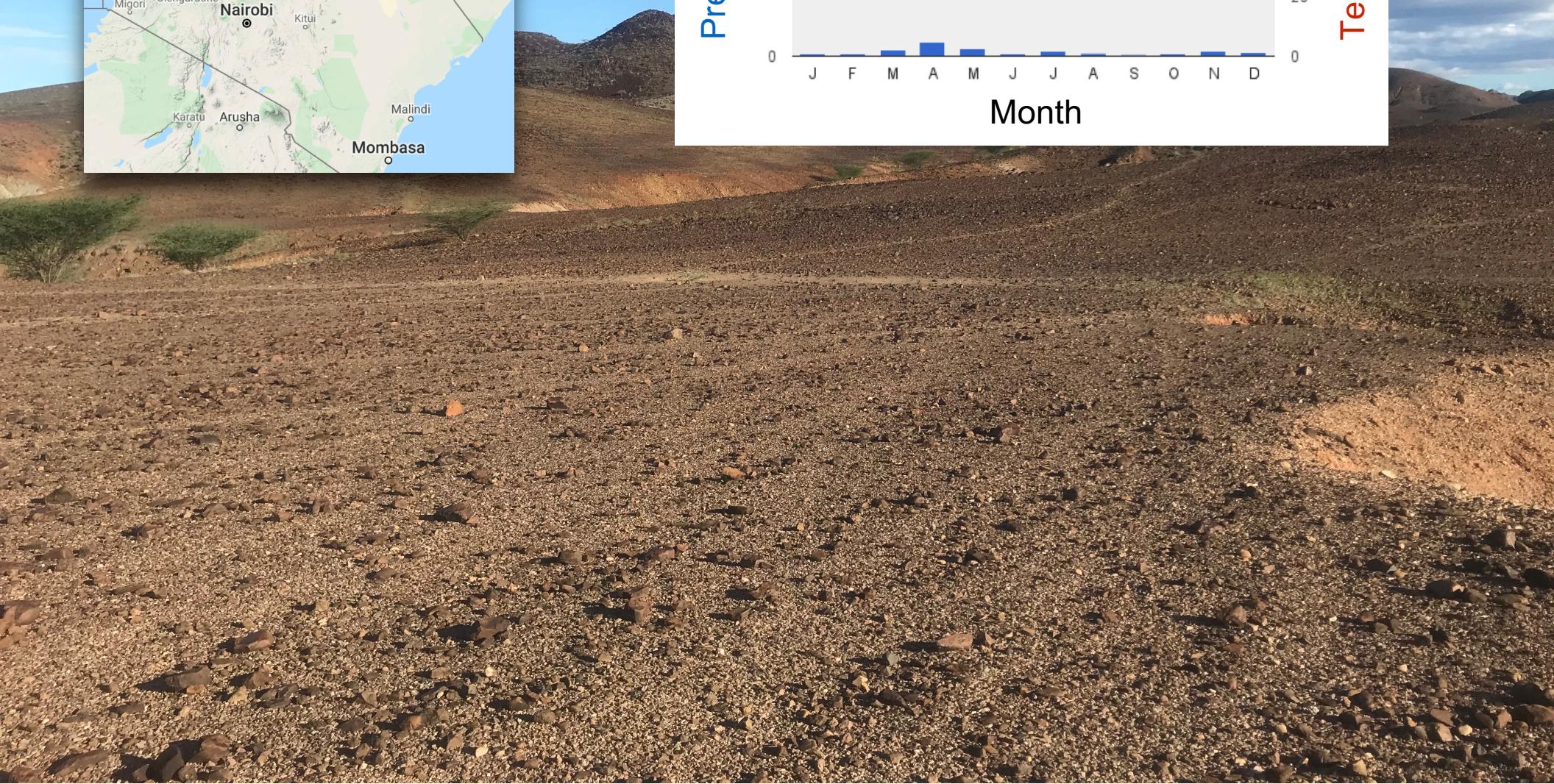
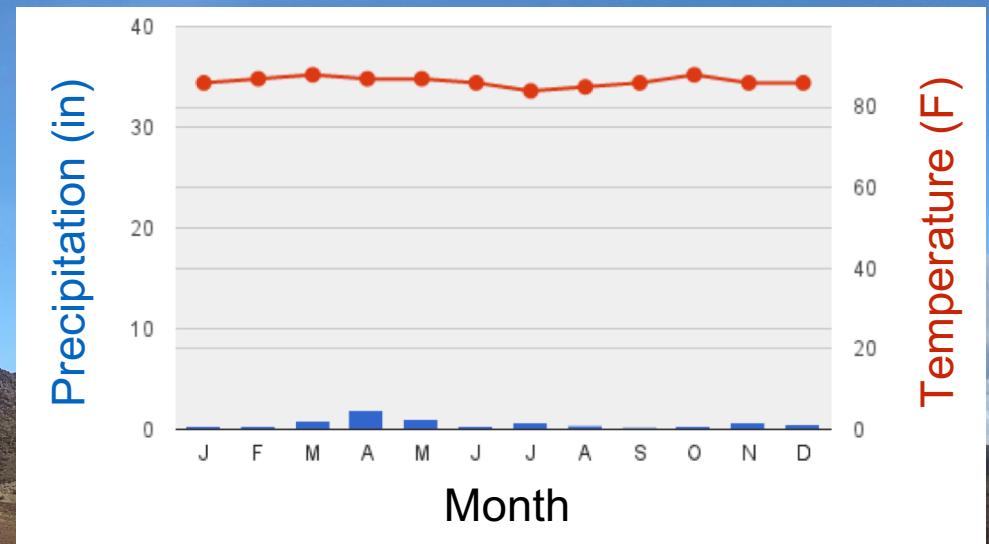


Evolutionary genomics and natural selection in the Turkana people of northern Kenya



