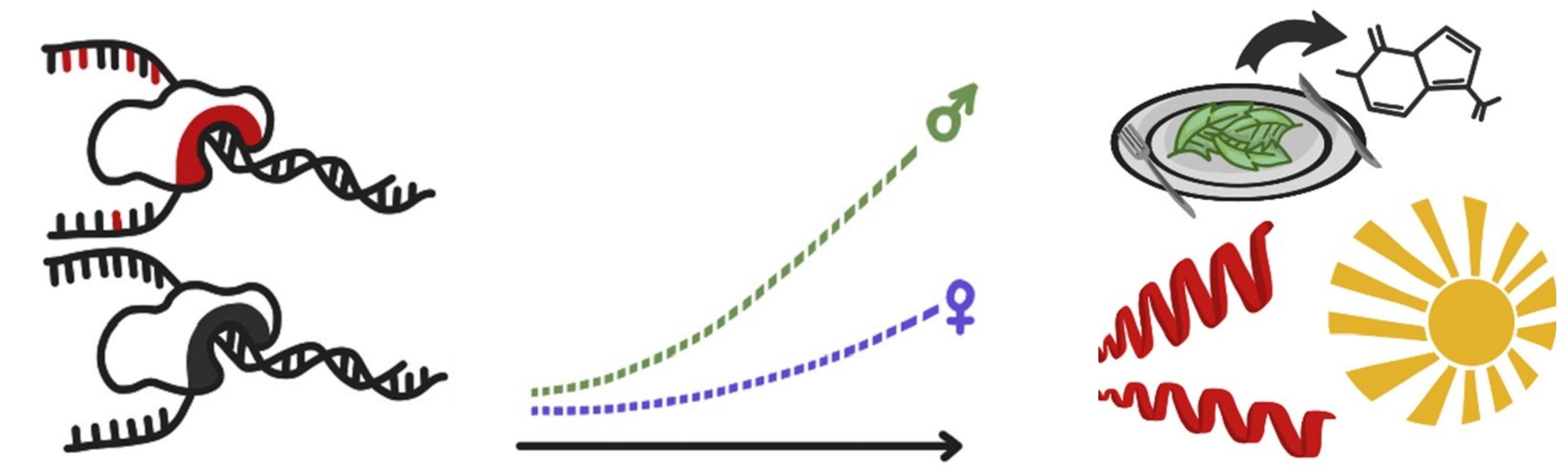


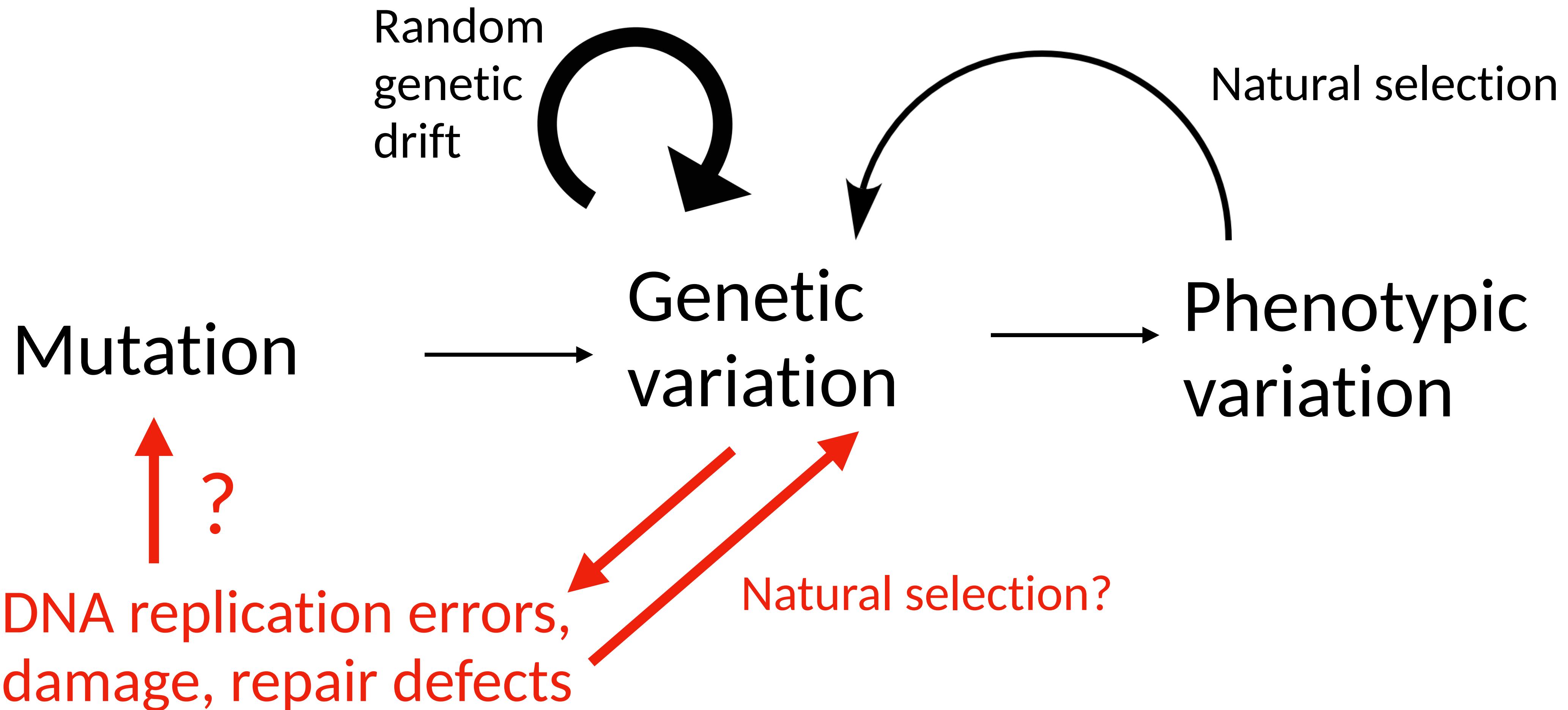
# The evolution of mammalian mutagenesis

Kelley Harris

University of Washington  
Department of Genome Sciences



# What role do mutations play in evolution?



# Significance of mutation rate variation within and among species

Sturtevant (1937)

“A body of data has been gradually accumulating that shows that different [fly] strains may deviate significantly from the usual values for mutation rate.”

- Most mutations are deleterious, so natural selection should disfavor high mutation rates
- Despite this selection, “mutations are accidents, and accidents will happen.”
- Different strains likely evolve different mutation-rate lowering mechanisms, which might be more effective in flies that adapt to mutagenic environments



# Do large, long-lived species require more effective DNA repair to avoid cancer and age-related degradation?

Peto's Paradox (1977)

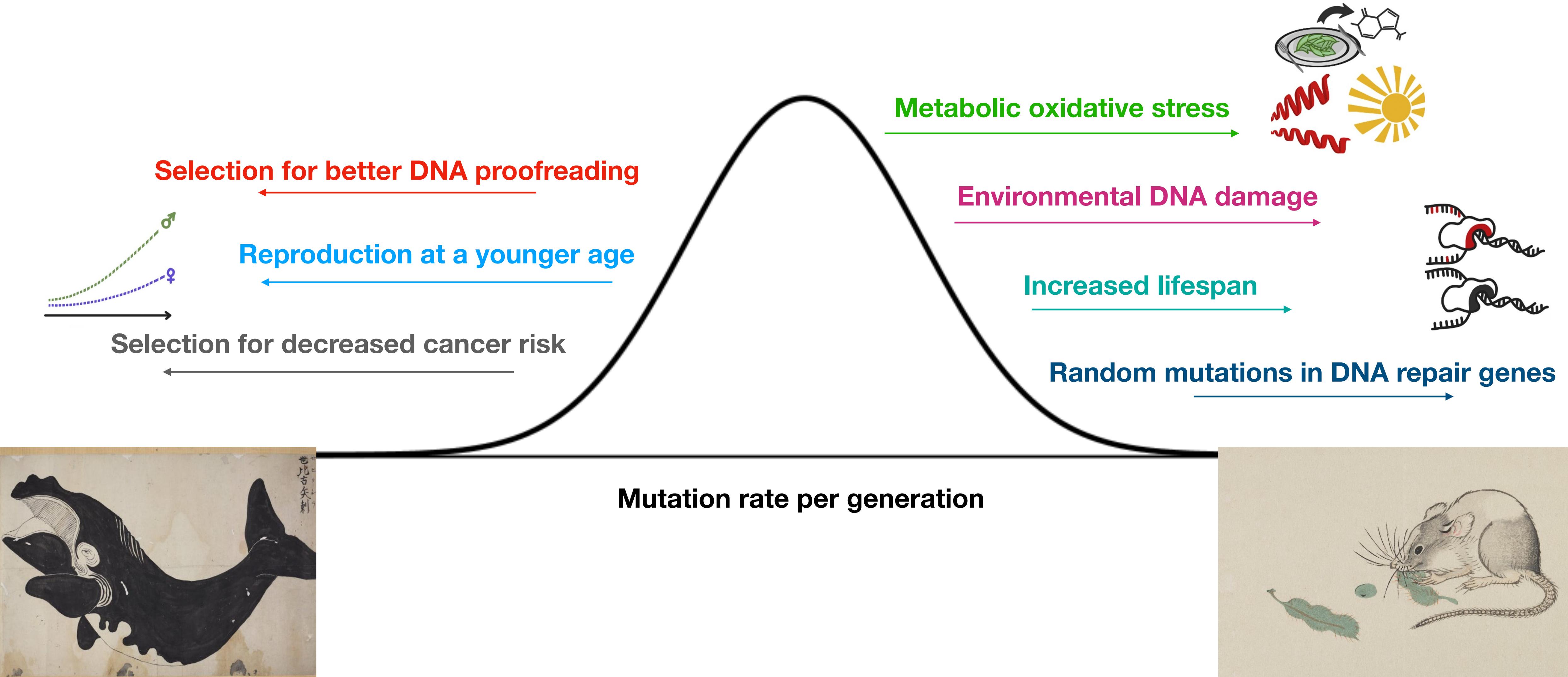


"Geigyo Hinshu Zukan" (Fourteen Varieties of Whales) (1760) (New Bedford Whaling Museum)



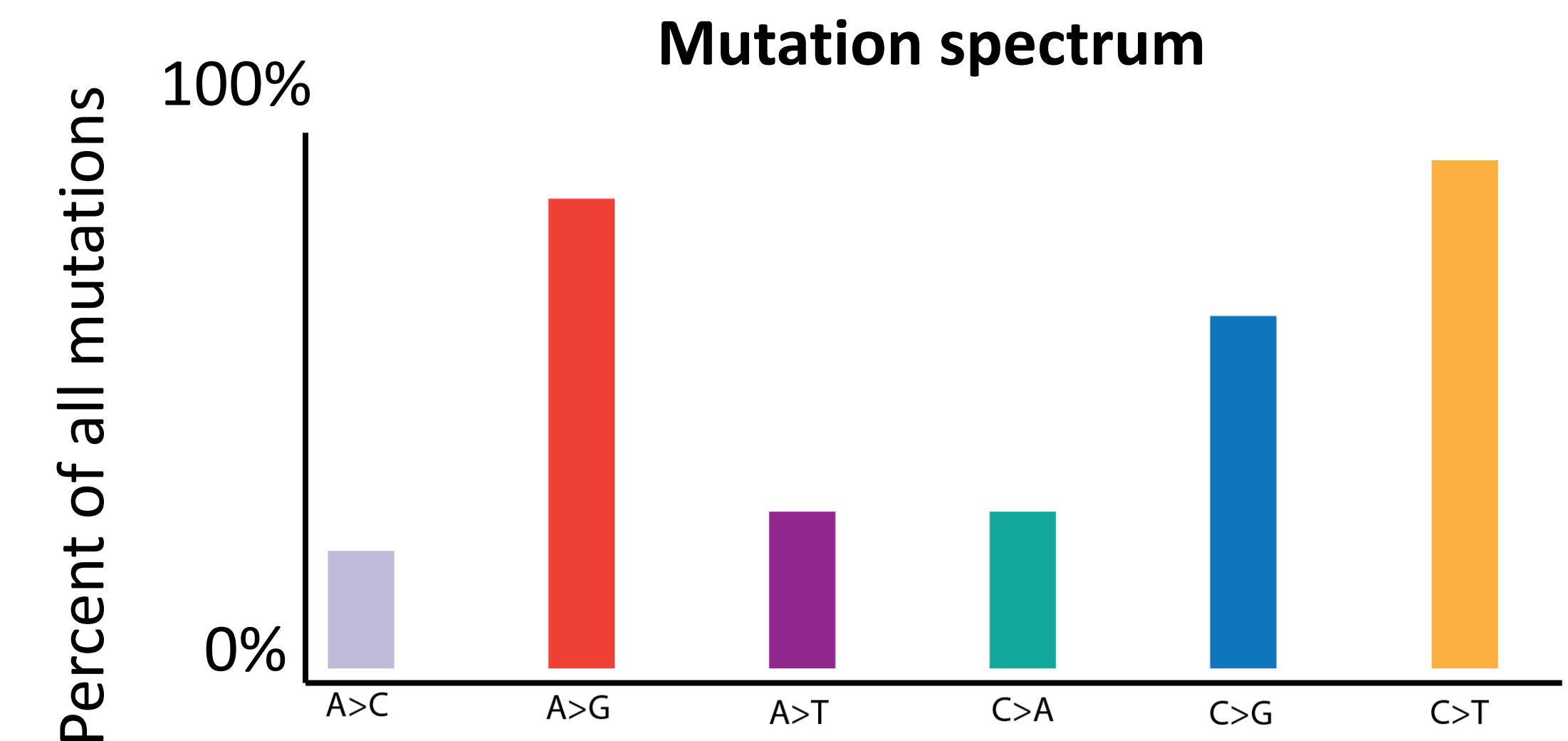
"Mouse eating peapods," Gyokusen (1850), Brooklyn Museum

DNA repair, cancer avoidance, reproductive life history, and many other factors might cause mutation rates to evolve over time

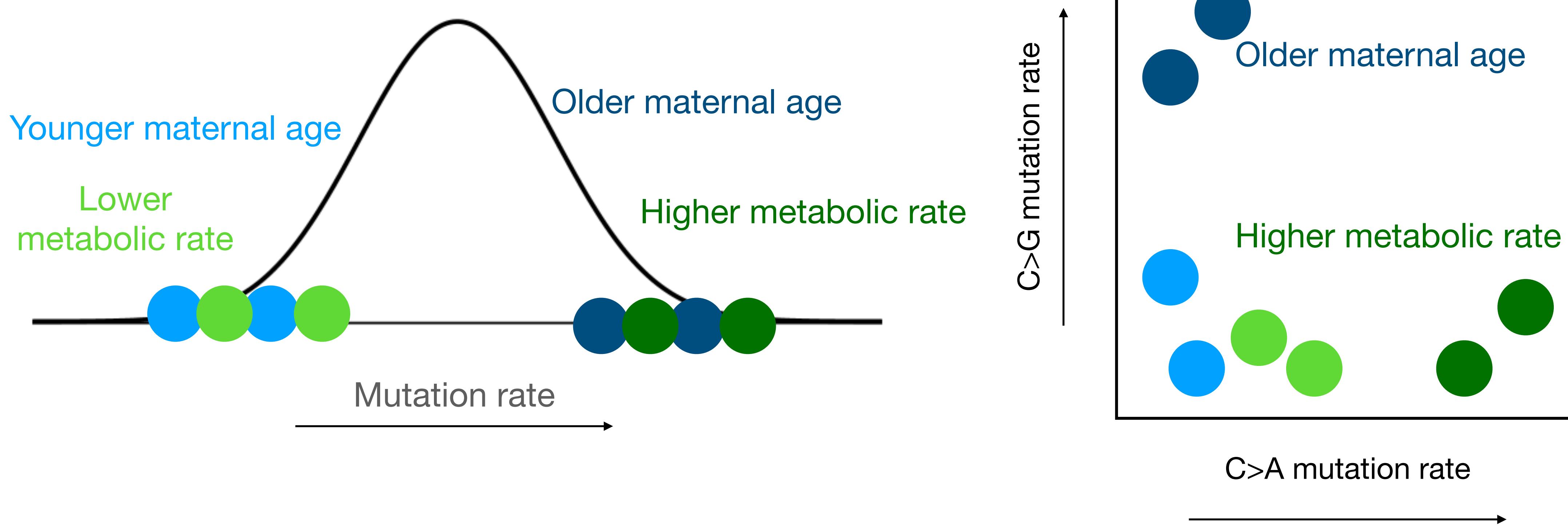


# Different mutational processes can create different types of mutations

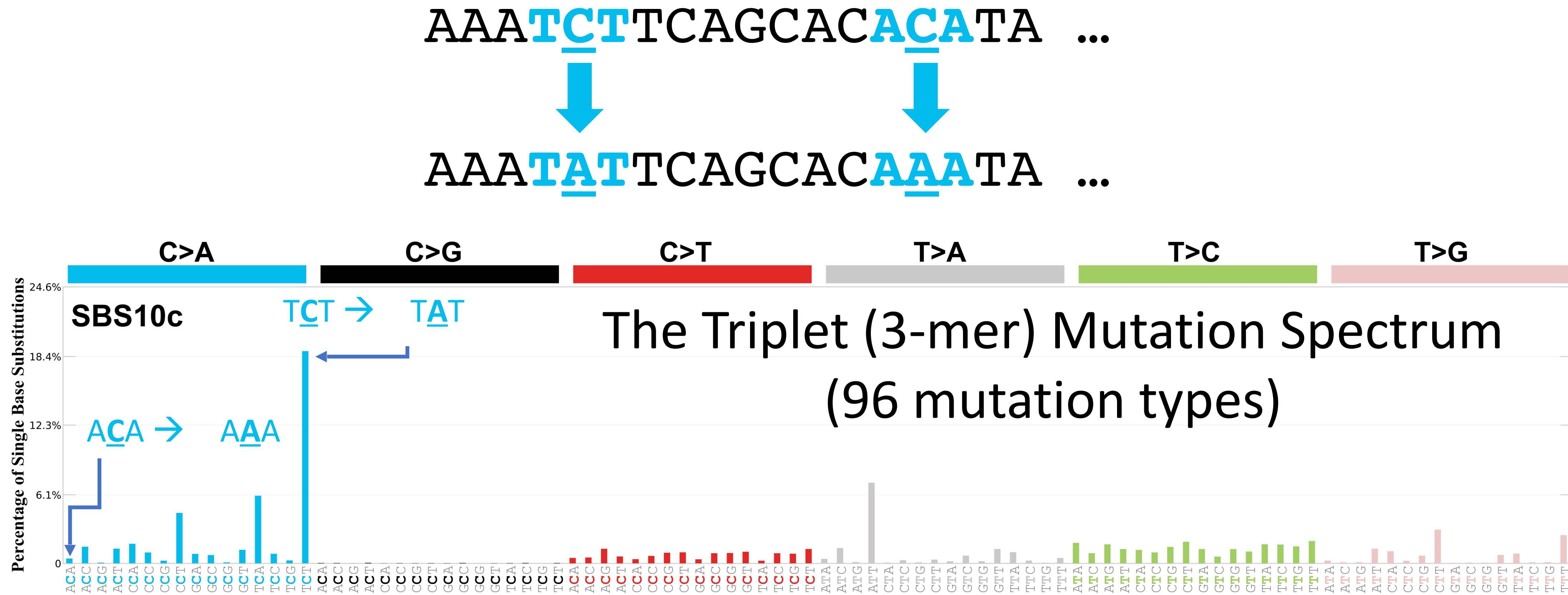
AAAATTTCAGCACTGATA ...  
AAAGTTTCAGTACTGCTA ...



