Epigenetic variation in *Arabidopsis*

Magnus Nordborg
Gregor Mendel Institute
Vienna

The model

- Rapid generation time
- Drosophila-sized genome
- Readily crossable, transformable, etc.

The model *au naturel*...

Natural habitats (Sweden)

October 2013

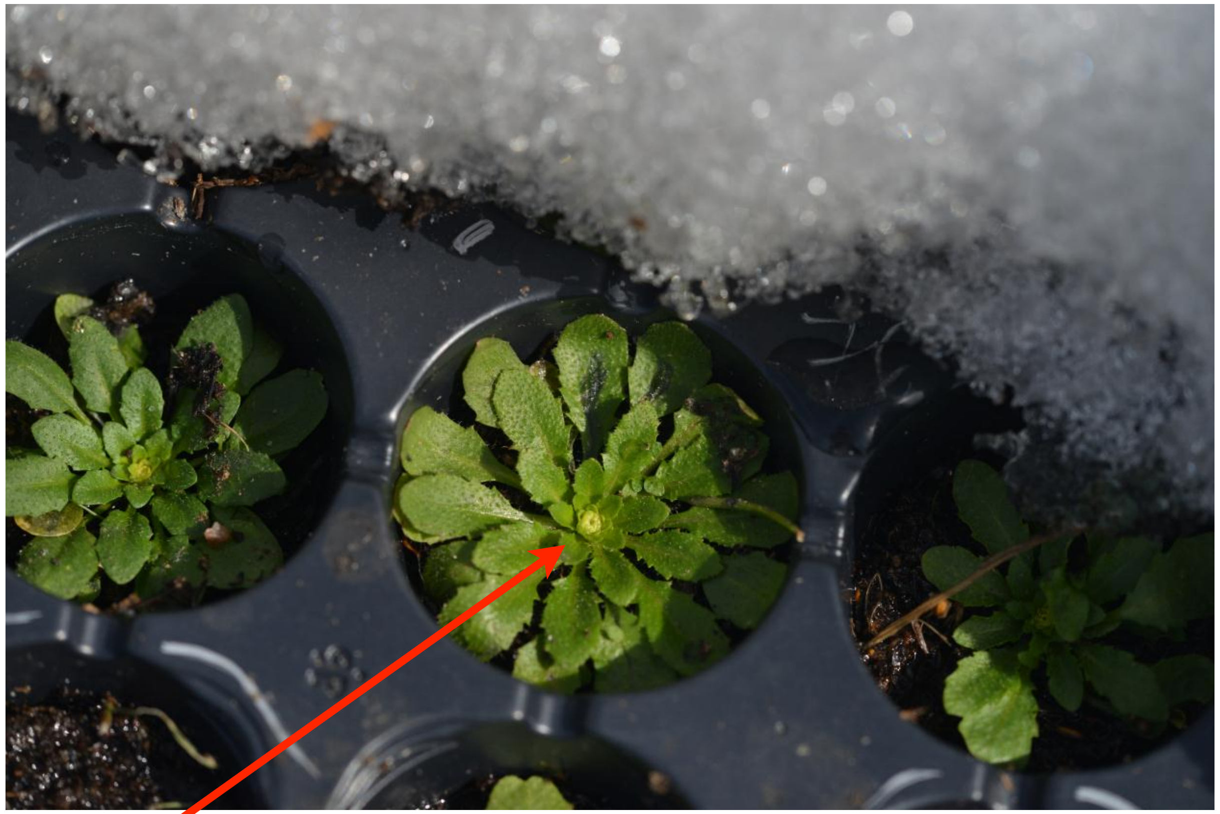
- Seedlings have emerged, and are happily growing at temperatures very different from those in typical laboratories...
- They will overwinter as rosettes (under snow, if they are lucky), and flower early in spring

April 20th, 2014

- Snow is melting rapidly
- Plants in our "common garden" experiments are ready to flower...

April 20th, 2014

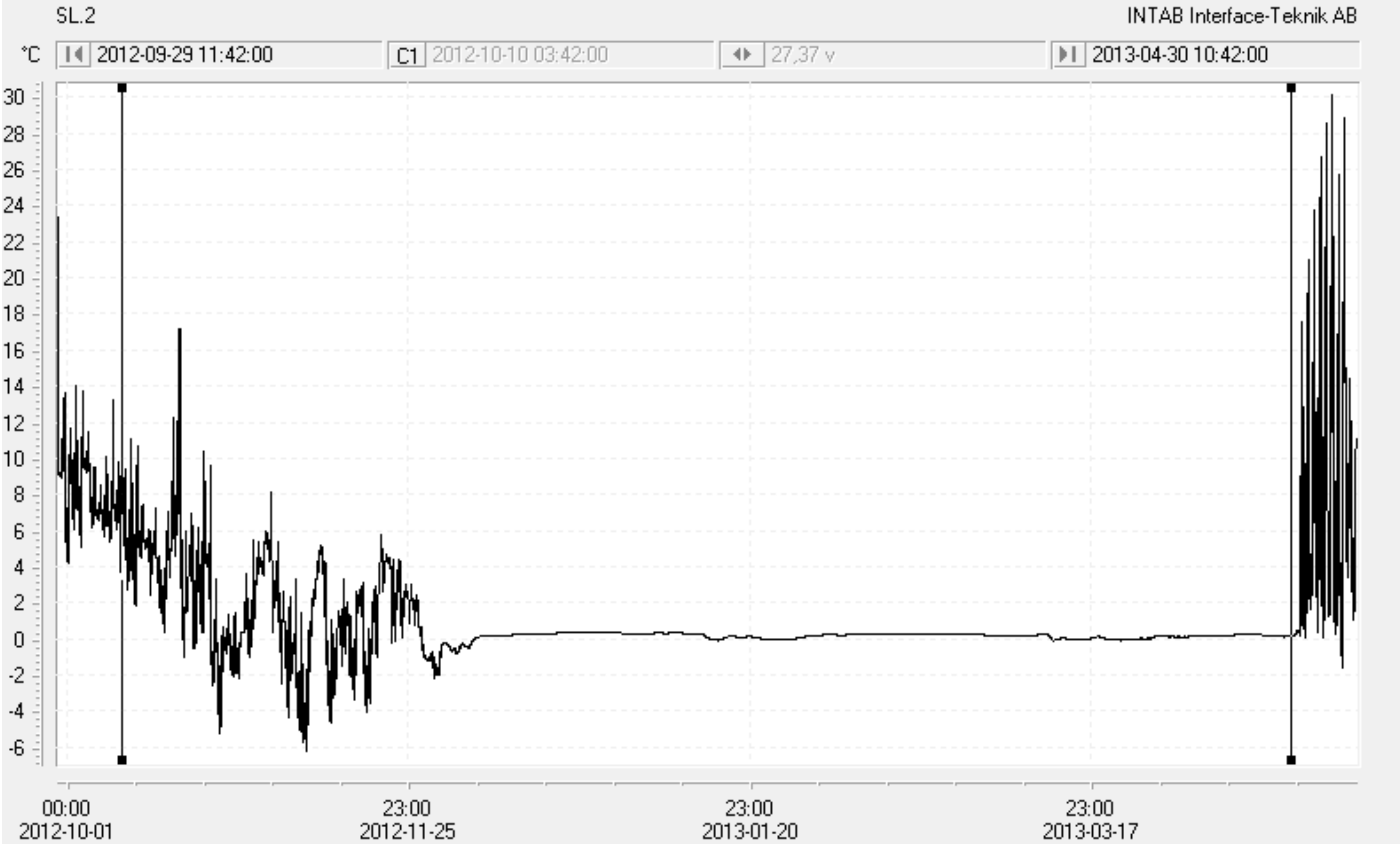
- Snow is melting rapidly
- Plants in our "common garden" experiments are ready to flower...



!

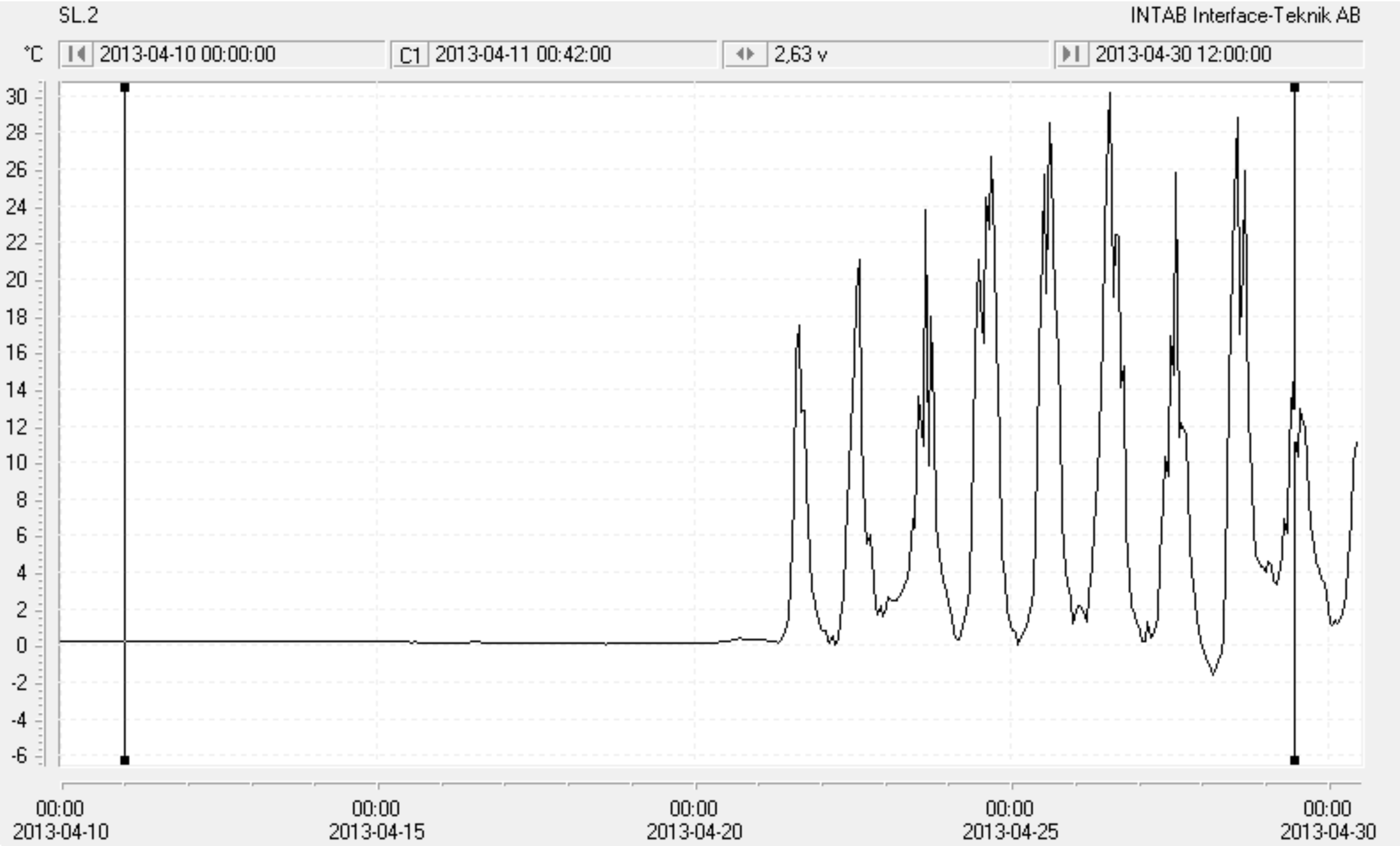
A week later...

