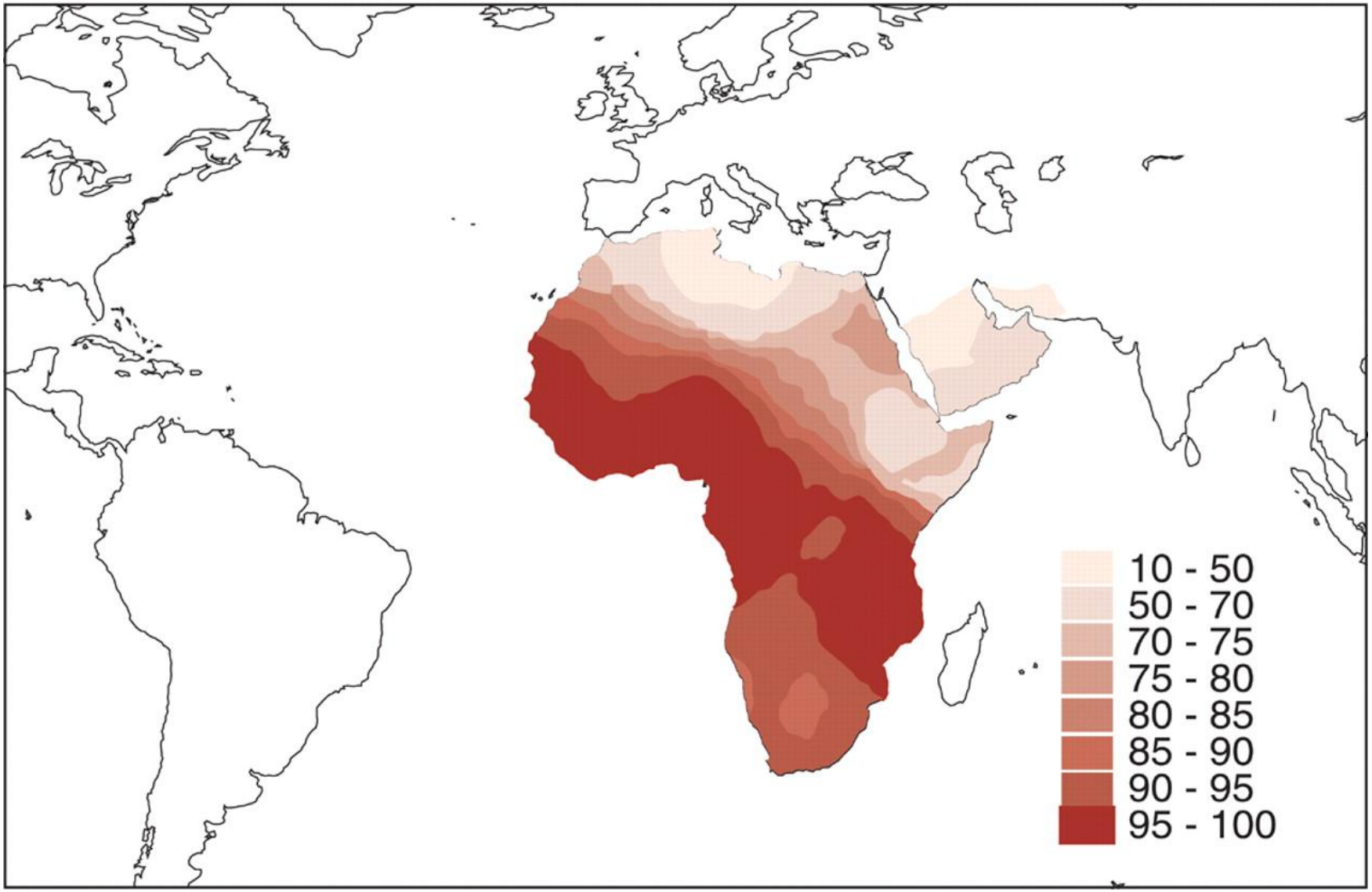
P C Sabeti et al. Science 2006;312:1614-1620

Ways to detect genes under positive selection
4. Differences between populations

Different populations may have different selective pressures (domestication of cattle, resistance to malaria etc.)

Different alleles may be selected in different populations.

Extreme population differences in FY*O allele frequency.