# Exchanging a 1-dimensional mutation rate distribution for a higher-dimensional mutation spectrum distribution

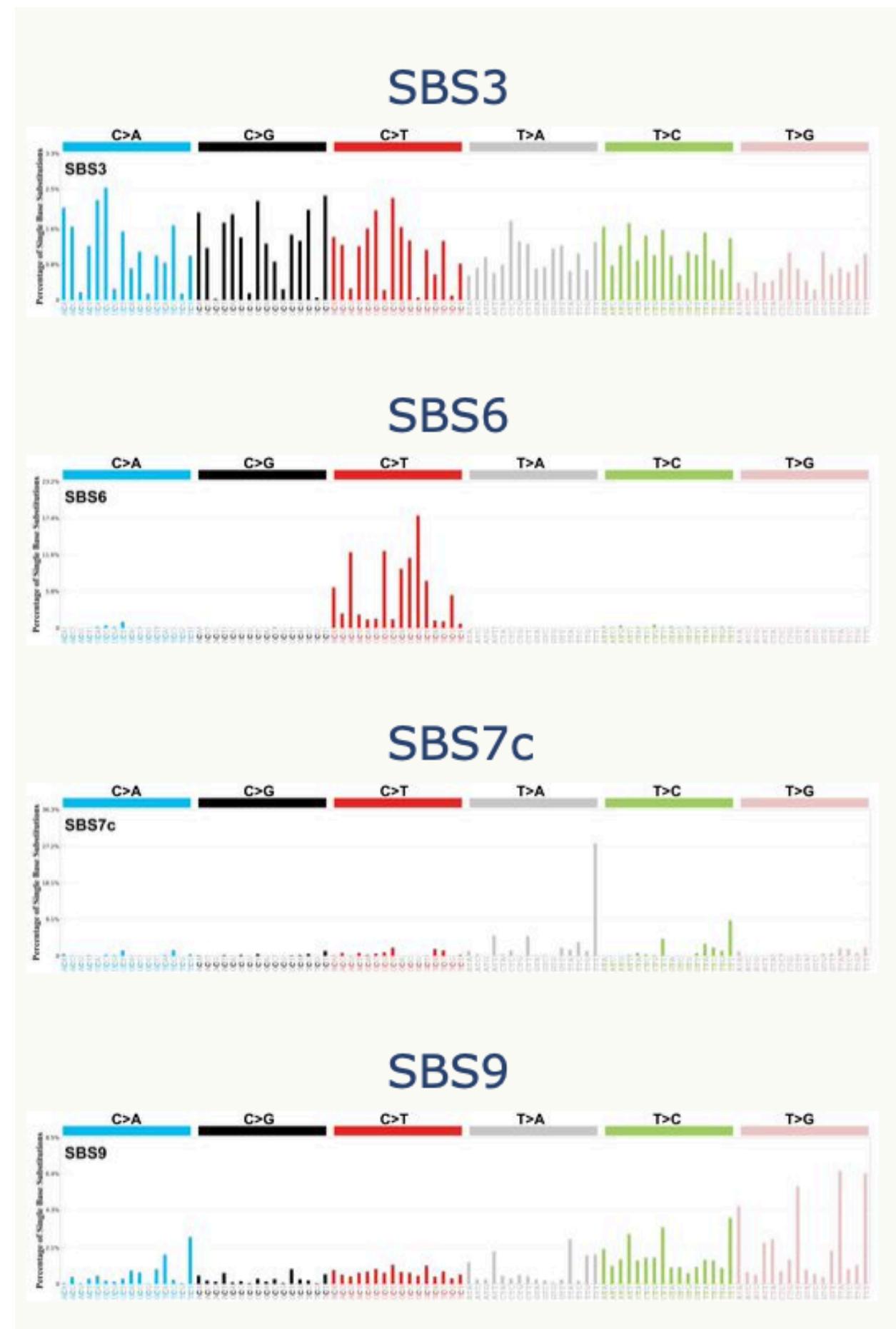


# A context-dependent mutation spectrum can resolve specific DNA repair defects

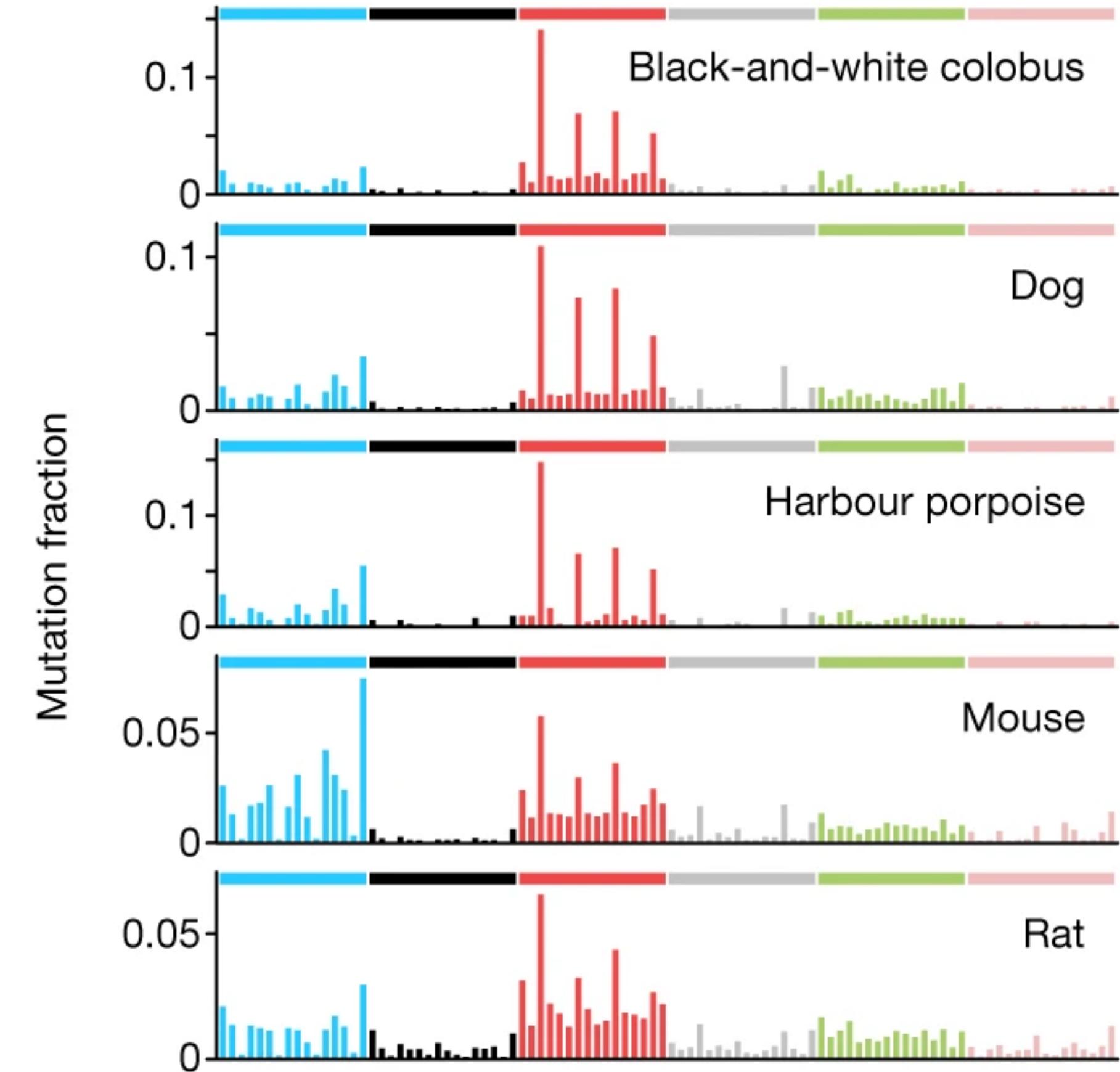


Mutation signature associated with defective *POLD1* proofreading in cancer

# Both pathological and normal tissues exhibit mutation spectrum variation

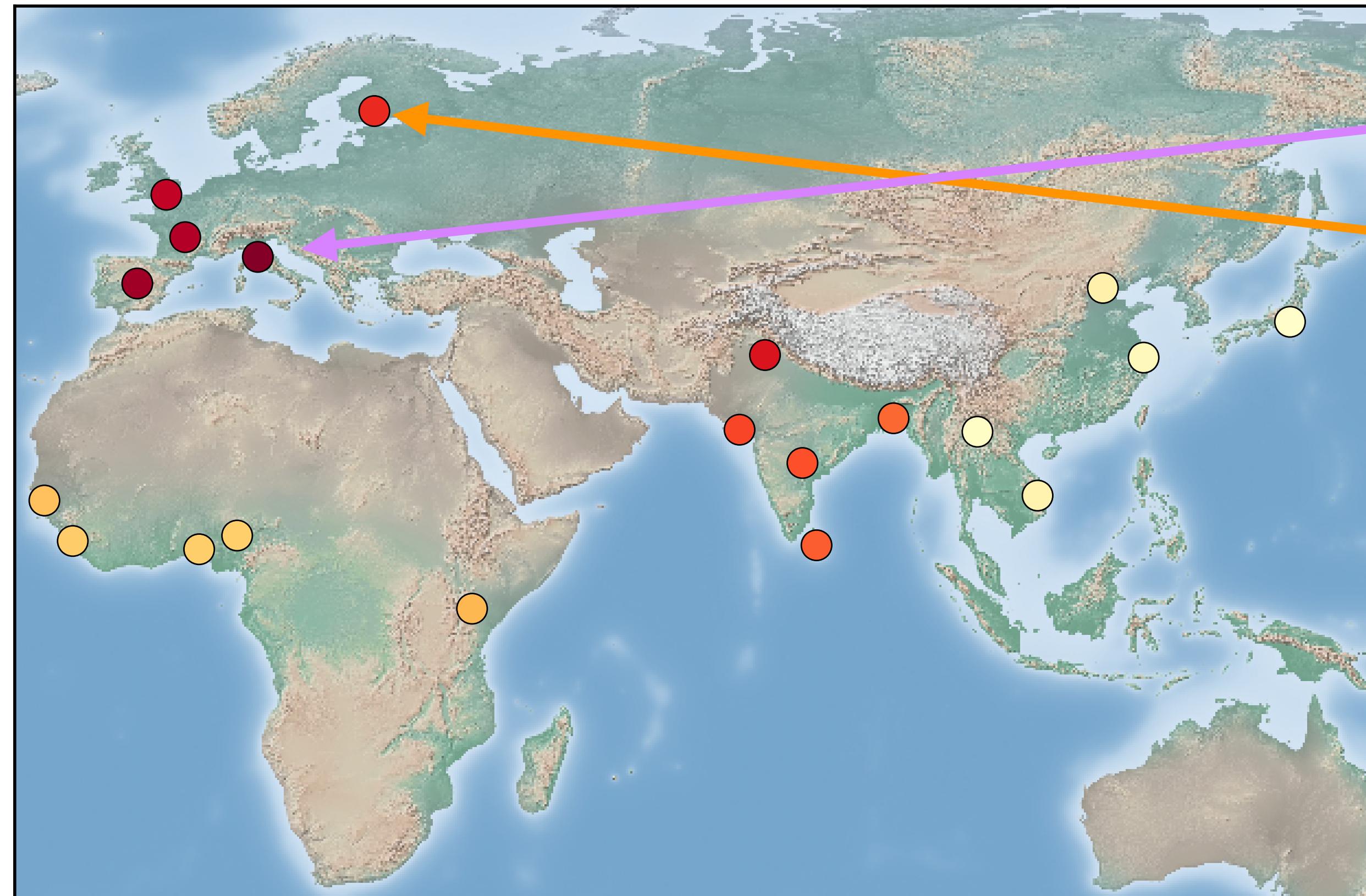


Pathological components of tumor spectra  
(COSMIC database of mutational signatures)

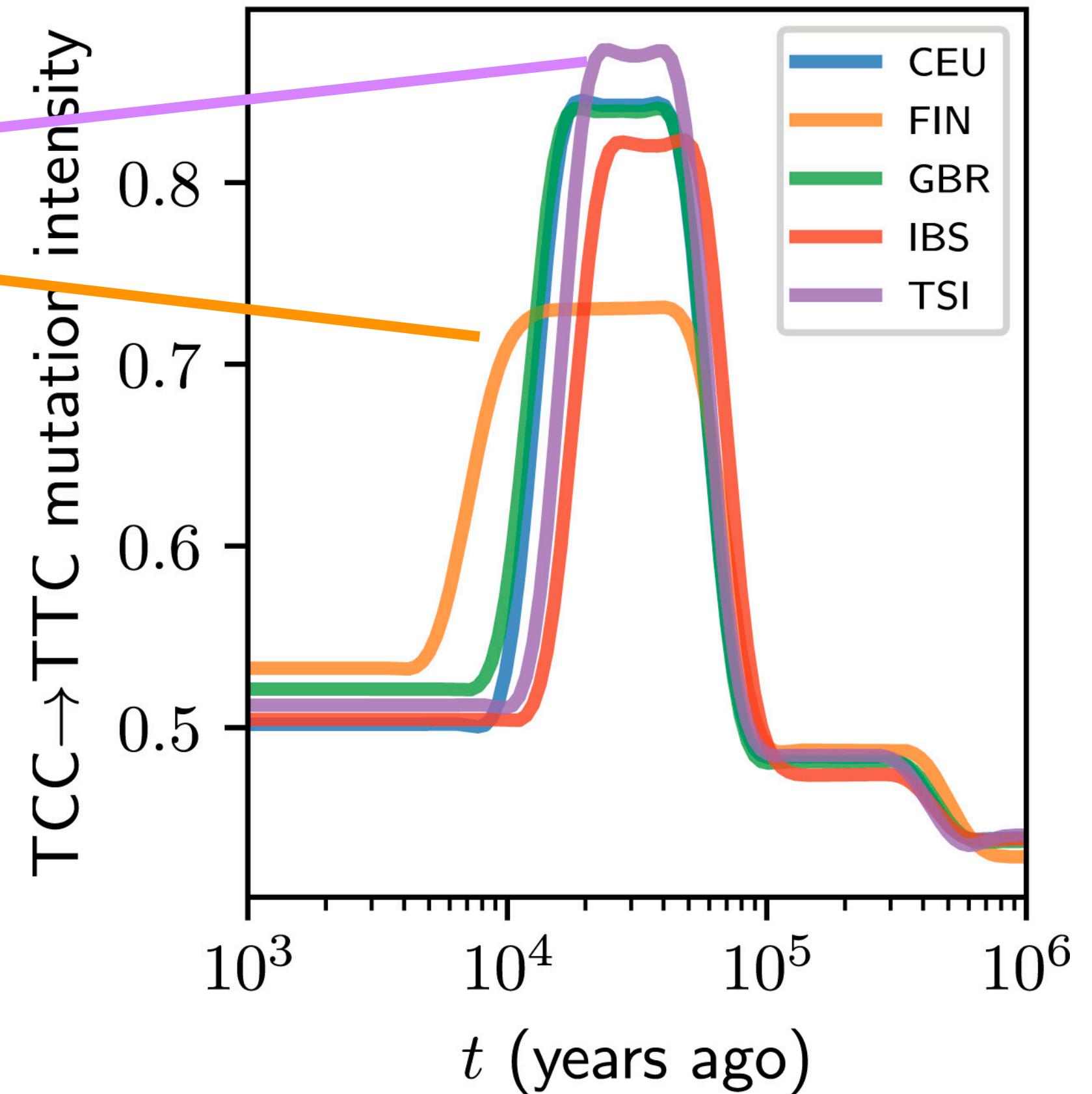


Somatic mutations in normal intestinal crypts  
(Cagan, et al. 2022)

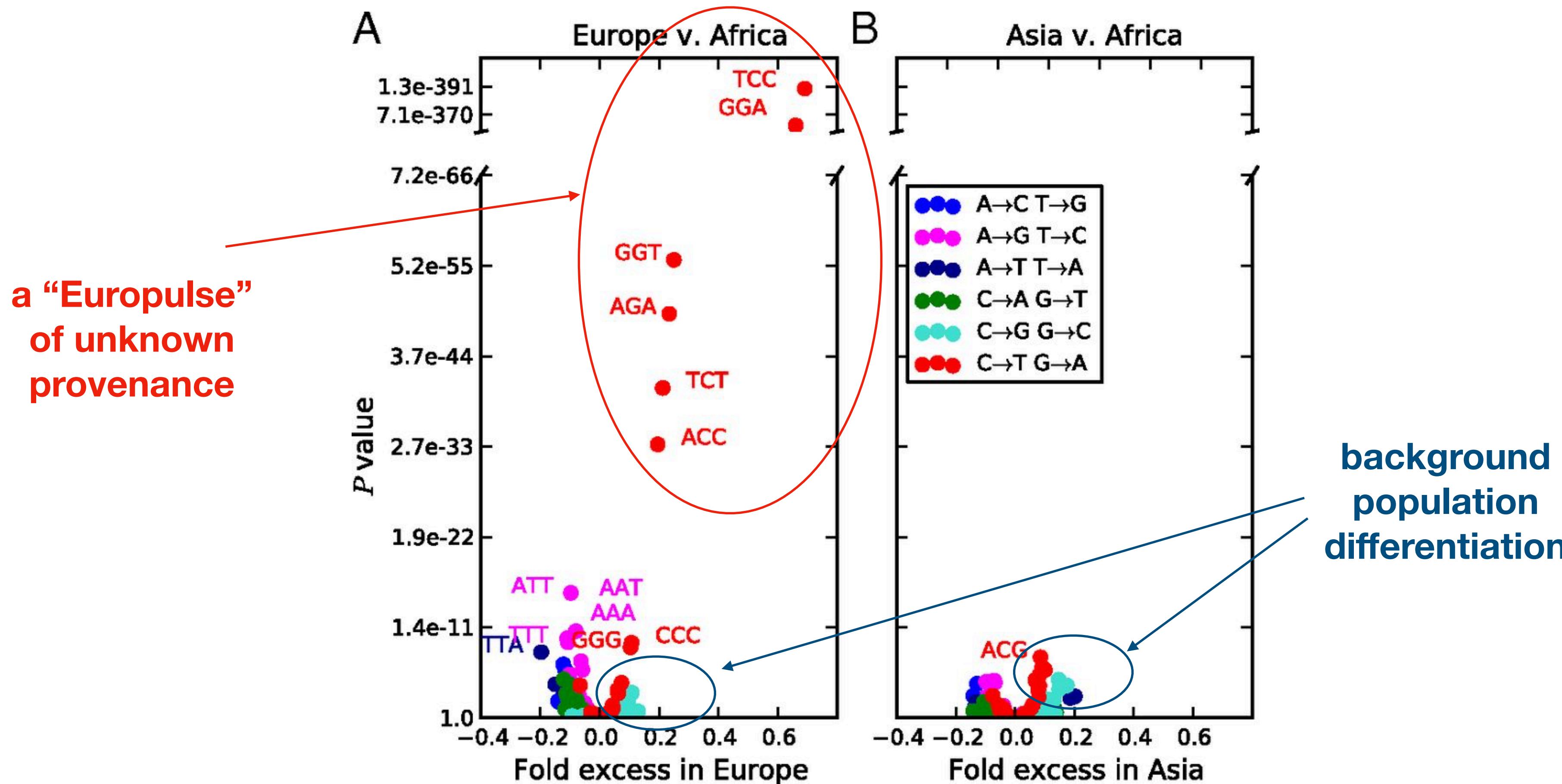
# Normal human genetic variation in Europe and South Asia is enriched for a mutational signature of unknown etiology



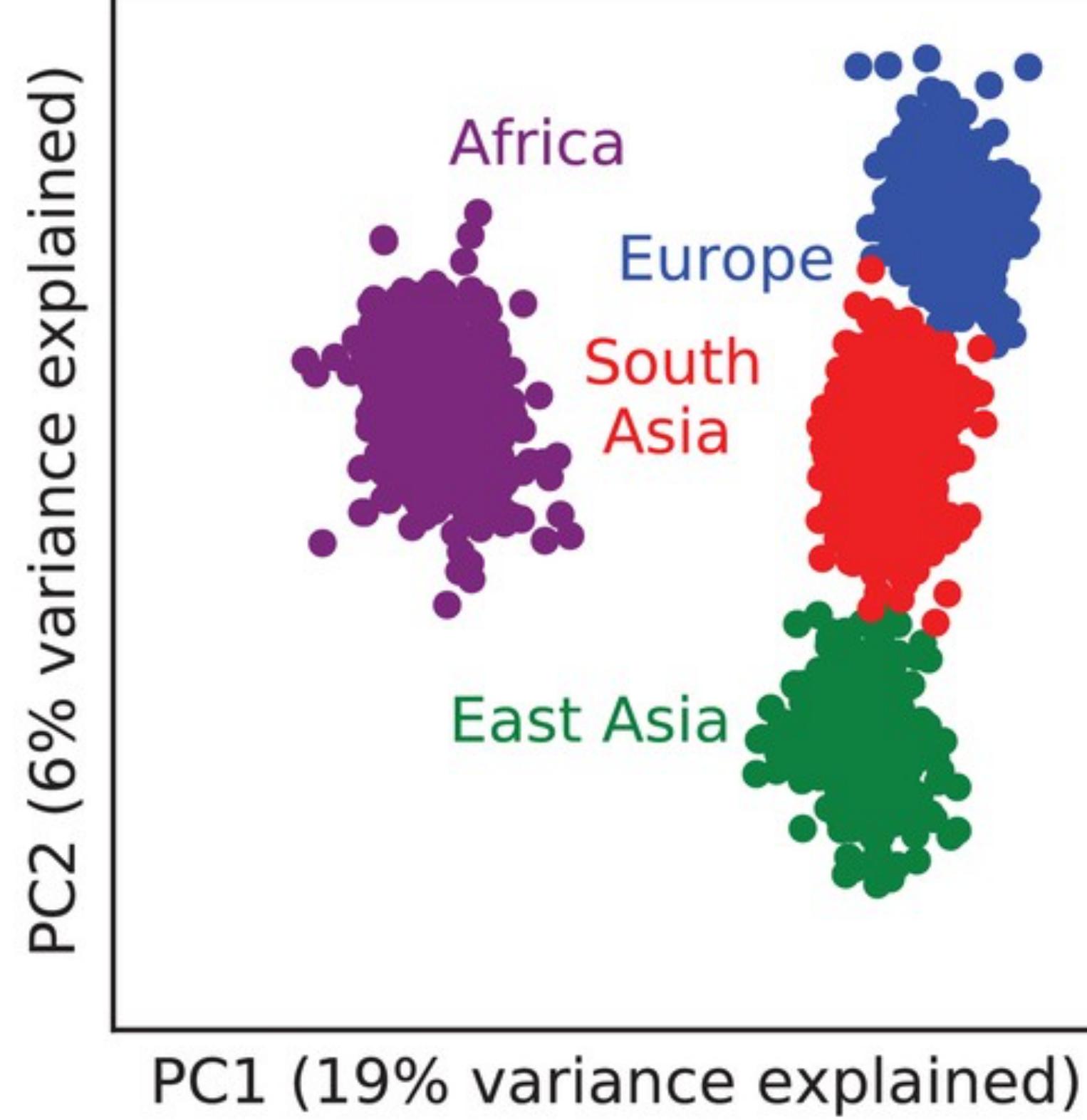
0.017      0.018      0.019  
TCC $\rightarrow$ TTC Mutation Fraction



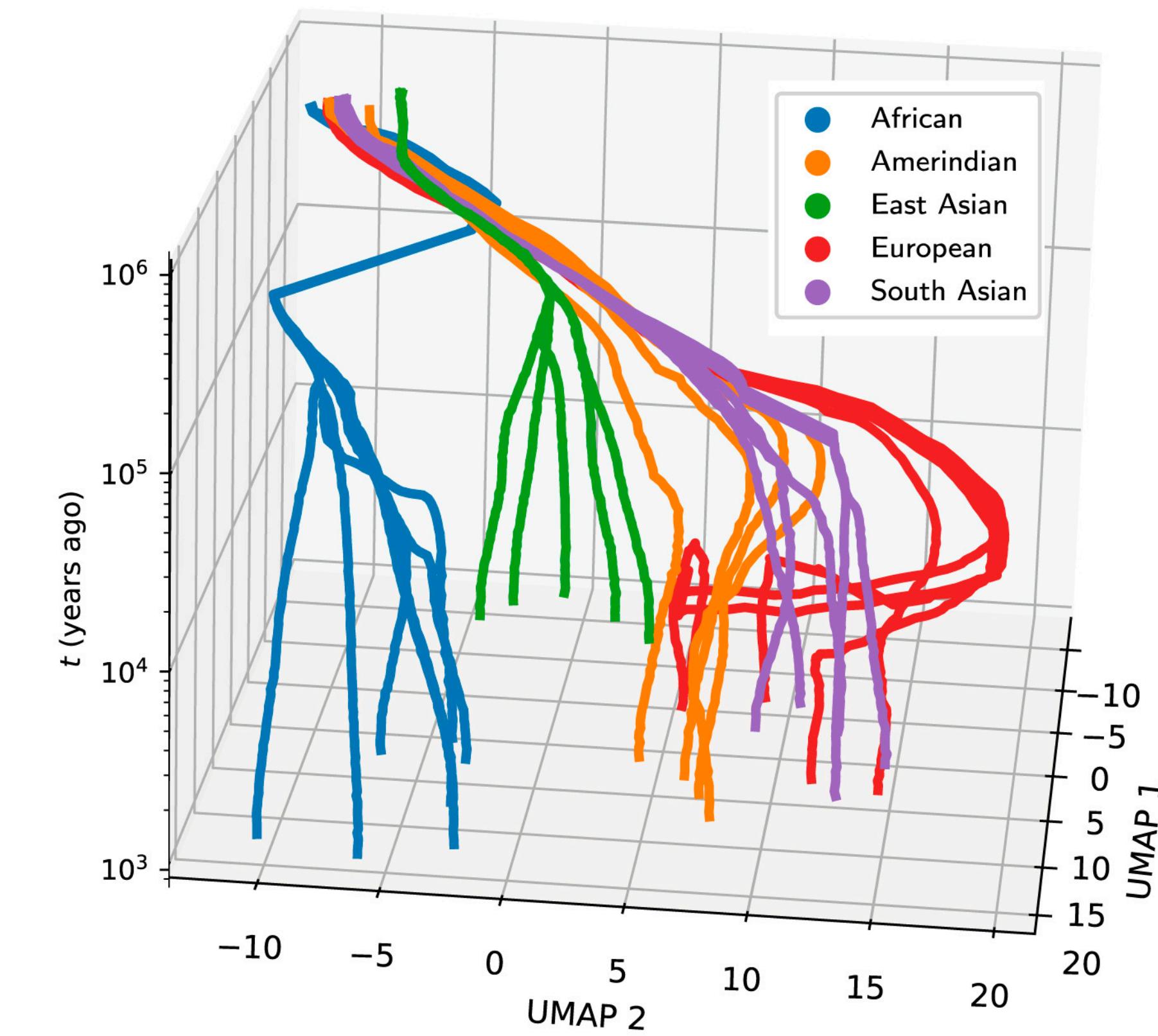
# Other mutation types differ more subtly in abundance between human populations