Lake Turkana, Kenya

Study population: the Turkana



Study population: the Turkana

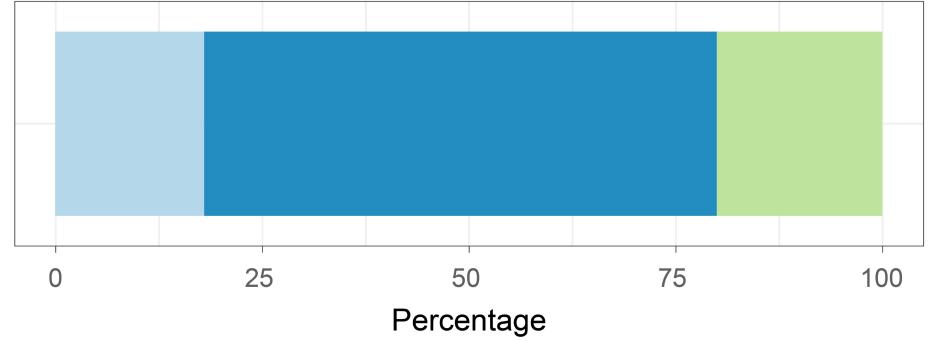


Study population: the Turkana

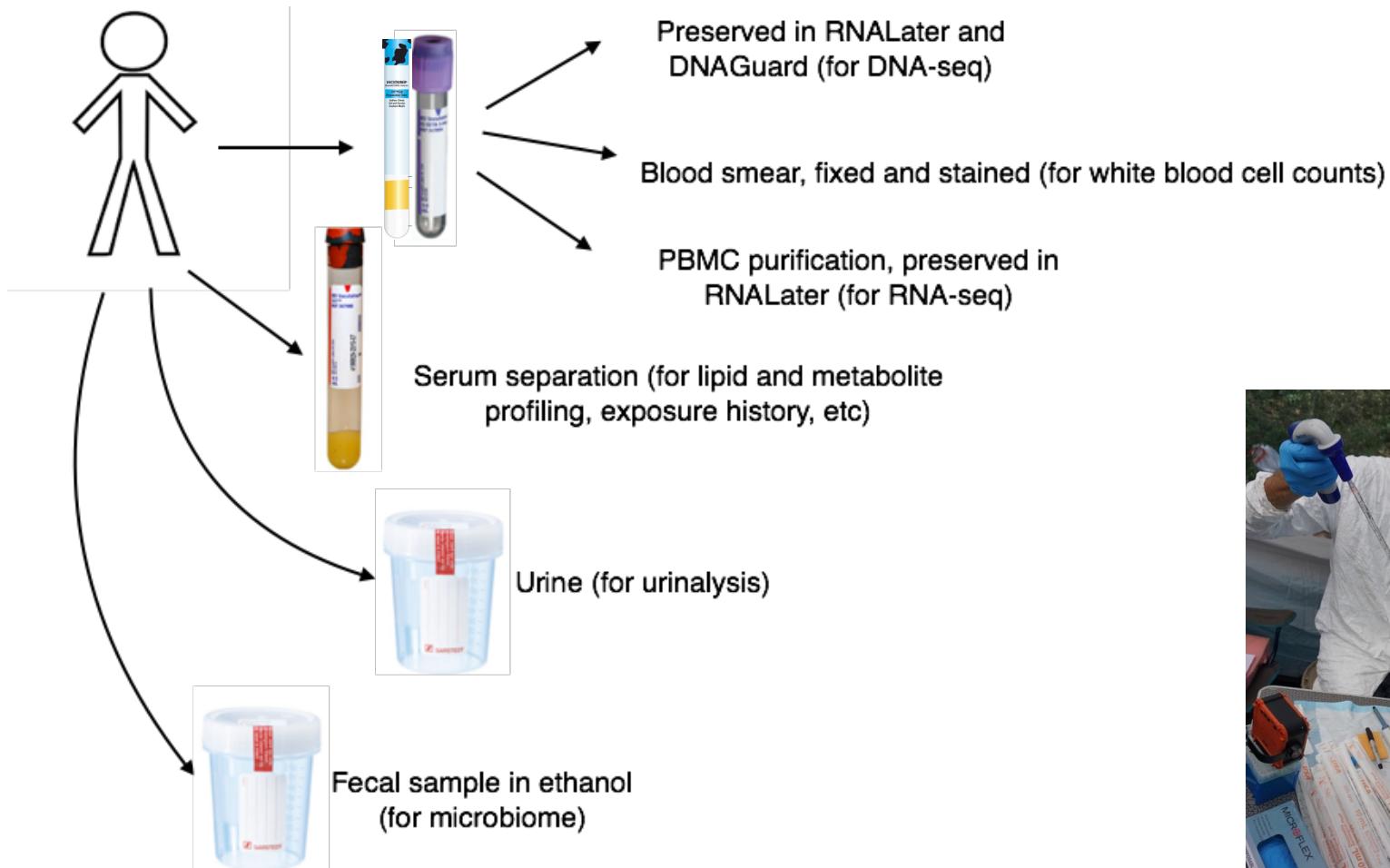


Dietary item

- Animal meat, blood, or fat
- Animal milk or milk products
- Other



Sampling

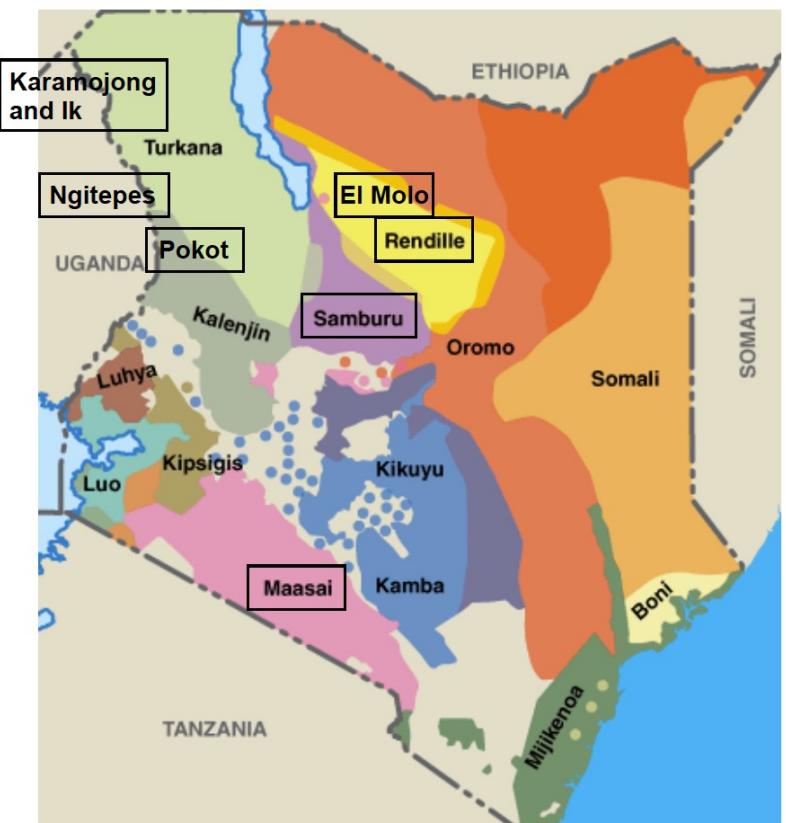


Signature of natural selection in Turkana genomes

Number of Genomes Sequenced



| Group | Total |
|------------|-------|
| Turkana | 301 |
| EI Molo | 7 |
| Karamojong | 7 |
| Masaai | 7 |
| Ngitepes | 8 |
| Pokot | 7 |
| Rendille | 8 |
| Samburu | 7 |
| Ik | 5 |



How have the Turkana adapted to their arid, pastoralist lifestyle?

| Group | Whole genome sequences |
|------------|------------------------|
| Turkana | 301 |
| EI Molo | 7 |
| Karamojong | 7 |
| Masaai | 7 |
| Ngitopes | 8 |
| Pokot | 7 |
| Rendille | 8 |
| Samburu | 7 |
| Ik | 5 |



*Generate a high coverage reference panel
(n=106; ~12M variants called from ~20x data with GATK)*

TCTATTGTTCAA
TCTATTGTTCAC
CCCTCCTCCCTCC
CCCTCCTCCCTCC
TCCTCCTCCCTCC
TCCATTTTCAA
TCCATTTTCAC

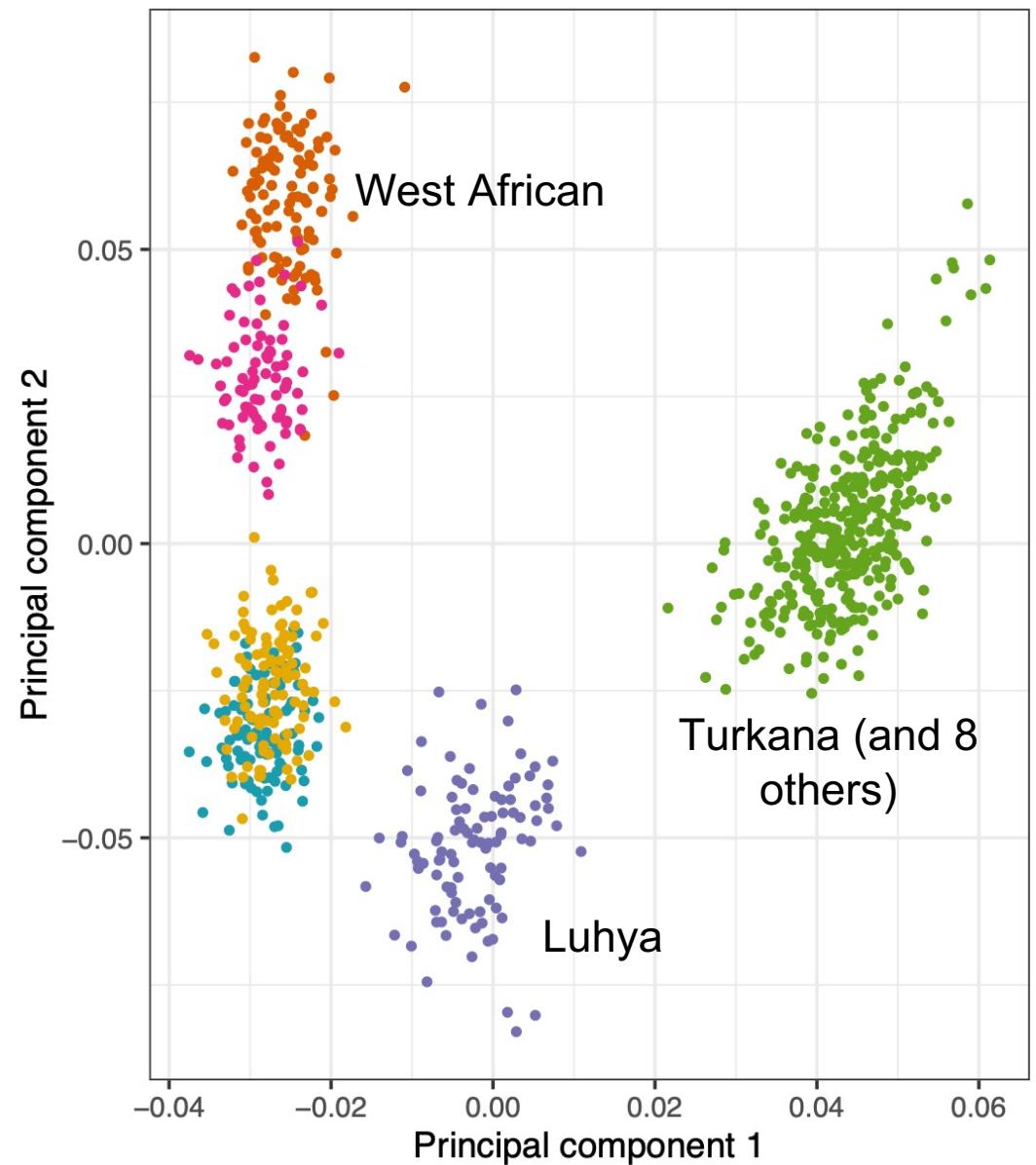
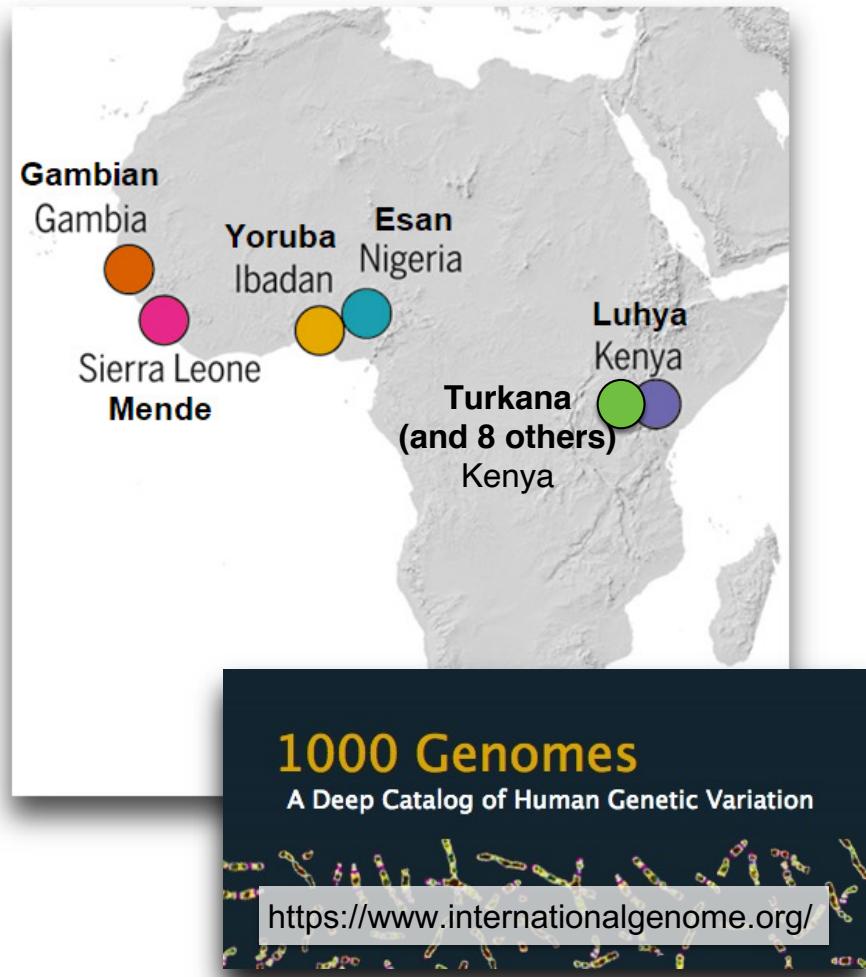
*Sequence remaining samples at low coverage
(n=261; ~6x data)*