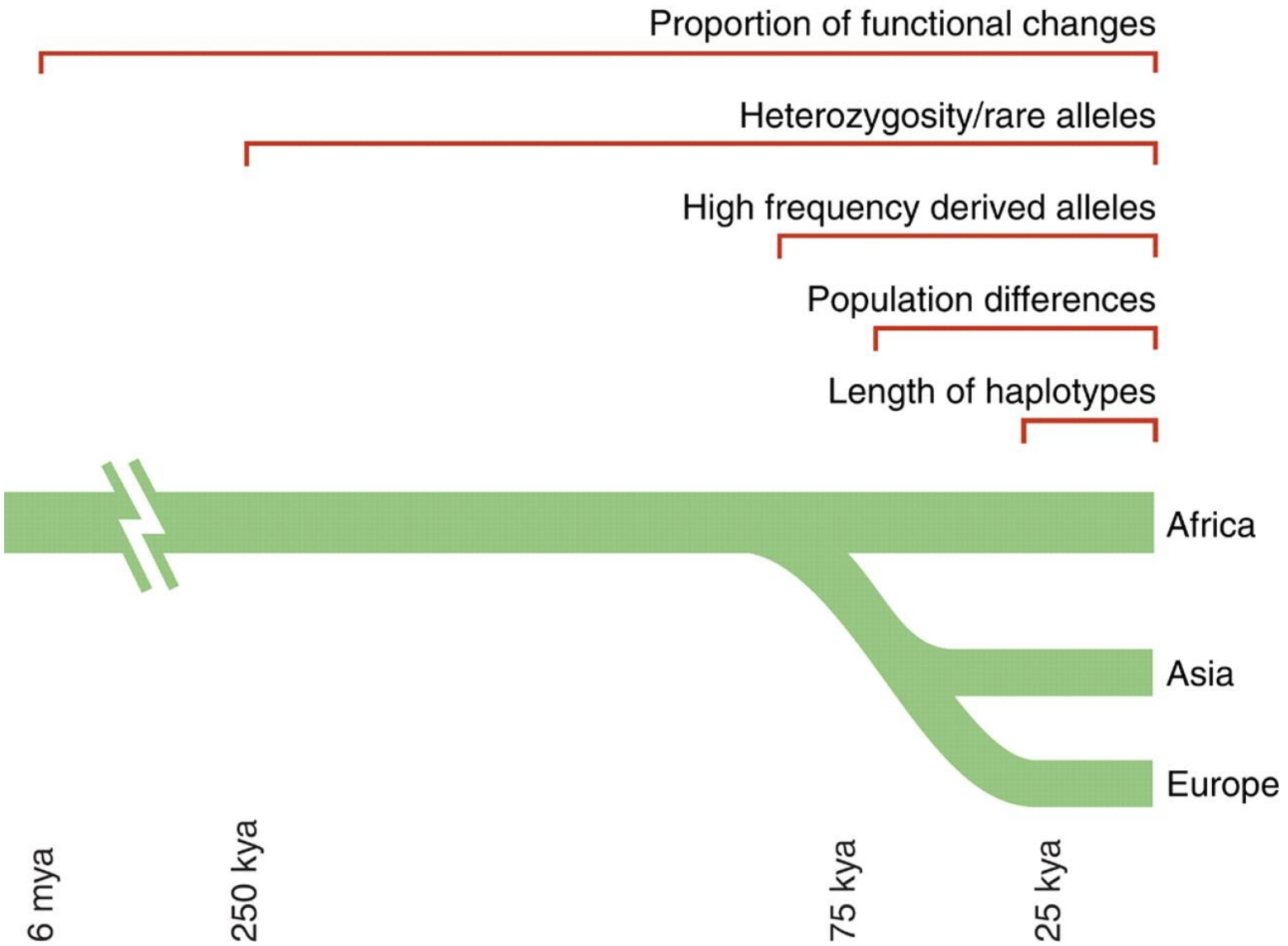


P C Sabeti et al. Science 2006;312:1614-1620

Ways to detect genes under positive selection

5. Long Haplotype

Fig. 1. Time scales for the signatures of selection.



P C Sabeti et al. Science 2006;312:1614-1620

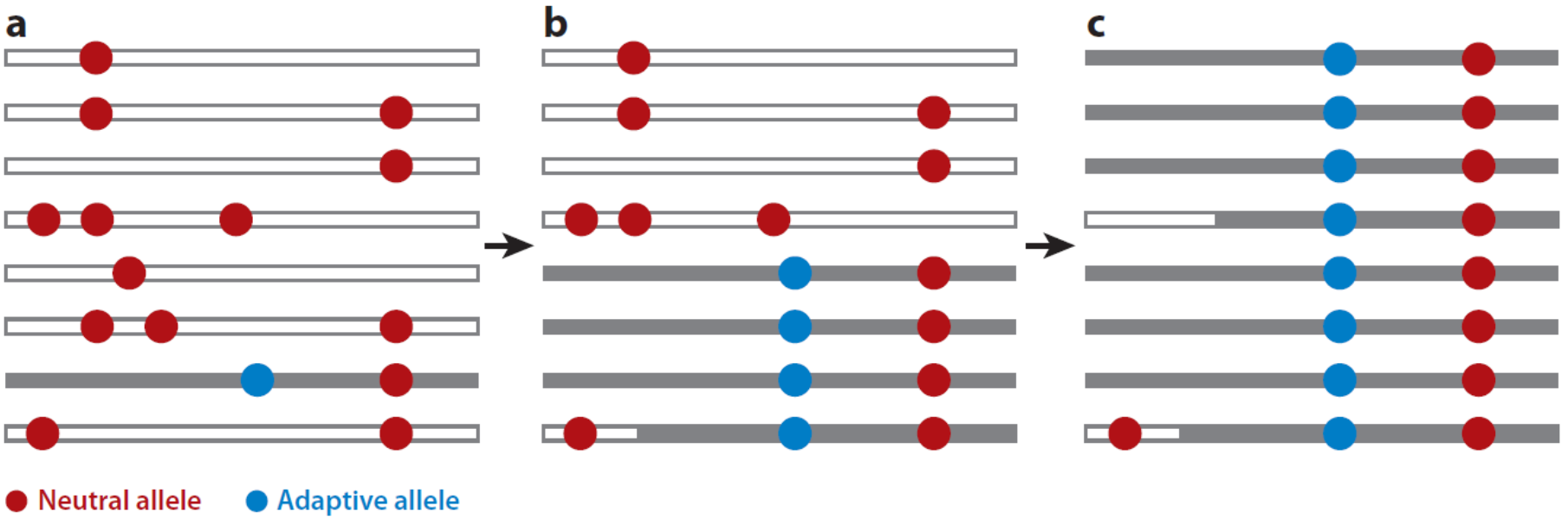
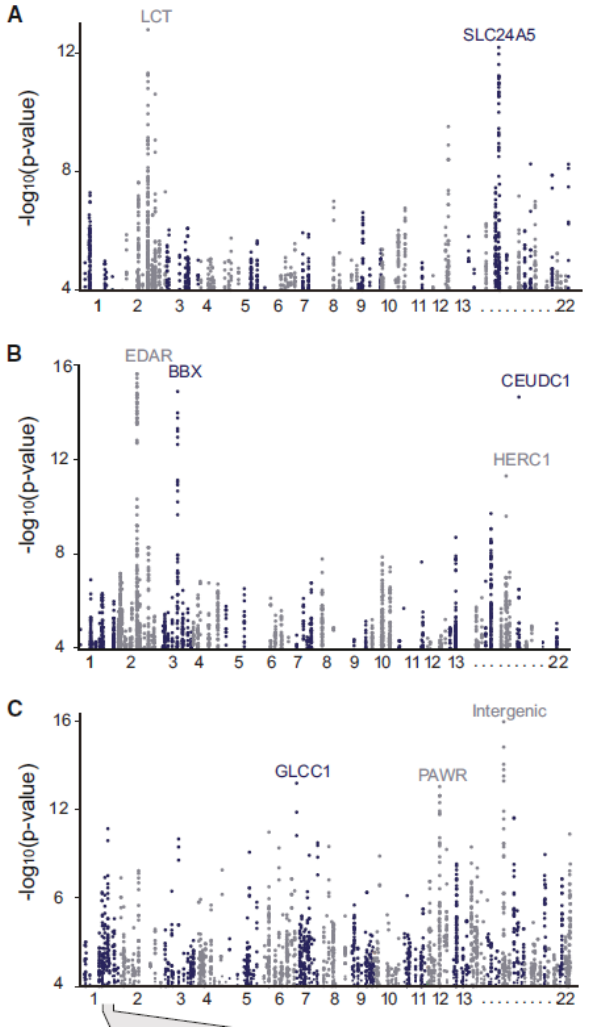