Ground temperatures at a northern site



Svante Holm

Ground temperatures at a northern site



SL.2

INTAB Interface-Teknik AB

2013-04-10 00:00:00

C1 2013-04-11 00:42:00

2,63 v

2013-04-30 12:00:00

Svante Holm

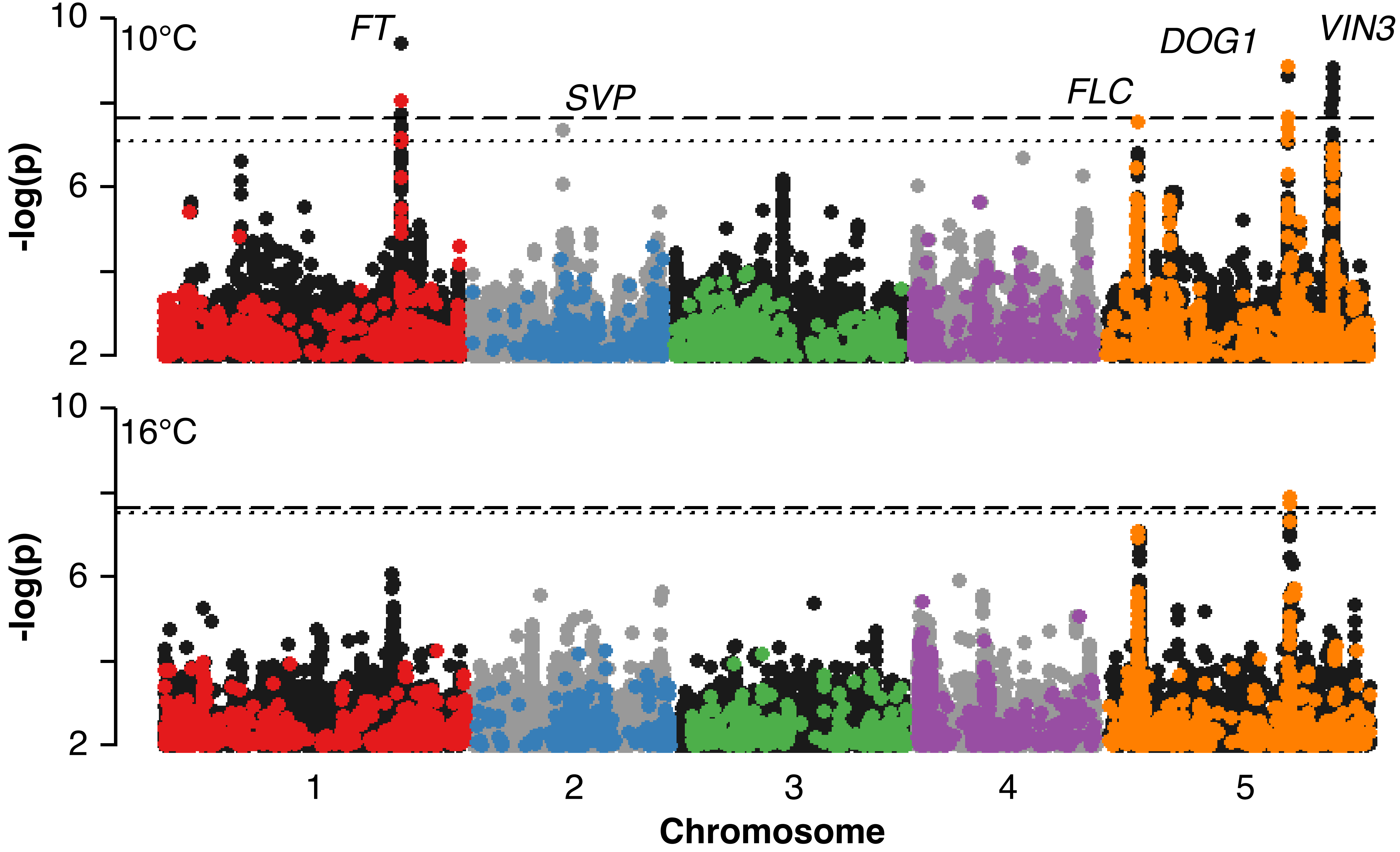
1001 Genomes

- Full* sequences of 1135 accessions, coupled with transcriptomes and methylomes
- Genomes, seeds, and phenotypes
- Tools (http://gwas.gmi.oeaw.ac.at)

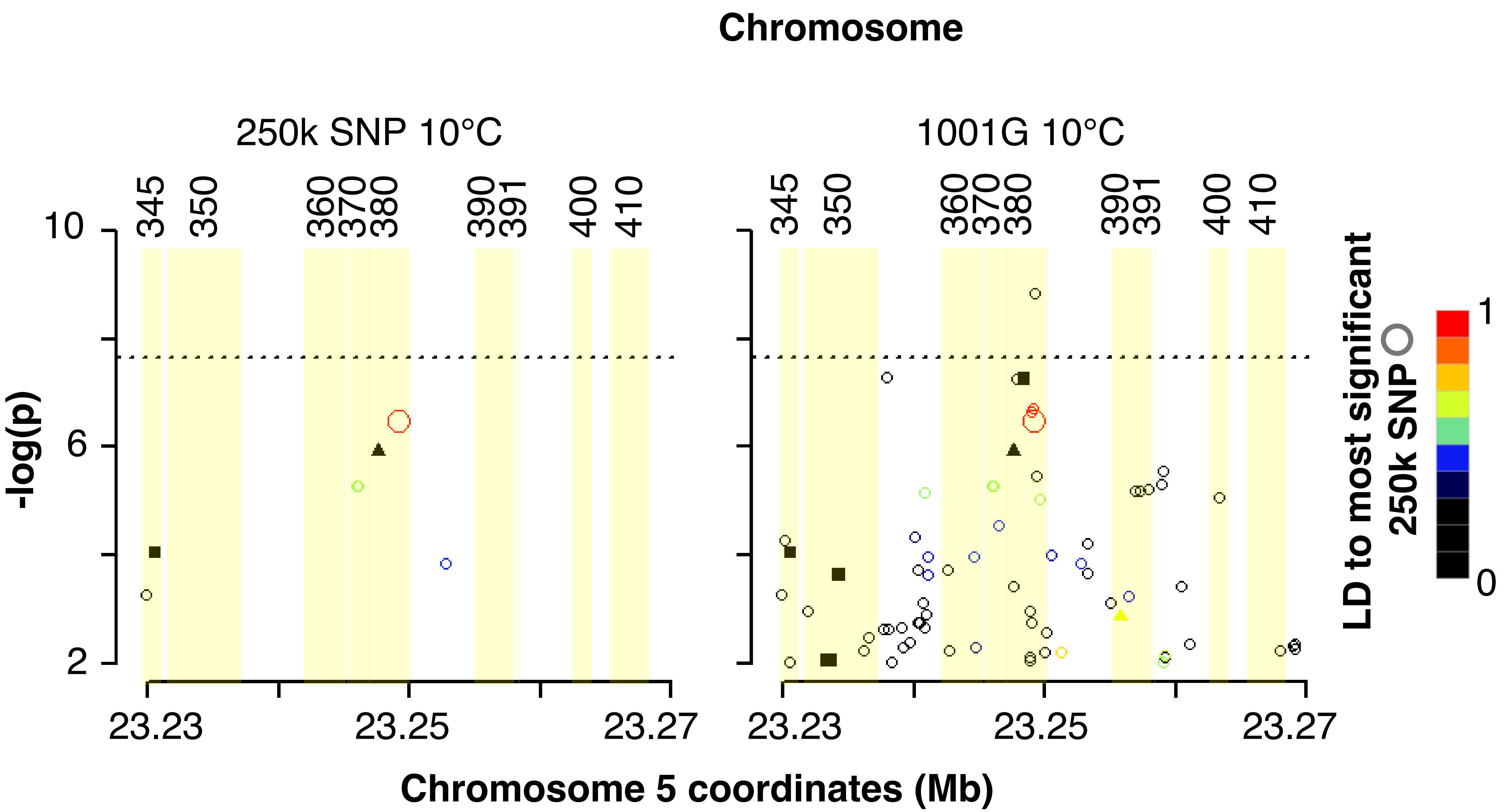
- 1001 Genomes Consortium, *Cell*, 2016
- Kawakatsu *et al*, *Cell*, 2016