---A_G_C---
---T_T_T---

*Impute missing data
(SHAPEIT2 and IMPUTE2)*

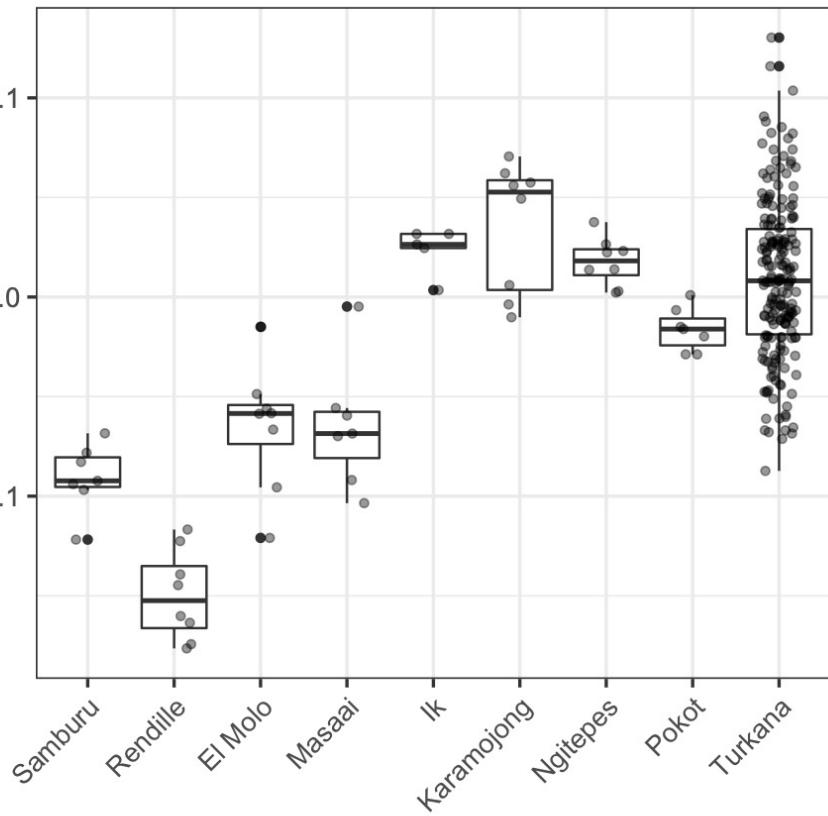
TCTATTGTTCA~~A~~
~~C~~CCCTCCTCCCTCC

Genetic variation in the Turkana and other African groups

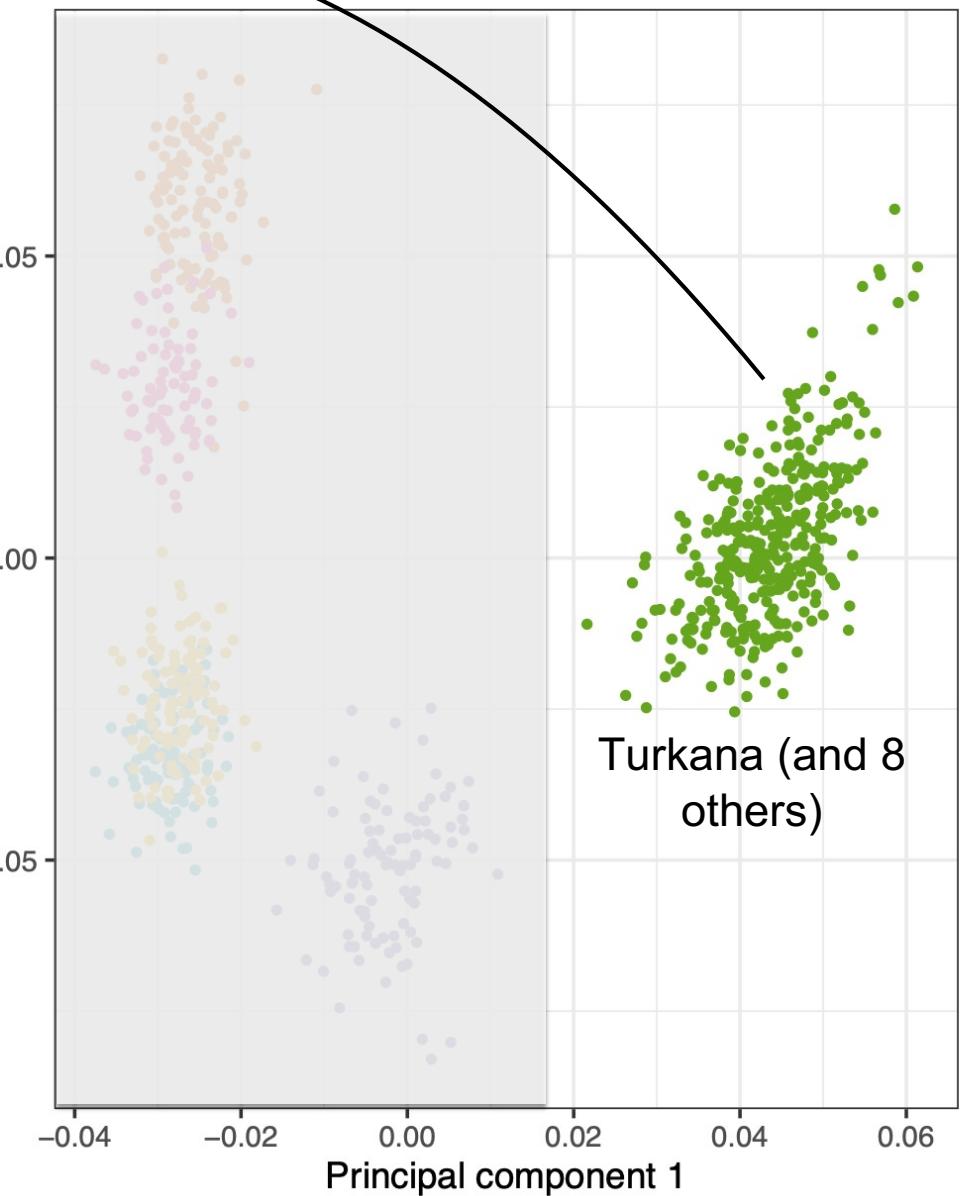


Genetic variation in the Turkana and other African groups

Principal component 2

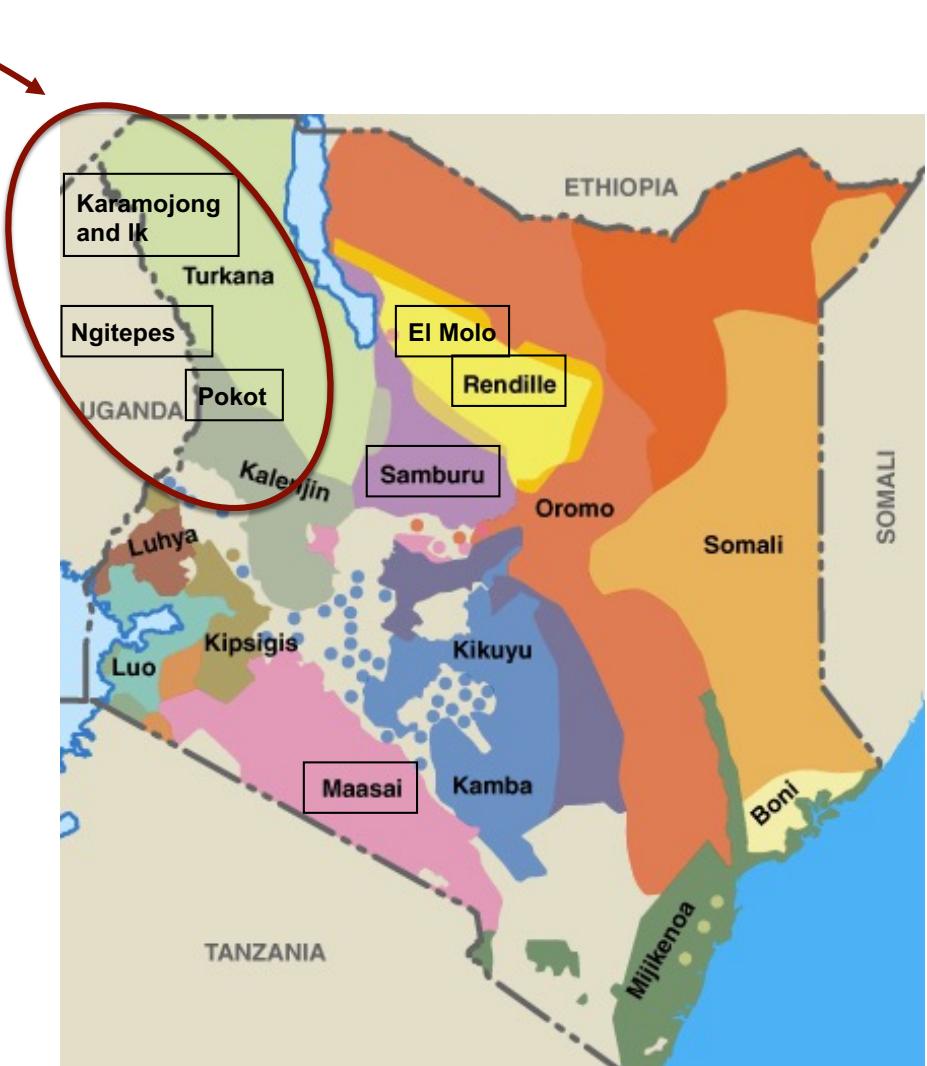
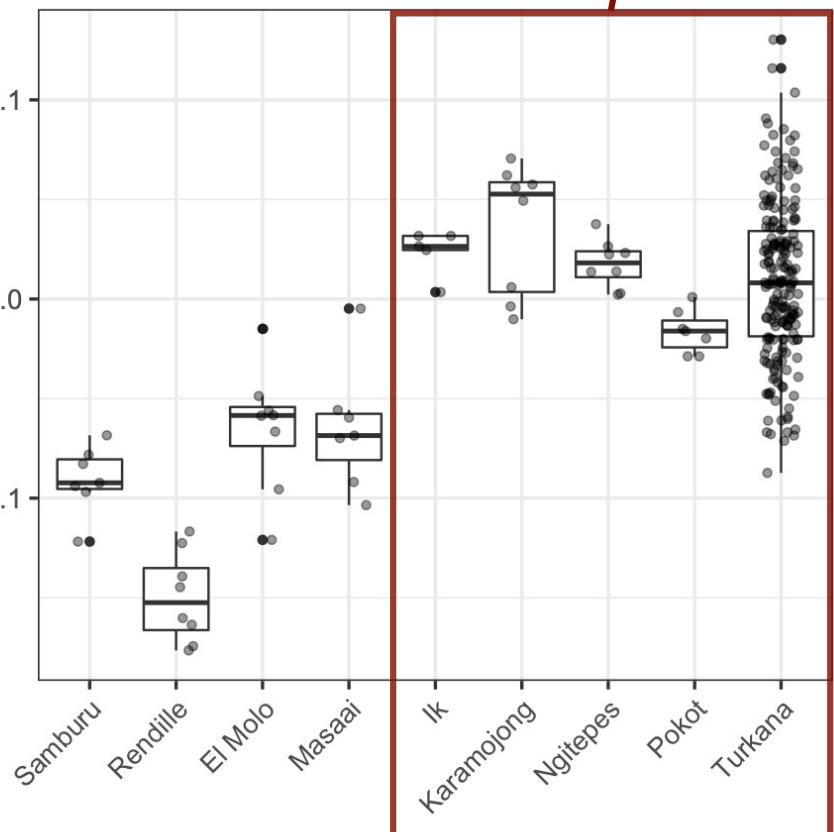