# Background differentiation explains the majority of human mutation spectrum variation

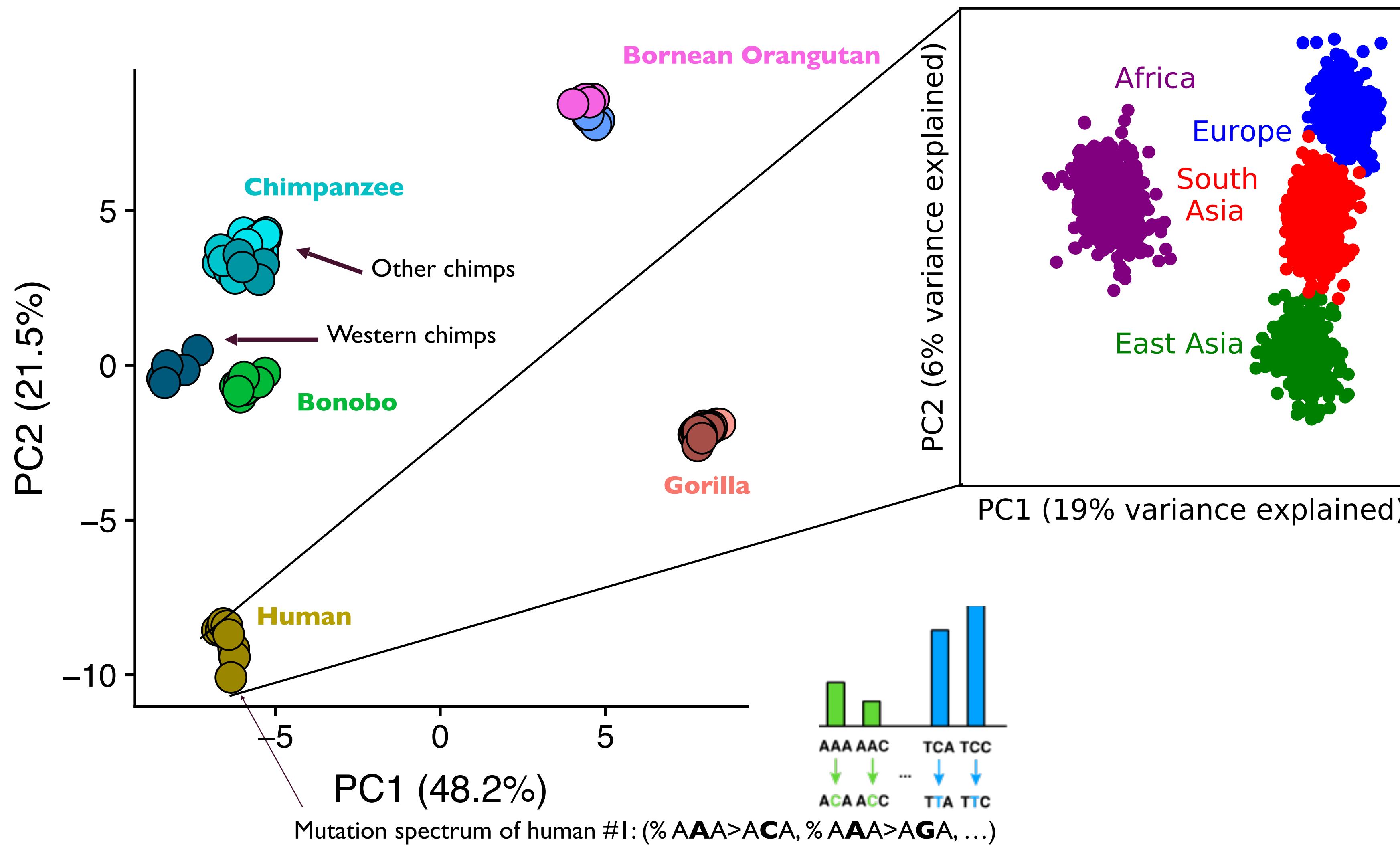


Harris and Pritchard 2017



Dewitt, et al. 2021

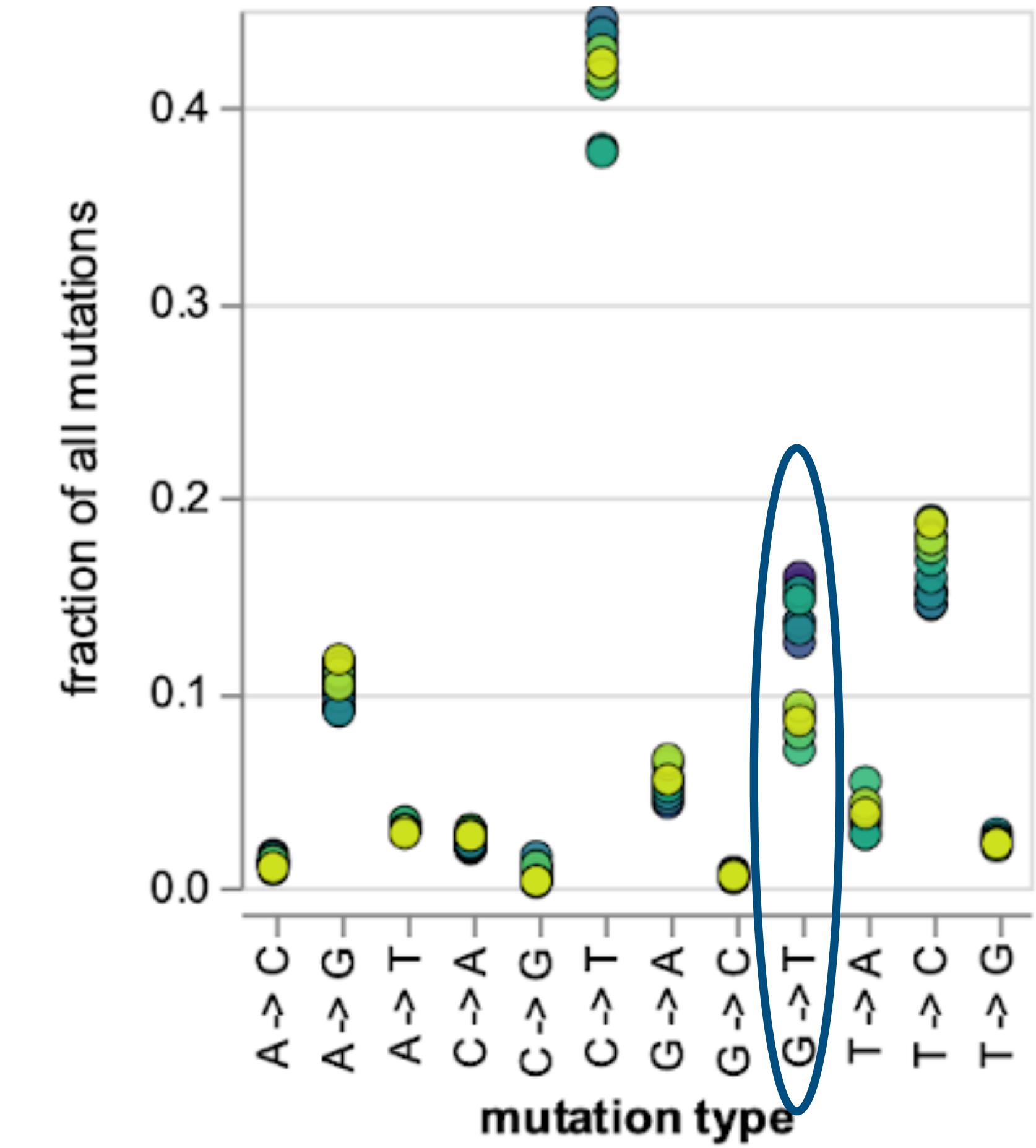
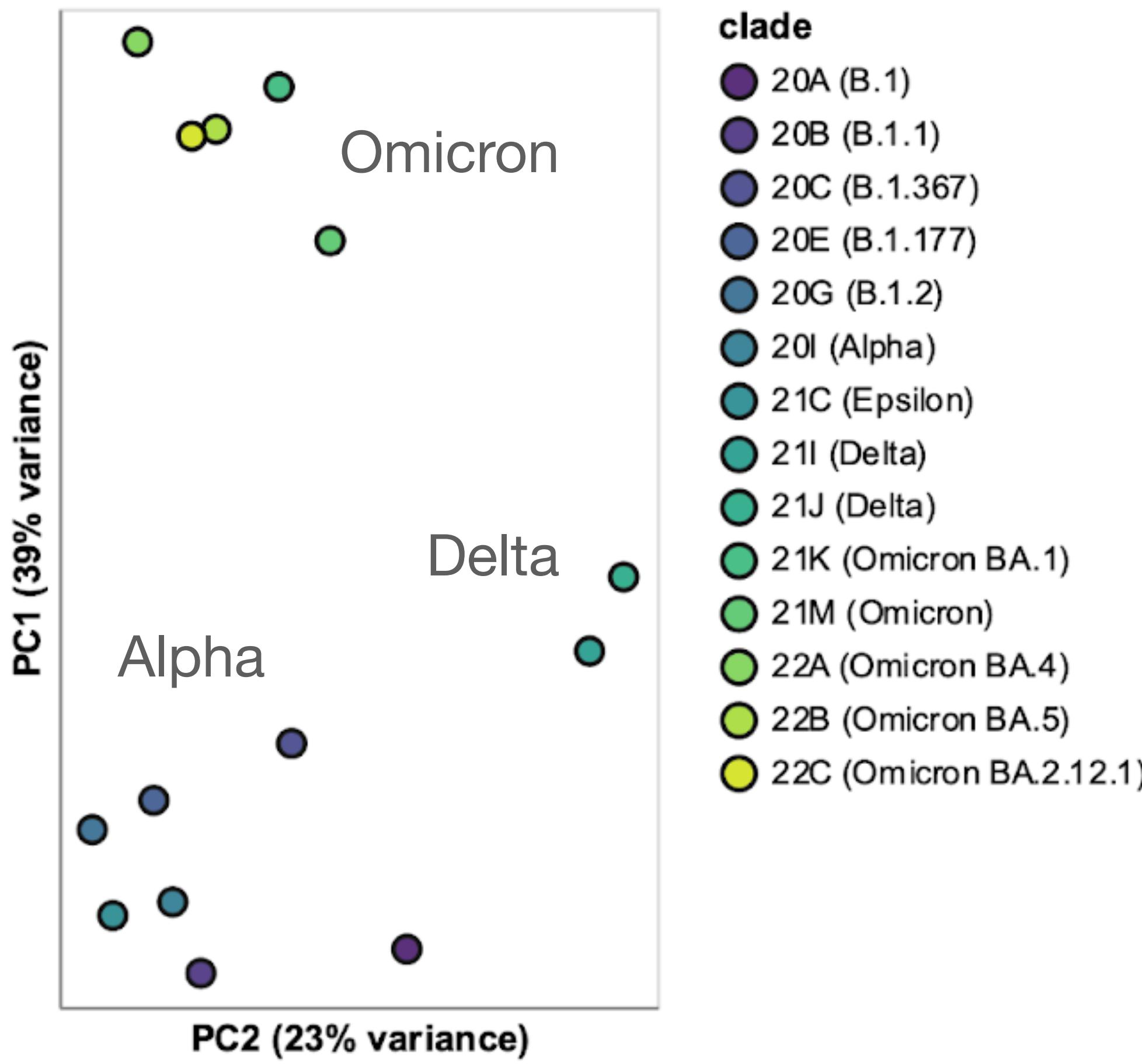
# All human populations and great ape species have distinctive mutation spectra



# Mutation spectrum divergence during 3 years of SARS-CoV-2 evolution



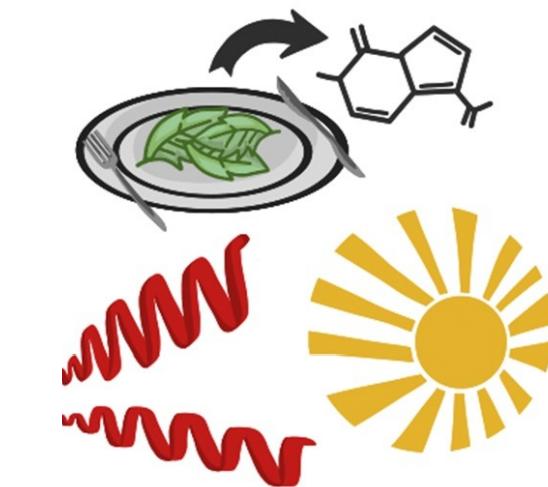
Jesse Bloom



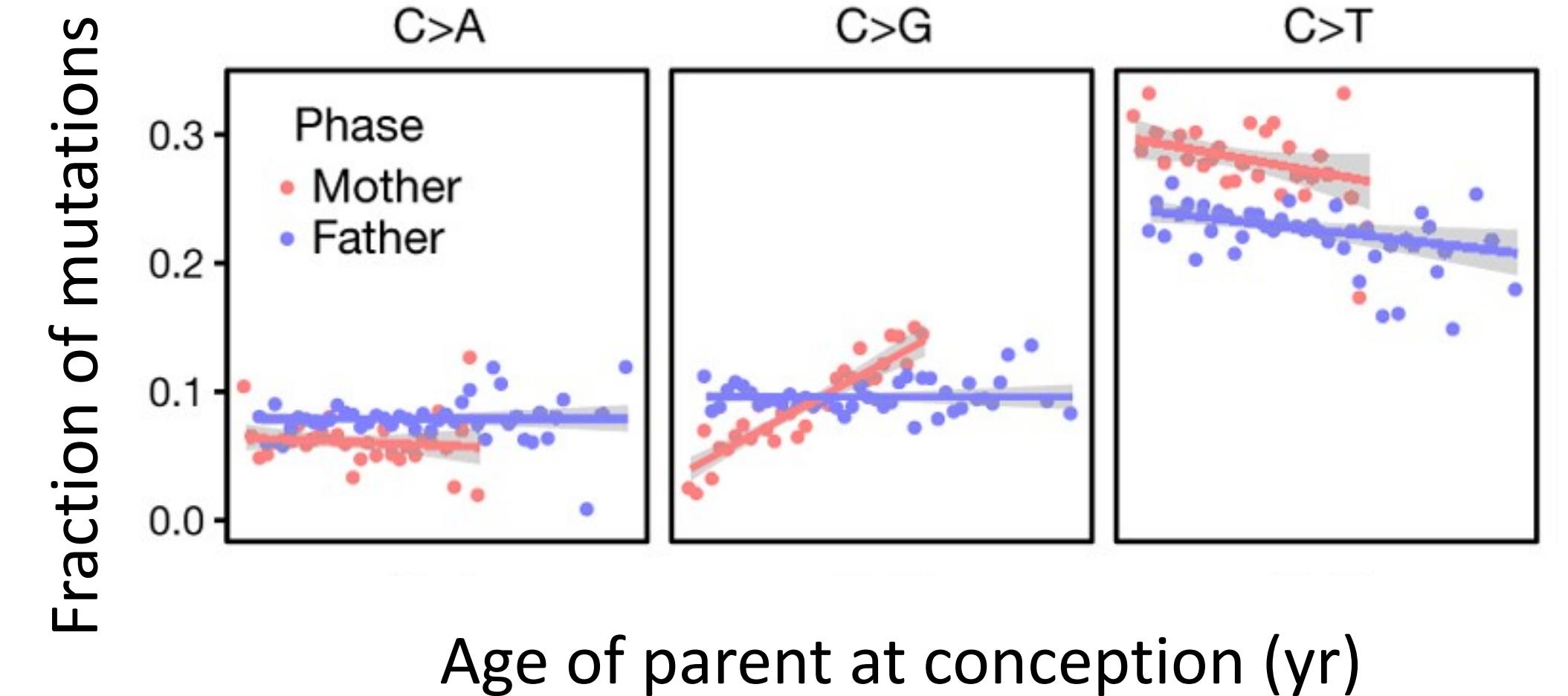
- Most populations and species have their own distinctive mutation spectrum phenotypes
- What might be driving these differences?



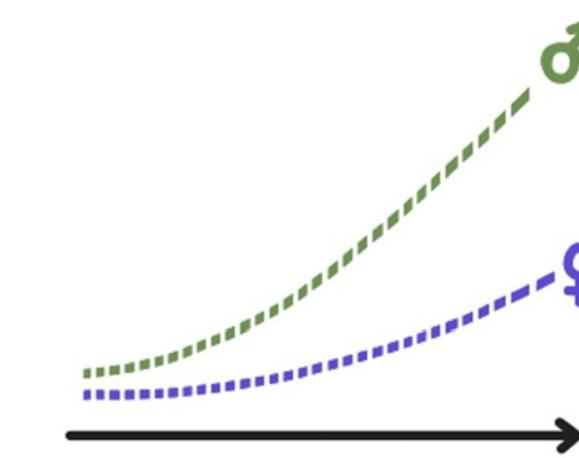
genetics?



environment?



Modified from Jónsson *et al. Nature*  
(2017), Figures 1 & 2, Annabel Beichman



reproductive  
age?

# Genome-wide Association Studies are currently underpowered to detect mutator alleles

Cost of one direct mutation rate measurement: ~\$3,000



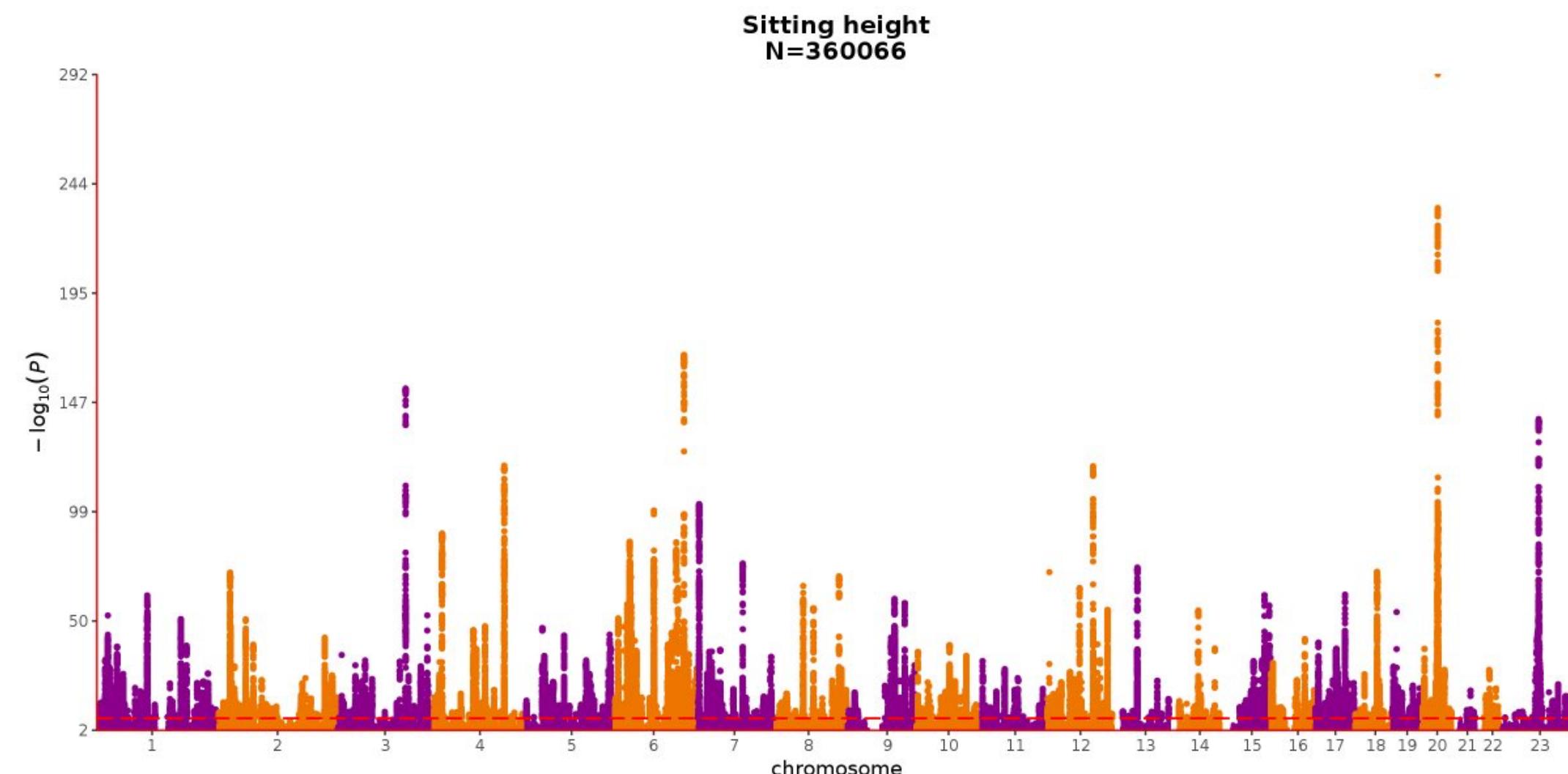
[romper.com](https://www.romper.com)

GWAS of height, cholesterol, etc: ~\$100 genotype chip plus phenotype measurement

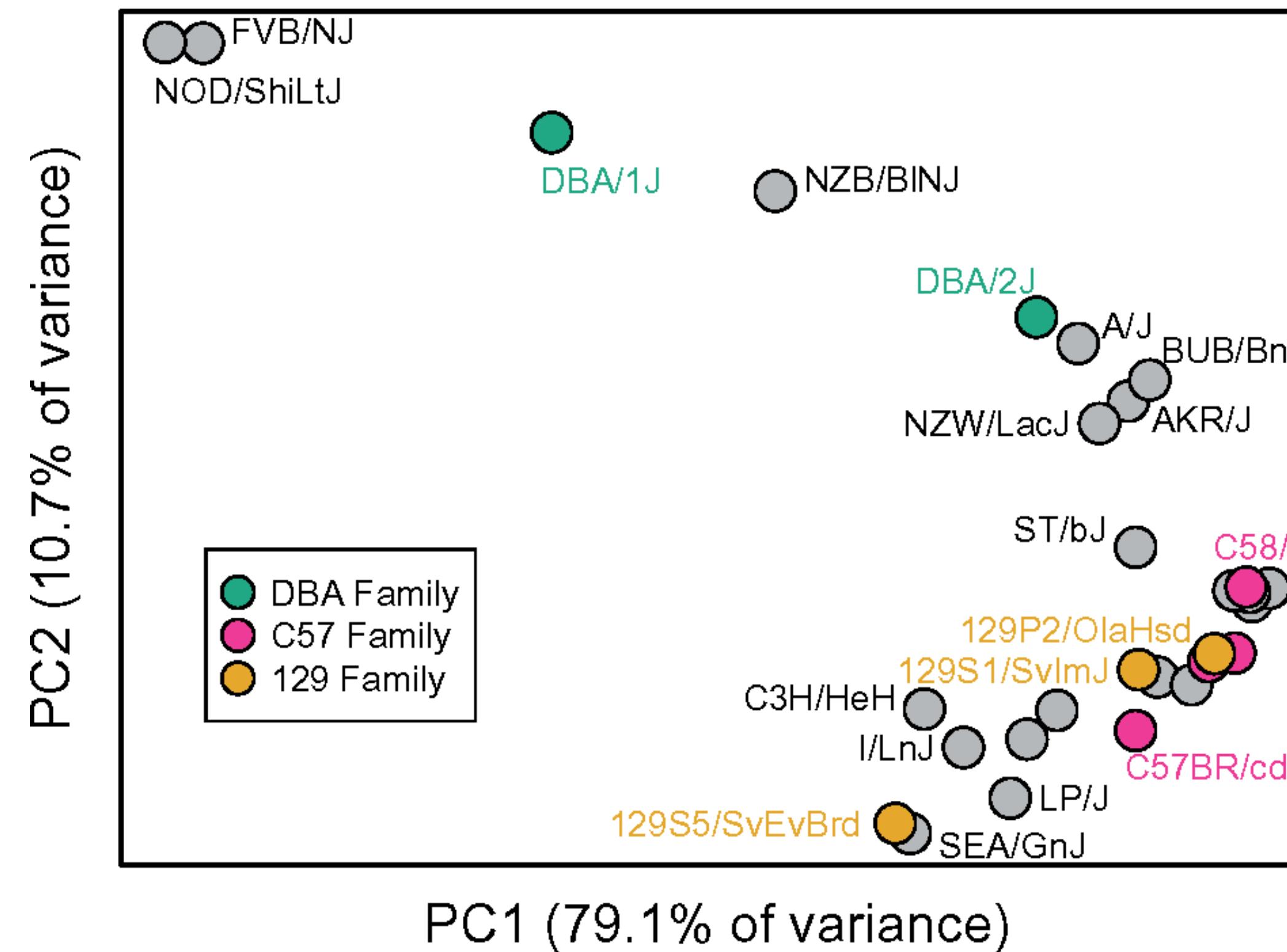
1,465 parent/child trios → zero mutation rate heritability detected

(Kessler, et al. 2020)

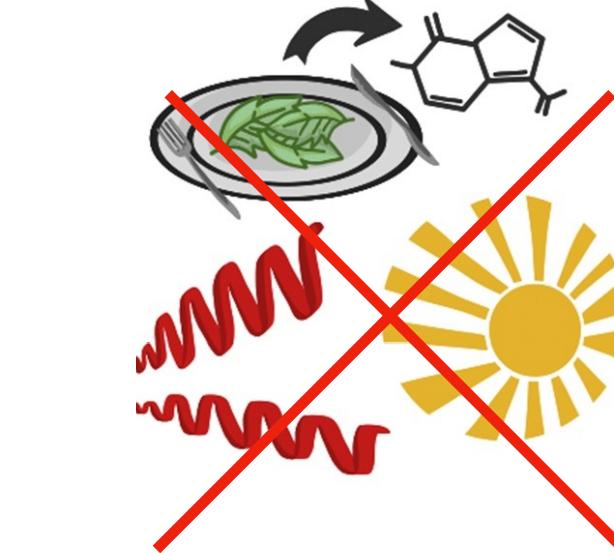
State of the art height, cholesterol GWAS include 500,000 individuals



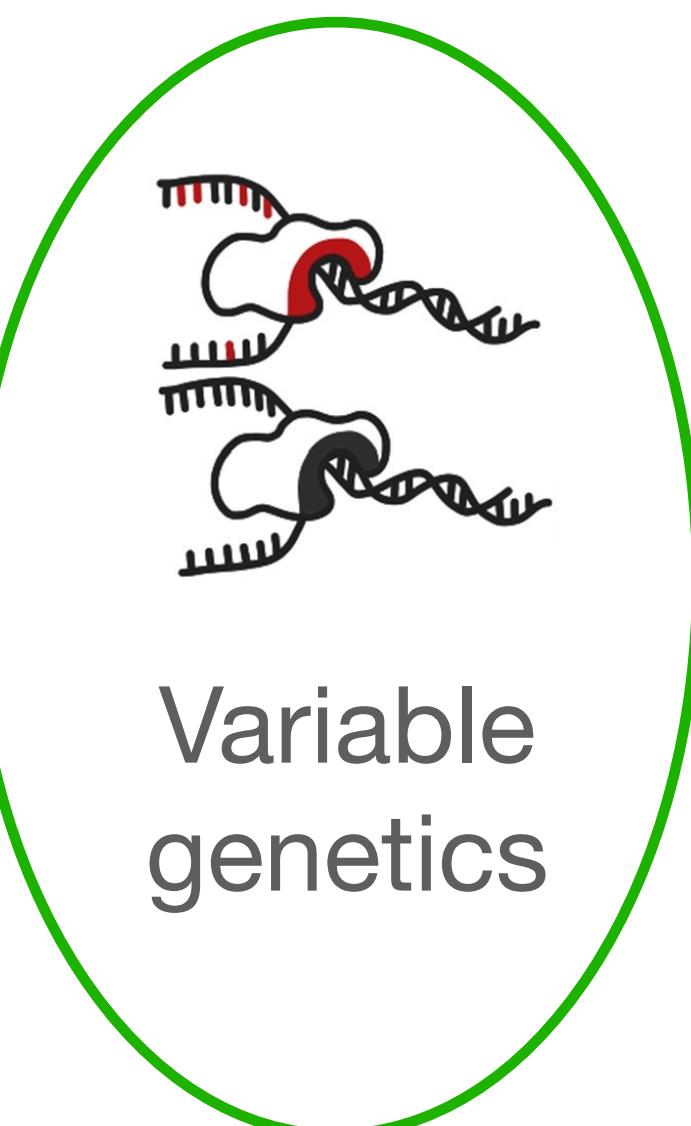
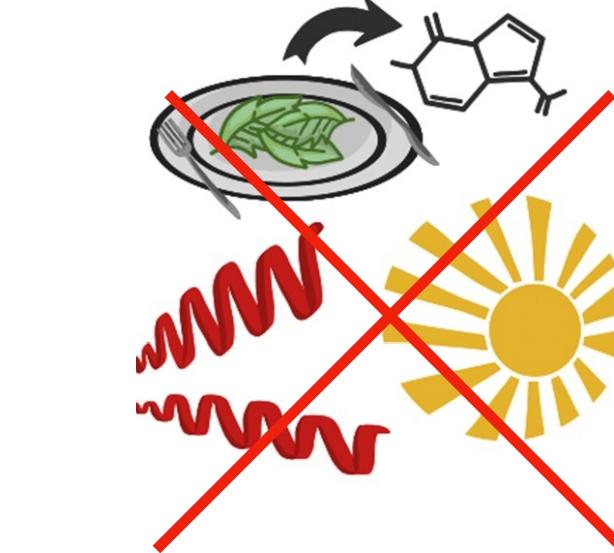
# Mutations in laboratory mice point toward genetic control of mutation spectra



Same reproductive history



Same environment



Variable genetics



Dumont 2019

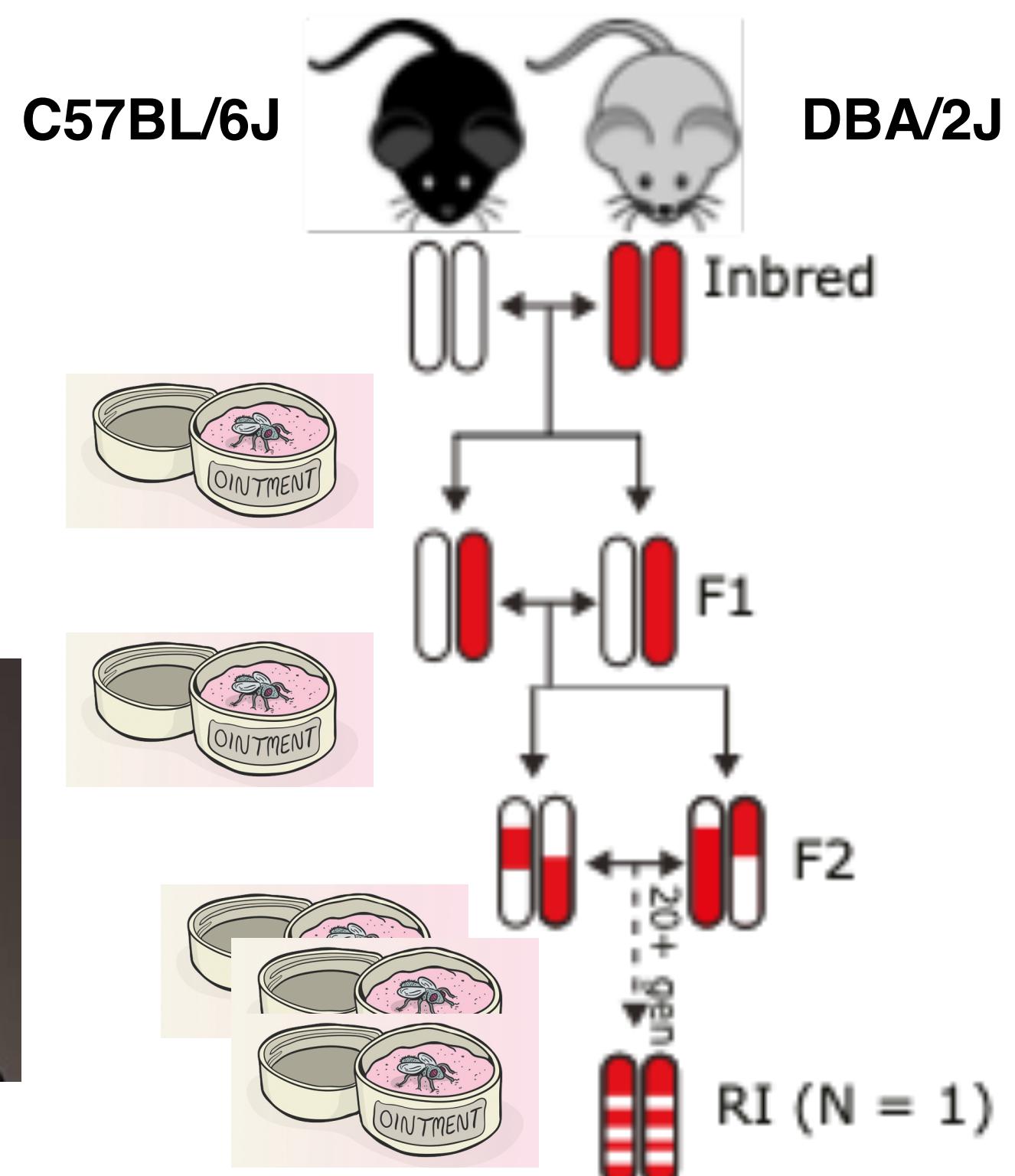
# The BXD mice: an exceptionally large, old collection of recombinant inbred lines (RIL)