**caveat emptor*

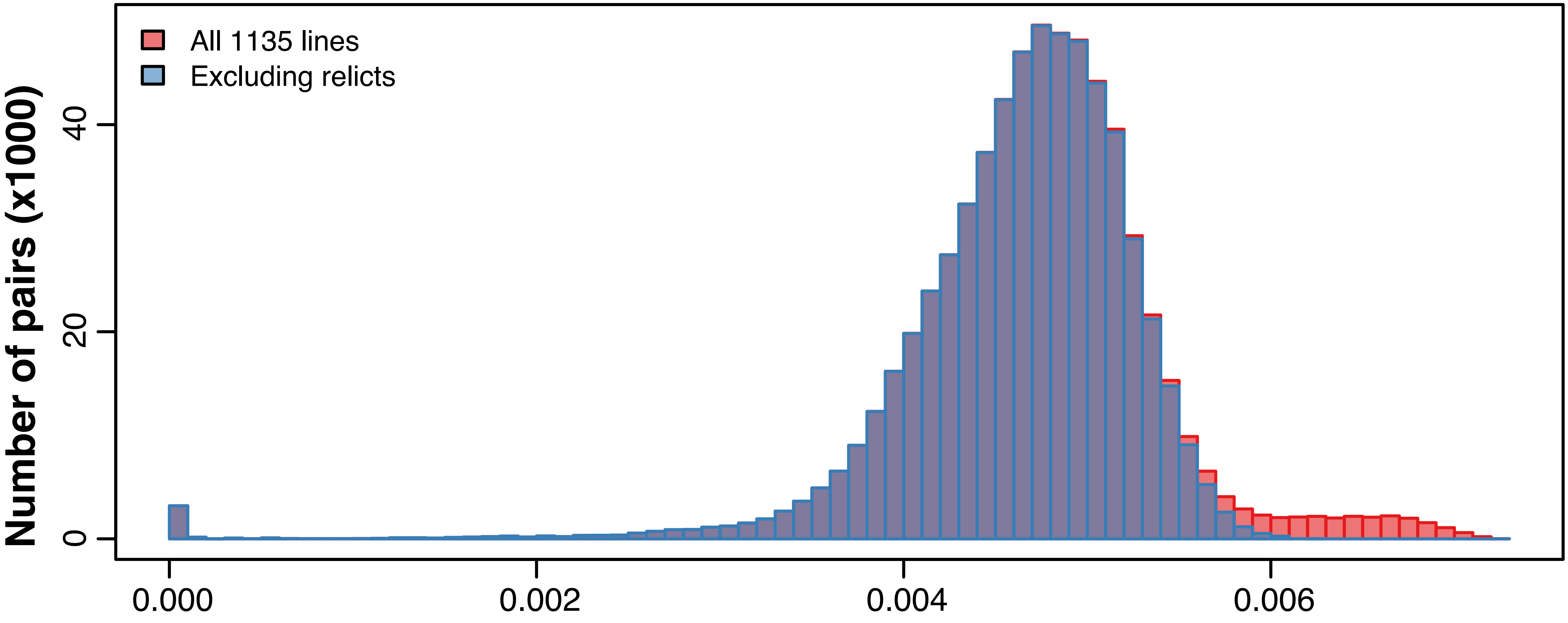
GWAS for flowering



Zoom-in on *SVP* region

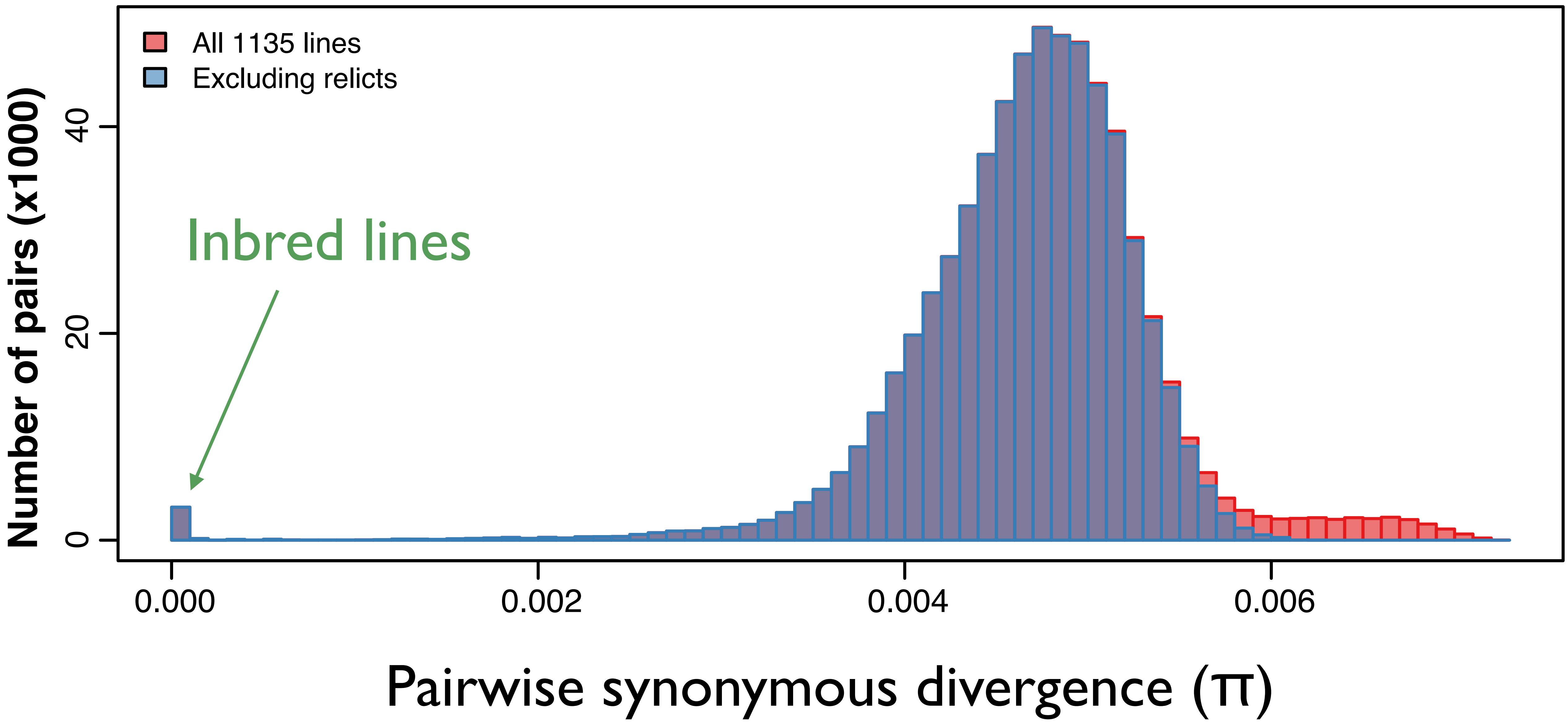


A peculiar distribution of pairwise divergence

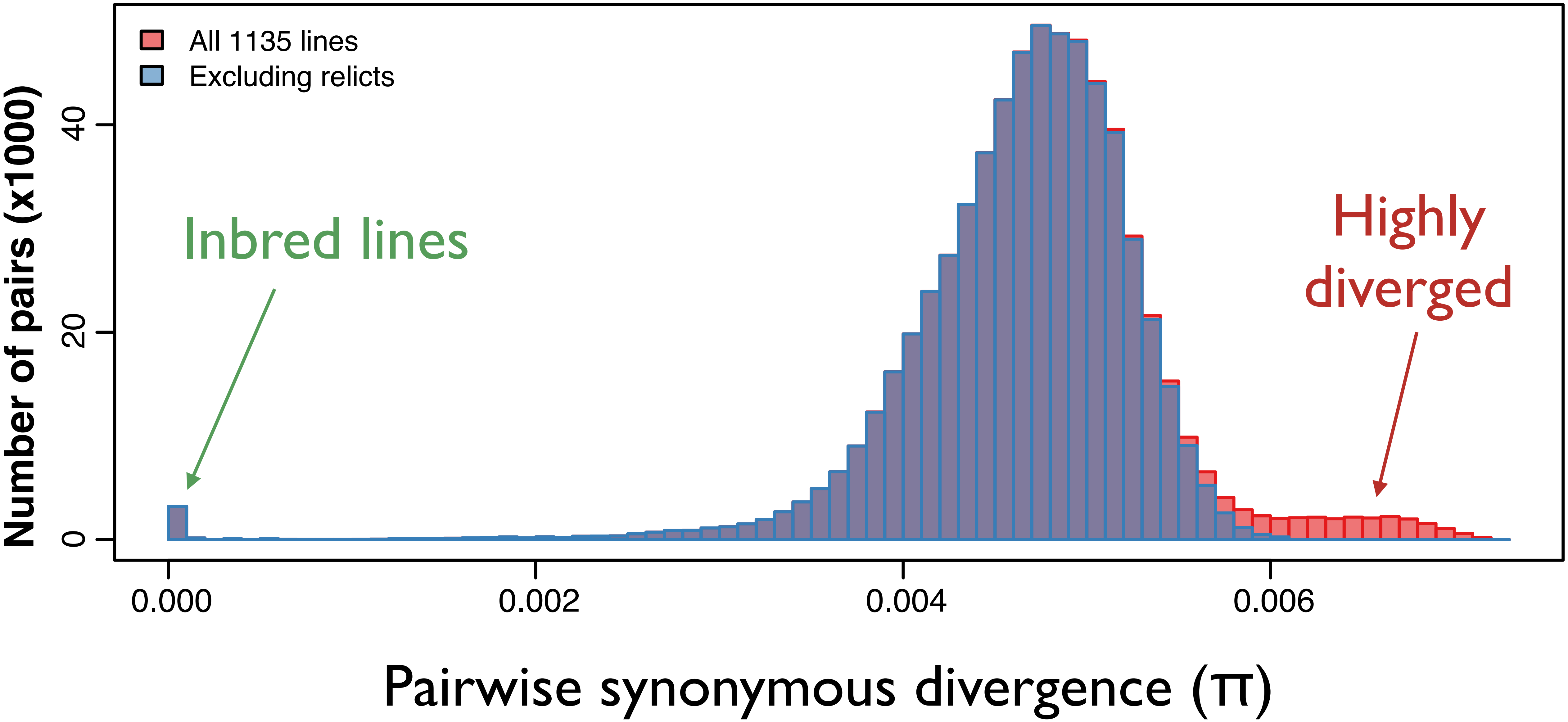


Pairwise synonymous divergence (π)

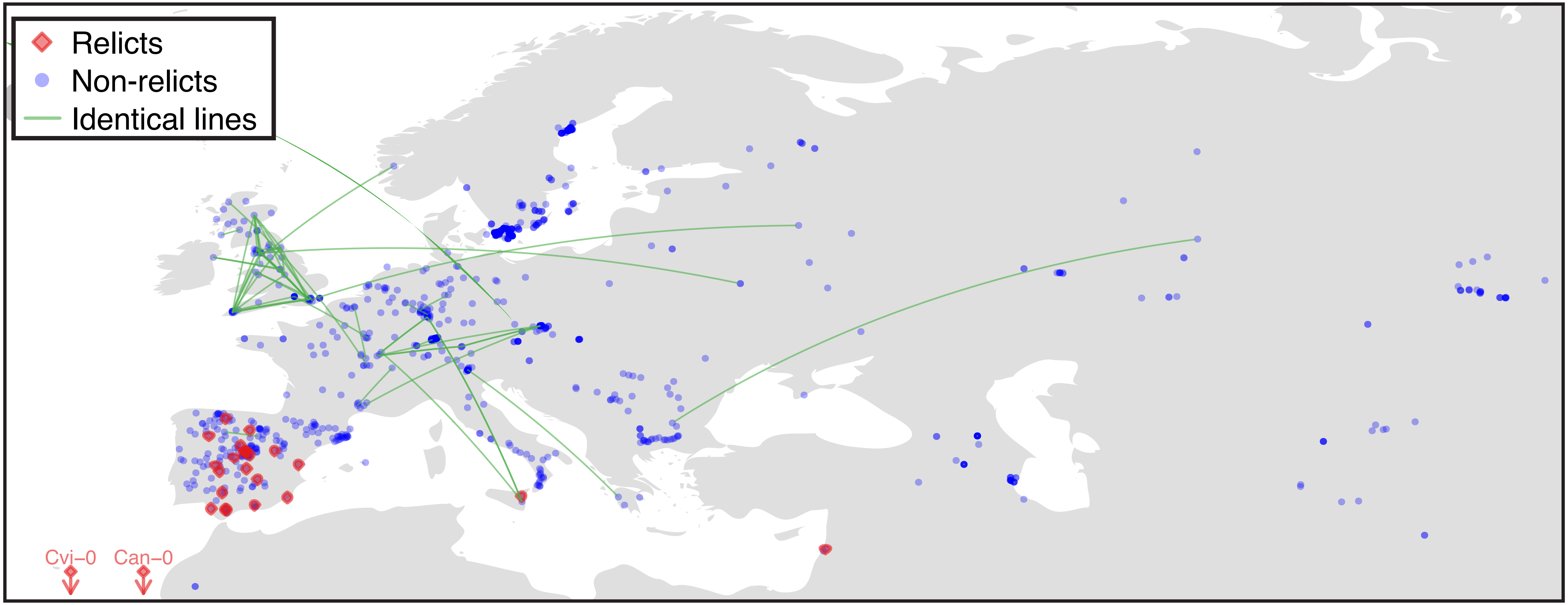
A peculiar distribution of pairwise divergence