Principal component 2



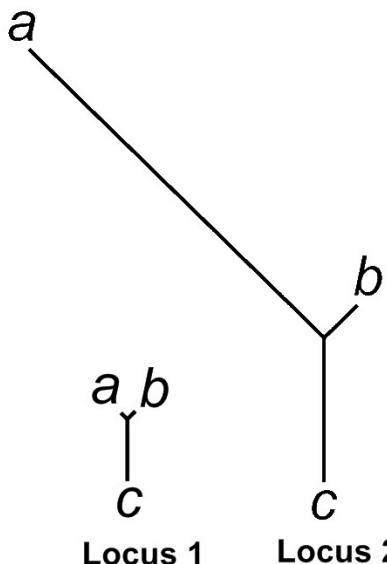
Genetic variation in the Turkana and other African groups

Principal component 2

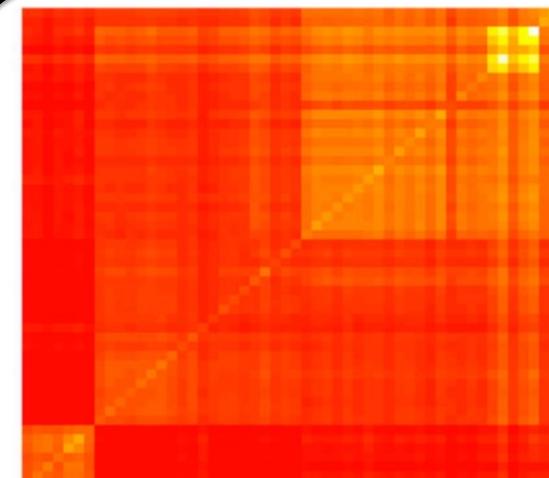


Measuring for positive selection

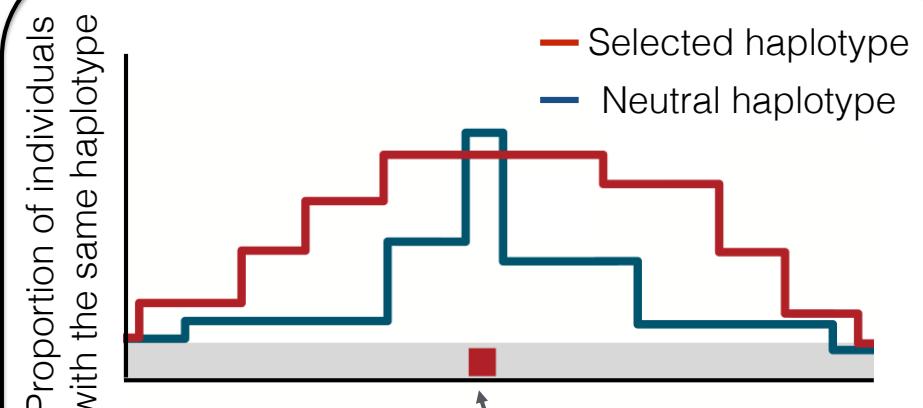
Fisher's combined score (FCS)



Population branch statistic (PBS):
Yi et al. 2010, *Science*



BayEnv2 XtX:
Günther and Coop 2013, *Genetics*



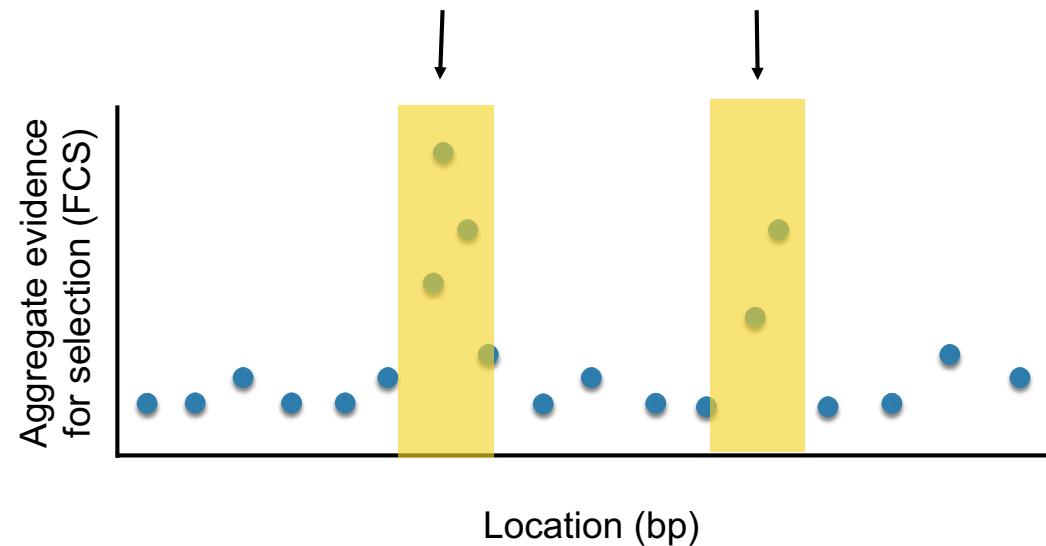
Integrated haplotype score (iHS):
Voight et al. 2006, *PLoS Biology*

Testing for polygenic adaptation

Genes involved in a given molecular function

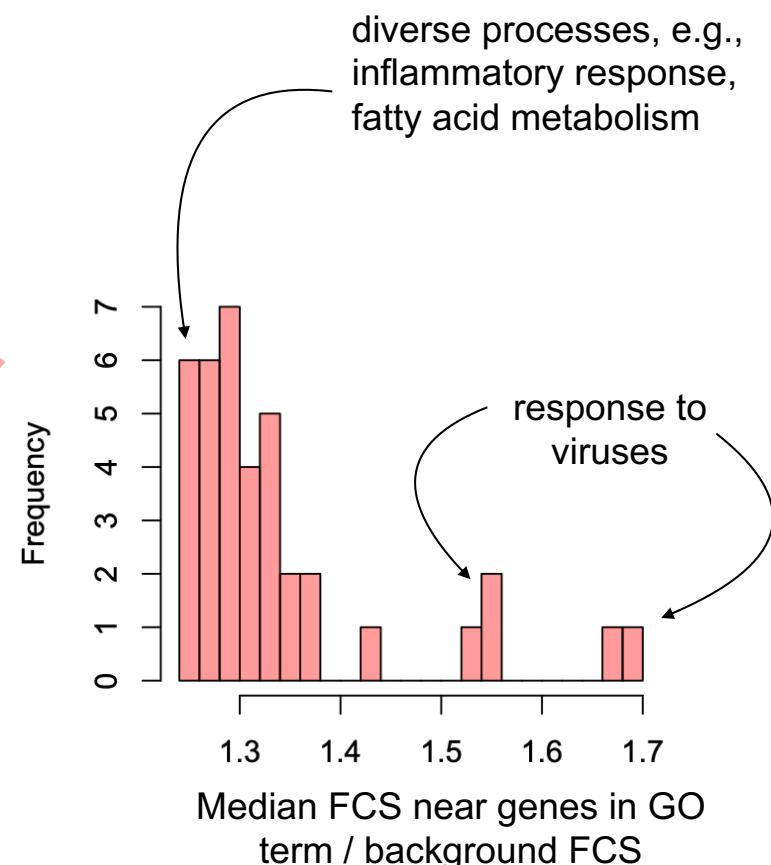
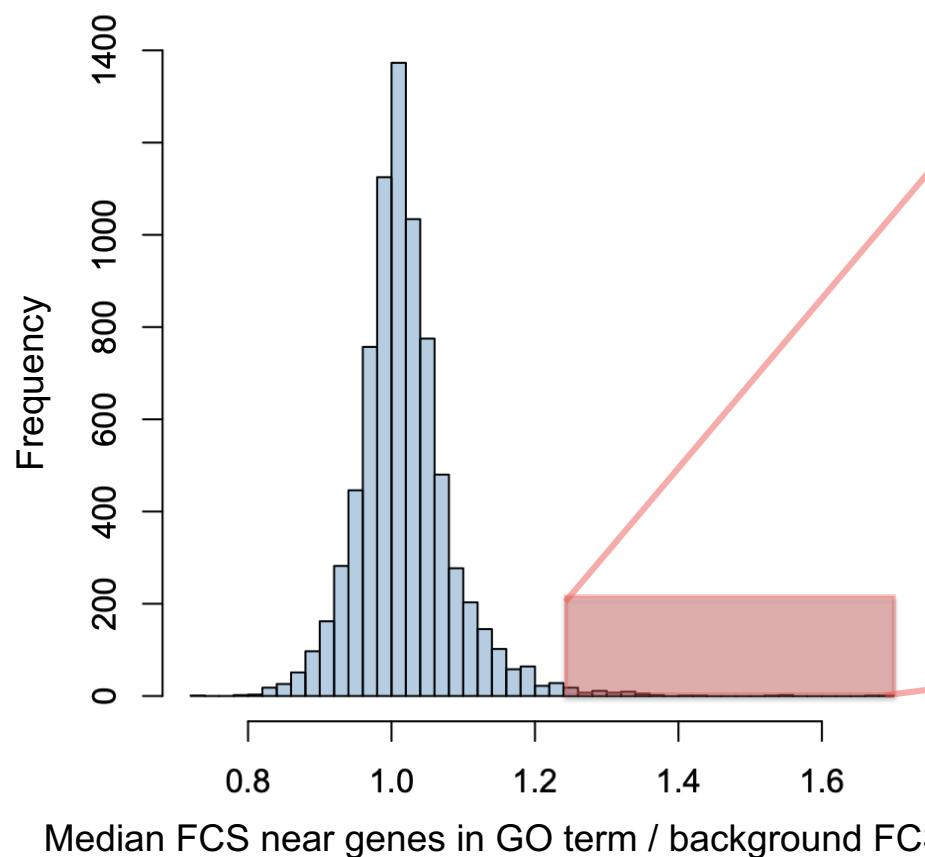