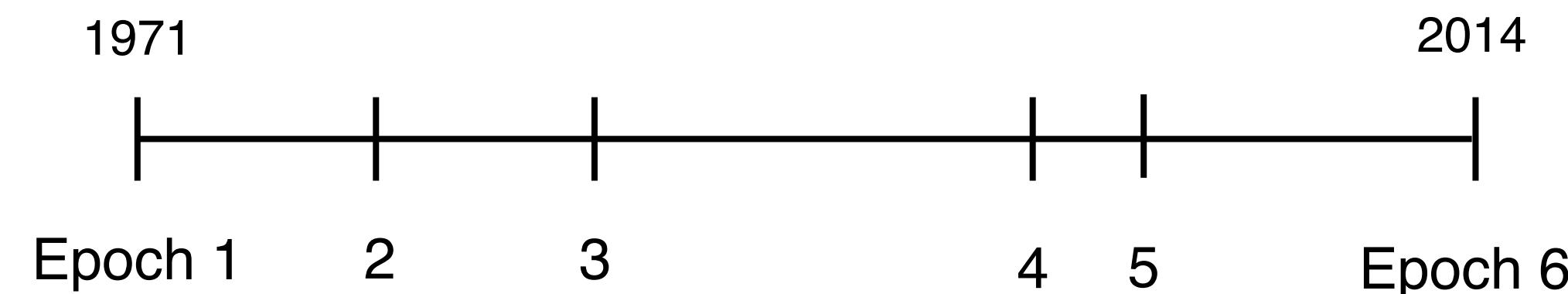
**David Ashbrook**  
University of Tennessee  
Health Science Center



**Jonathan Pritchard**  
Stanford



Taylor, et al. 1973  
Ashbrook, et al. 2019



151 recombinant inbred BXD lines across 6 epochs, with ~60 recombinations per genome

Oldest lines propagated by brother/sister mating  
for up to 188 generations

Used to QTL-map the architecture of complex  
traits like drug and alcohol dependence



**Tom Sasani**

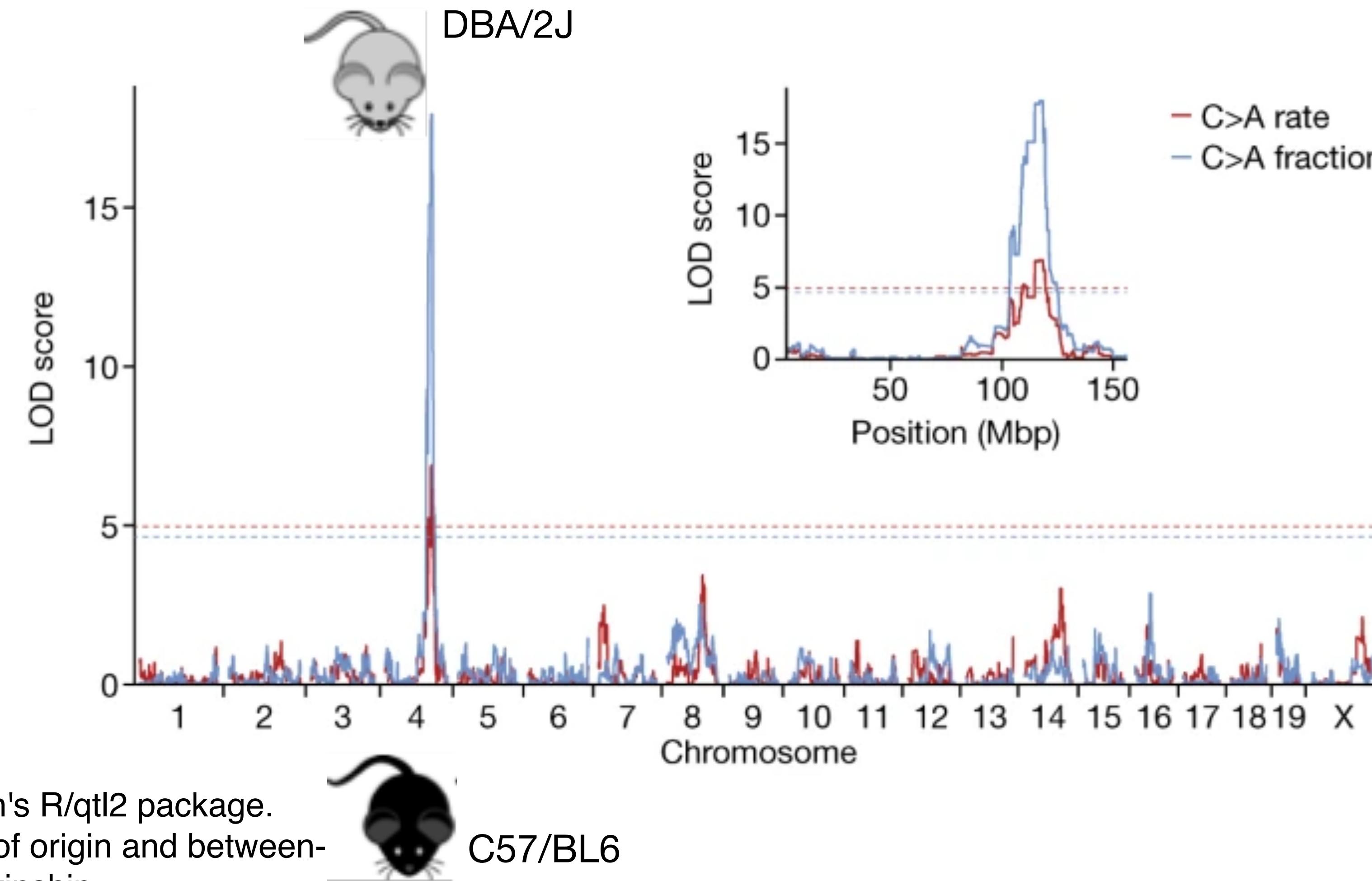


**Rob Williams**  
University of Tennessee  
Health Science Center



**Abraham Palmer**  
UCSD

# DBA/2J ancestry at the QTL is associated with a higher C>A mutation fraction genomewide



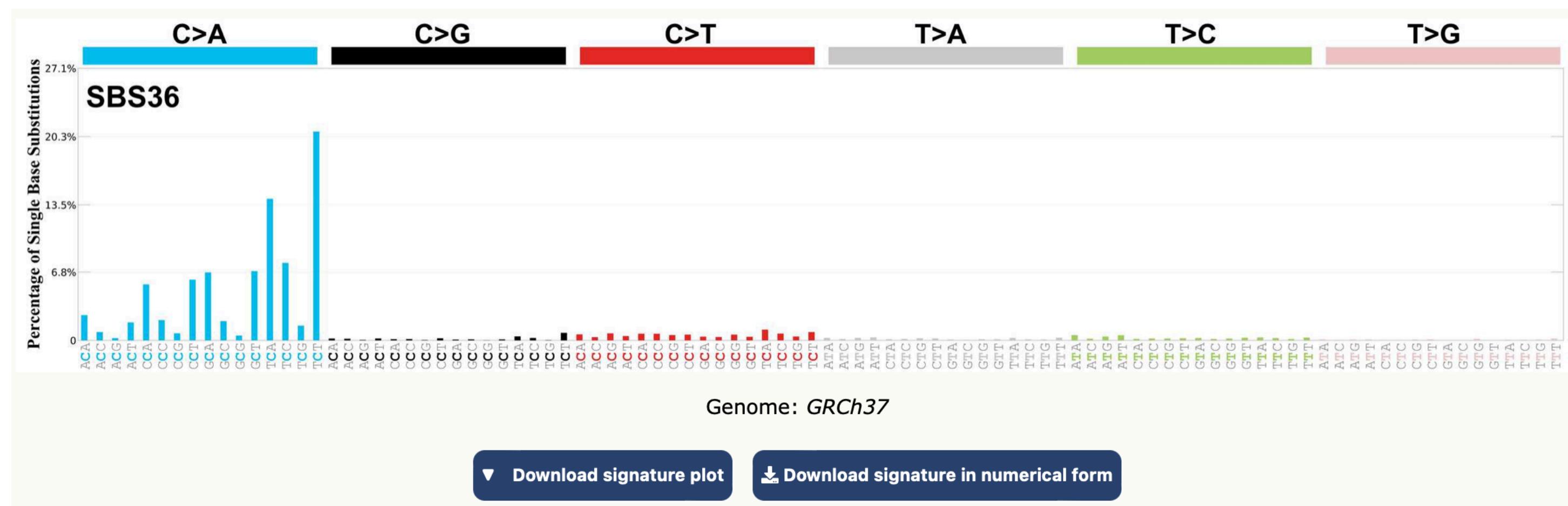
Using Karl Broman's R/qtl2 package.  
Controlling for epoch of origin and between-strain kinship.



**Tom Sasani**

Sasani, et al., *Nature* 2022

# Mice harbor a mutator allele in the DNA repair gene *Mutyh*, which causes heritable colorectal cancers in humans



Mutational profile using the conventional 96 mutation type classification. This classification is based on the six substitution subtypes: C>A, C>G, C>T, T>A, T>C, and T>G, as well as the nucleotides immediately 5' and 3' to the mutation.

[Help](#)

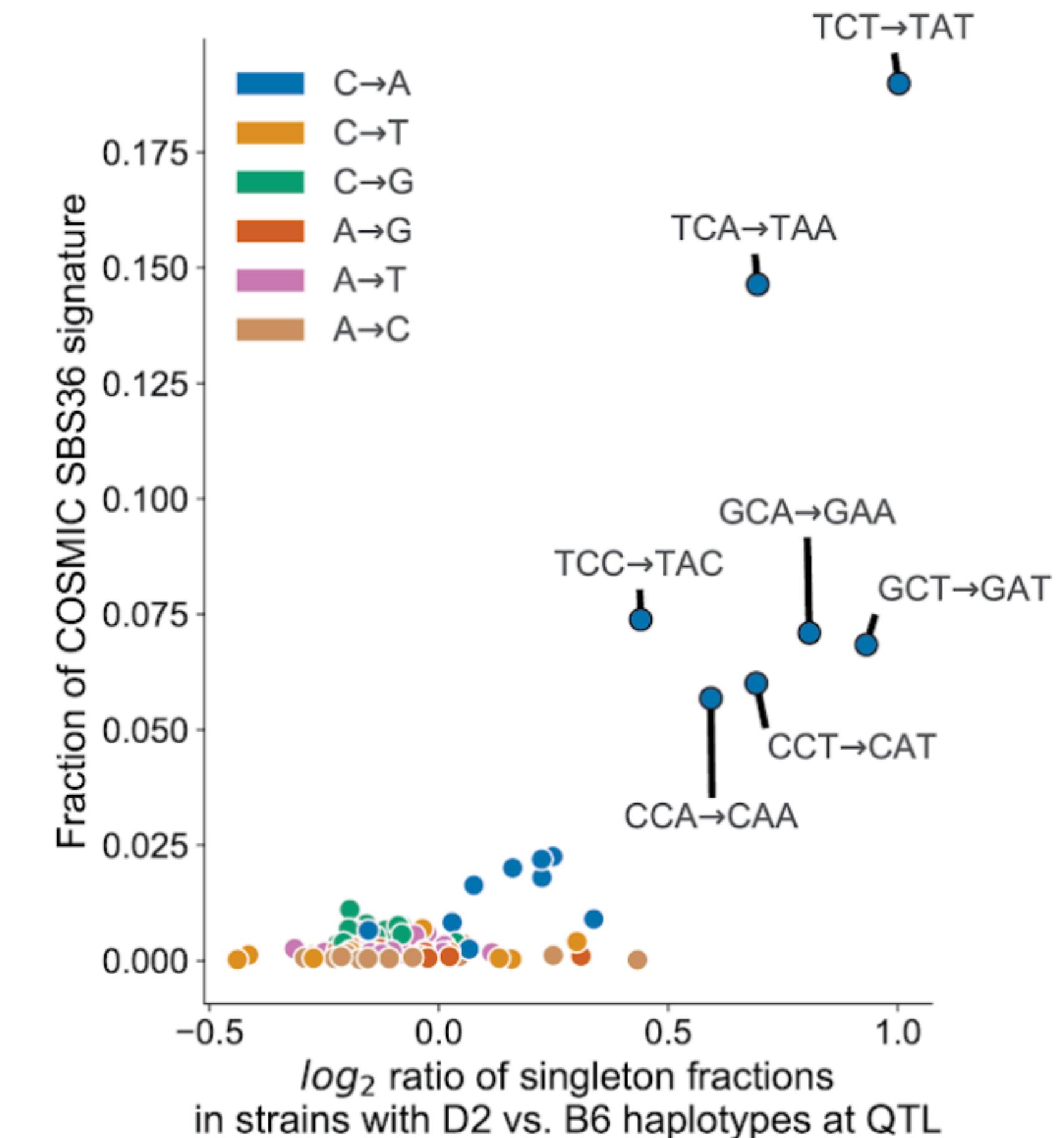
## Proposed aetiology

Defective base excision repair, including DNA damage due to reactive oxygen species, due to biallelic germline or somatic *MUTYH* mutations.



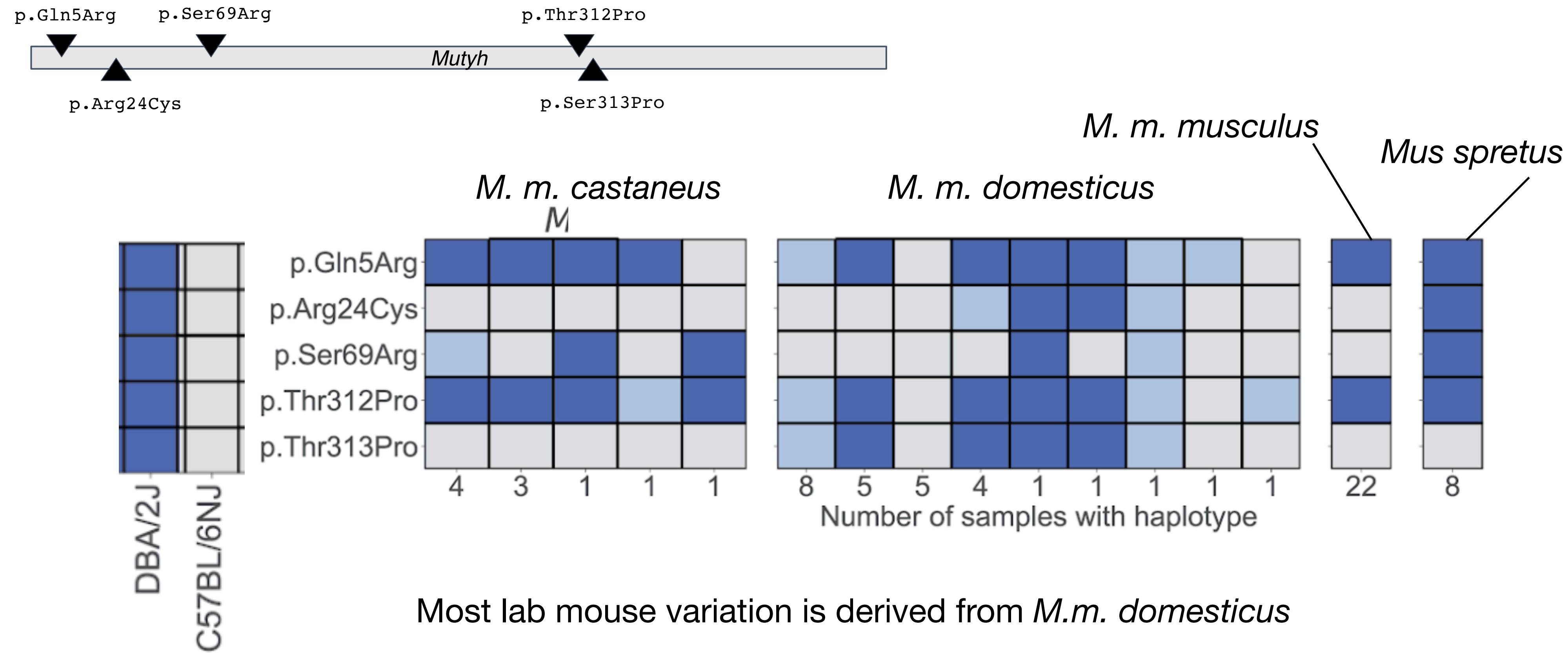
Projects ▾ Data ▾ Tools ▾ News ▾ Help ▾ About ▾ Search COSMIC

Mutational Signatures (v3.3 - June 2022)

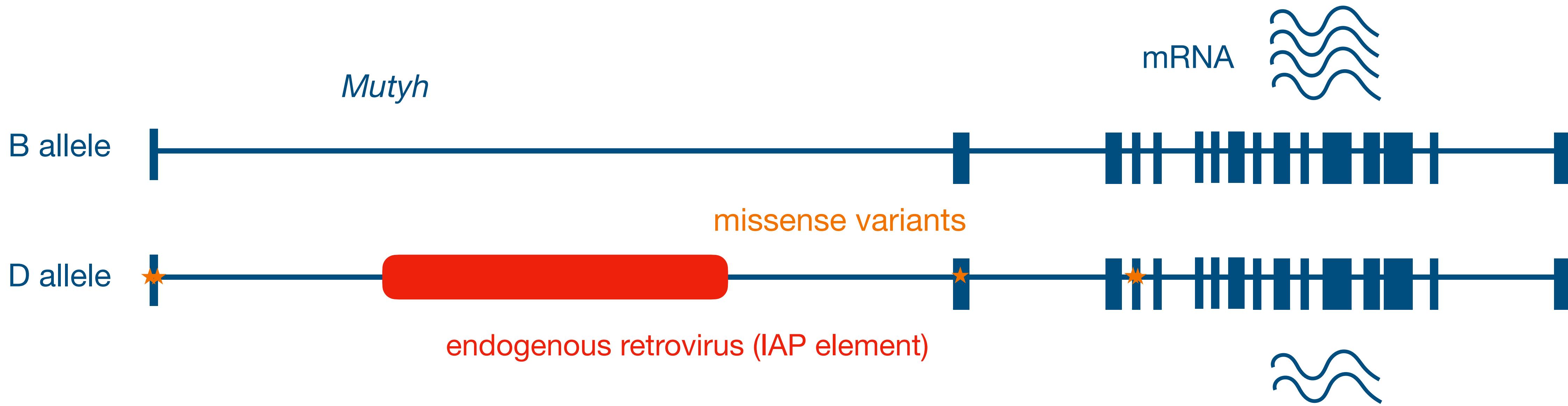


Al-Tassan, et al. 2002 Pilati, et al. 2017 Sasani, et al., *Nature* 2022

# C57BL/6 and DBA/2J differ at 5 non-synonymous *Mutyh* sites that segregate in wild mice



# An ERV appears to suppress *Mutyh* expression in DBA/2J



**Resolution of structural variation in diverse mouse genomes reveals chromatin remodeling due to transposable elements**

Ardian Ferraj<sup>1,2</sup>, Peter A. Audano<sup>2</sup>, Parithi Balachandran<sup>2</sup>, Anne Czechanski<sup>3</sup>, Jacob I. Flores<sup>2</sup>, Alexander A. Radecki<sup>1,2</sup>, Varun Mosur<sup>2</sup>, David S. Gordon<sup>4</sup>, Isha A. Walawalkar<sup>1,2</sup>, Evan E. Eichler<sup>4</sup>, Laura G. Reinholdt<sup>3</sup>, Christine R. Beck<sup>1,2\*</sup>

	<b>Germline mutator</b>	<b>Somatic mutator/ Cancer driver</b>
	 <b>Rare, pathogenic human <i>Mutyh</i> variants</b>	
	 <b>Common mouse <i>Mutyh</i> variants</b>	

# Duplex sequencing of colon and blood from aged BXD mice



Candice Young



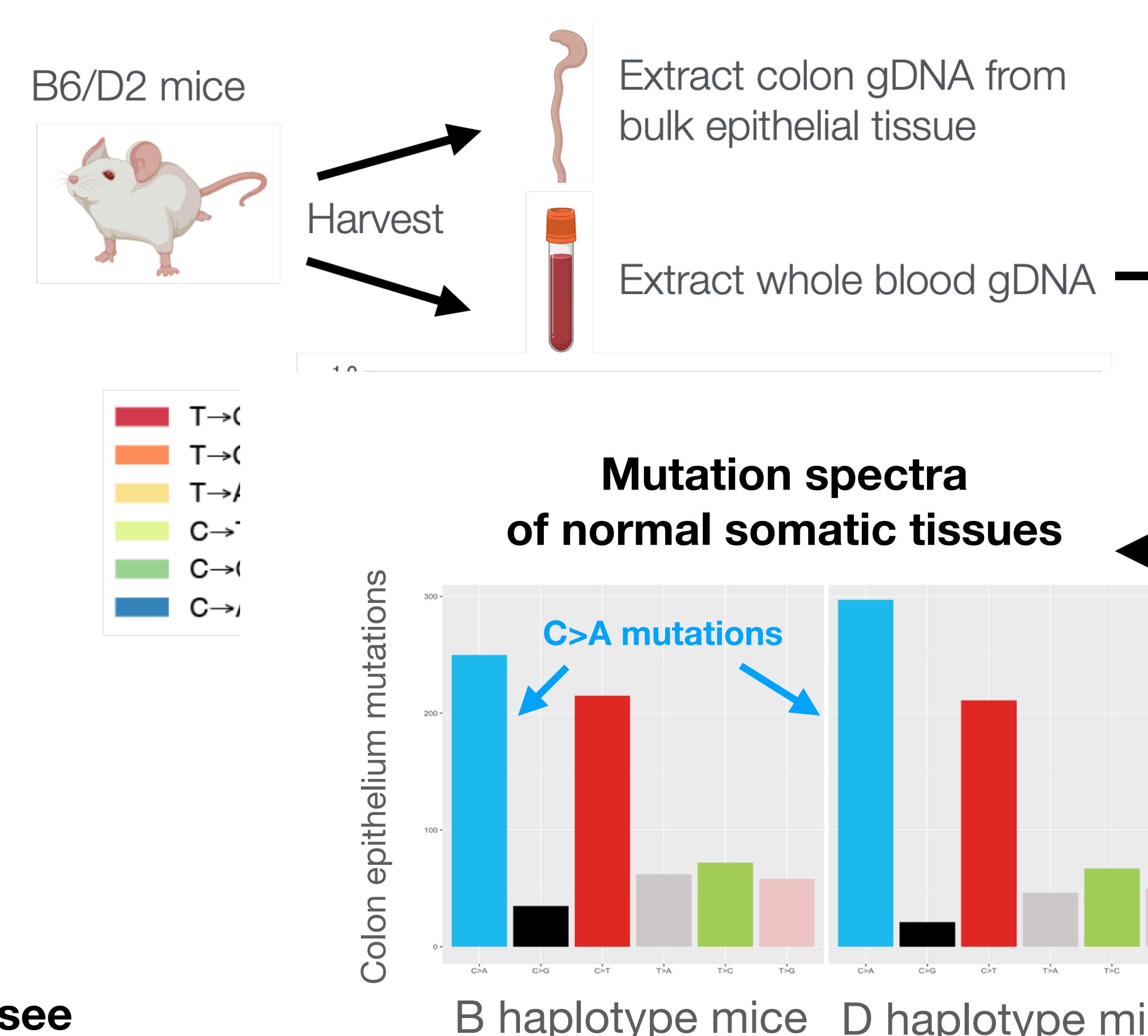
David Ashbrook  
University of Tennessee