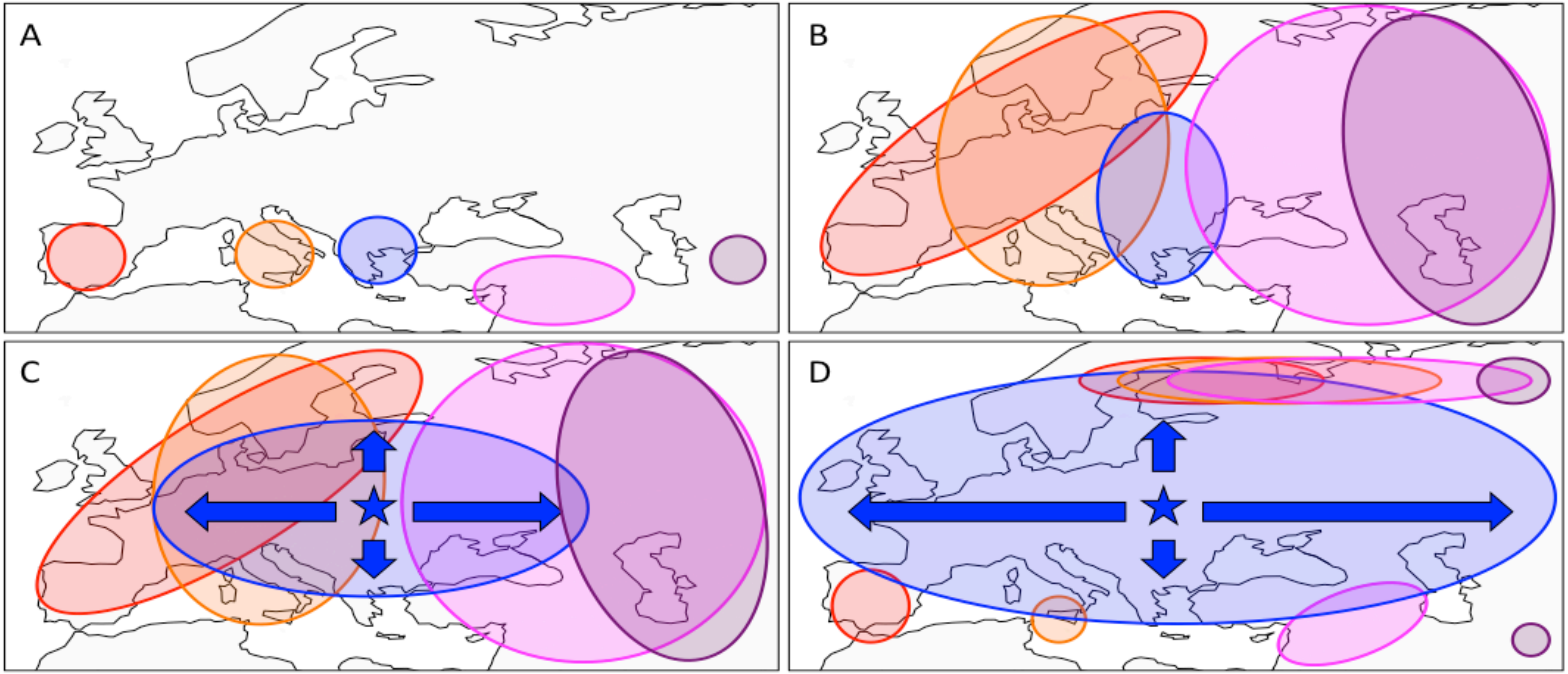
A peculiar distribution of pairwise divergence



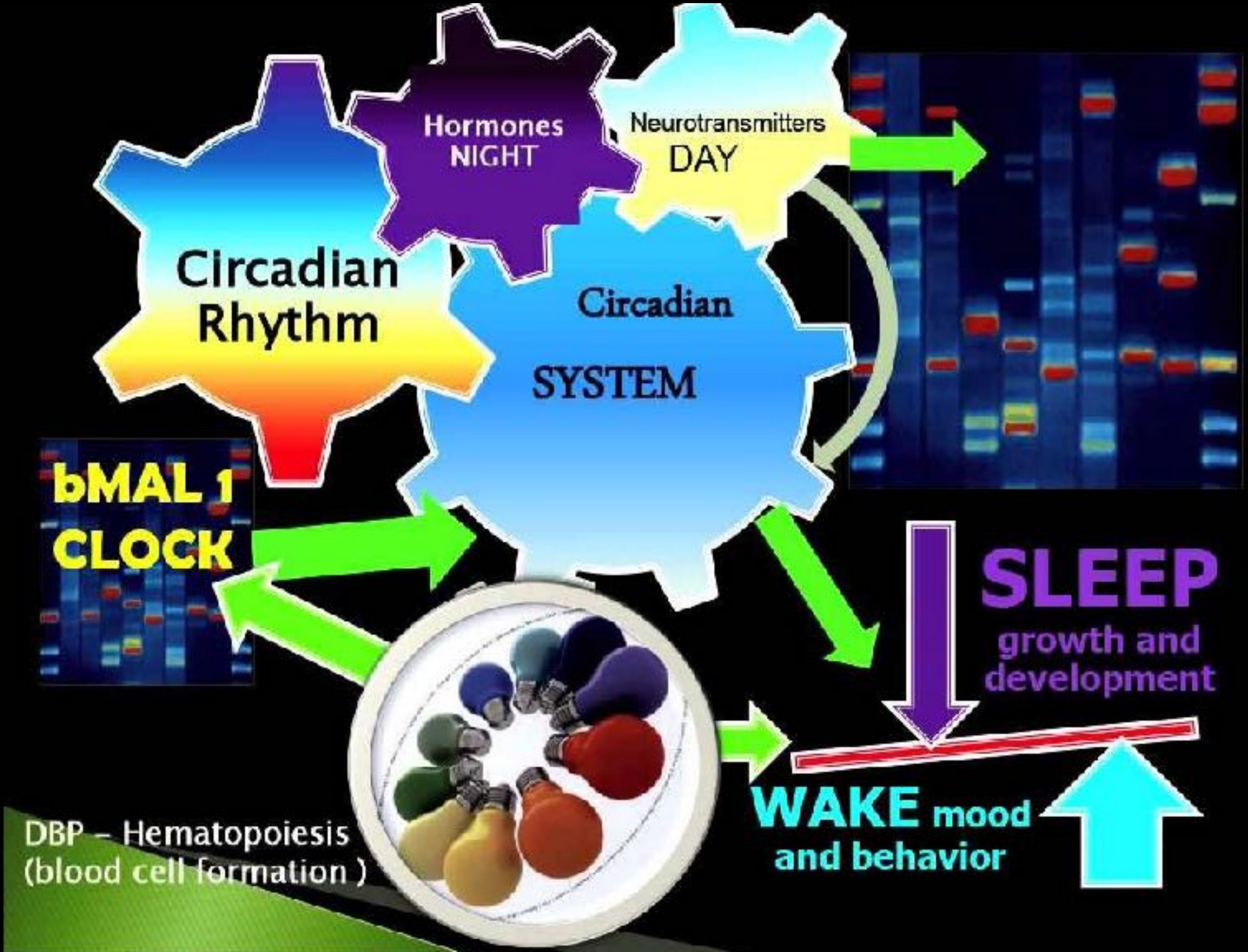
Ice Age relicts
(aka Neanderthals)