SNPs associated with biomarkers of cardiometabolic and renal system function

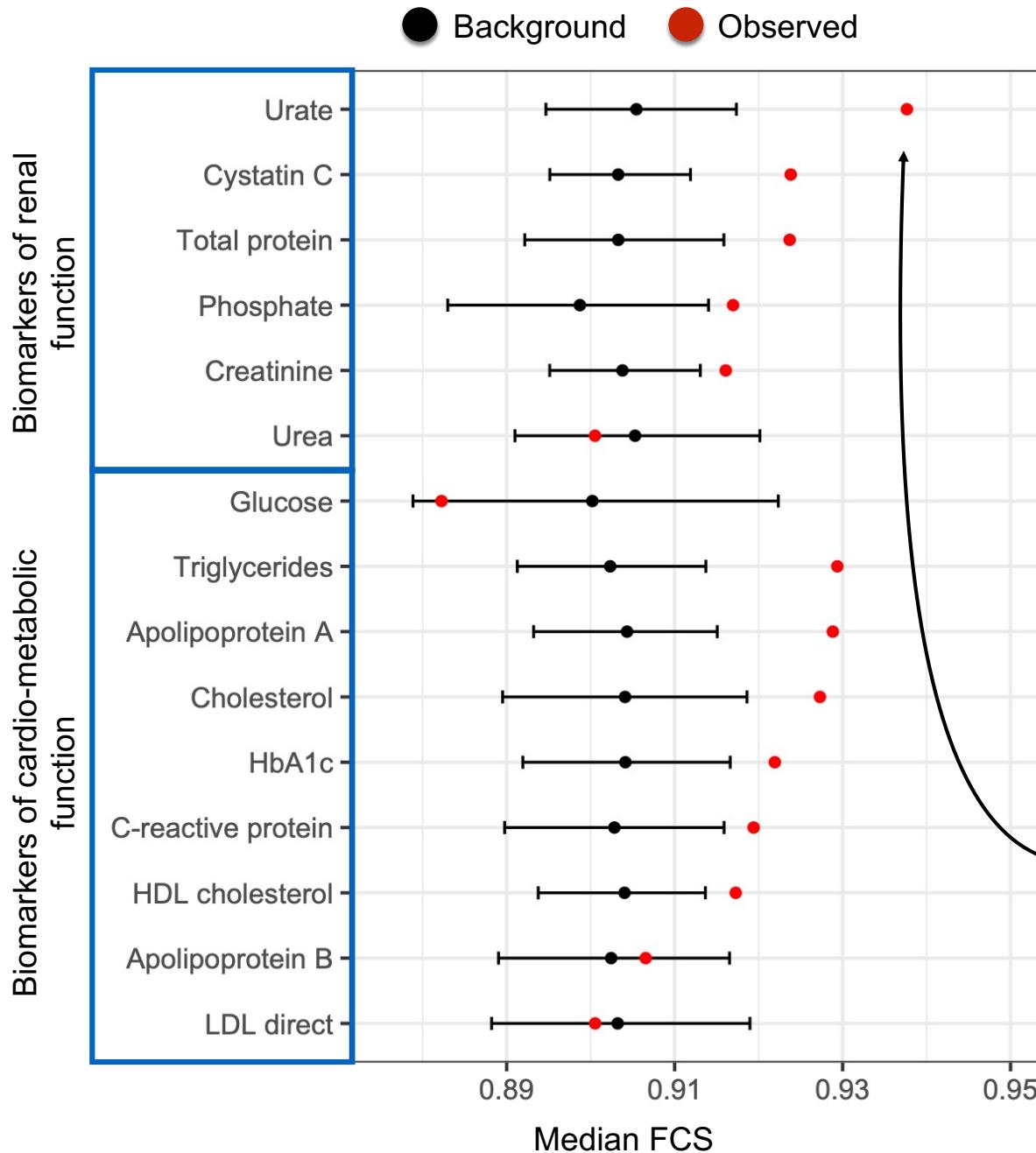


Testing for polygenic adaptation

Genes involved in a given molecular function



Testing for polygenic adaptation



SNPs associated with biomarkers
of cardiometabolic and renal
system function

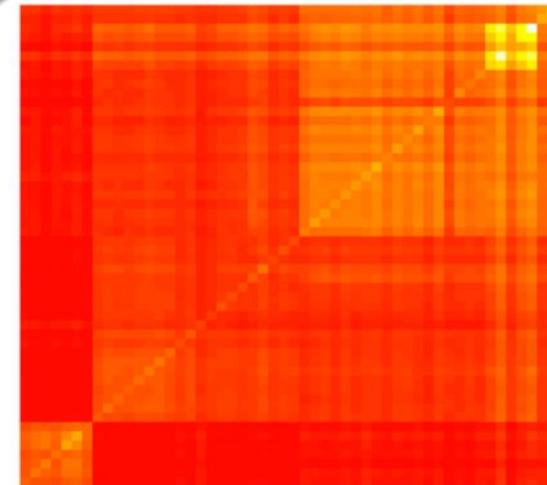


Testing for positive selection

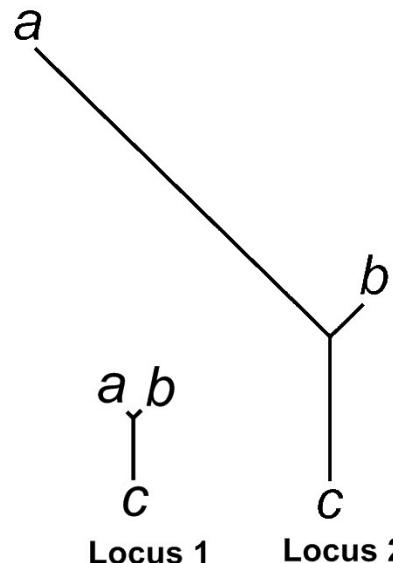
50kb regions with many outliers (>99th percentile):