Figure 1

Positive Selection Regions



Europe

Asia

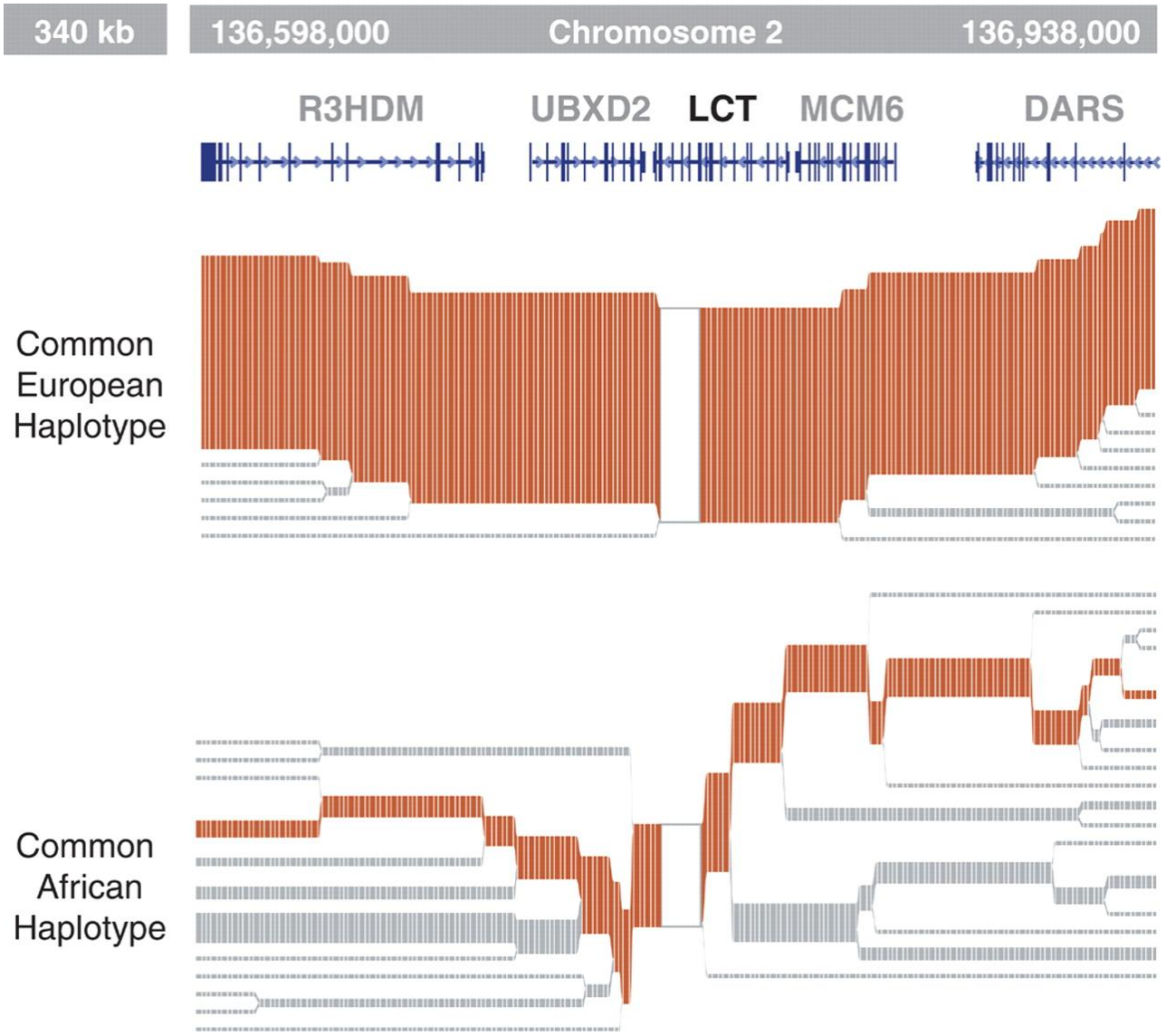
Africa

Grossman et al., Cell 2013

Look up lactase
rs4988235
A – derived, can drink milk
G – ancestral, lactose intolerant

Stuart = GG

Long haplotype surrounding the lactase persistence allele.



P C Sabeti et al. Science 2006;312:1614-1620

Europeans

Lactase

a mutation in a regulatory region near the gene for lactase (LCT) that allows lactose tolerance to persist into adulthood. This particular variant was apparently selected in parts of Europe after the domestication of cattle.

Frequency of the C/T-13910 allele in Lactase

The *LCT* region appears to have undergone a selective sweep 2000–20,000 years ago (4), coinciding with the domestication of cattle.

The high selection coefficient (between 0.014 and 0.15) distinguishes *LCT* as one of the most strongly selected loci in the human genome.

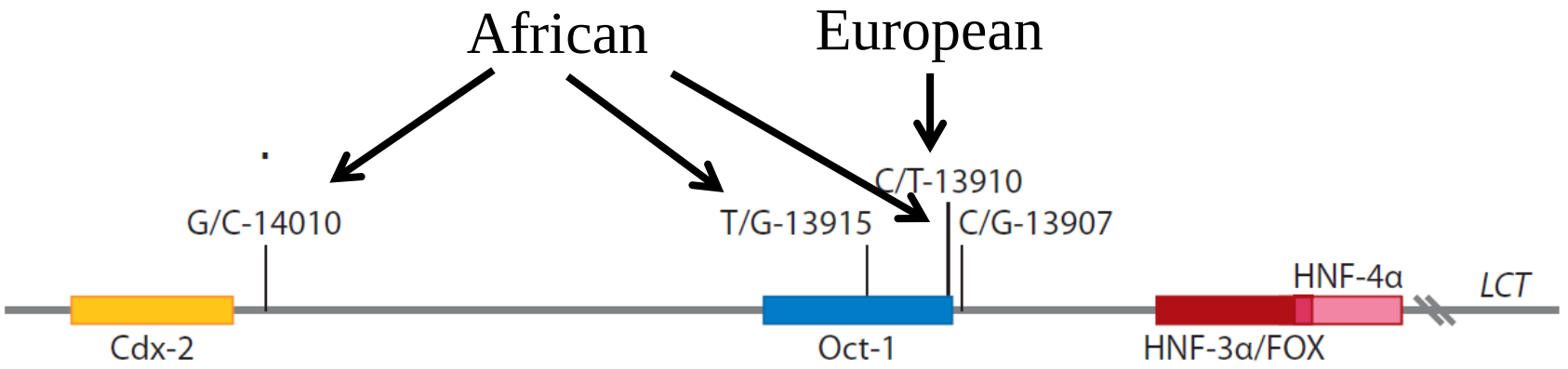


Figure 3

Locations of transcription factor-binding sites and predicted adaptive alleles upstream of *LCT*, the lactase gene. Three alleles were identified as potentially causal alleles in the African pastoral populations, whereas C/T-13910 was predicted to be the causal allele in Northern Europeans. Additionally, the T/G-13915 allele is correlated with lactase persistence in the Saudi Arabian population. The transcription factors and the sequence they bind in a supershift assay (48) are: HNF-4α (−13854 to −13830), HNF-3α and FOX (−13872 to −13848), Oct-1 and GAGA (−13933 to −13909), and Cdx-2 (−14040 to −14016).