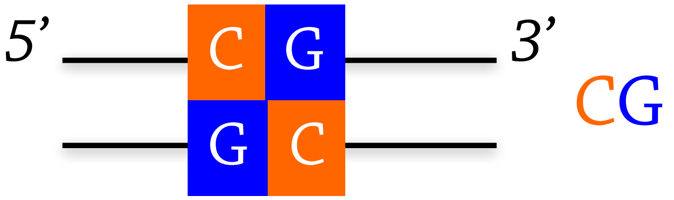


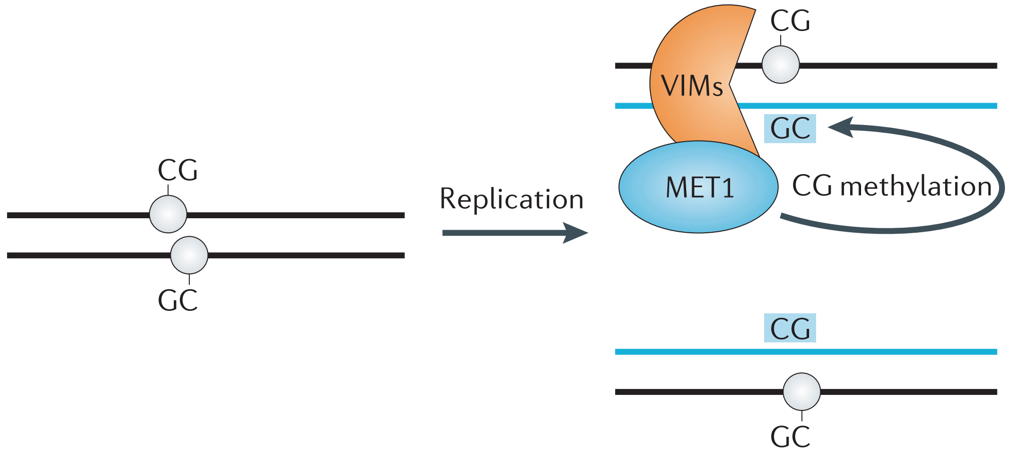
A possible scenario

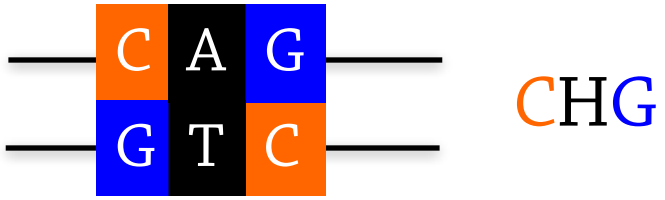


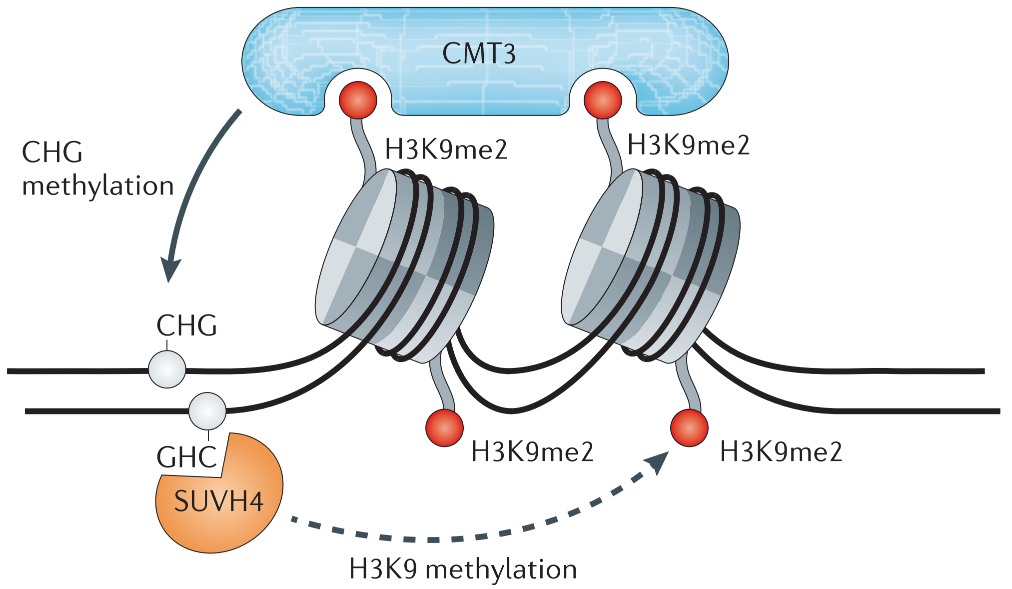
Epigenetics…



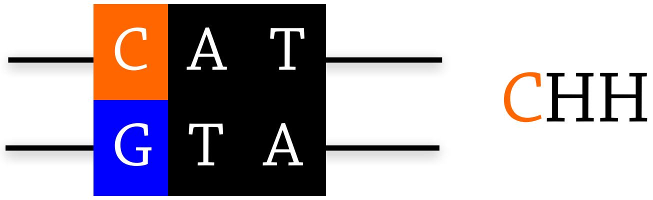




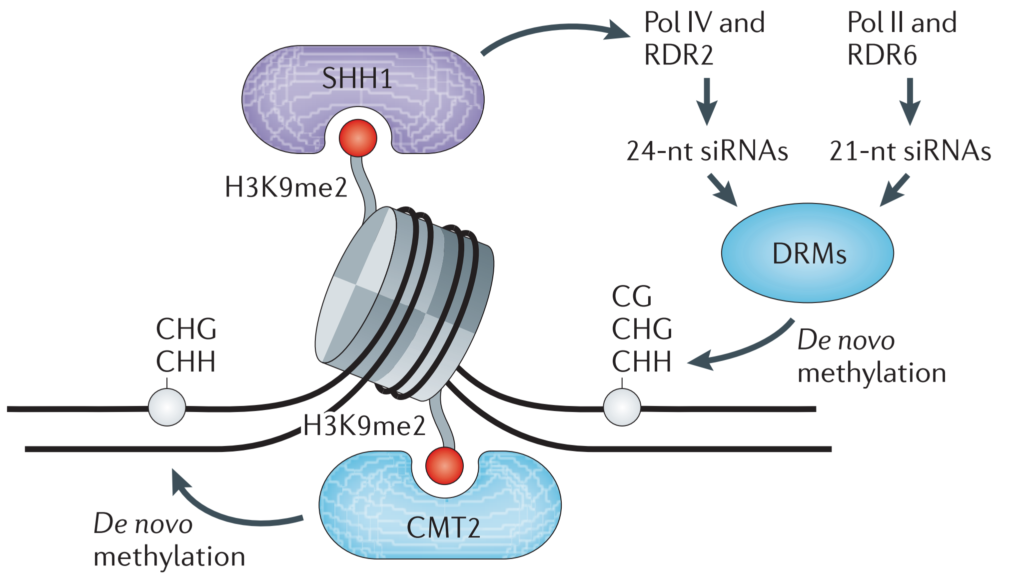




(RdDM)

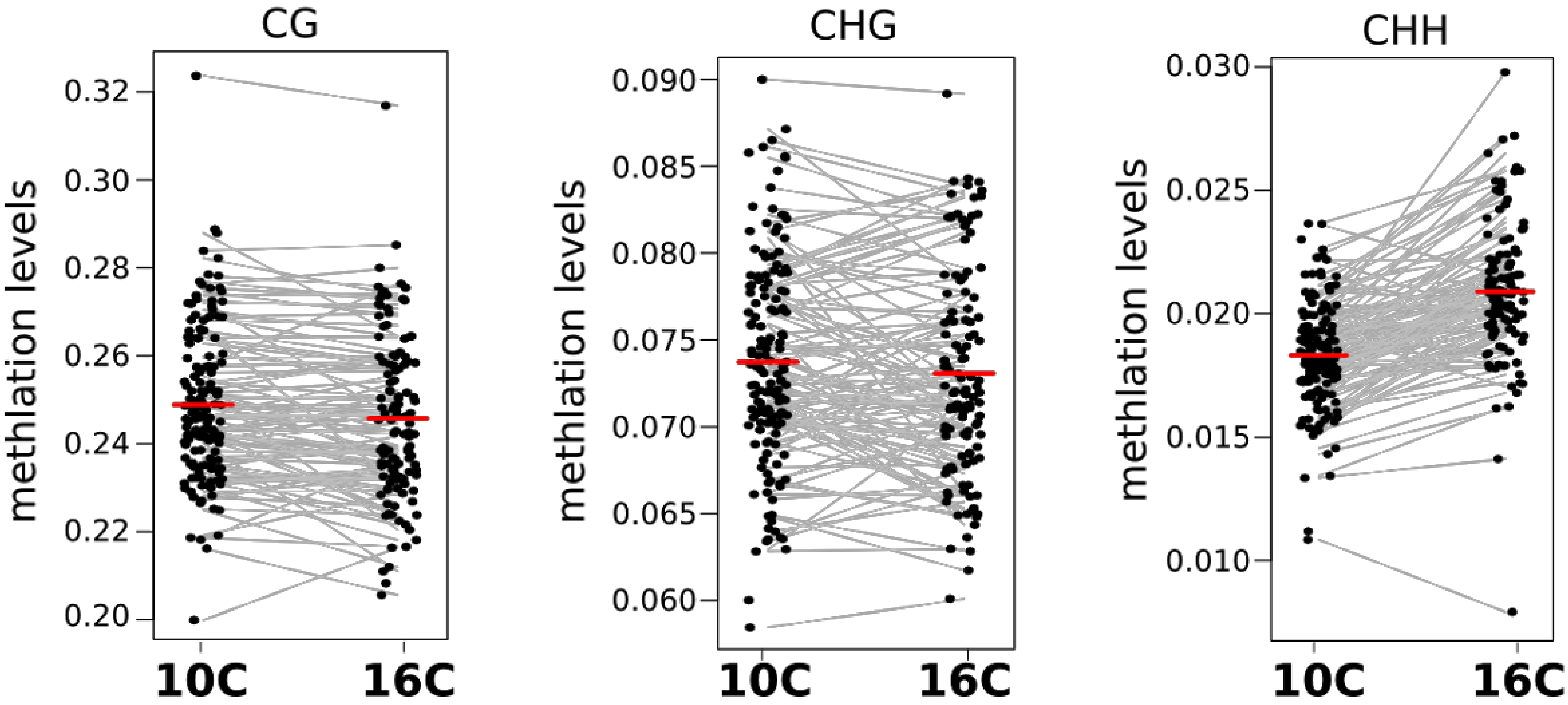


H is A, C or T



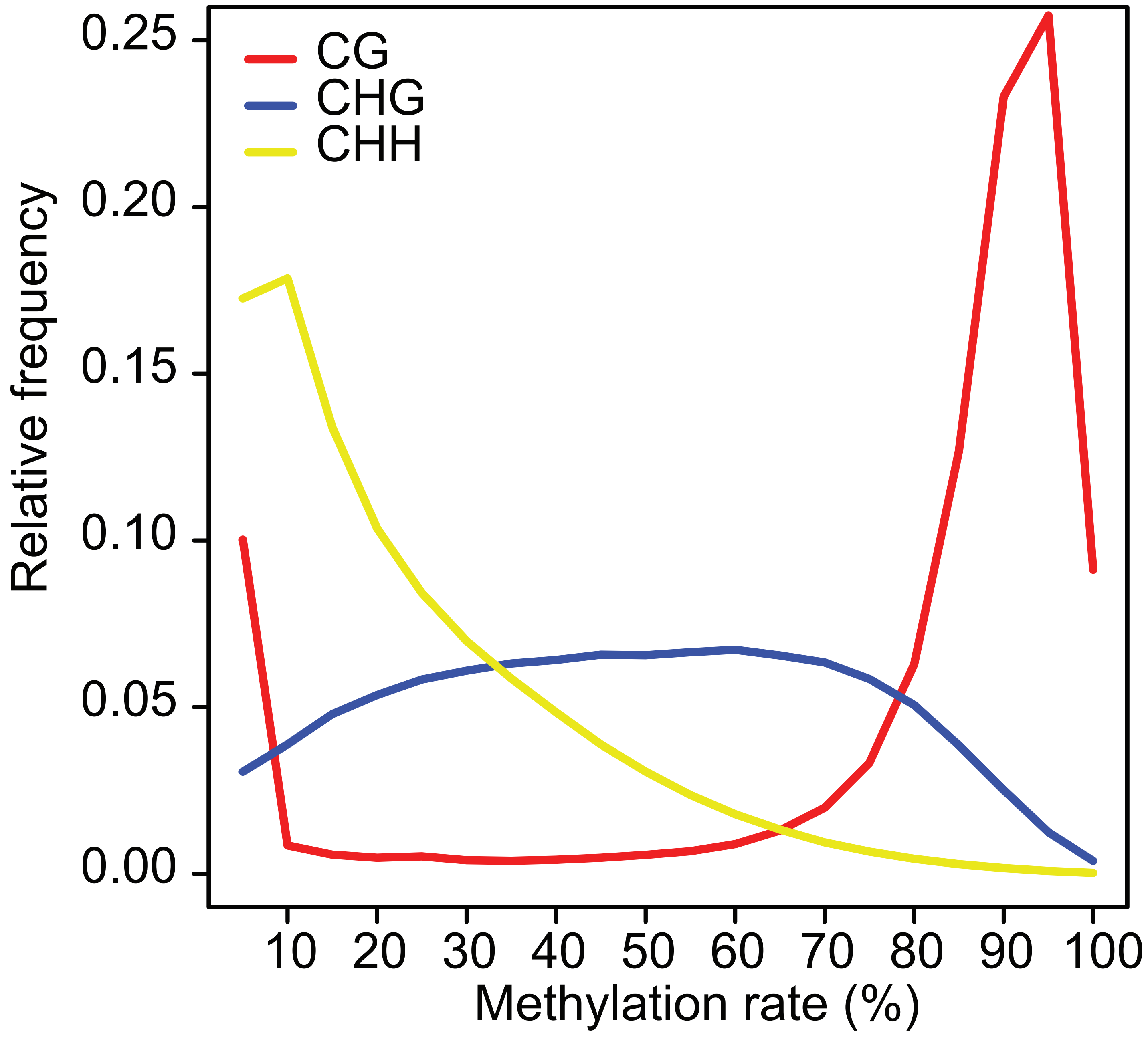
from Kawashima & Berger *Nature Rev Genet* 2014

CHH methylation is temperature-dependent



- Dubin *et al*, *eLife* 2015
- Meng *et al*, *PLoS Genet* 2016

Genotype or phenotype?