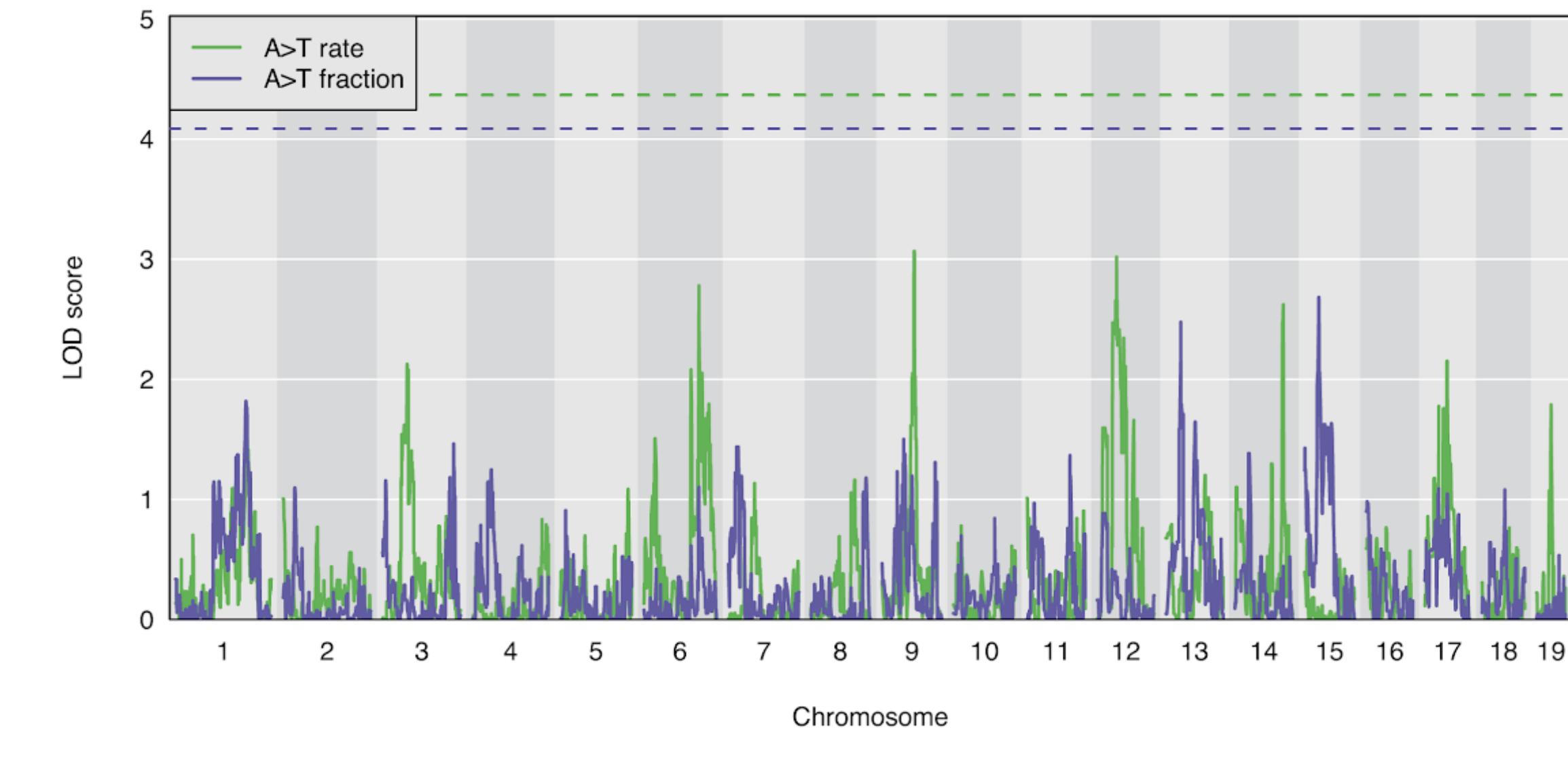
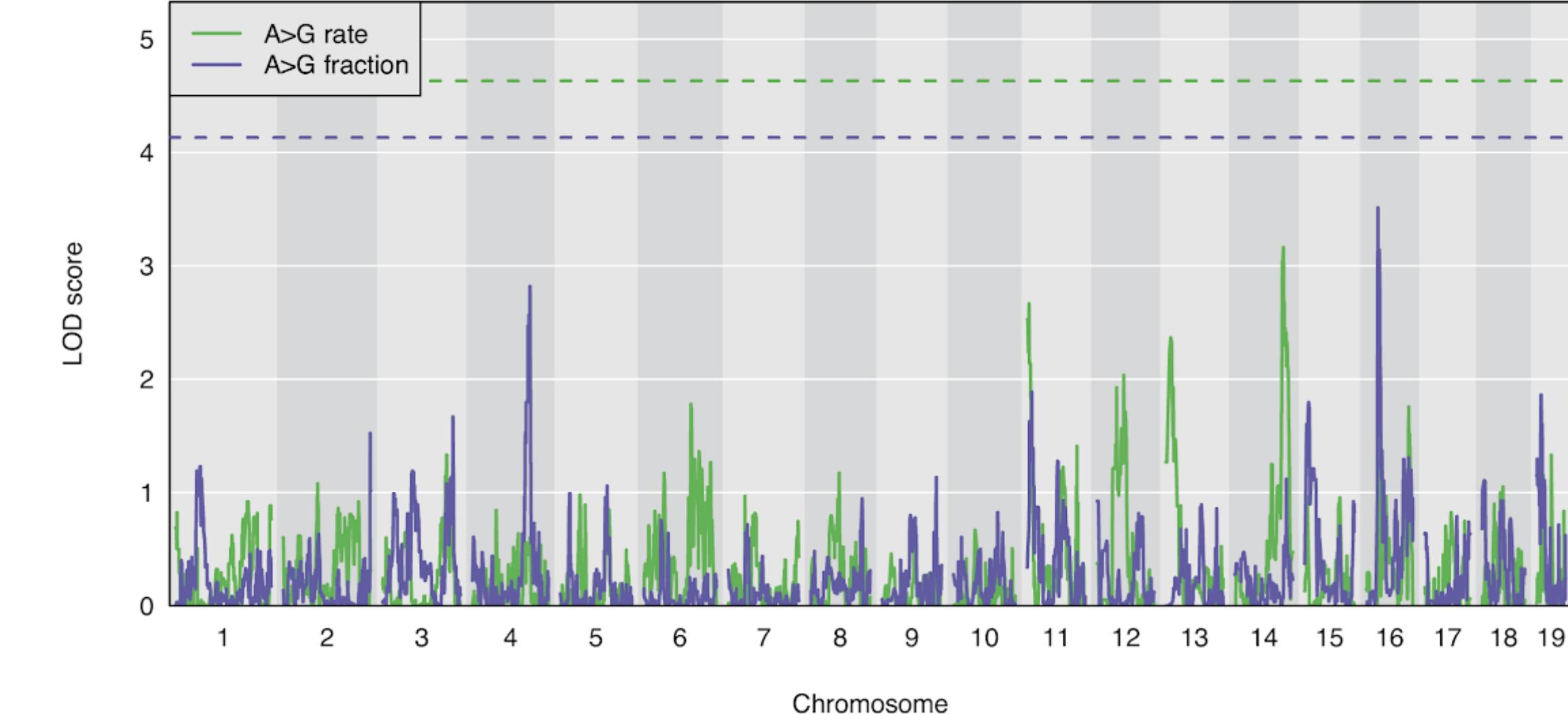
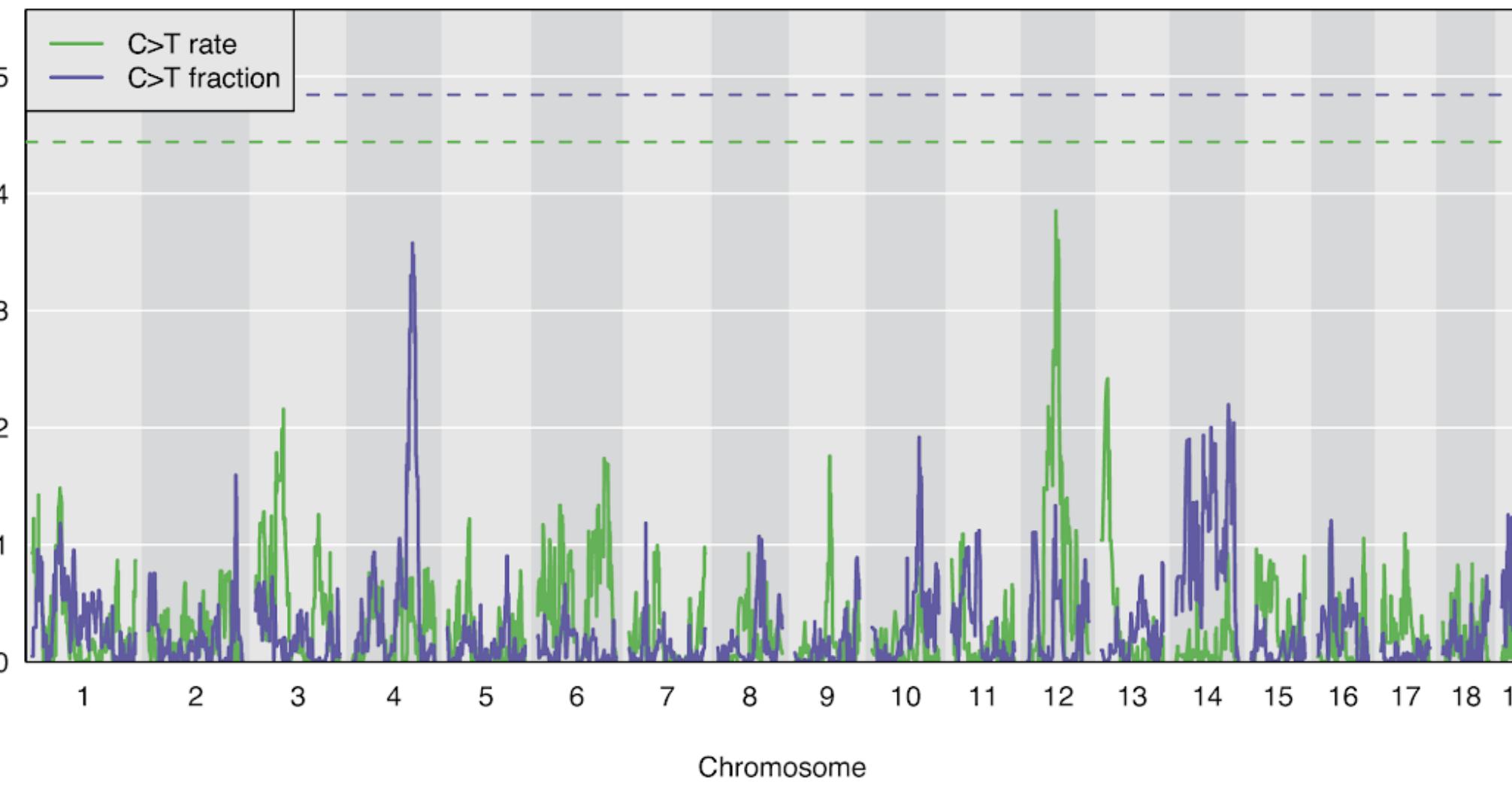
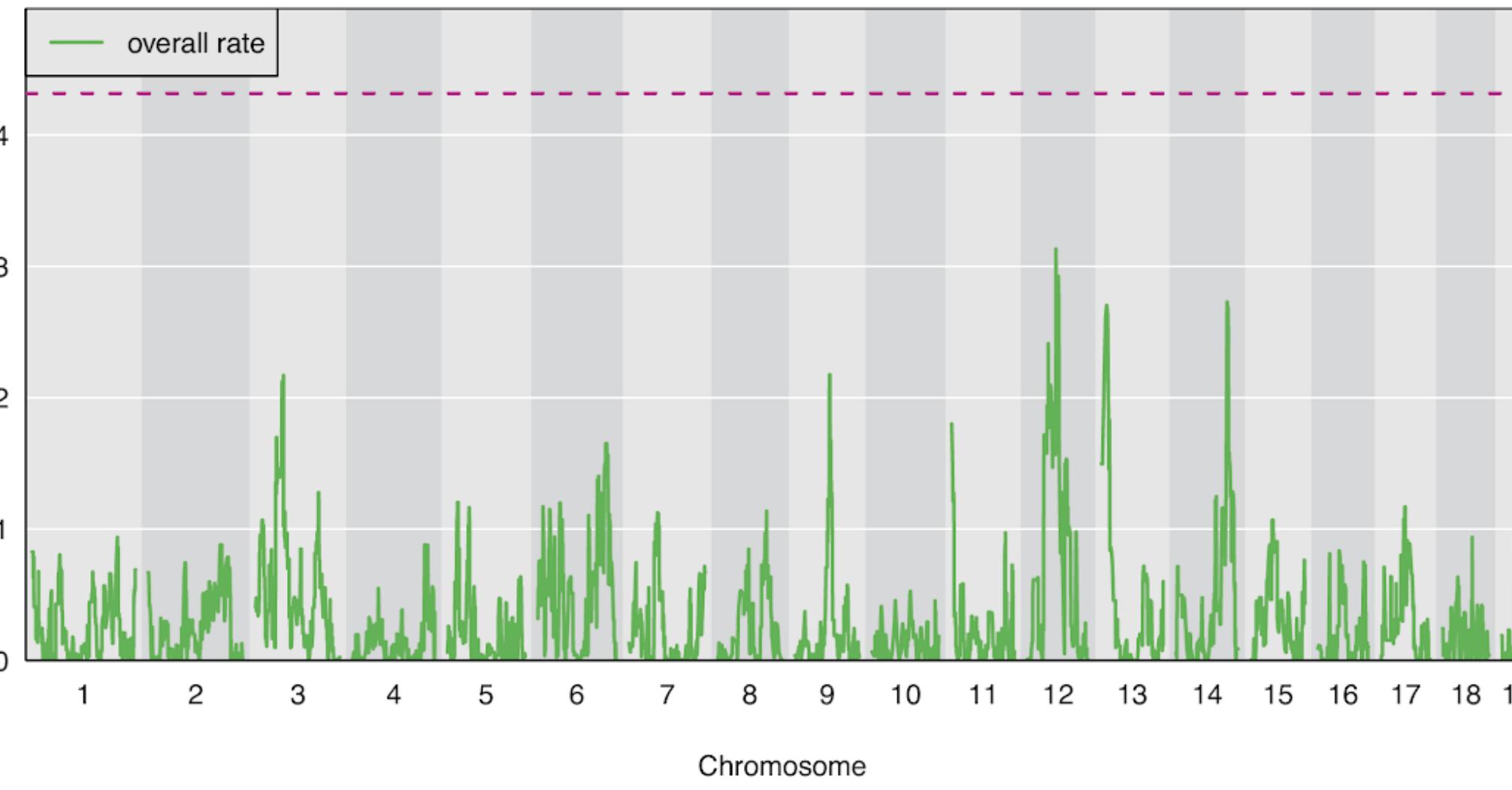
Bill Grady  
Fred Hutchinson  
Cancer Center



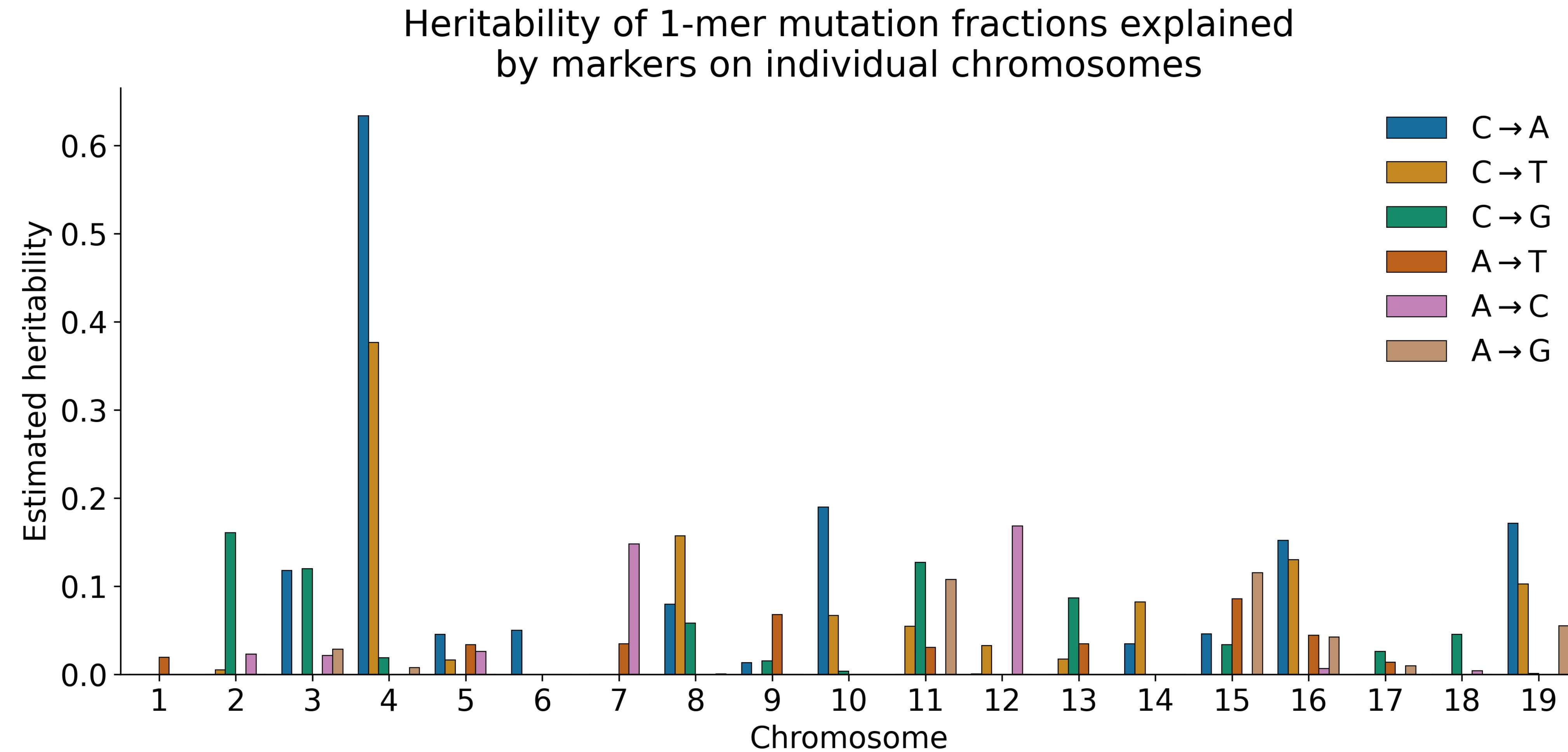
Rosana Risques  
UW Pathology



# Is *Mutylh-D* the only BXD mutator allele?



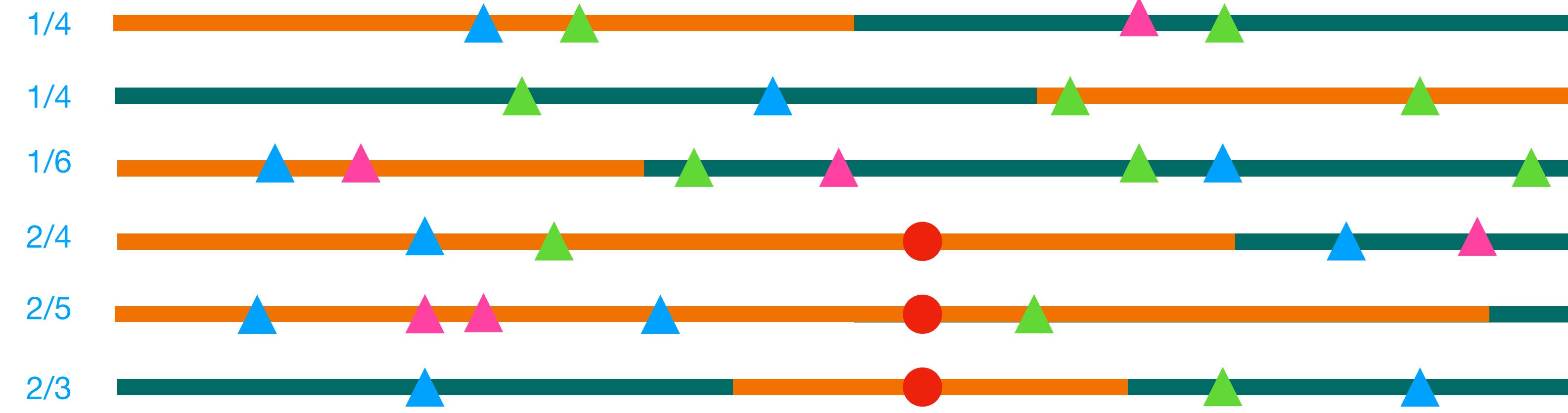
# Polygenic signal of mutation spectrum heritability outside C>A and chromosome 4



# Can we find more mutator loci by improving QTL scan power?



Aaron Quinlan



Classical QTL mapping:

1. Calculate  $\Delta / (\Delta + \Delta + \Delta)$  for each sample
2. Test whether  $\Delta / (\Delta + \Delta + \Delta)$  is significantly higher or lower on the ● background
3. Separately test  $\Delta / (\Delta + \Delta + \Delta)$  for association with ●



Tom Sasani

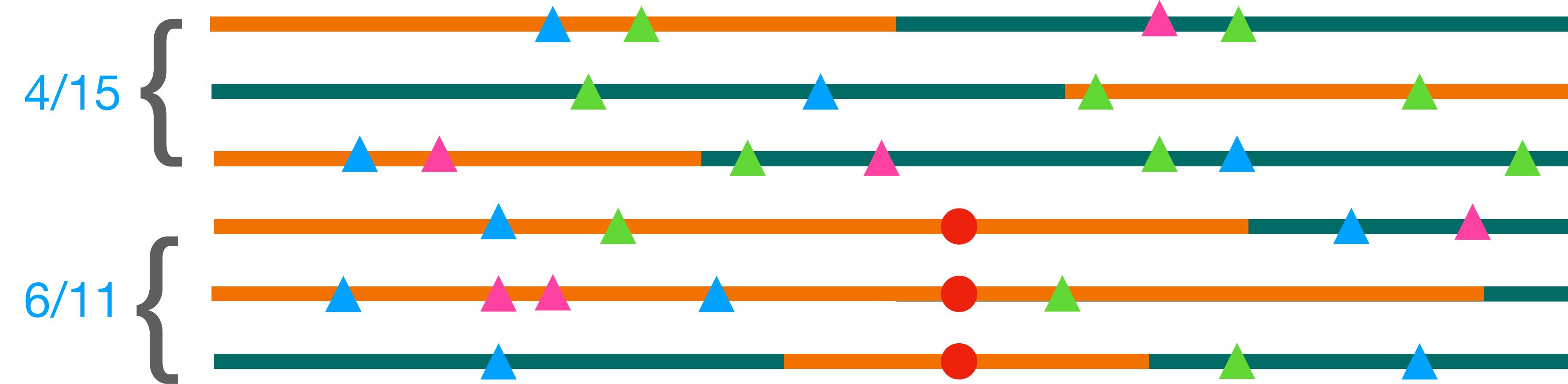
# Reducing estimator variance and multiple testing burden



Aaron Quinlan



Tom Sasani

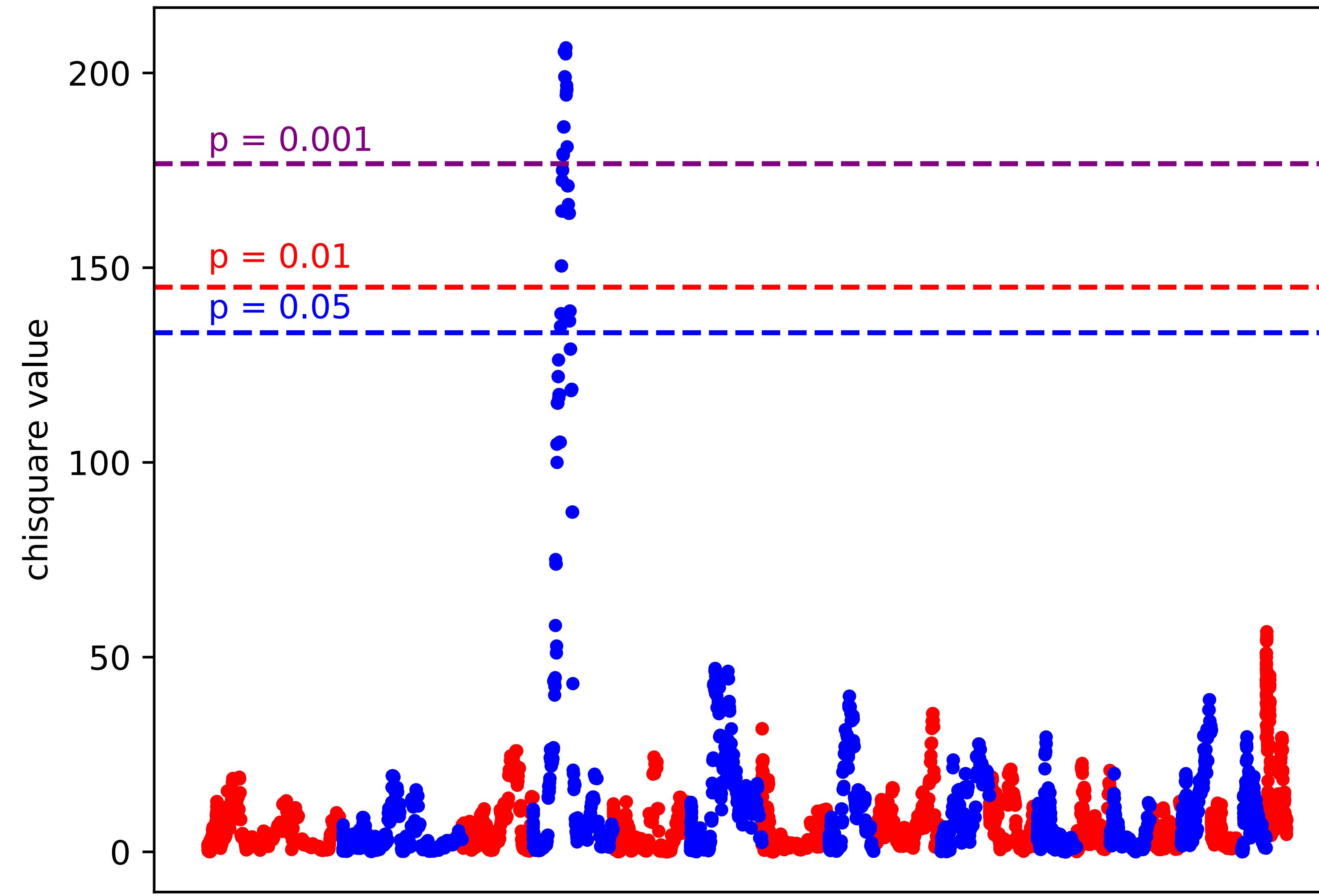


New method:

1. Calculate aggregate values of  $\Delta / (\Delta + \Delta + \Delta)$ ,  $\Delta / (\Delta + \Delta + \Delta)$ ,  $\Delta / (\Delta + \Delta + \Delta)$  on the ● background and the non-● background
2. Jointly test whether these low-variance proportions depend on ● via chi-square