SLC24A5: skin color

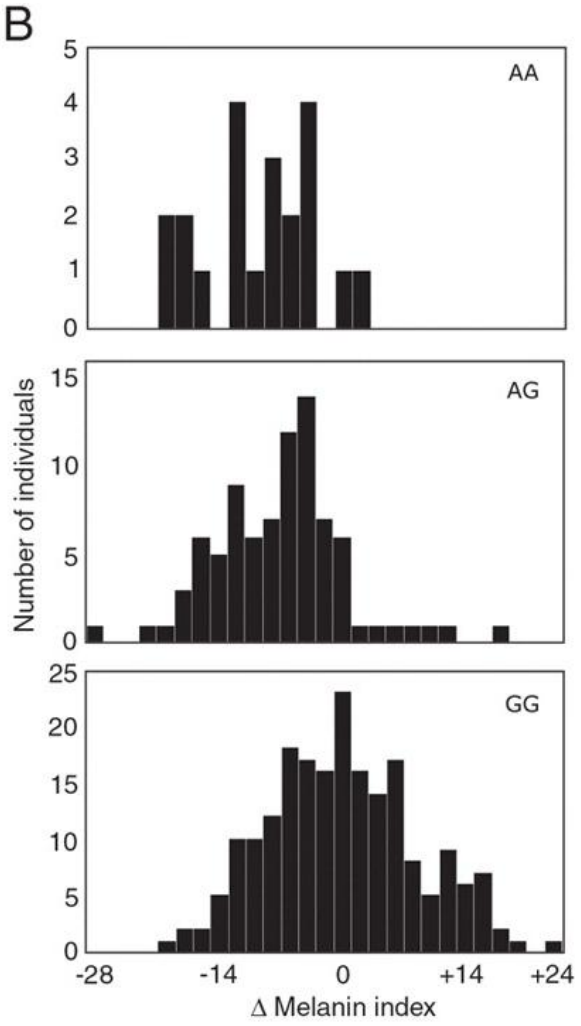
look up: rs1426654

(A), light-skinned european ancestry

(G), ancestral, dark skin

Stuart = GG

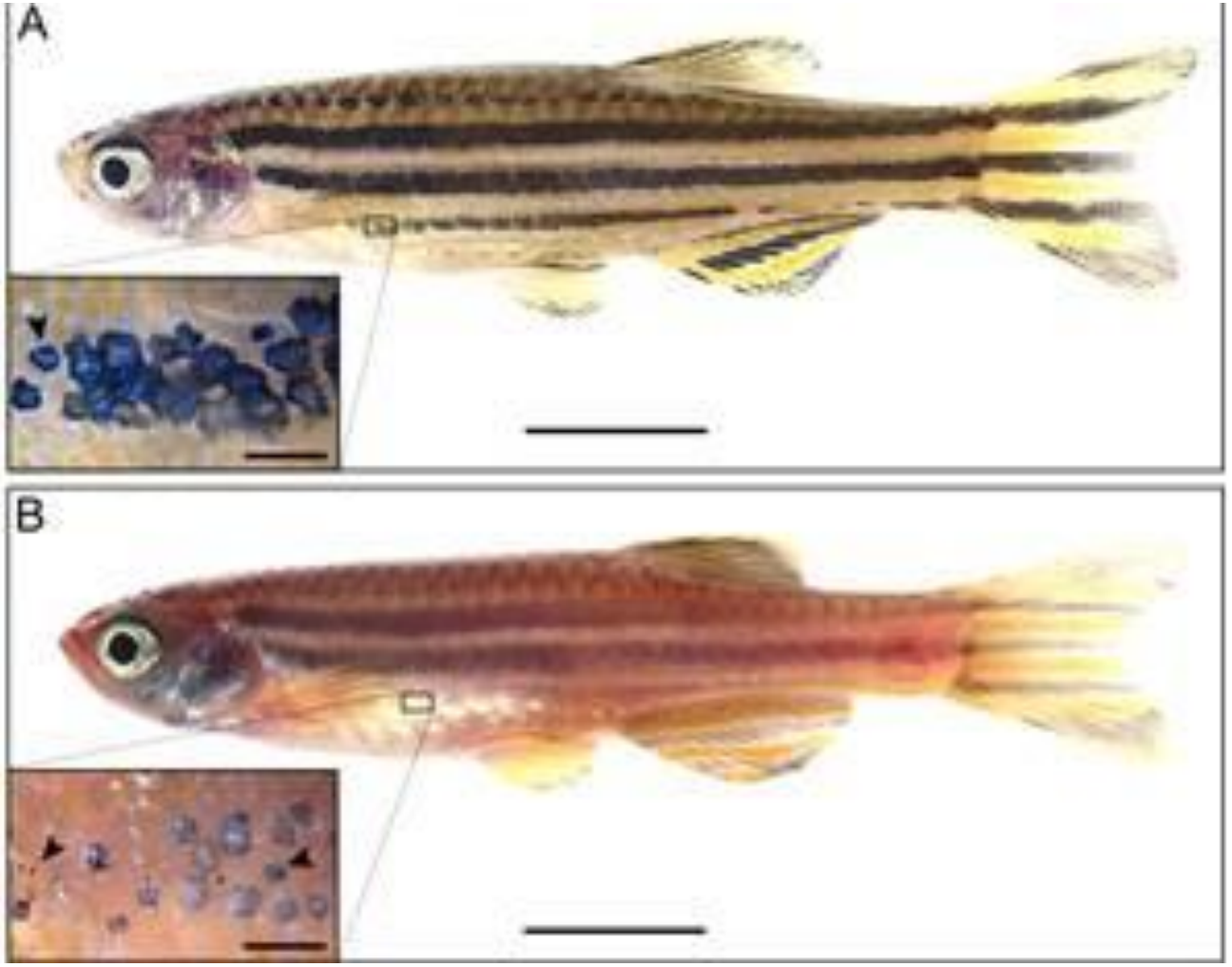
Fig. 6. Effect of SLC24A5 genotype on pigmentation in admixed populations.