From Becker *et al*, *Nature* 2011

Gene-body methylation (gbM)

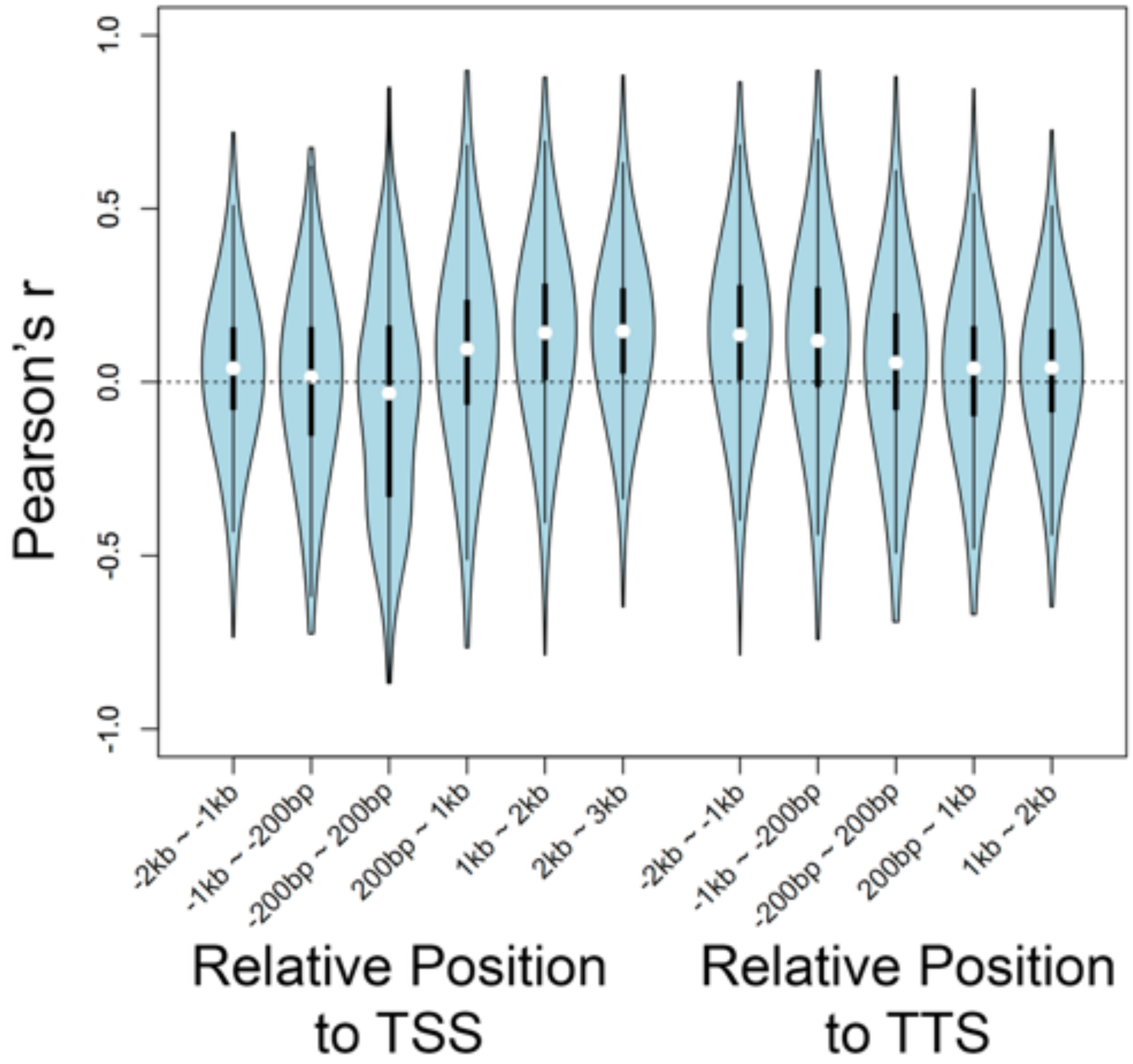
- Sparse CG methylation of housekeeping genes
- Evolutionarily conserved
- Mechanism and function totally unknown
- Generally positively correlated with expression

“TE-like” methylation (TEM)

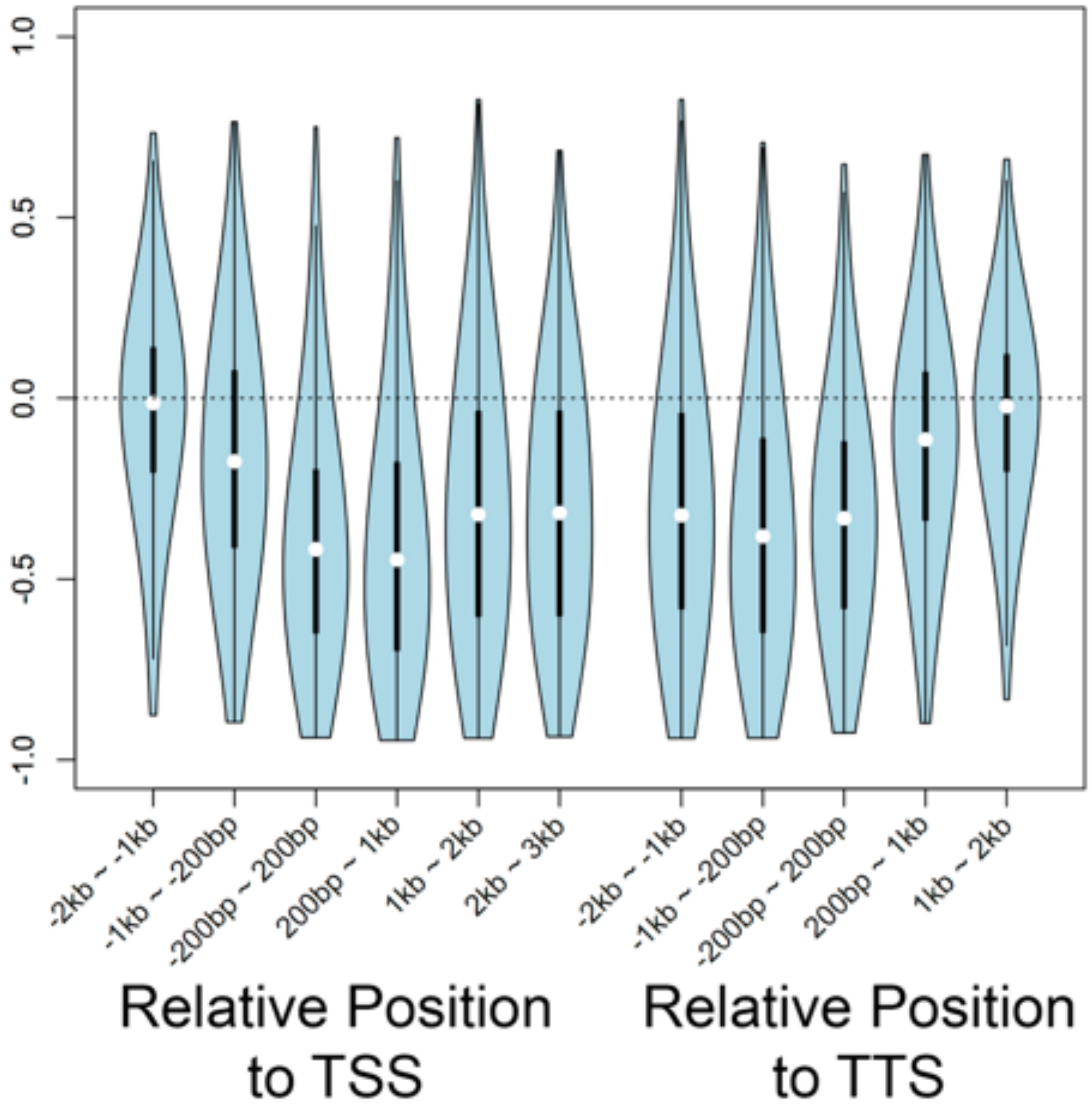
- Methylation in all contexts, CG, CHG, and CHH, mostly of TEs
- Associated with RdDM or CMT2 pathways
- Generally repressive

Effect on expression

gbM



TEM

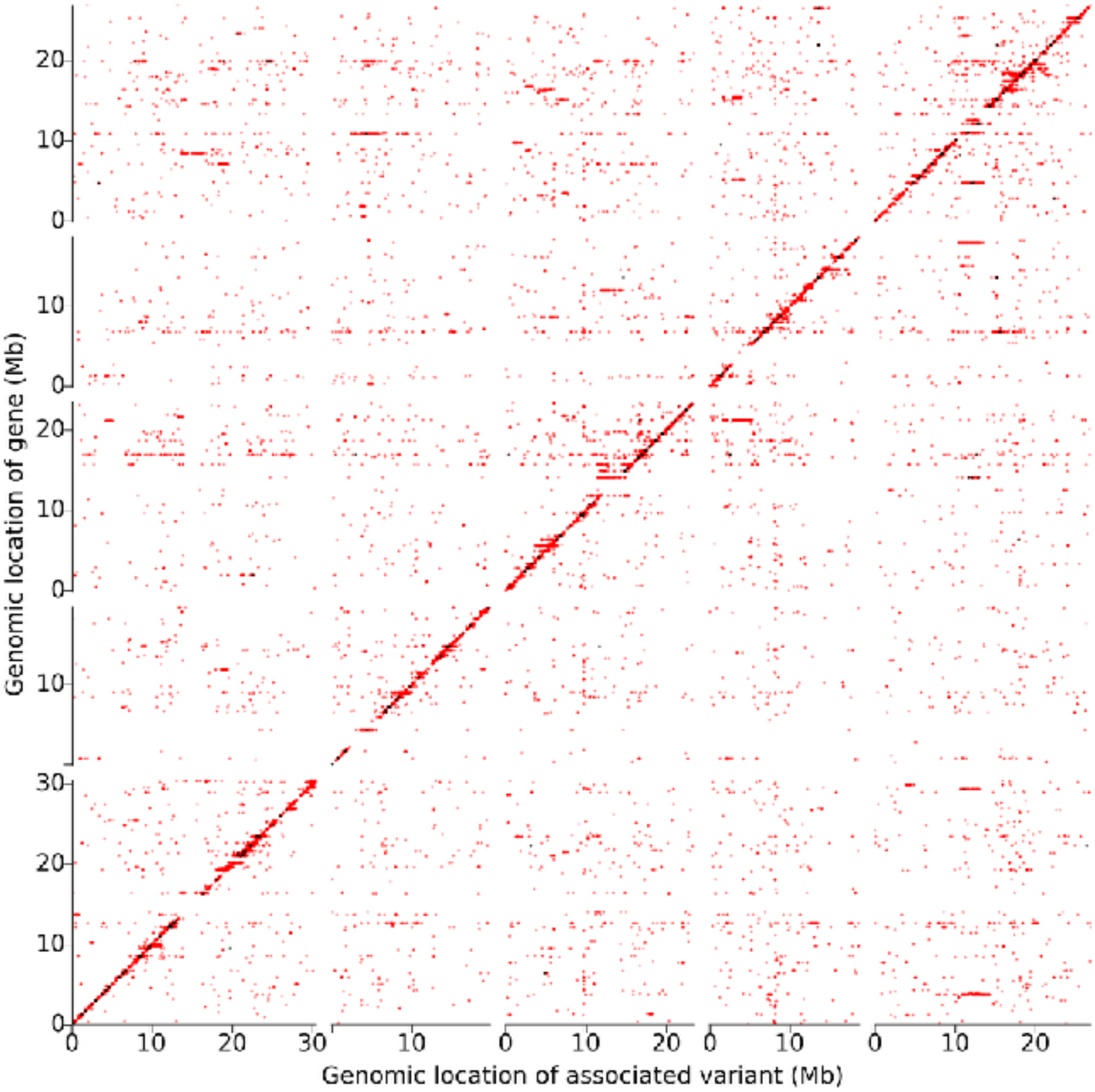


…but does it matter?