CCDC102B, SEMA6A*, TMEM132D, ZNF385D, SGCZ, CSMD1, STC1*

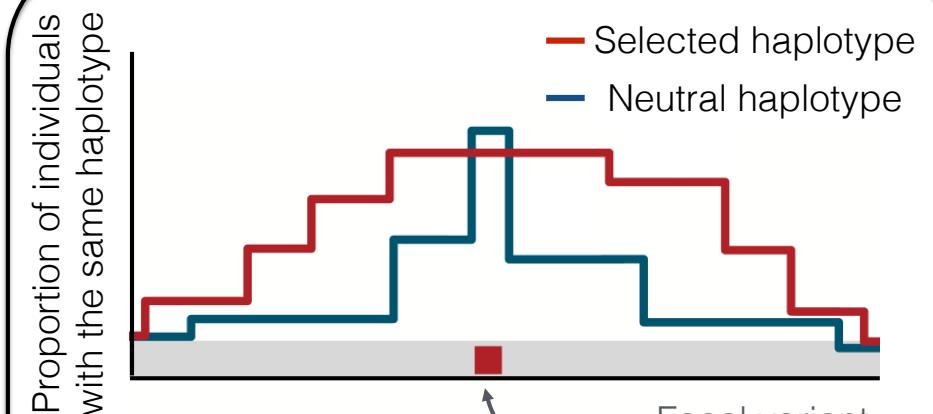
*Wagh et al. 2012, *PLoS One*



BayEnv2 XtX:
Günther and Coop 2013, *Genetics*

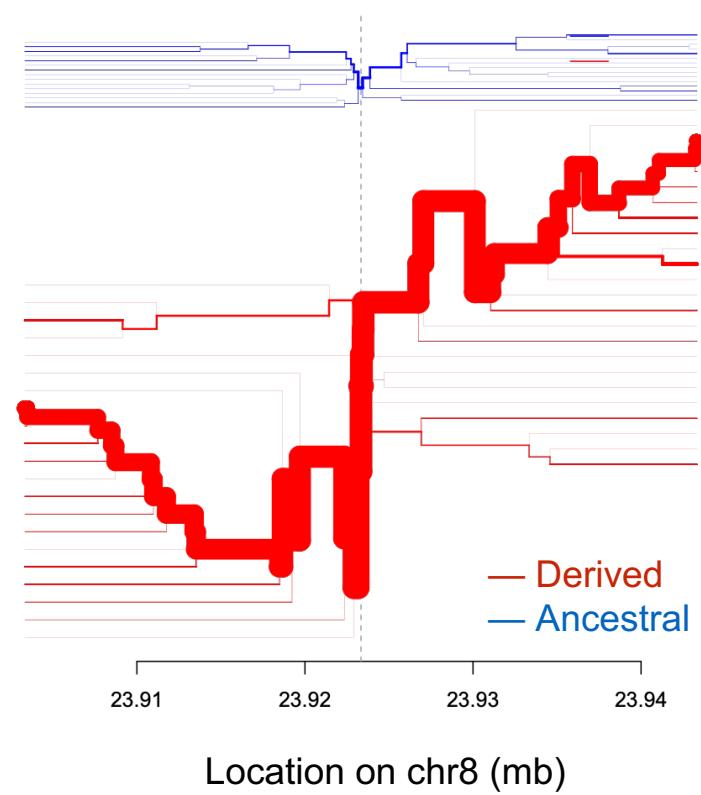
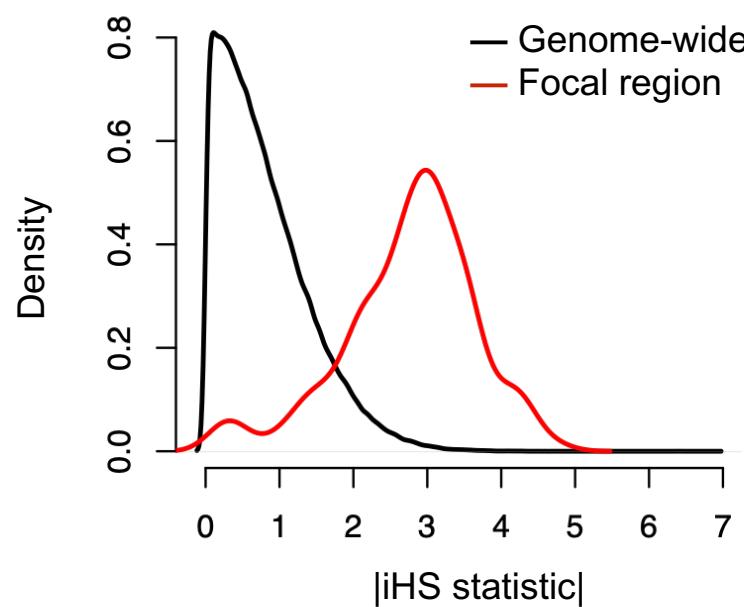
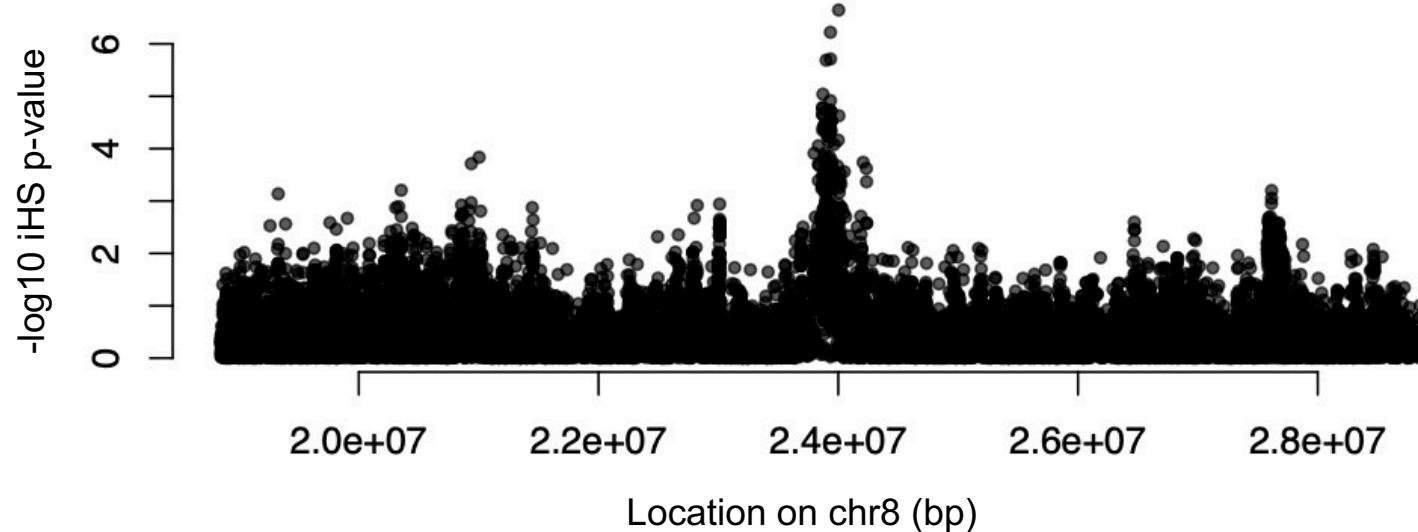


Population branch statistic (PBS):
Yi et al. 2010, *Science*

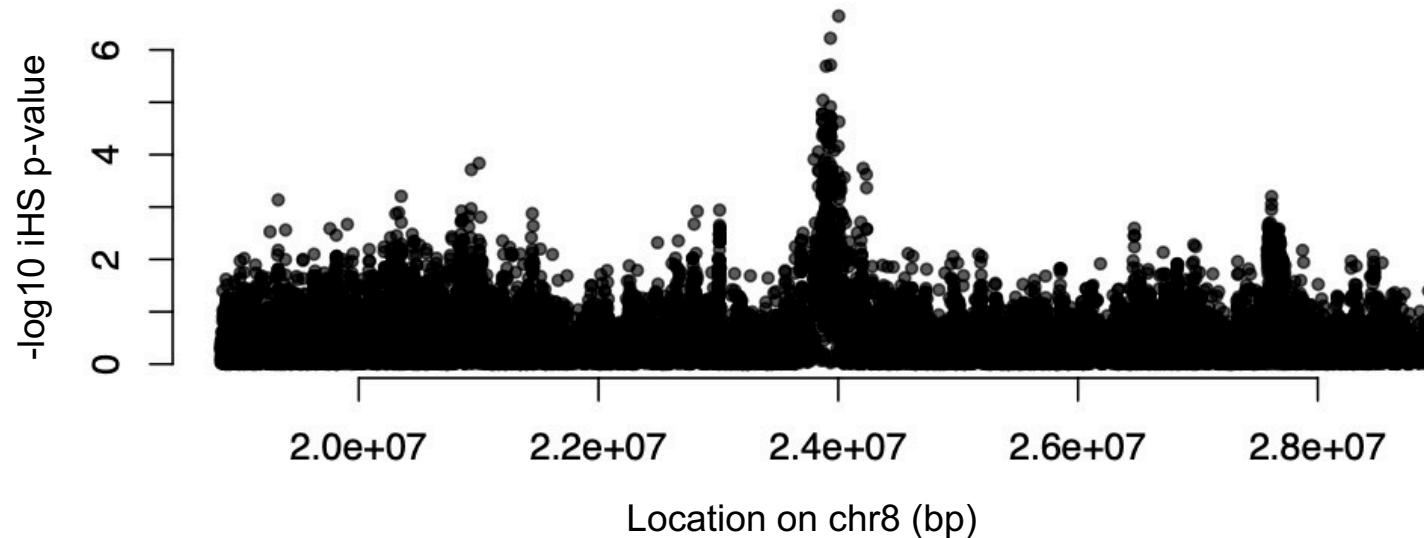


Integrated haplotype score (iHS):
Voight et al. 2006, *PLoS Biology*

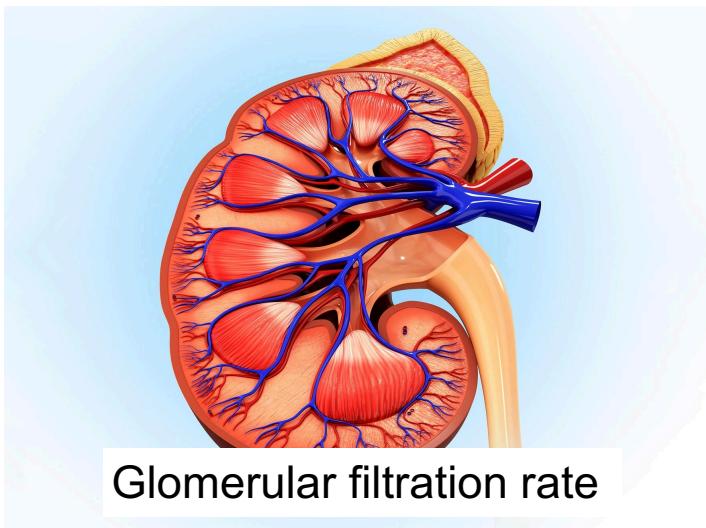
Best candidate region for positive selection: STC1



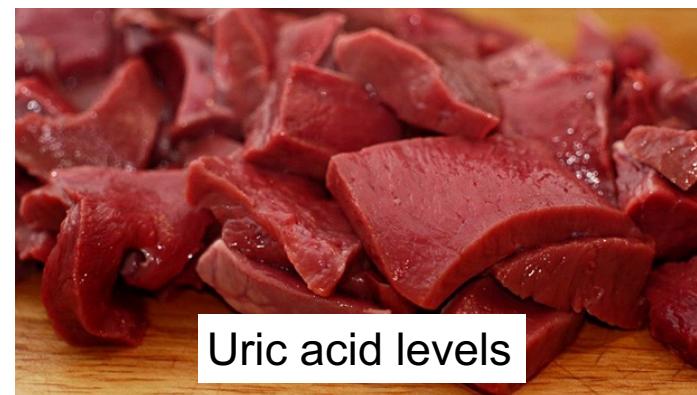
Best candidate region for positive selection: *STC1*



In GWAS, *STC1* is associated with:

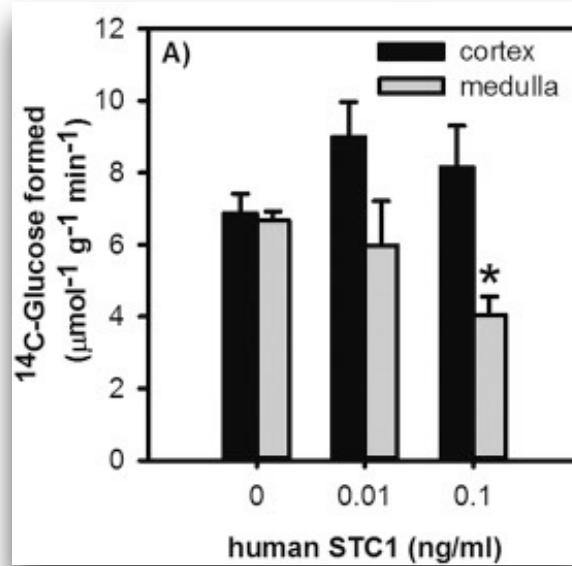


Glomerular filtration rate

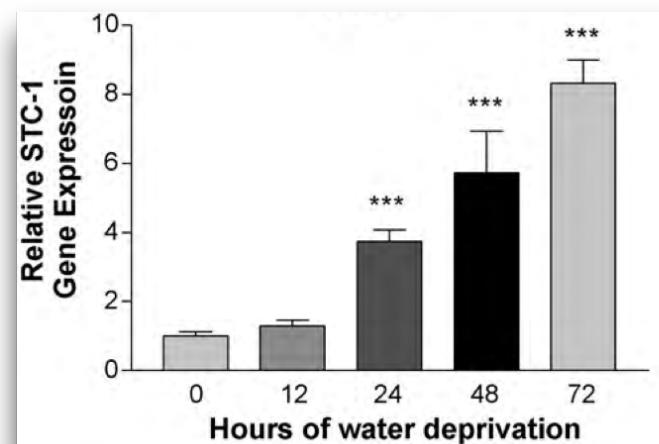


Uric acid levels

Best candidate region for positive selection: *STC1*

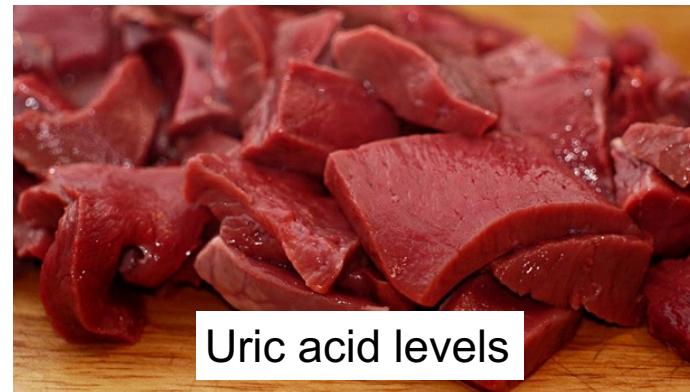
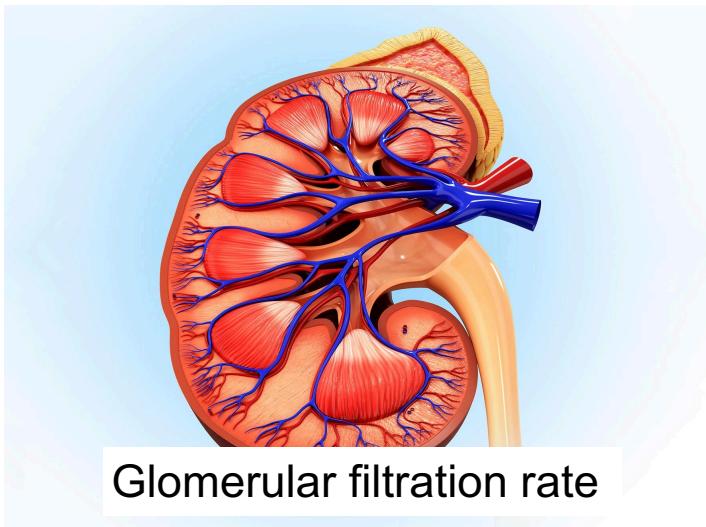


Schein et al. 2015, *Molecular and Cellular Endocrinology*

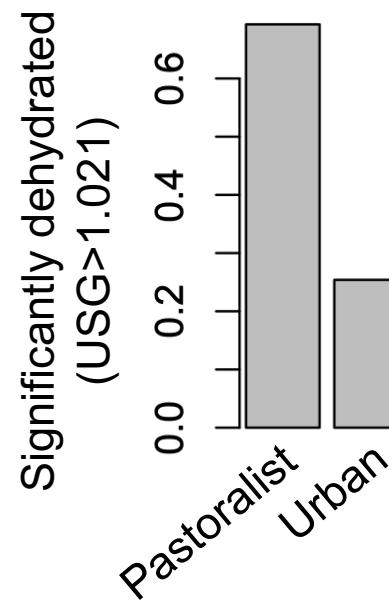


Turner et al. 2010, *Molecular and Cellular Endocrinology*

In GWAS, *STC1* is associated with:

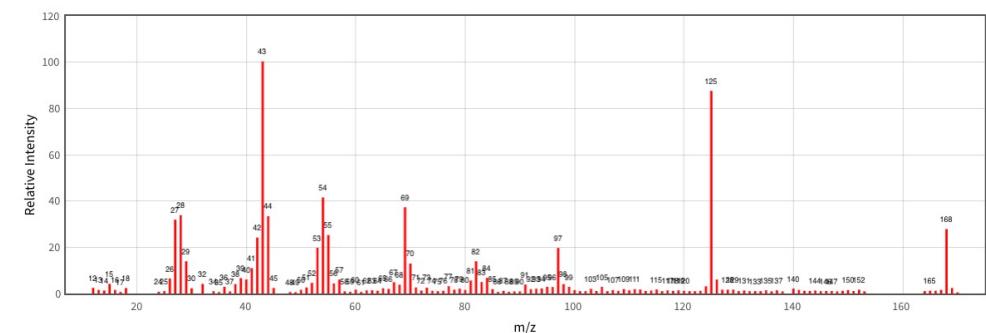
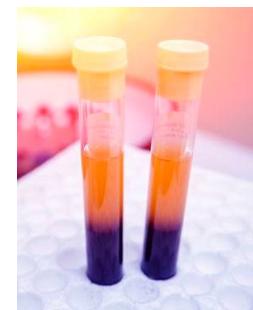


Study population: the Turkana



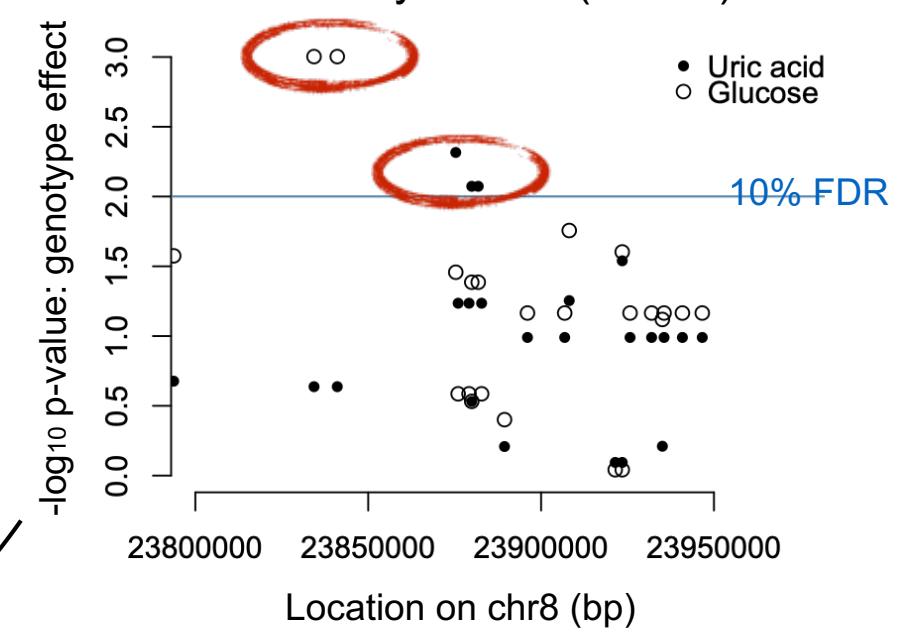
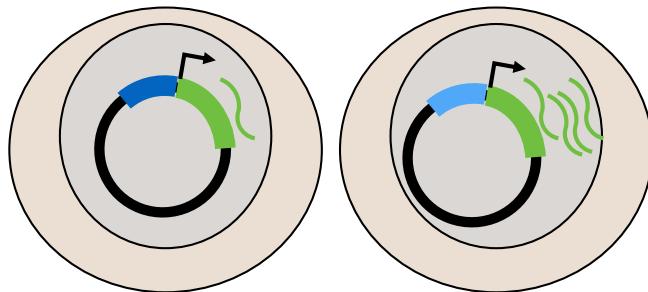
n=905

Functional validation: work in progress



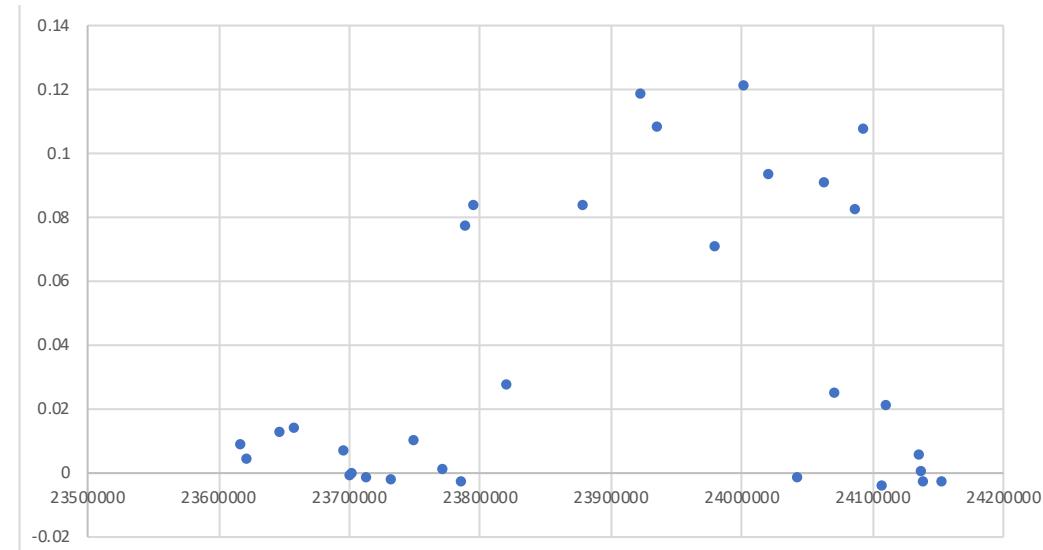
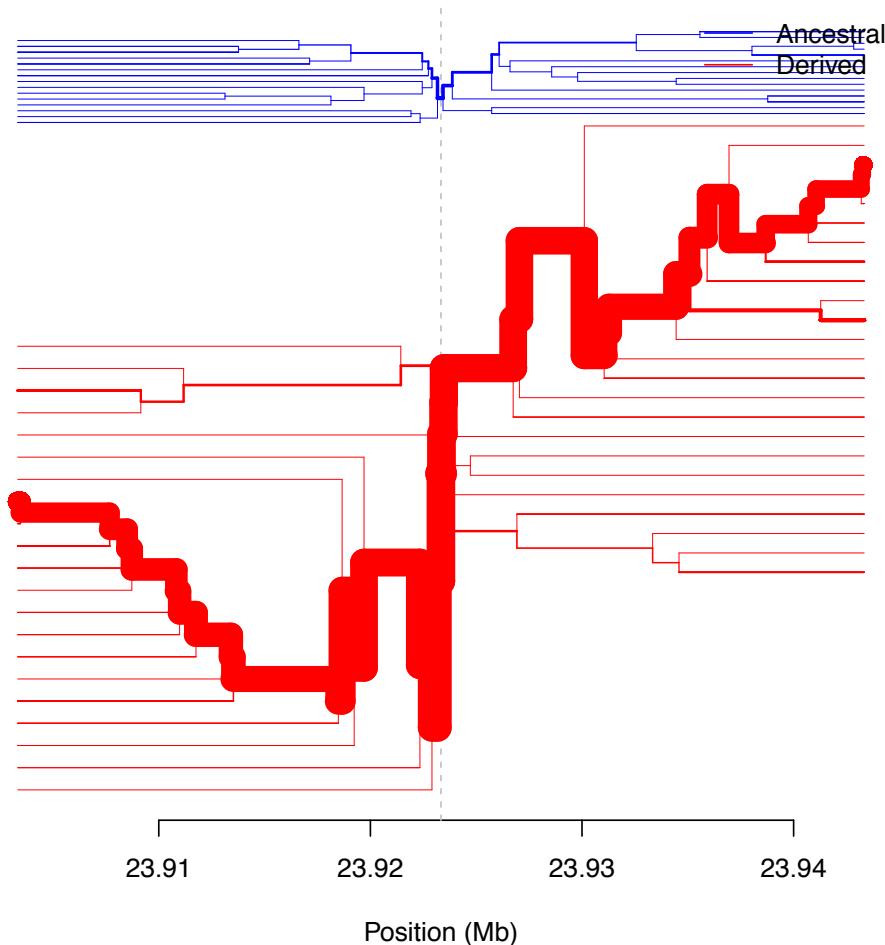
1) Association mapping: genotypes + serum & mass spectrometry