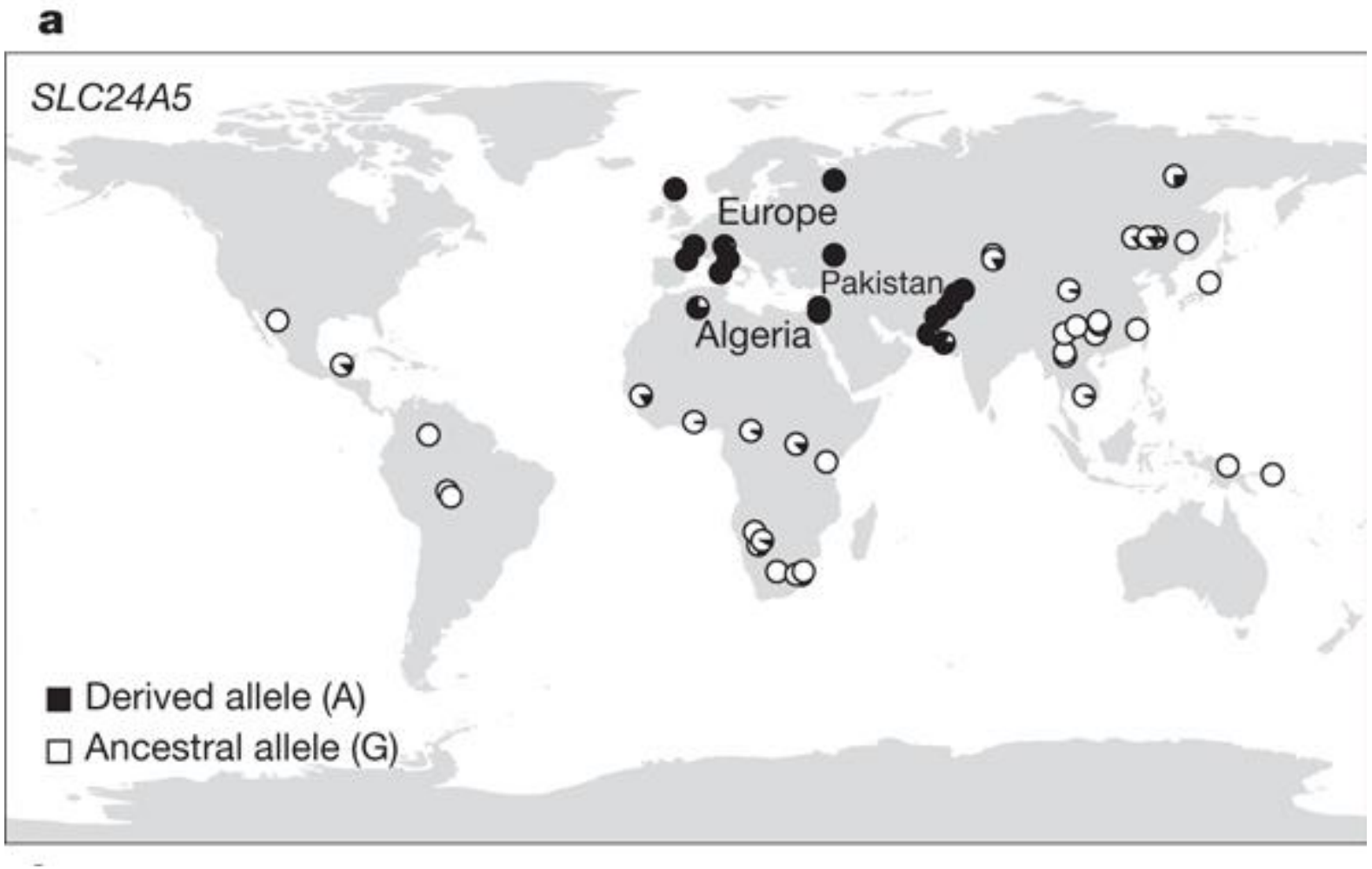


Stuart = GG

R L Lamason et al. Science 2005;310:1782-1786

The zebrafish SLC24A5 ortholog (golden) controls skin color

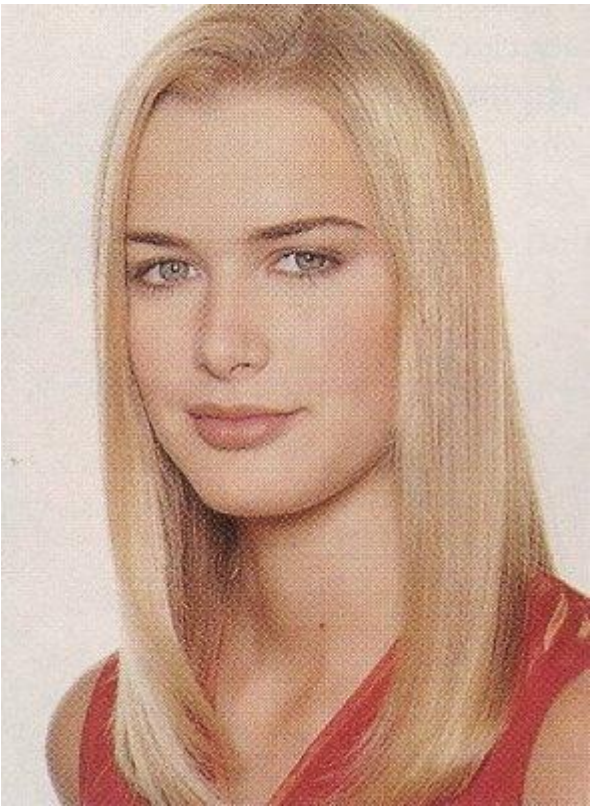




SLC24A5 involved in skin color. A111T .

SLC45A2: Also, an L374F substitution is at 100% frequency in the European sample, but absent in the Asian and African samples.

EDAR: Hair thickness



EDAR (rs3827760)
A = thin hair, 370Valine
G = thick hair, 370Alanine

Stuart = GG

Asian

EDAR and EDA2R

Hair morphology

have a central role in generation of the primary hair follicle

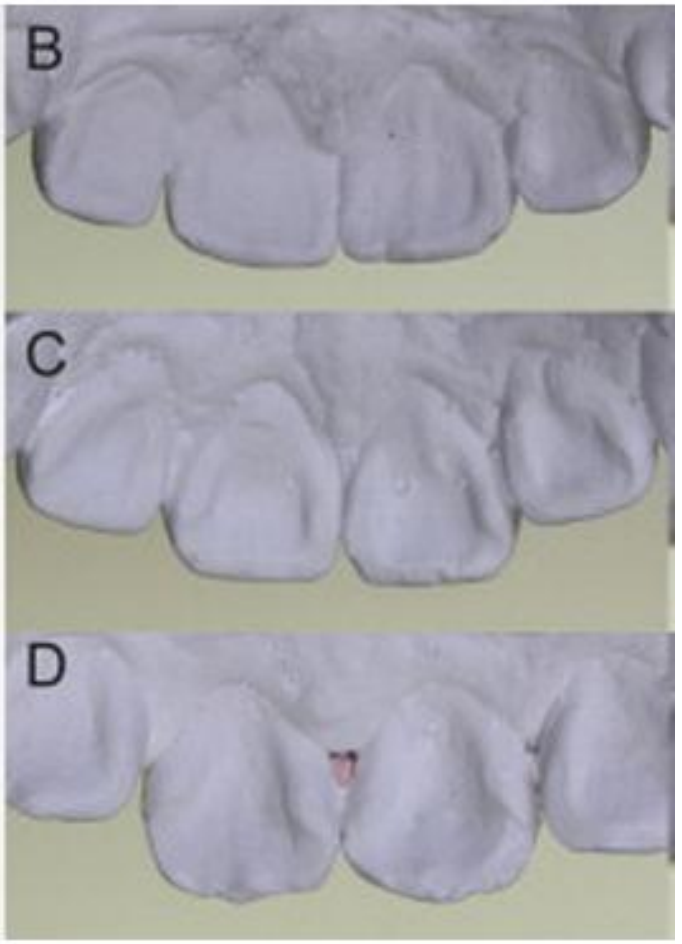
A mutation encoding a V370A substitution in *EDAR* is near fixation in Asia and absent in Europe and Africa