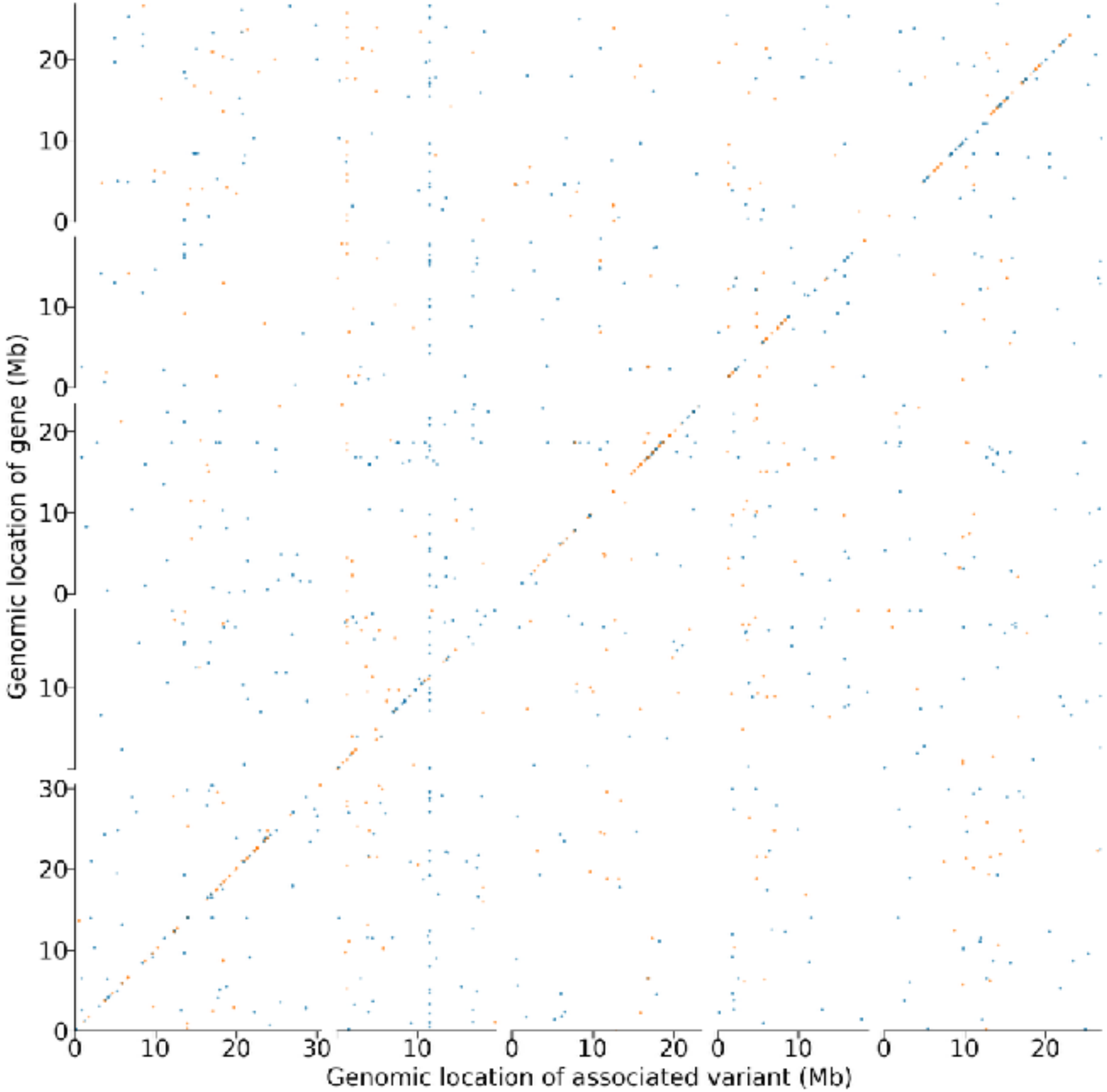
- Does knowing the epigenome help us explain expression variation if we already have the genome?

(Meng *et al.*, *PLoS Genet.* 2016)