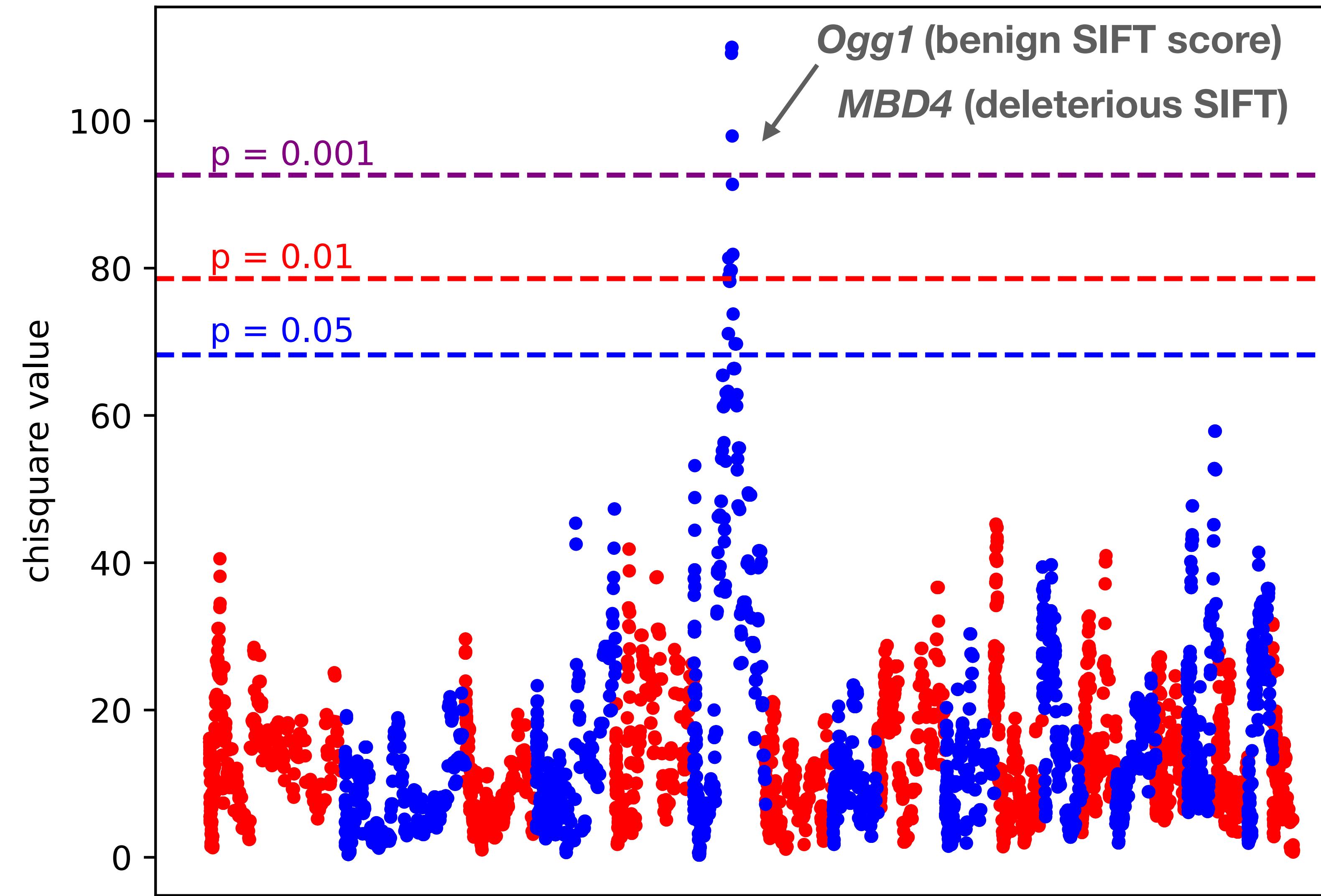
▲	▲	▲
● ▲	● ▲	● ▲

# Global associations between strain background and mutation spectrum

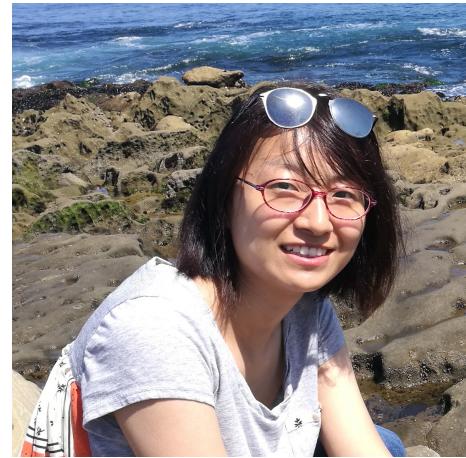


Sasani, et al. in prep.

# Conditioning on *Mutyh-D* reveals a mutator signal on chromosome 6



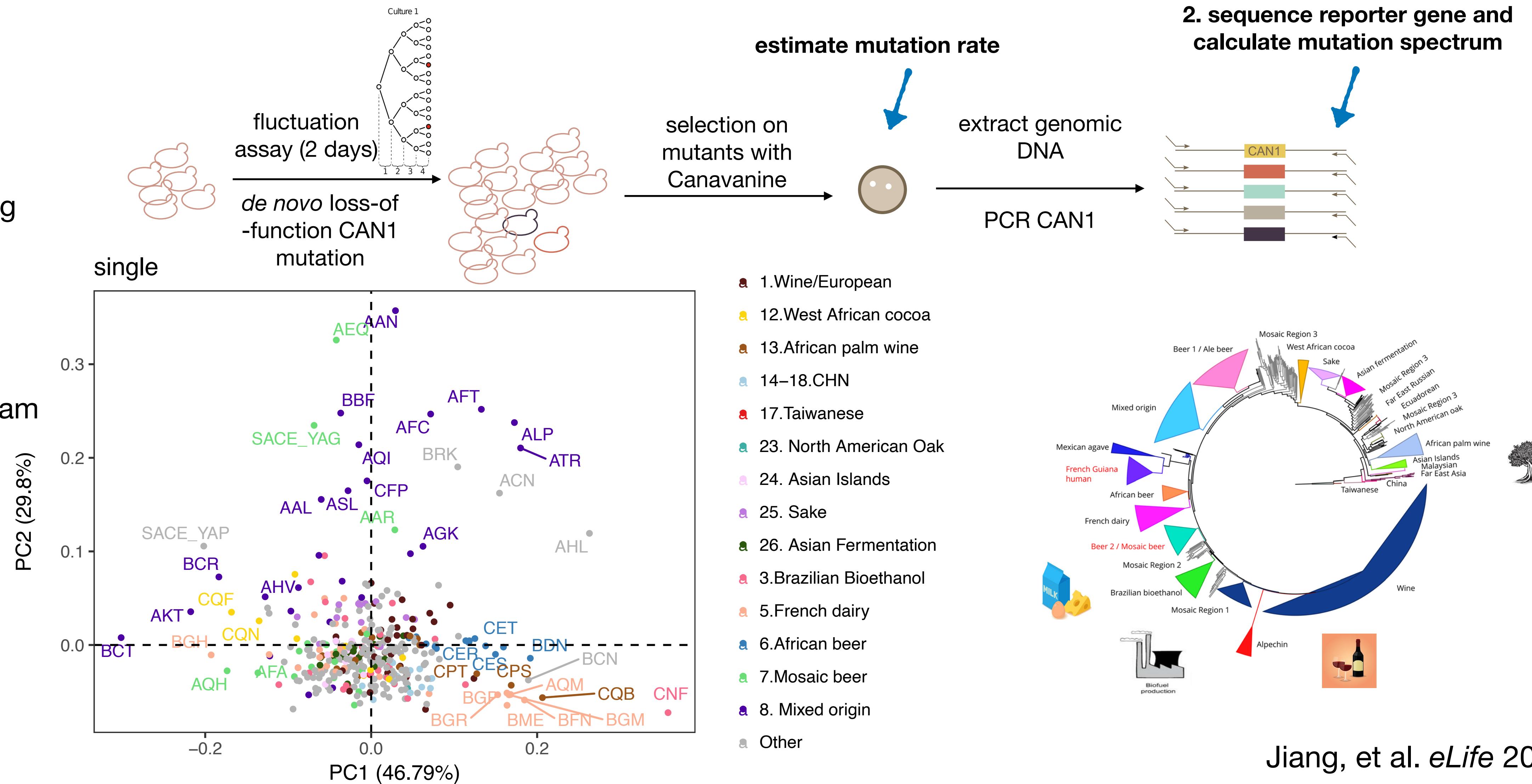
# A mutator allele of *OGG1* (an interaction partner of *Mutyh* in 8-oxo-G repair) elevates the C>A mutation rate in mosaic beer yeast



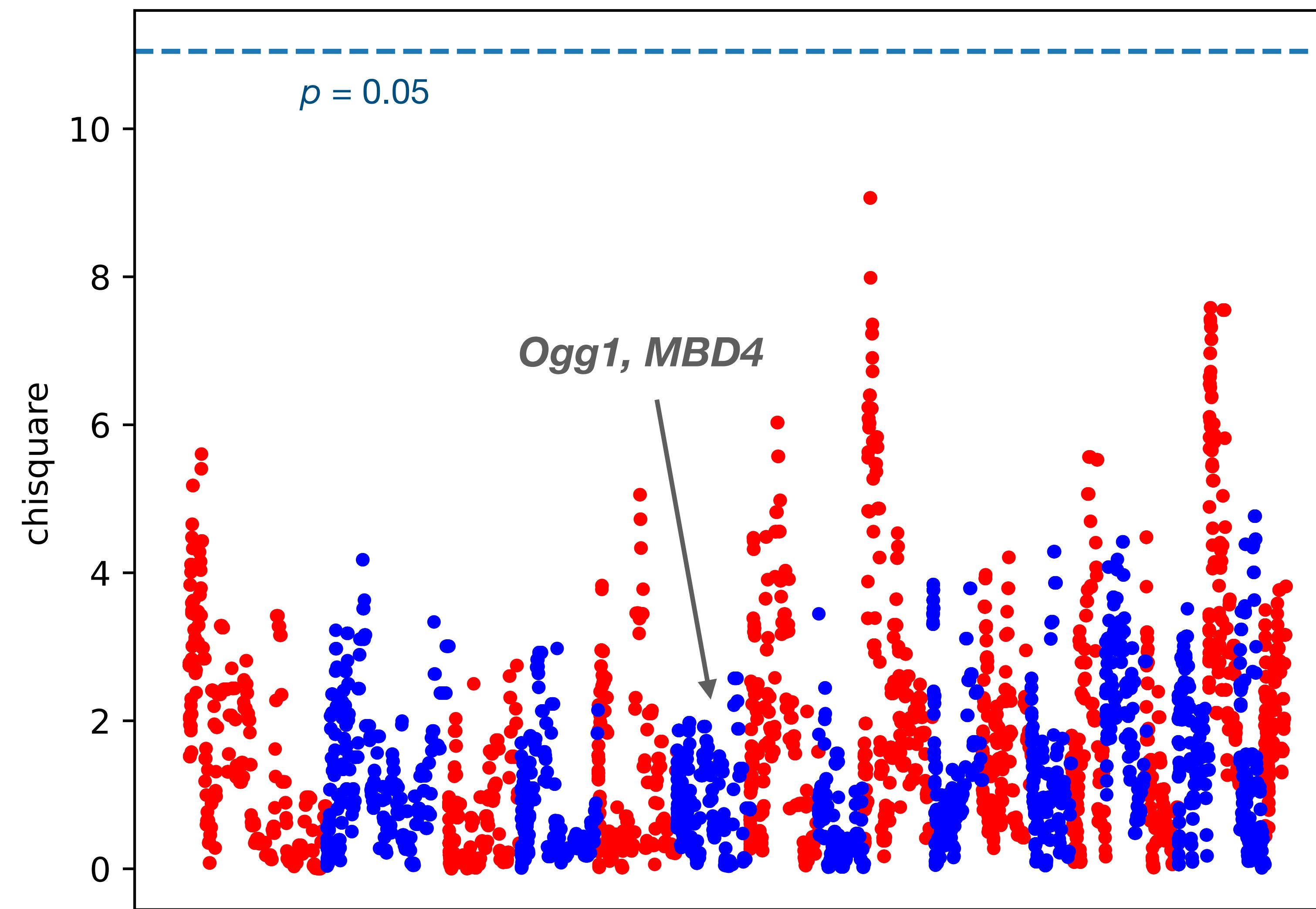
Pengyao Jiang



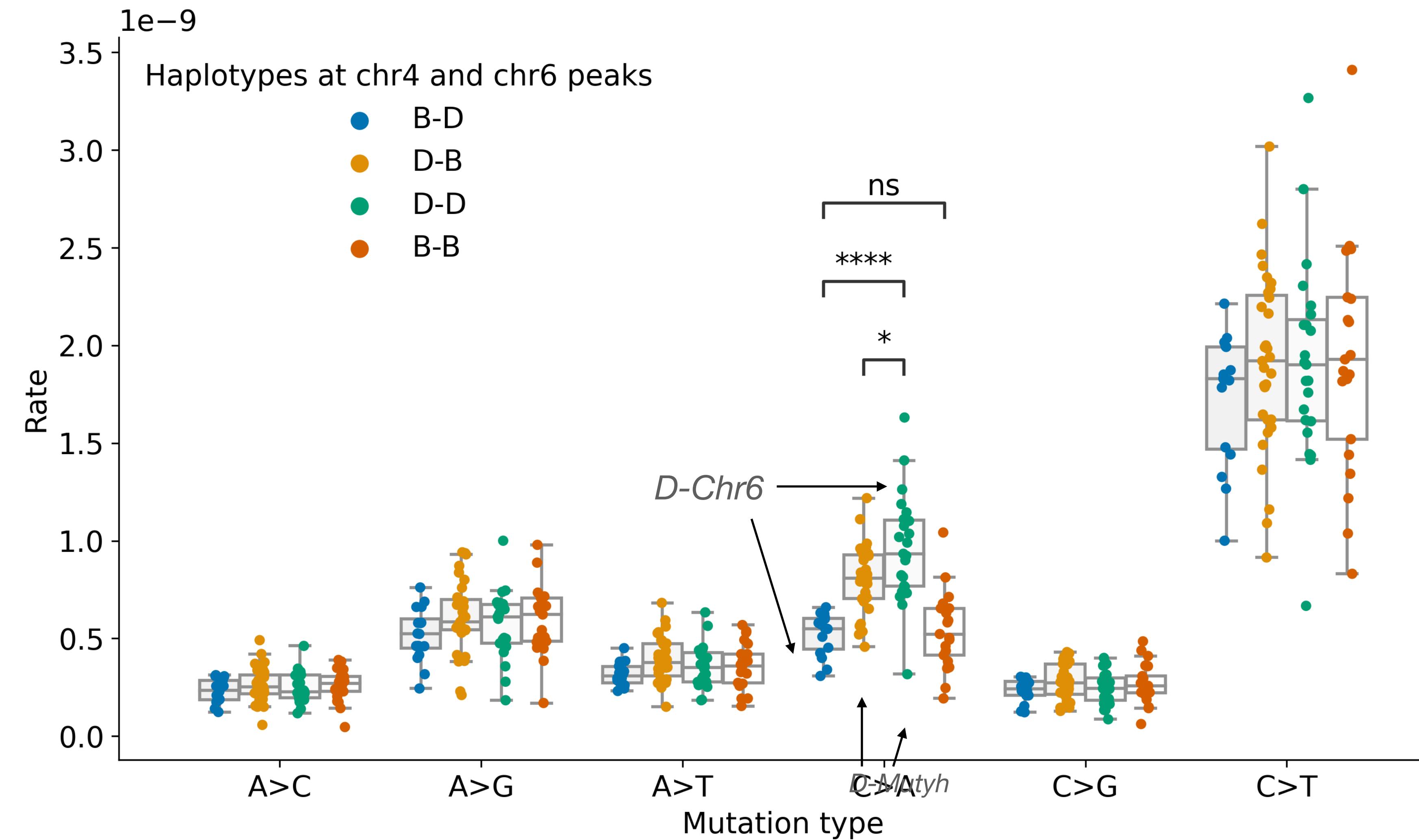
Maitreya Dunham



# Chr6 association is absent from *B-Mutyh* mice, suggesting epistasis

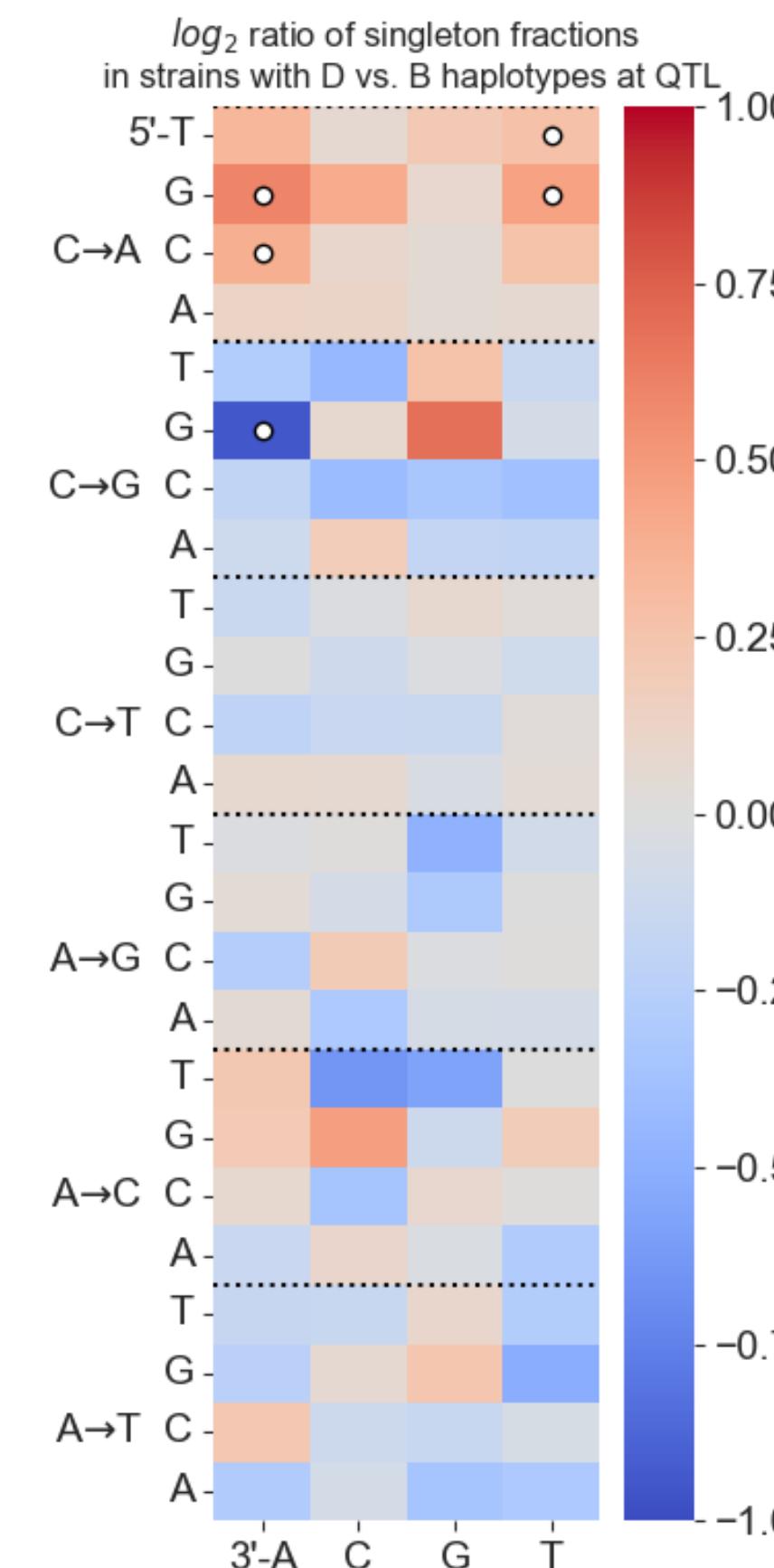


# *D-Chr6* further elevates the C>A mutation rate in *D-Mutyh* mice

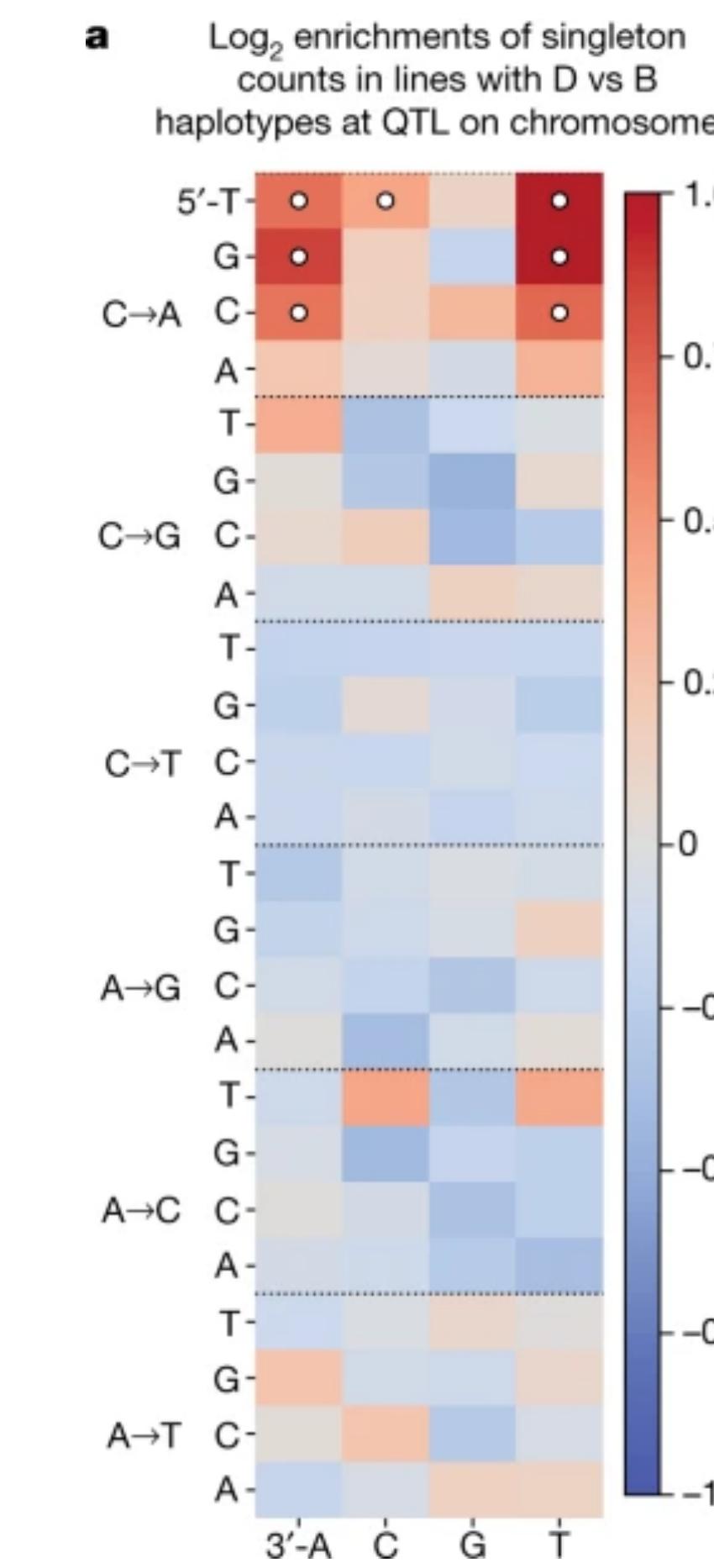


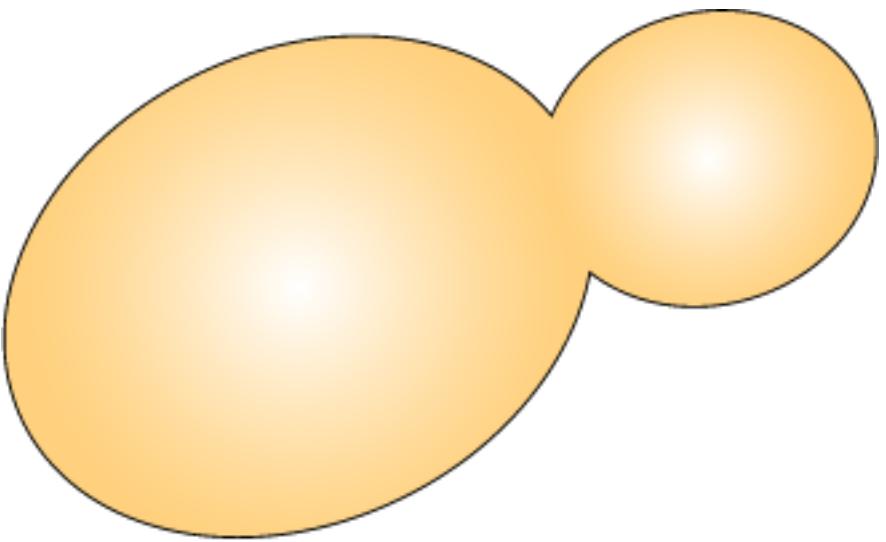
# Sequence context bias is consistent with reactive oxygen species damage