2) Reporter assays in kidney cell lines



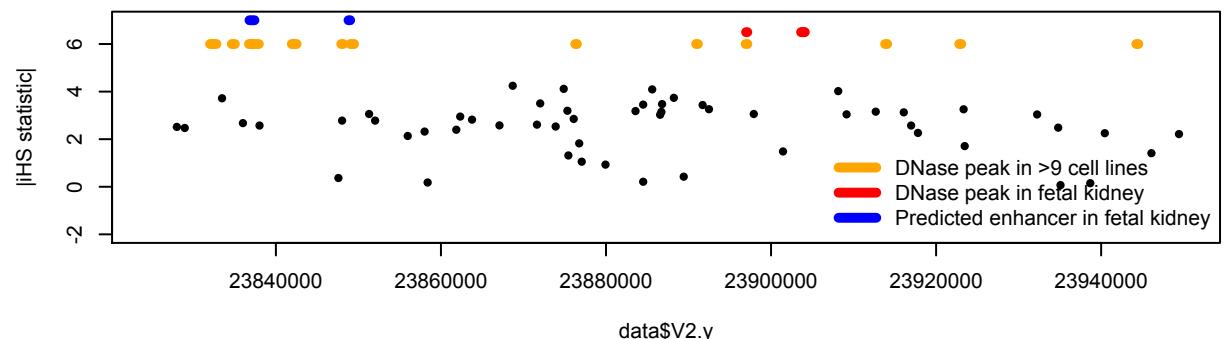
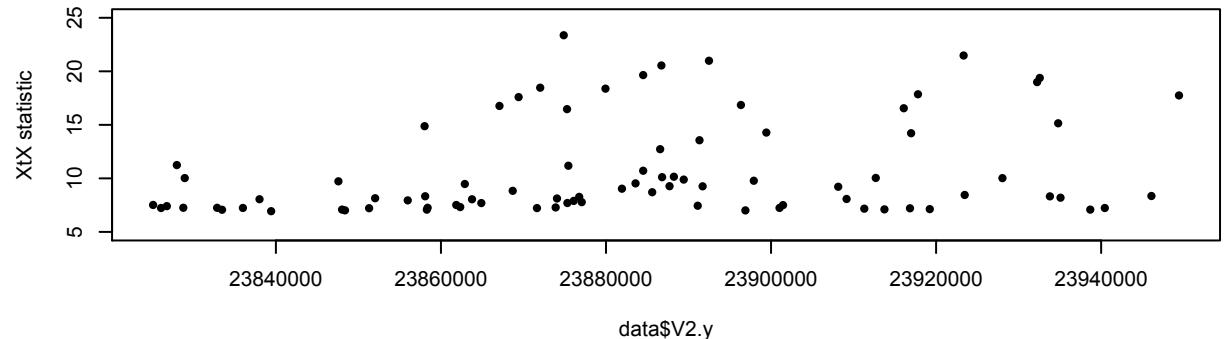
Selection scan results - STC1

- Bifurcation plot showing that many individuals carry a similar haplotype in the STC1 region (centered on SNP with strongest iHS signal)
- PBS statistics for the same candidate region using array data (to check that low coverage sequencing didn't induce weird results)

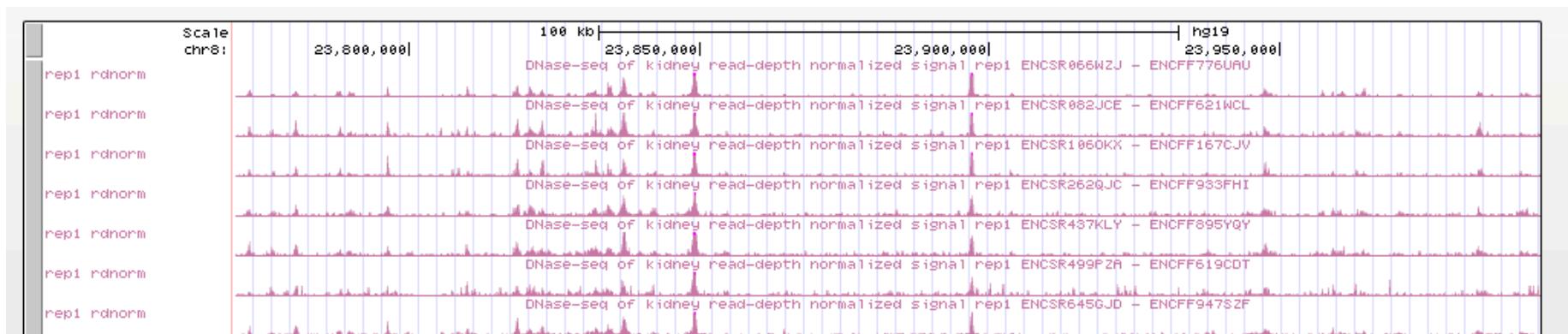


Functional data - STC1

- Same stats as first slide on the y-axis but overlaid with annotation data from ENCODE



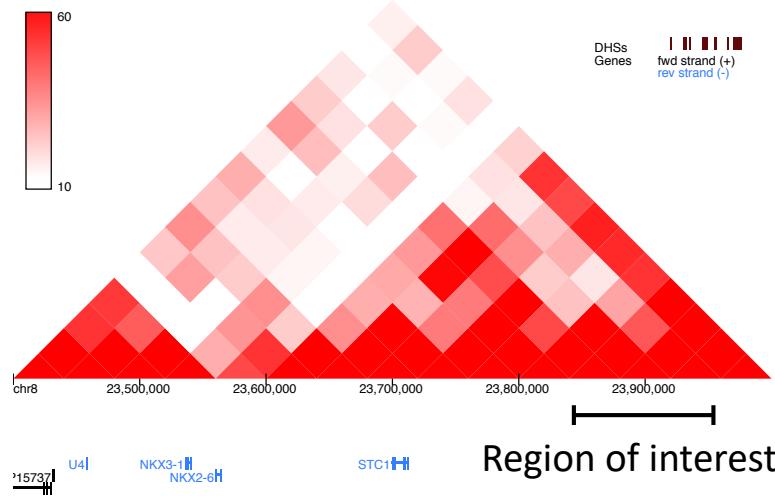
- DNase data from fetal kidney samples from ENCODE (http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr8%3A23770064%2D23995054&hgsid=861346719_CcnTRCz9BqlqJdai6Nn0XcQH1gy4)



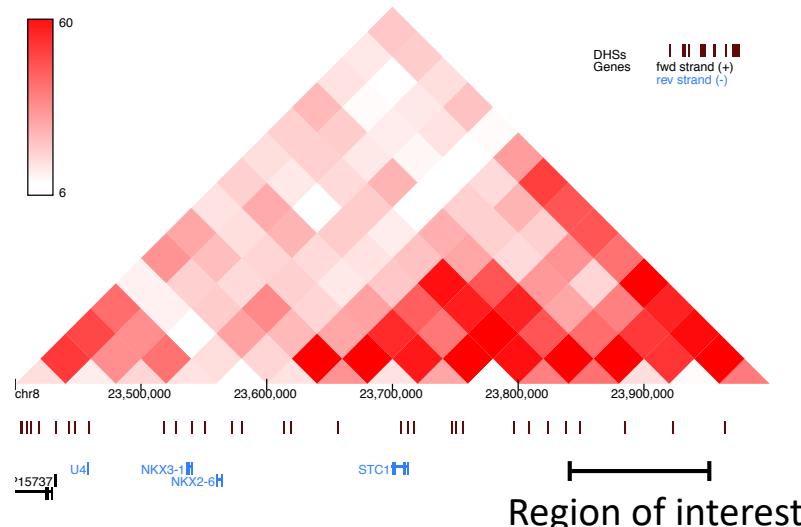
Functional data - STC1

- HiC data from human primary cells (<http://promoter.bx.psu.edu/hi-c/view.php>)
- Plotted are a few cell types closely related to kidney and/or where STC1 is highly expressed (kidney data isn't available from this study)

Liver



Adrenal gland



Heart (left ventricle)