100% in Pima Indians and in parts of China, and 73% in Japan

EDAR (rs3827760)
hair thickness in Asians

G/G G/A A/A

EDAR (rs3827760)
shovel teeth in Asians



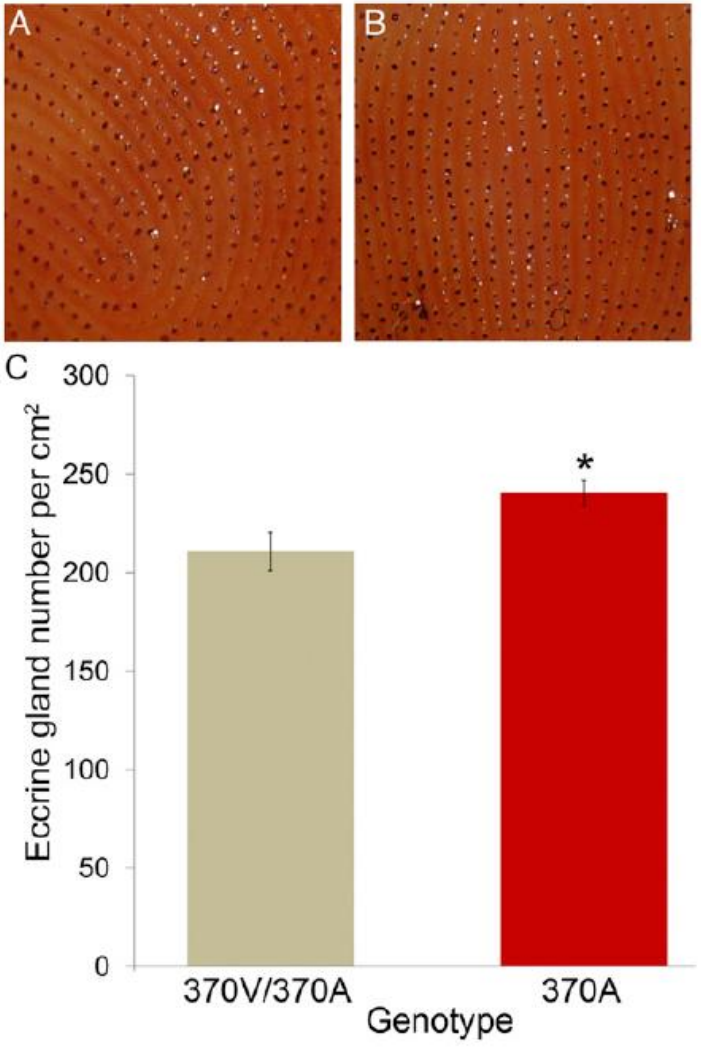
AA

GA

GG

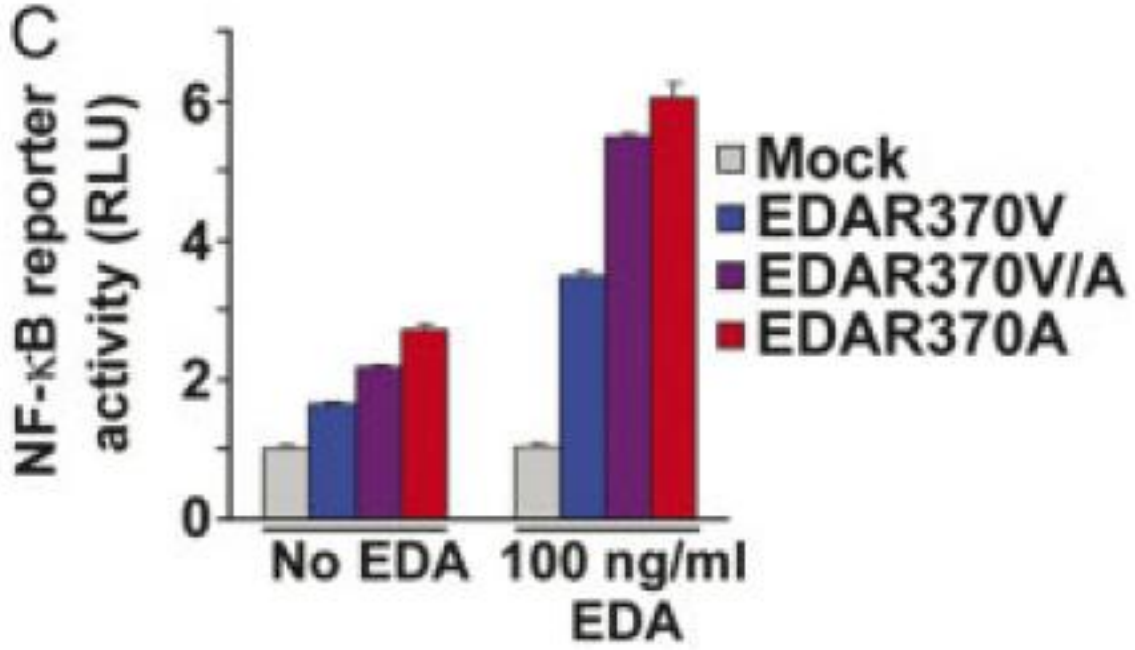
Kimura et al., AJHG 2009