*D v B at Ogg1  
on D-Mutyh background*

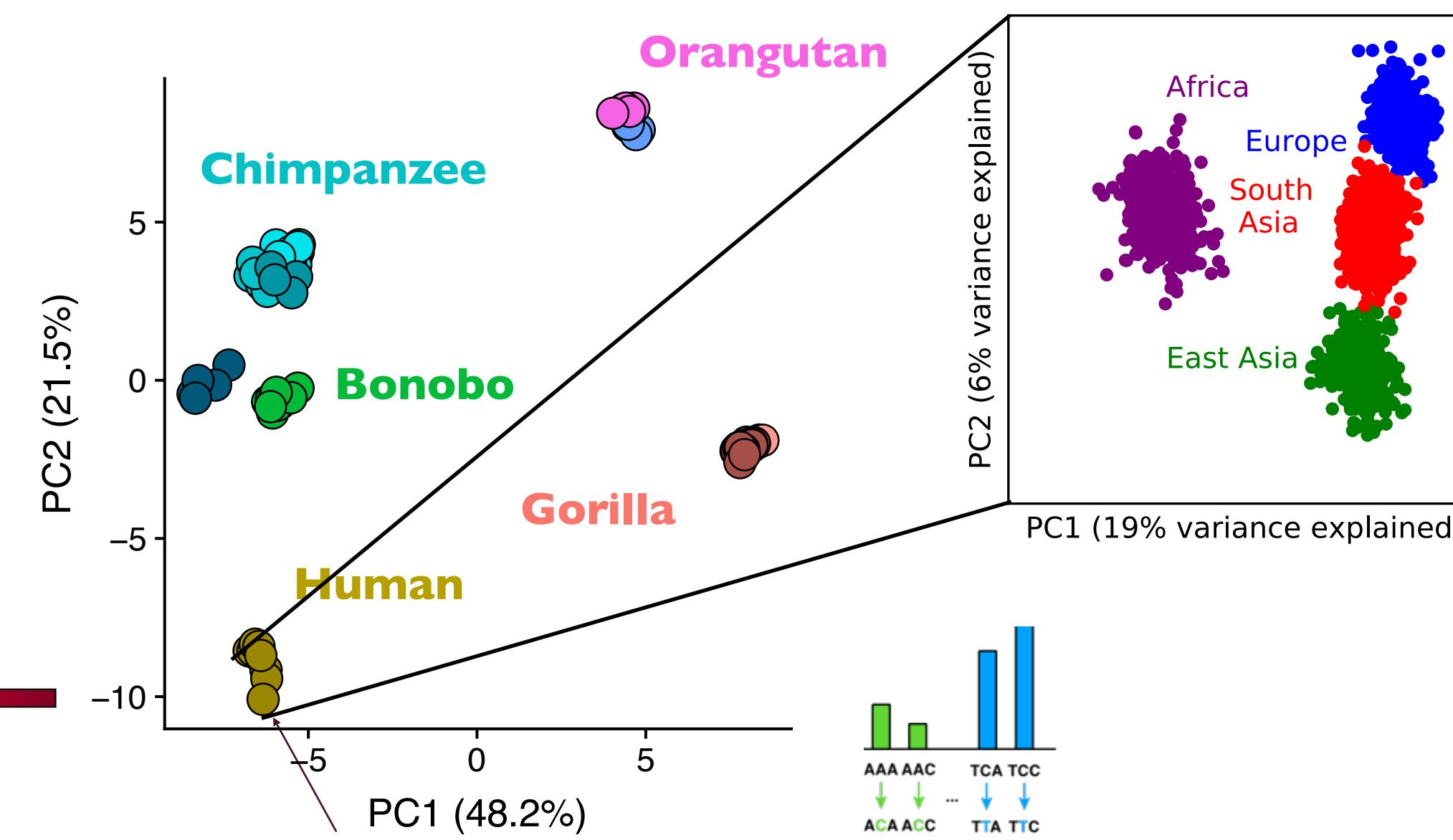
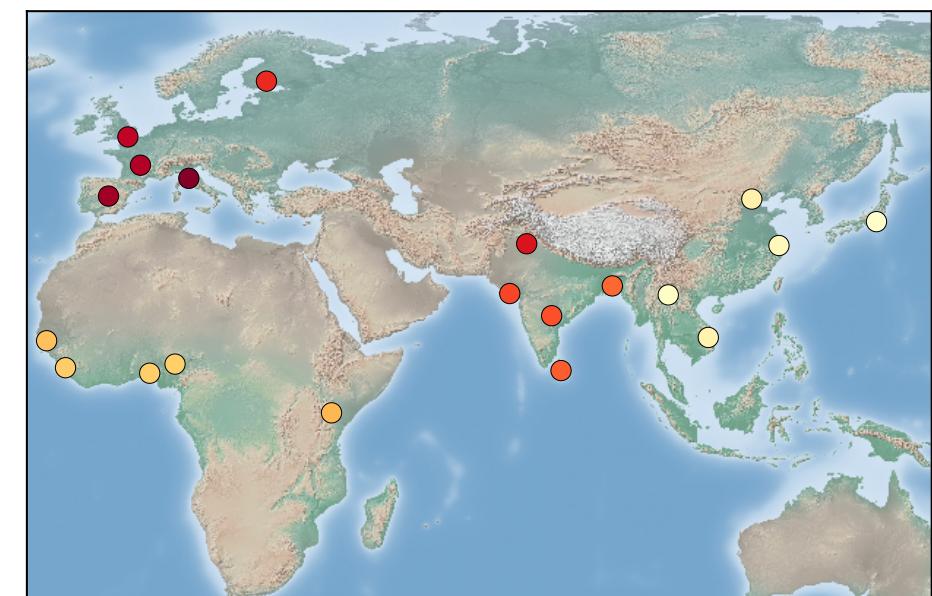


*D v B at Mutyh*





Mutation spectrum variation in mice and yeast are at least partly determined by genetic differences in DNA pathways



Is the same true of humans, apes, and other mammals?

# If mutator alleles arise often, we expect closely related species to have more similar mutation spectra

Mice:  
Harr et al. (2016)

Apes:  
Prado-Martinez et al. (2013)

Humans:  
Byrska-Bishop et al. (bioRxiv)

Bears:  
Miller et al. (2012)  
Cahill et al. (2013)  
Liu et al. (2014)  
Benazzo et al. (2017)  
Barlow et al. (2018)

Wolves:  
Broad Institute Canine Reference Panel (2019)

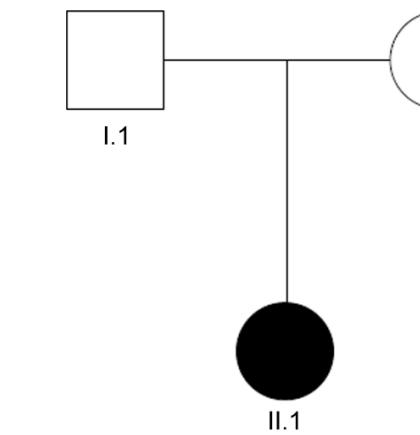
Vaquita:  
Robinson & Kyriazis et al. (2022)

Fin whale:  
Nigenda-Morales & Lin et al. (in revision)



# To estimate mutation spectra at scale, we use indirect polymorphism data instead of *de novo* mutation data

Polymorphism data						
chr1	300	0/1	1/1	0/0	0/0	0/0
chr1	875	0/0	0/1	0/0	1/1	0/0
chr1	1240	0/0	0/0	0/0	0/0	0/1
chr1	3040	0/0	0/1	0/1	0/0	0/0



- Large sample size ( $10^4$ — $10^6$  variants per species)
- Affected by selection and GC-biased gene conversion (gBGC), especially at high allele frequencies
- A mix of mutation spectra from many individuals living in many environments and time periods
- Small sample size (100—1000 variants per species)
- Not affected by selection or gBGC
- Pipelines and data quality cause large variation in Type I & II error, potentially biasing spectrum
- A snapshot of one contemporary family's mutation spectrum

If mutator alleles arise often and are tolerated by evolution,  
the mutation spectrum should have strong *phylogenetic signal*

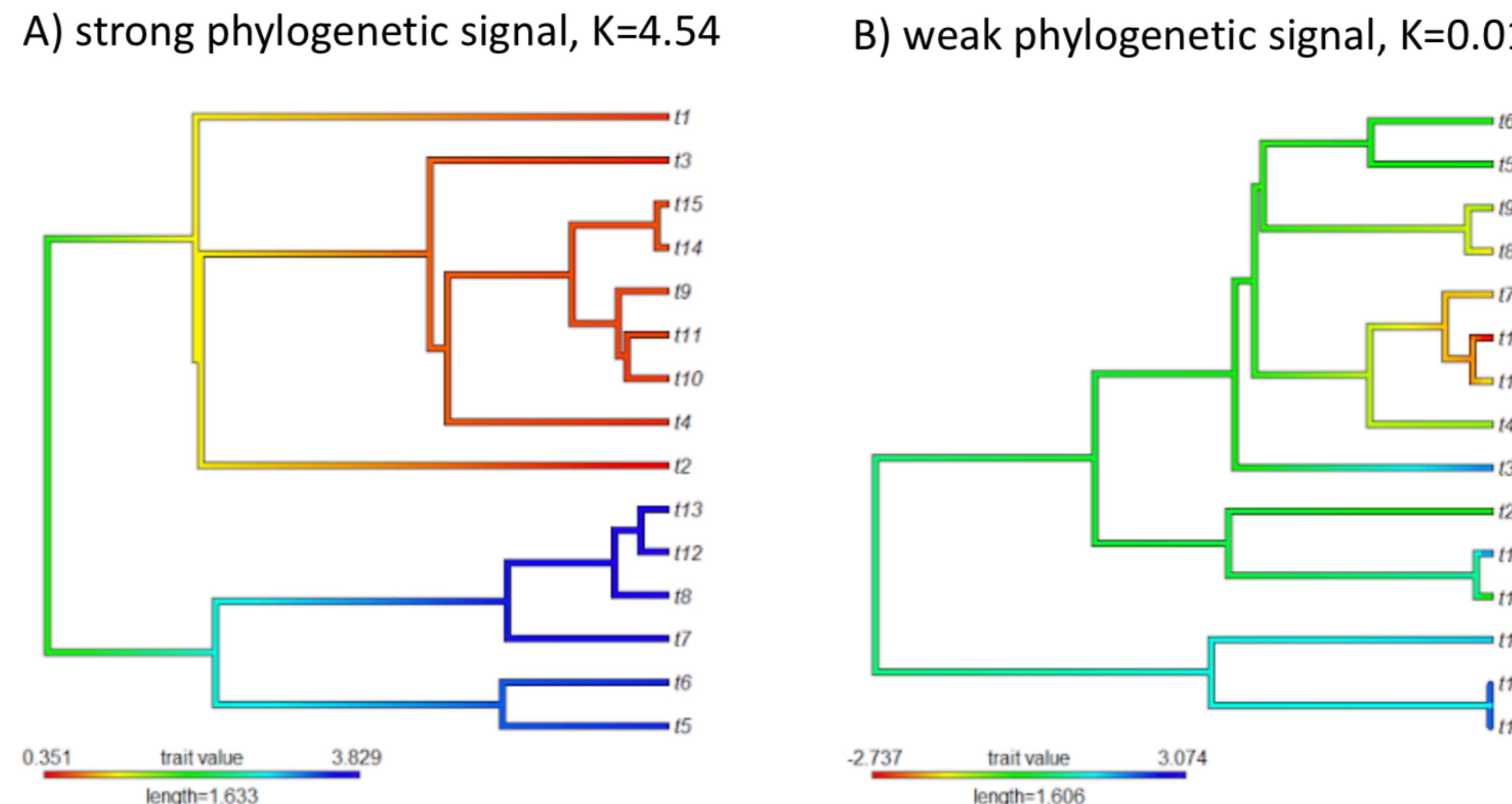
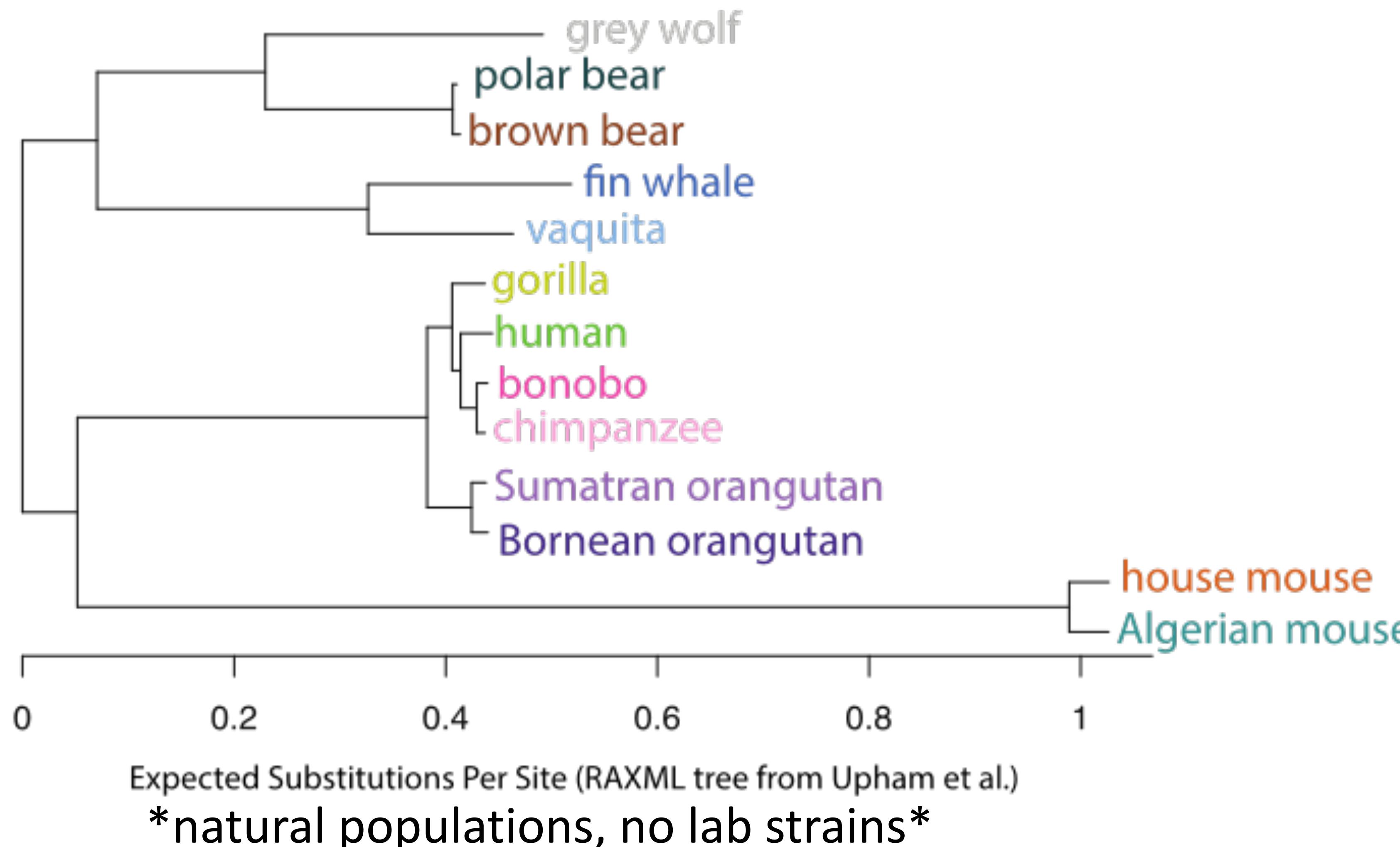
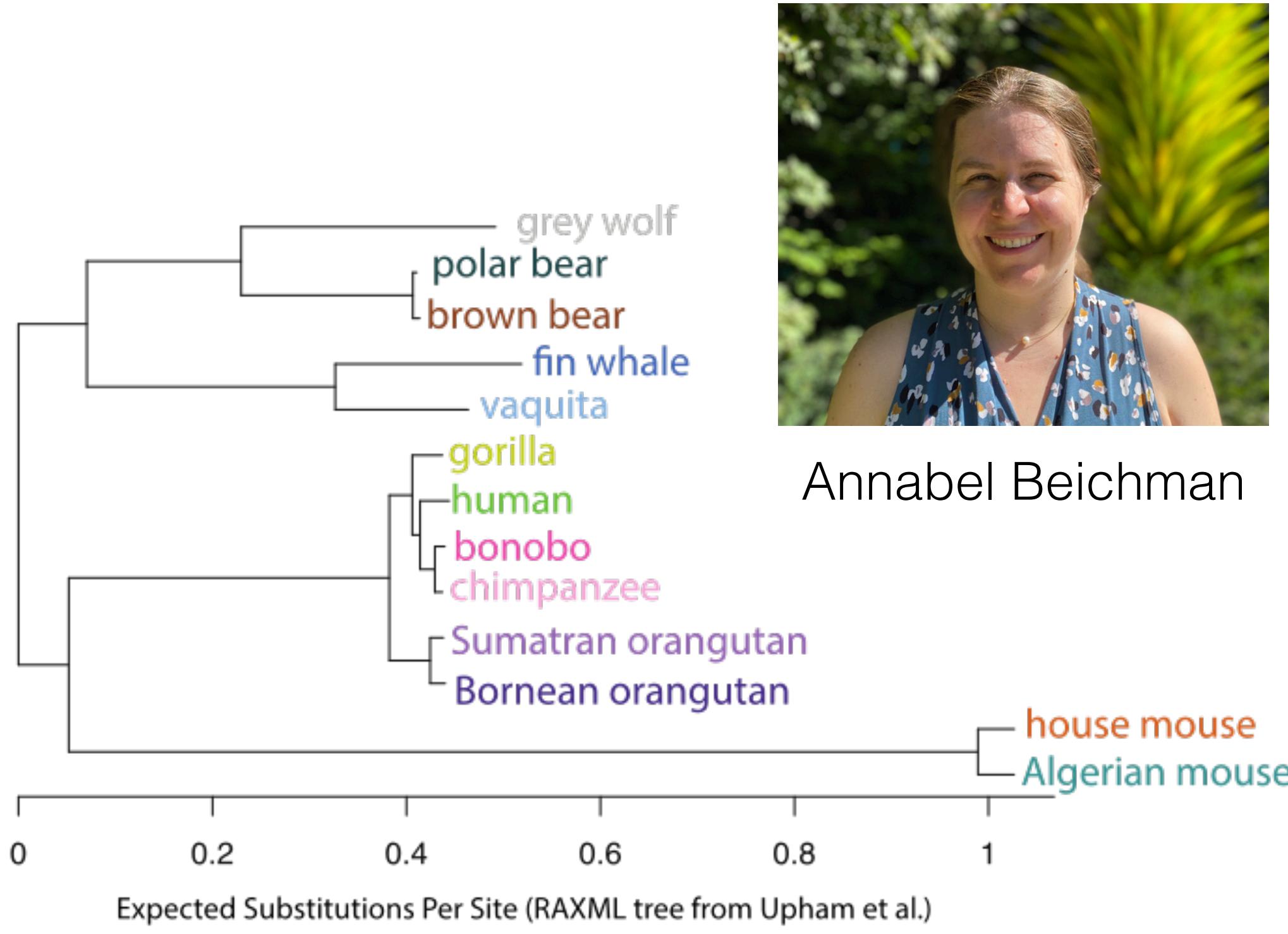


Figure 1. Mock examples of traits showing a strong (A) vs. a weak (B) phylogenetic signal.  
Examples were created using R packages “phytools”, “geiger” and “evotrait”  
(rfunctions.blogspot.com).

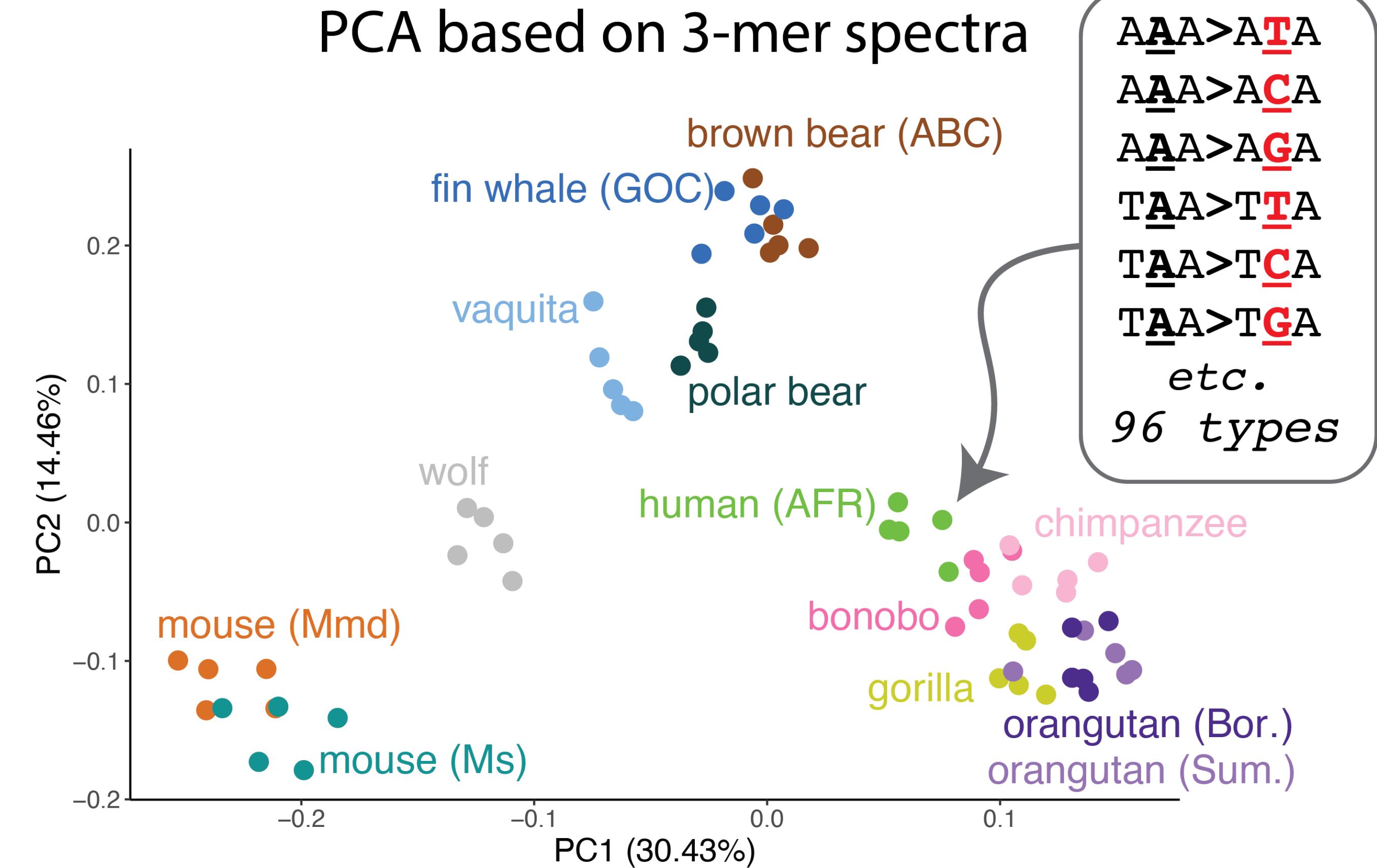
# Mammal phylogeny spanning over 100 million years of evolution



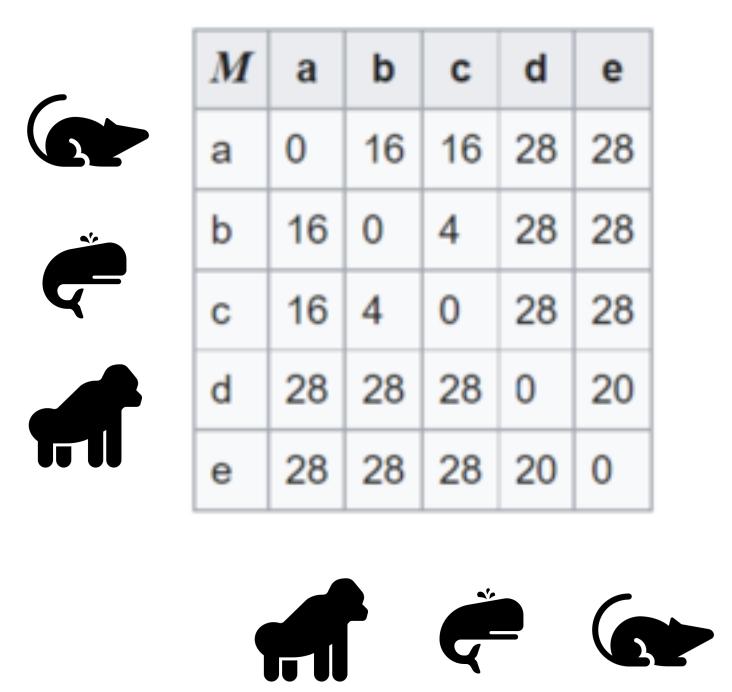
# Principal components of mutation spectrum variation appear to recapitulate mammalian phylogeny



Annabel Beichman

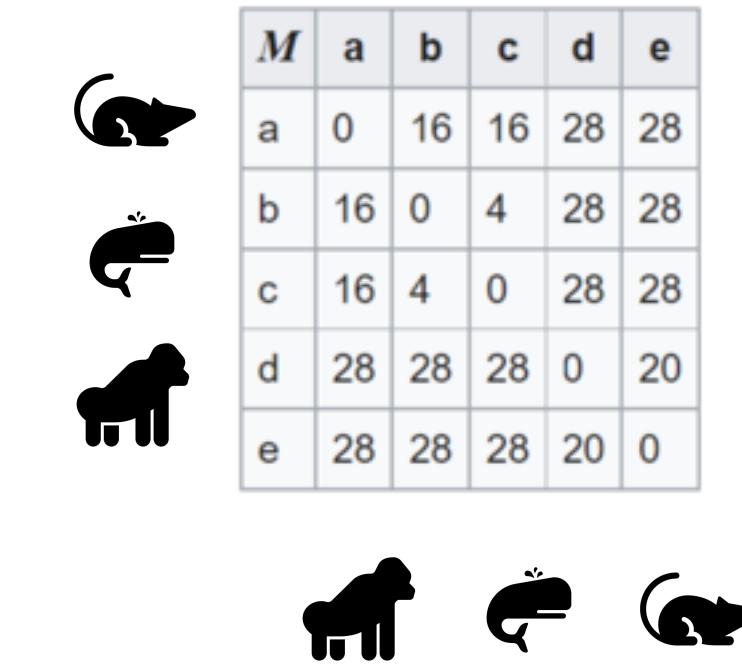


# Does the mutation spectrum divergence matrix correlate with the phylogenetic distance matrix?

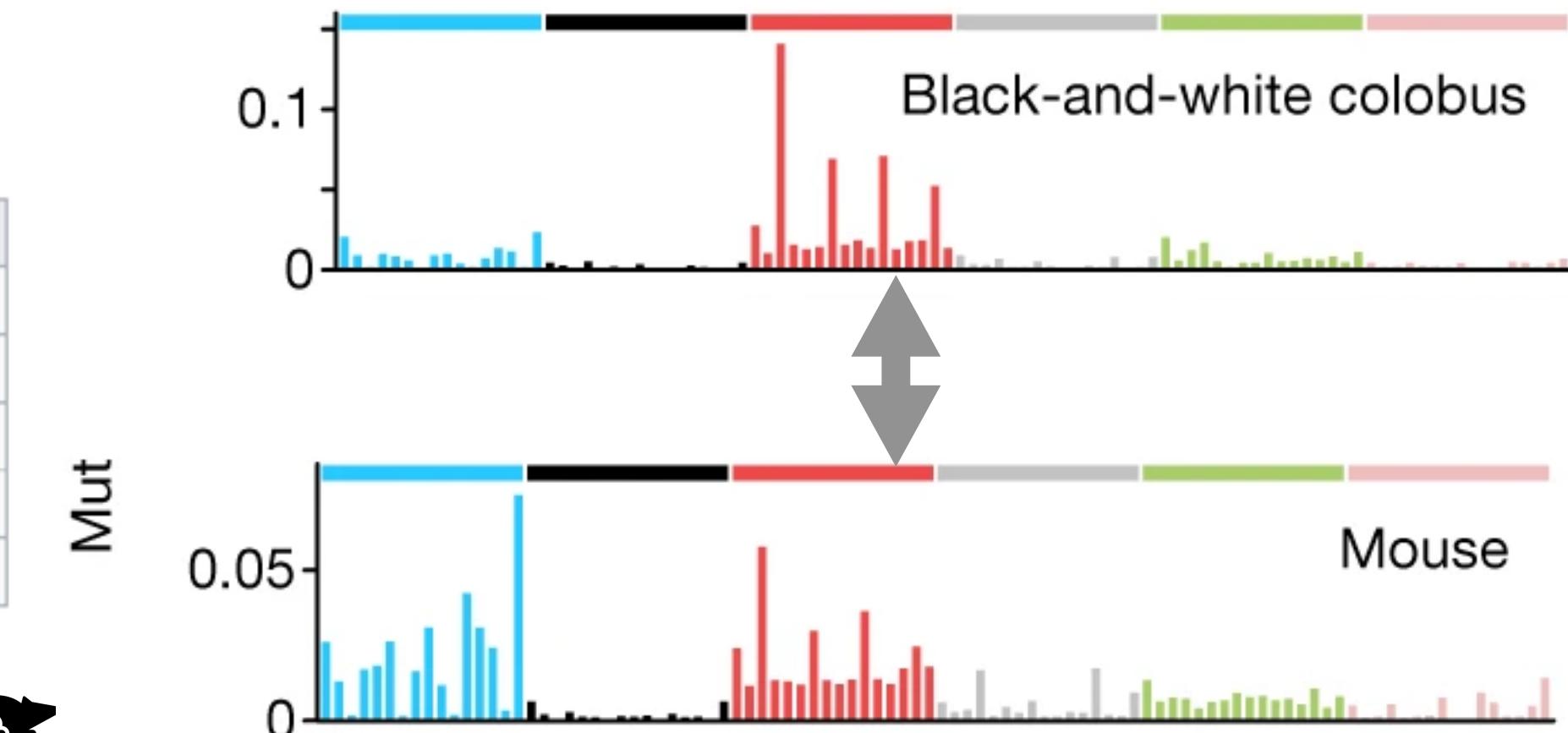


Phylogenetic distance matrix

~?

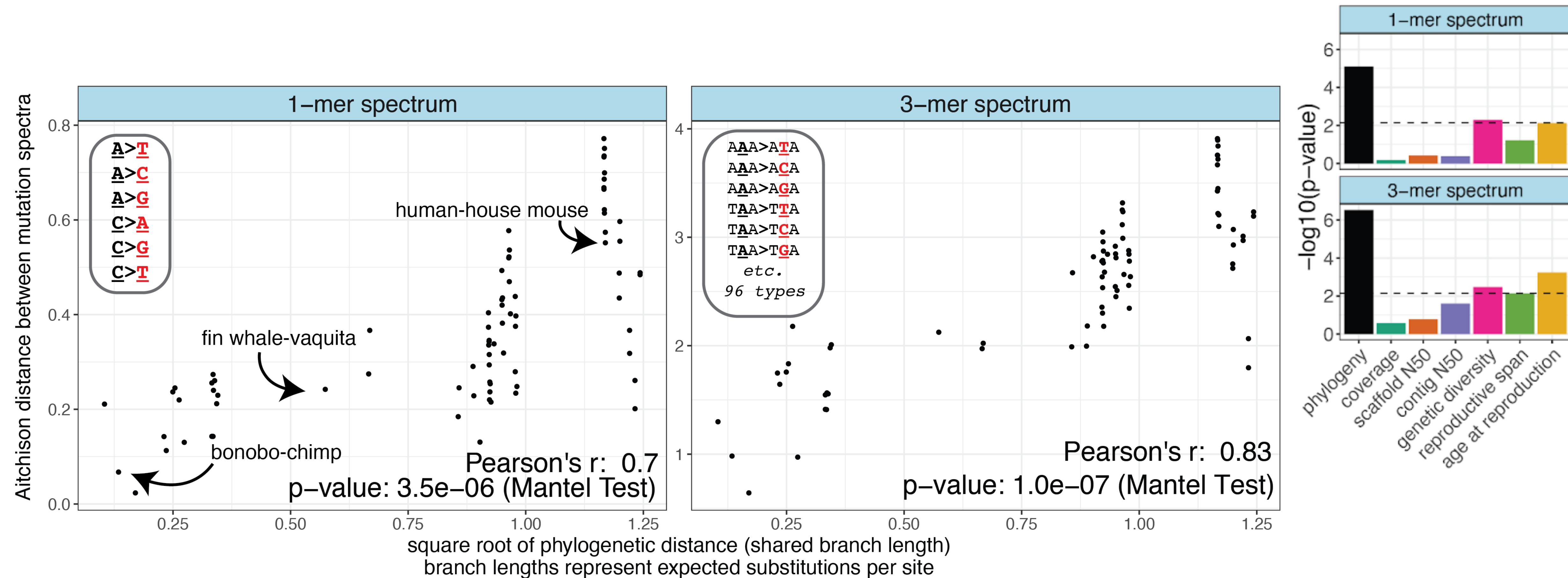


Mutation spectrum distance matrix  
(Aitchison-transformed CLR)

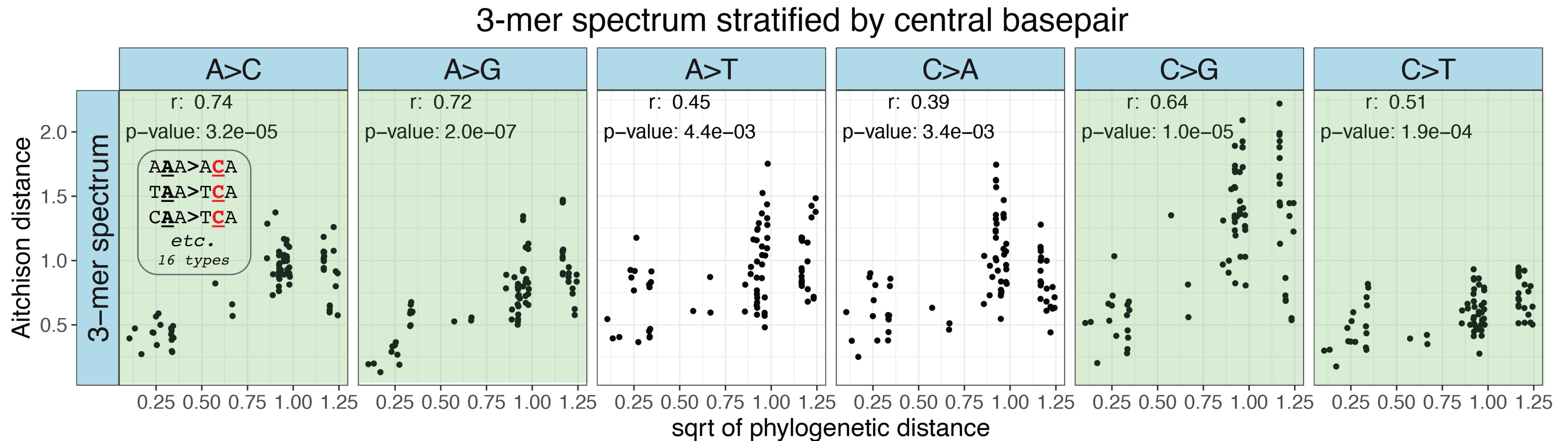


**Mantel test** compares this matrix correlation with 10 million permuted matrix correlations

# Phylogenetic distance matrix explains mutation spectrum variation better than reproductive age or reproductive lifespan does



Within most 1-mer mutation types, closely related species have more similar mutational context dependence



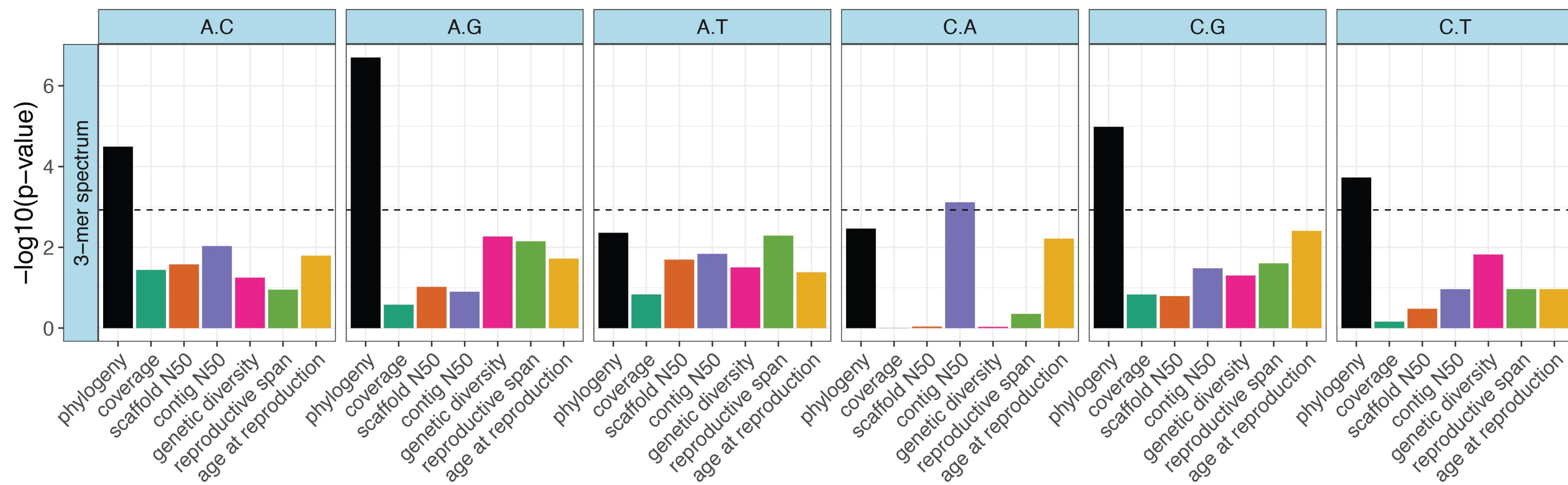
# Phylogeny explains more mutation spectrum variation than technical confounders or reproductive lifespan

## Technical confounders

- Coverage
- Scaffold N50
- Contig N50

## Biological confounders

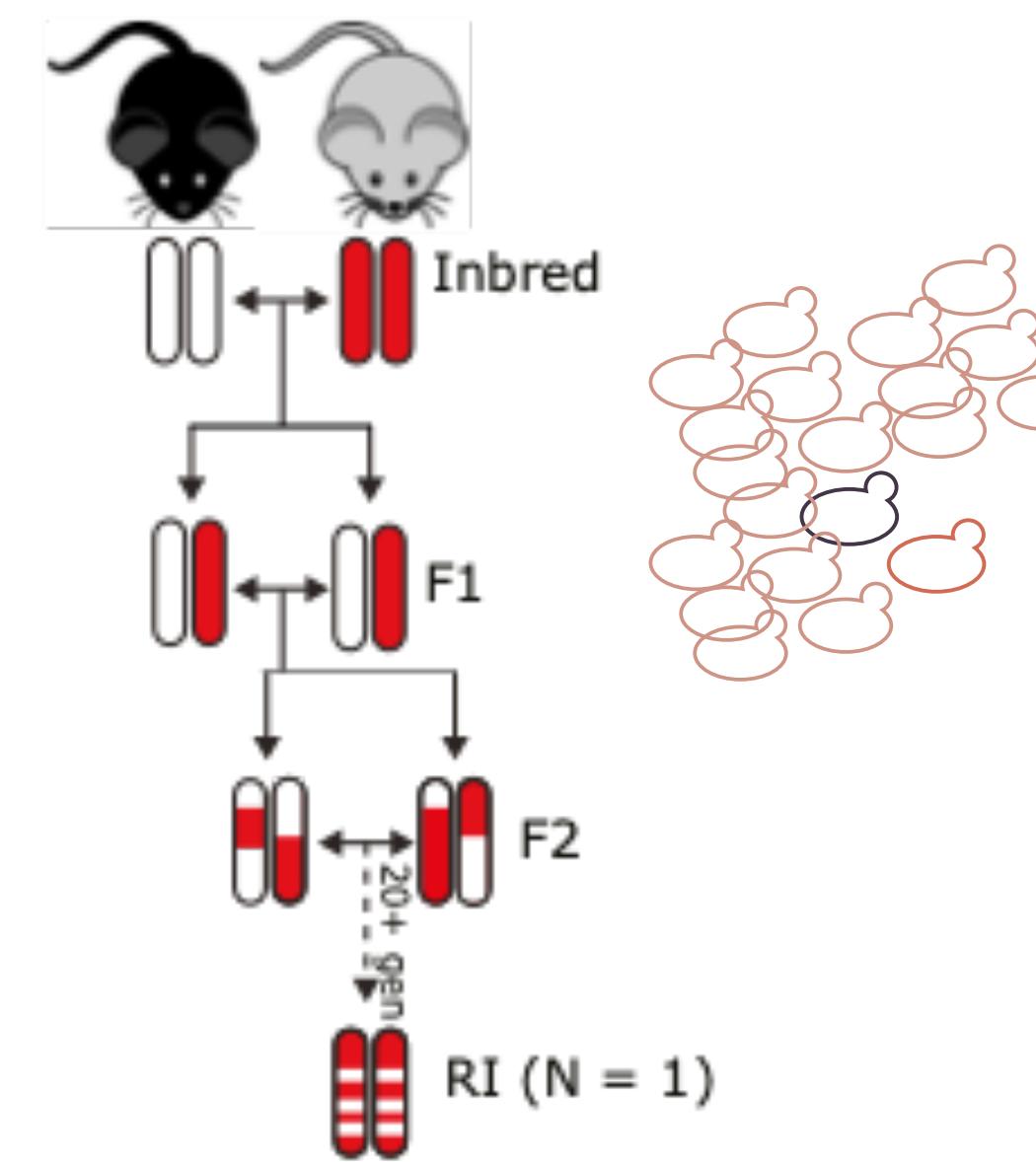
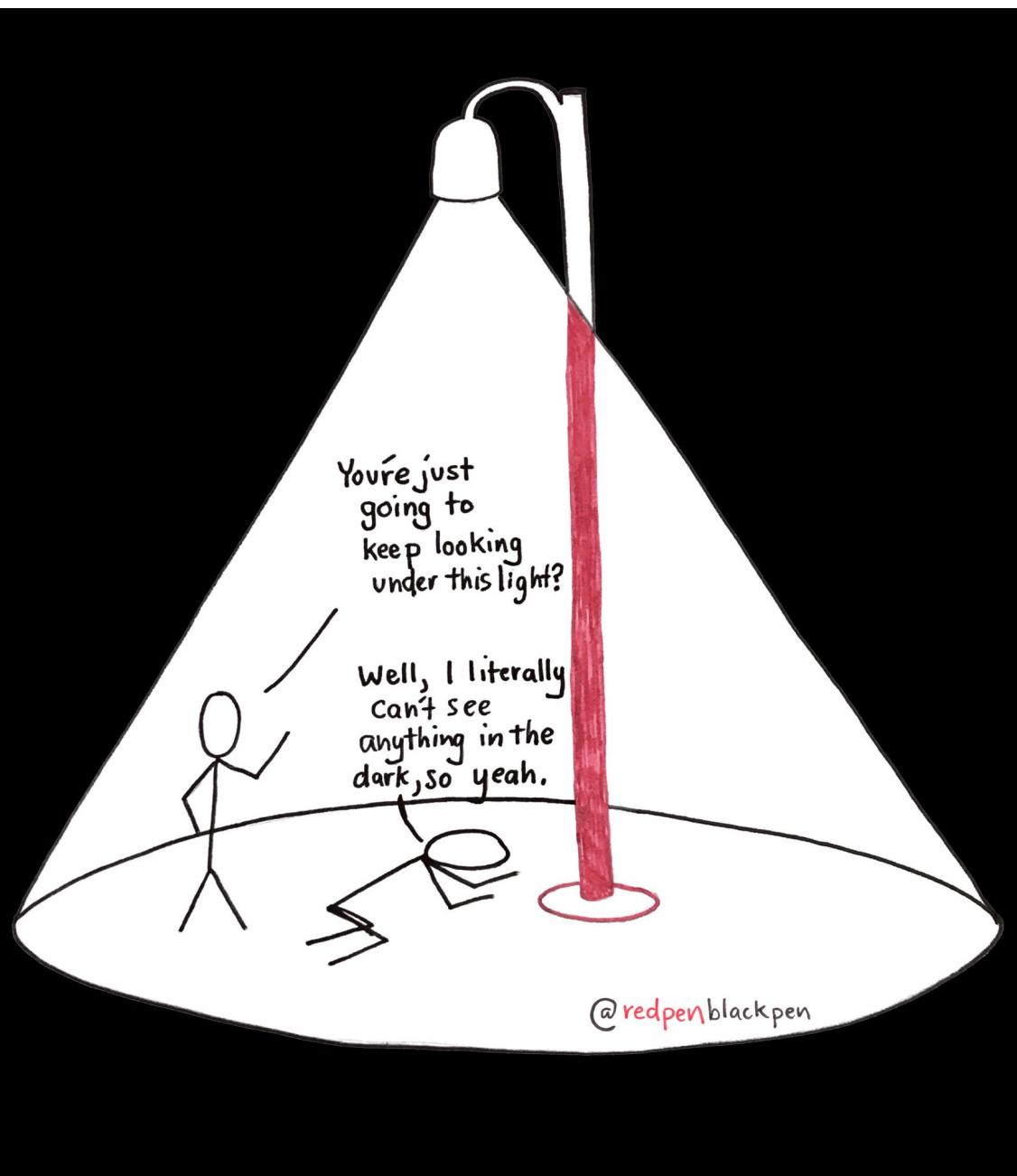
- Genetic diversity
- Reproductive lifespan
- Average reproductive age



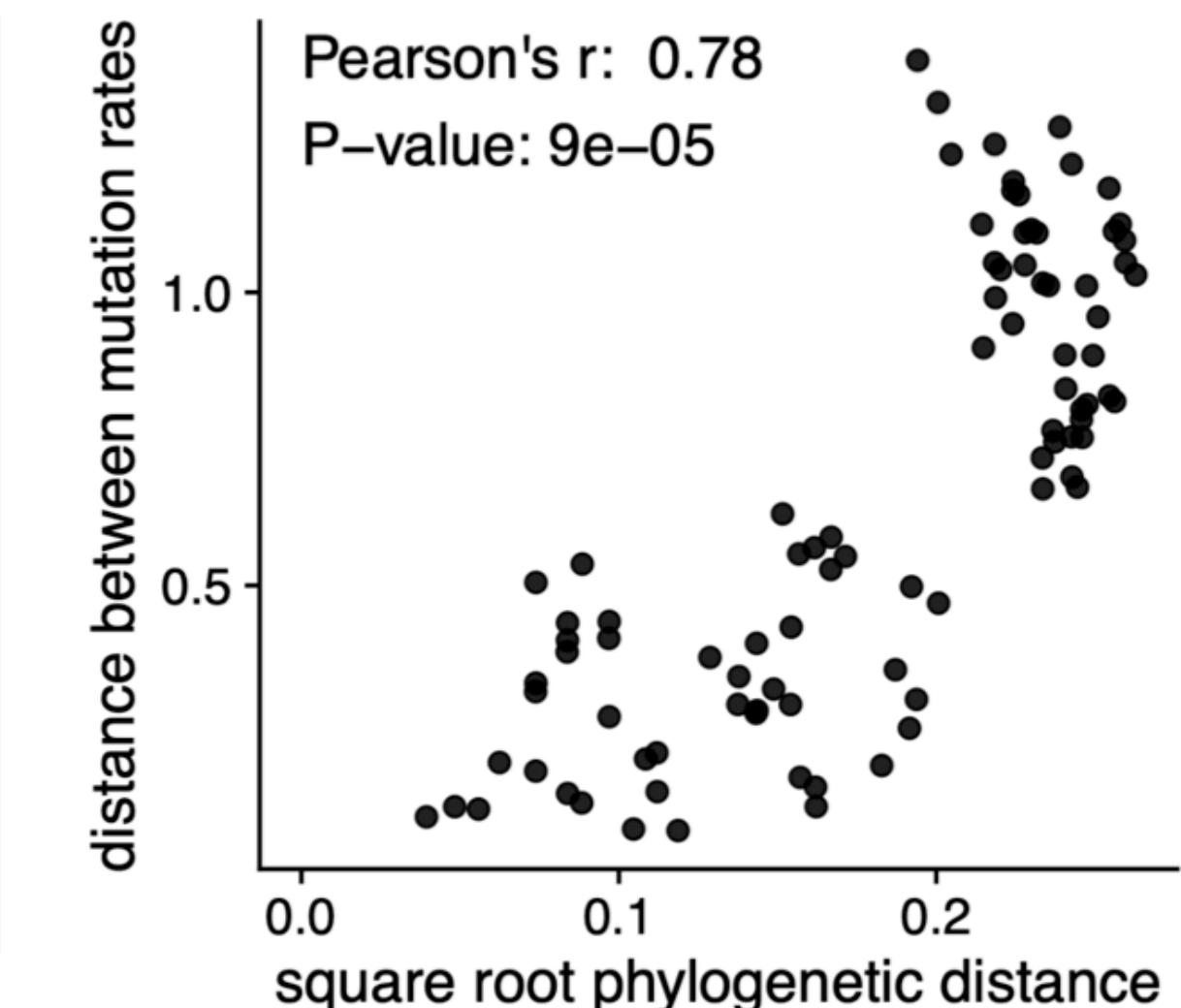
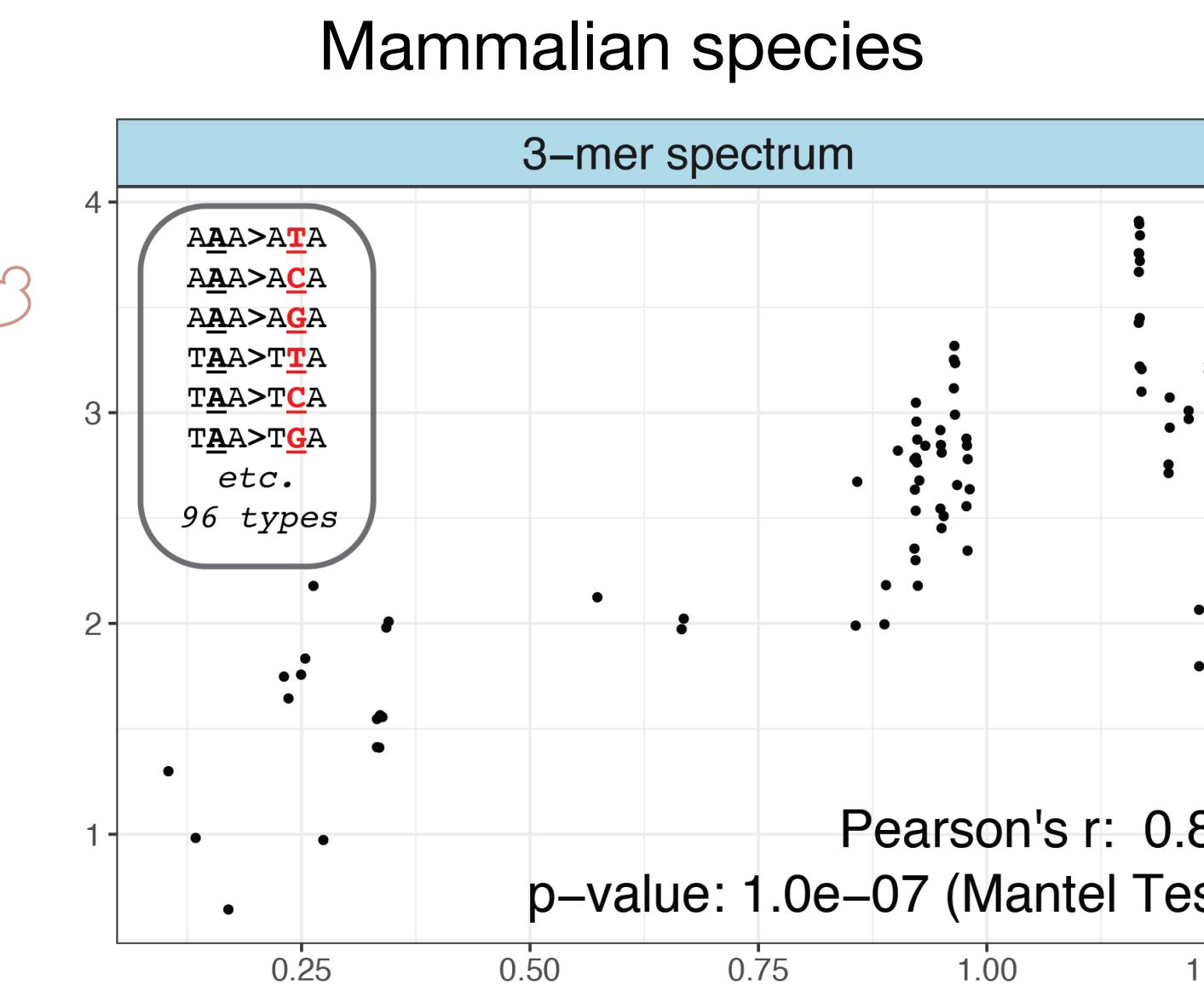
“It is probable that different strains will have accumulated different genes lowering the mutation rate, and that most strains (in cross-fertilizing organisms) will be somewhat heterogeneous for genes affecting that rate.”

—*On the effects of selection on mutation rate*, A.H. Sturtevant, 1937

A few model organism breeding structures have made it possible to identify these genes



In most cases, we still have little power to identify mutator alleles, but mutational processes appear consistently variable and heritable



# Acknowledgements



Tom Sasani



Jonathan Pritchard



Rob Williams



David Ashbrook



Pengyao Jiang



Candice Young



Maitreya Dunham



The Harris Lab



BURROUGHS  
WELLCOME  
FUND



SEARLE SCHOLARS