EDAR (rs3827760)
sweat gland number



A – 370V
G – 370C

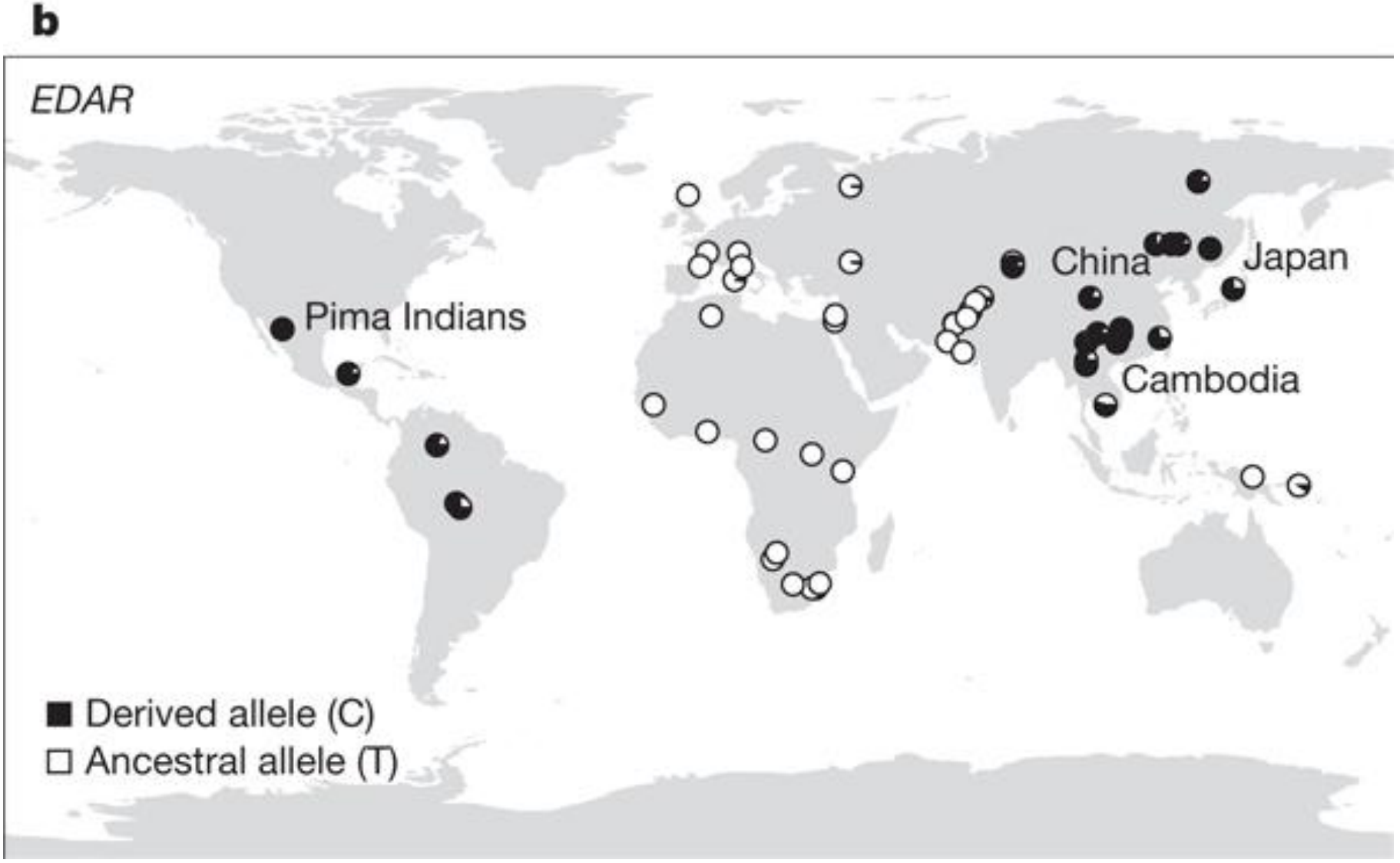
EDAR 370A is more active than 370V

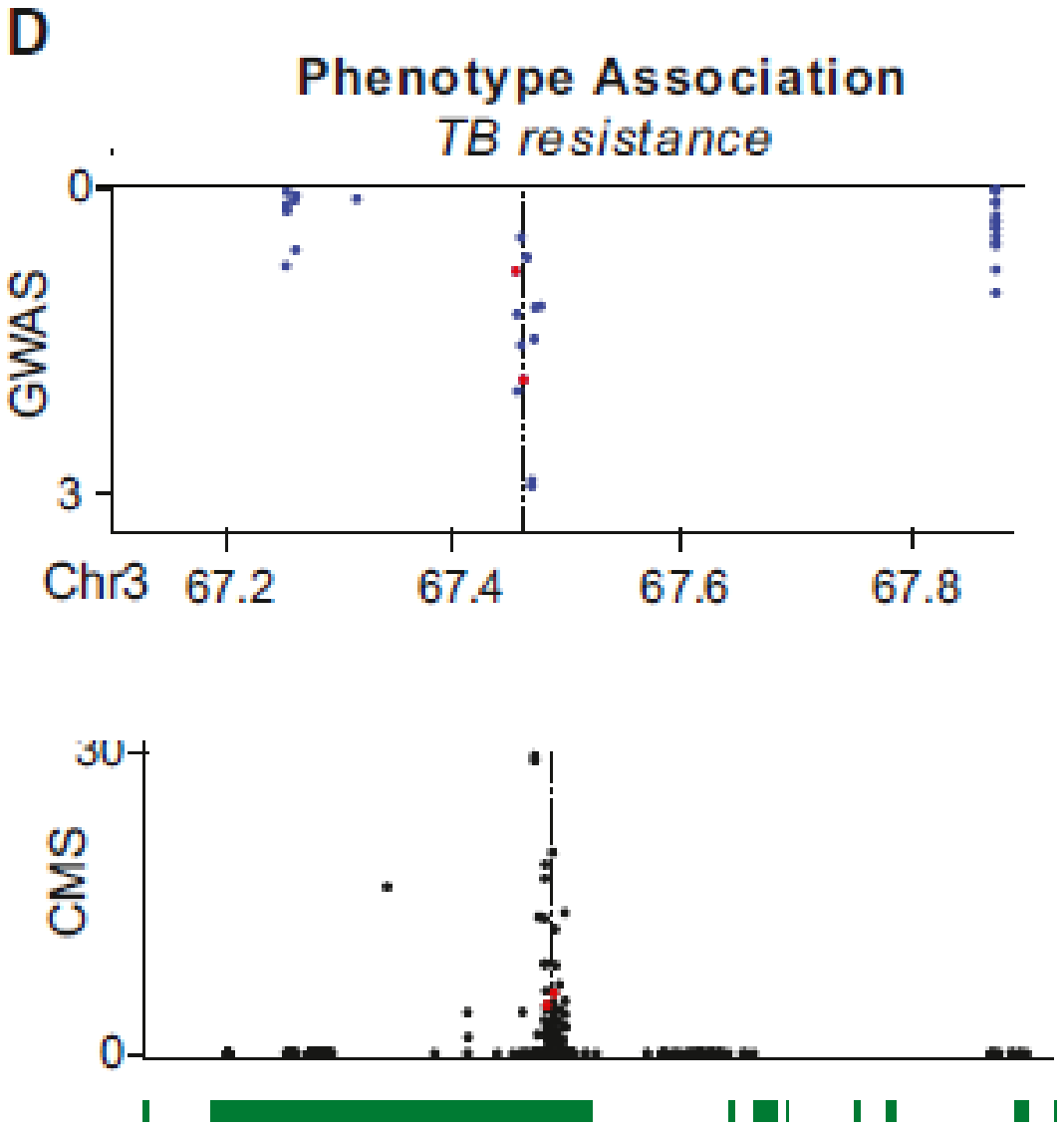


EDAR overexpression

A

Asian


b
EDAR
Pima Indians
China
Japan
Cambodia
Derived allele (C)
Ancestral allele (T)



Region previously known to be associated with TB resistance

Also shows strong signal for positive selection

Grossman et al., Cell 2013

Lookup TLR5

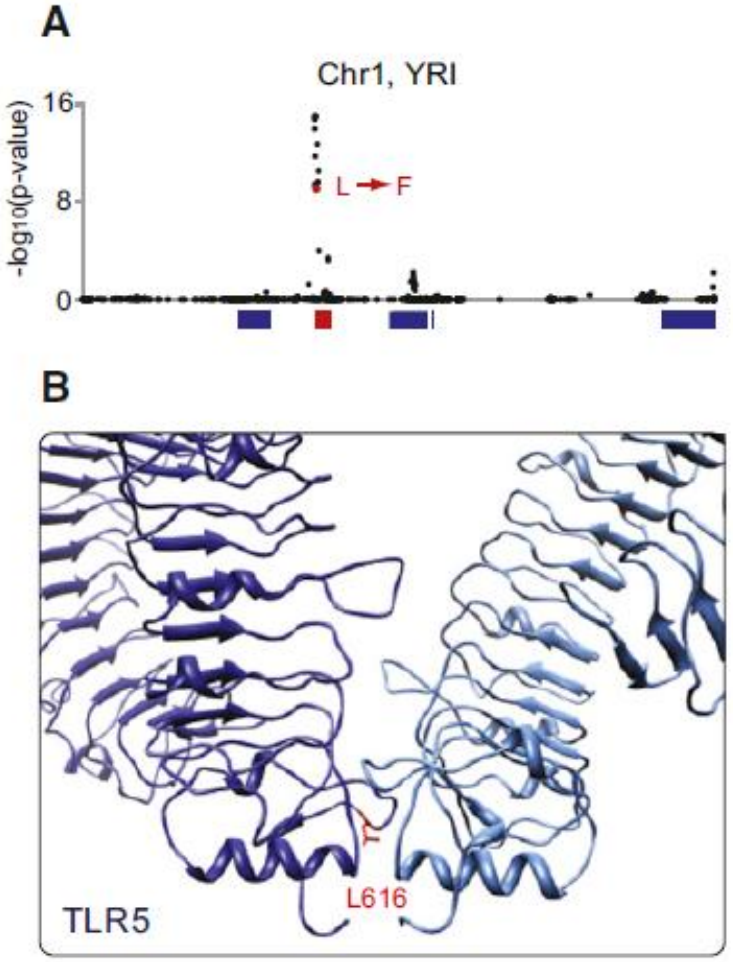
rs5744174

G = derived, resistant to Salmonella 616F

A = ancestral, sensitive to Salmonella 616L

Stuart = AA

Toll Receptor 5

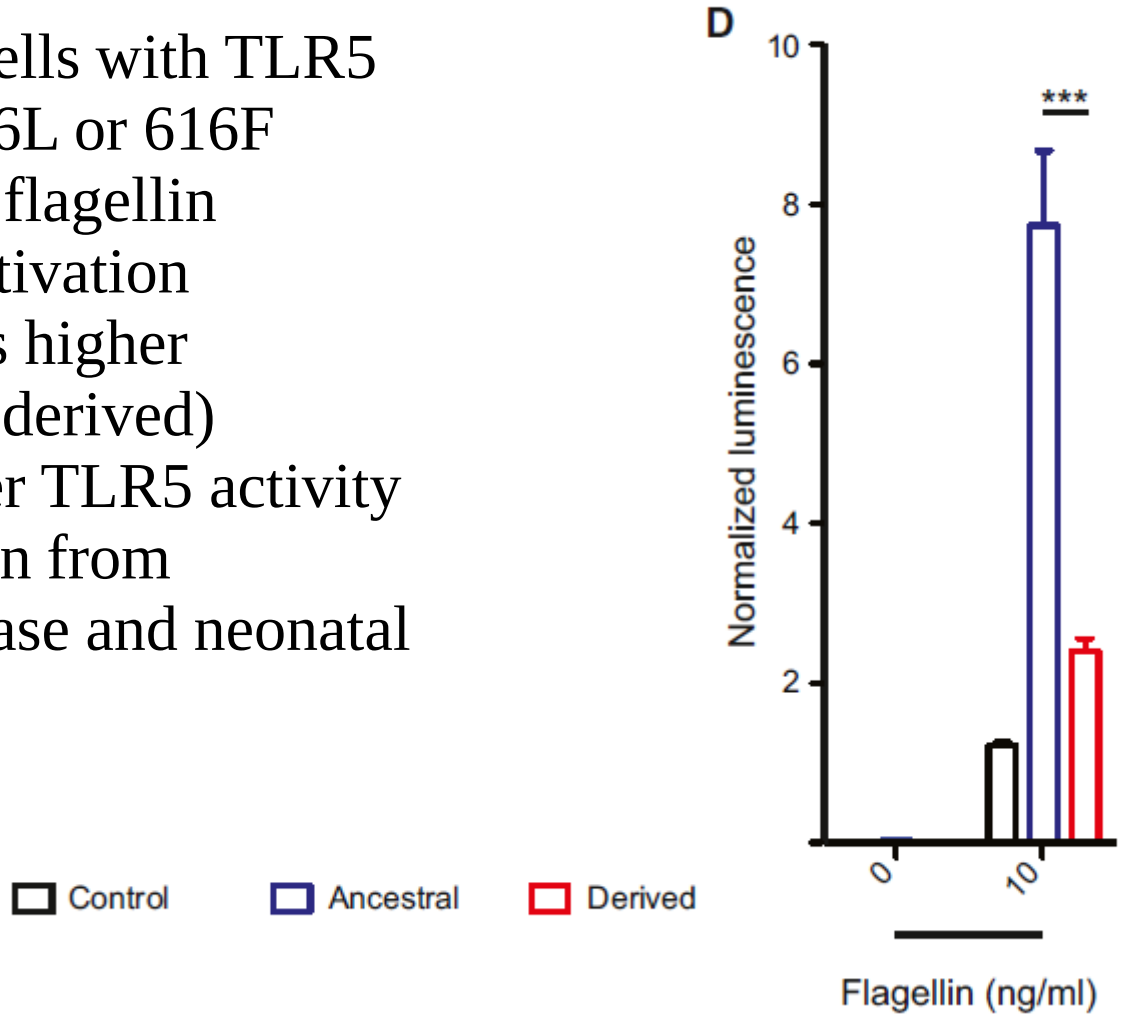


Shows positive selection in YRI
L616F in extracellular binding site
Receptor for bacterial flagellin
Activates NFkB
Proinflammatory response
Differences in TLR5 affect response to Legionnaires disease, neonatal sepsis and Salmonella infection

Grossman et al., Cell 2013

Functional Characterization of TLR5 (L616F)

Transfect Jurkat cells with TLR5 carrying either 616L or 616F
Activate bacterial flagellin
measure NFkB activation
L (ancestral) gives higher activation than F (derived)
suggests that lower TLR5 activity provides protection from Legionnaires disease and neonatal sepsis



Grossman et al., Cell 2013