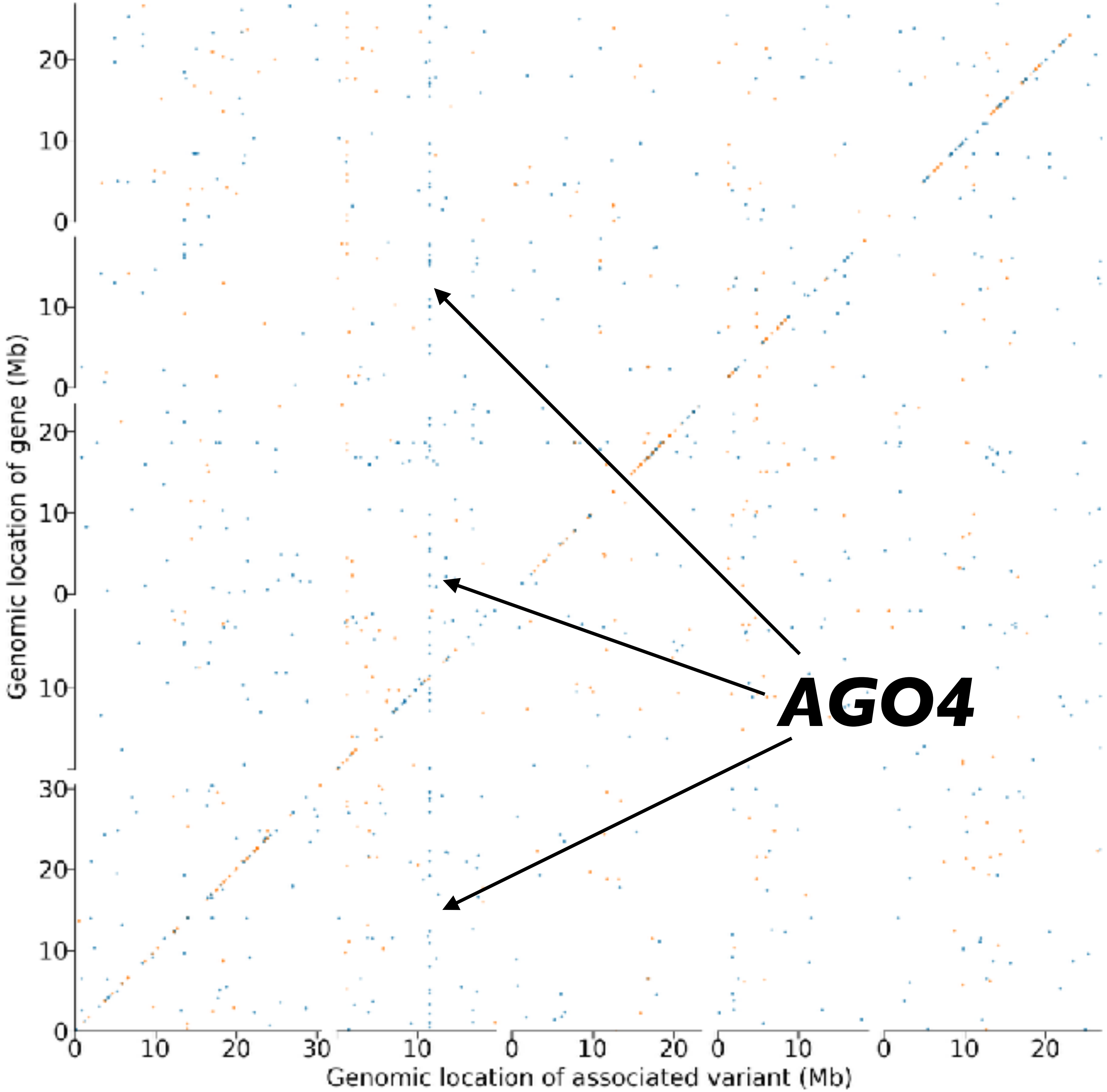
Expression and SNPs

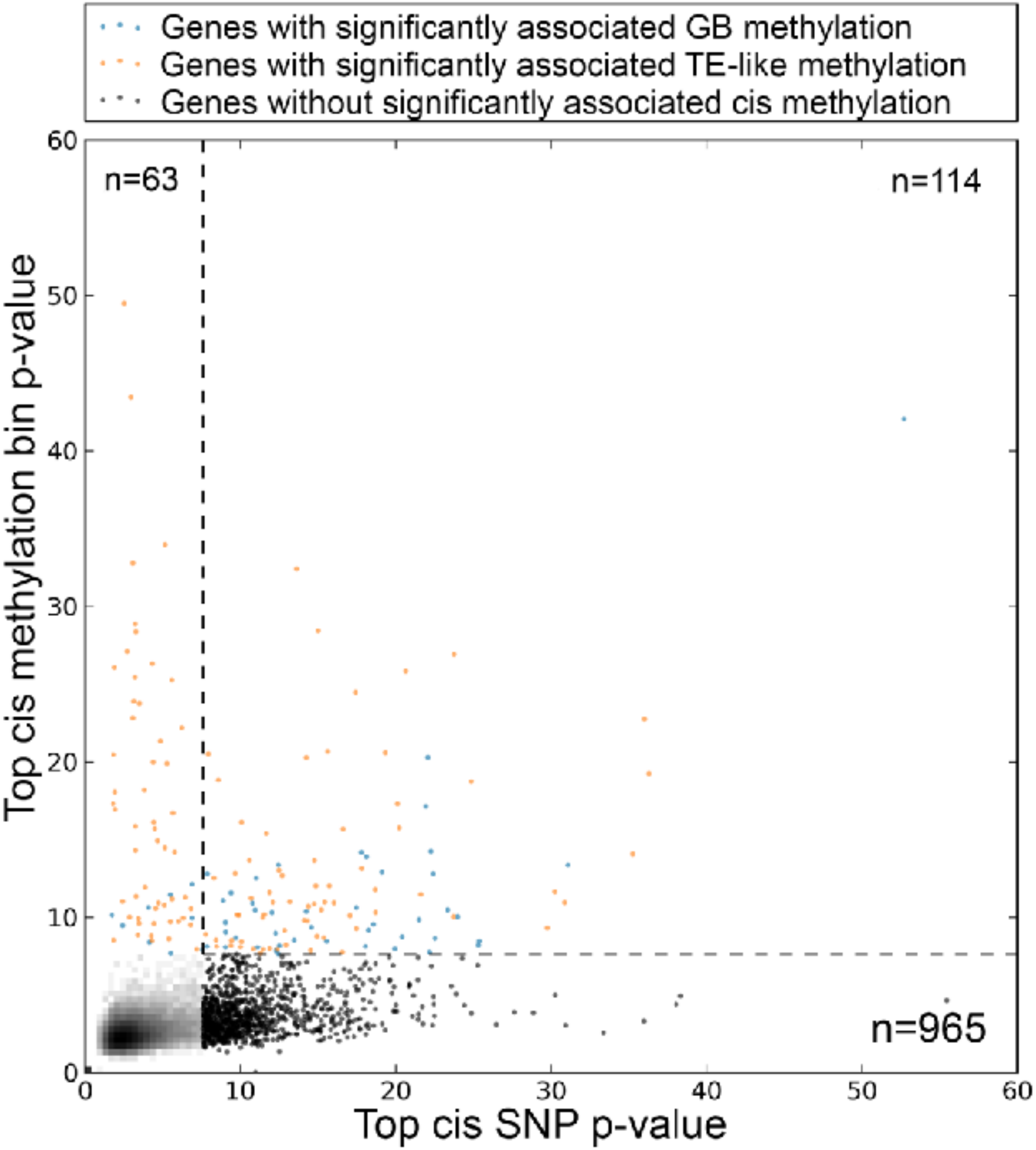


Expression and methylation

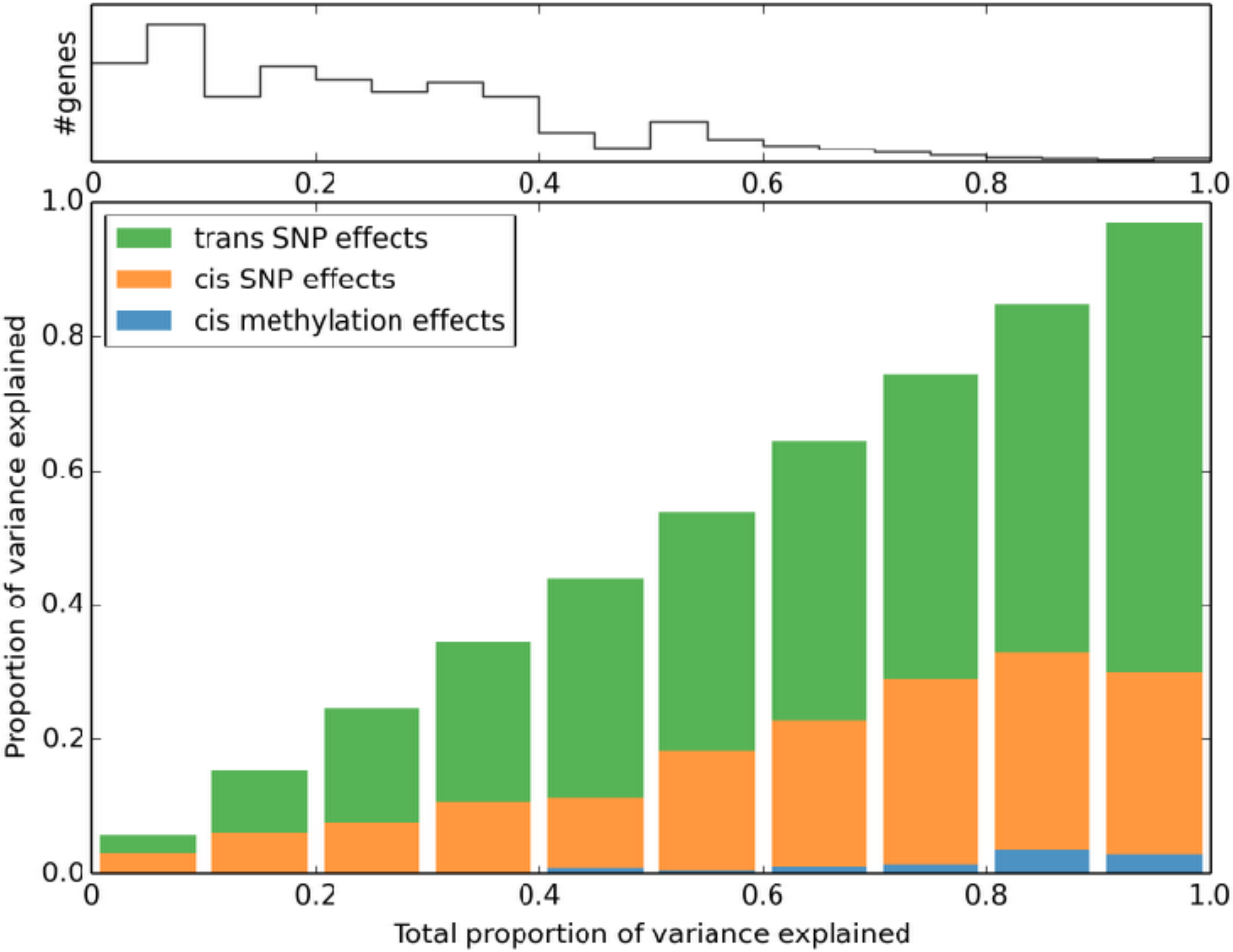


Expression and methylation

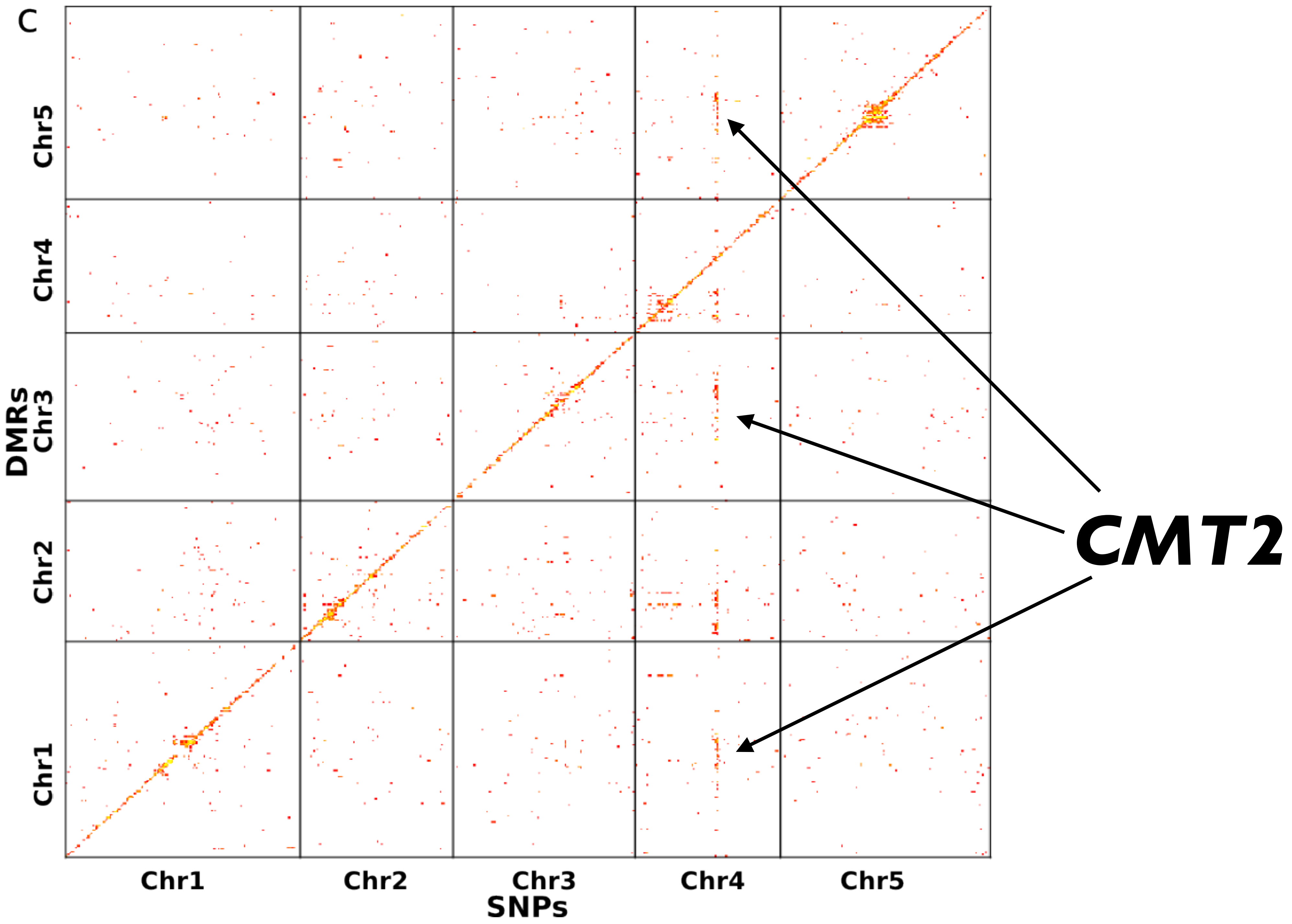



Genes with significantly associated GB methylation
Genes with significantly associated TE-like methylation
Genes without significantly associated cis methylation
n=63
n=114
n=965
Top cis methylation bin p-value
Top cis SNP p-value
0
10
20
30
40
50
60

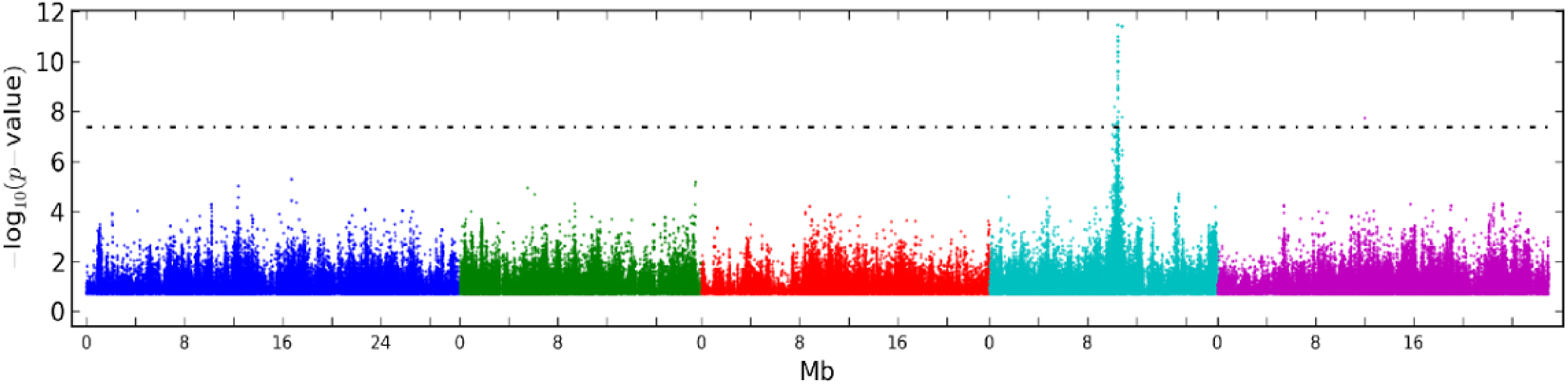
Methylation explains a tiny fraction of variation

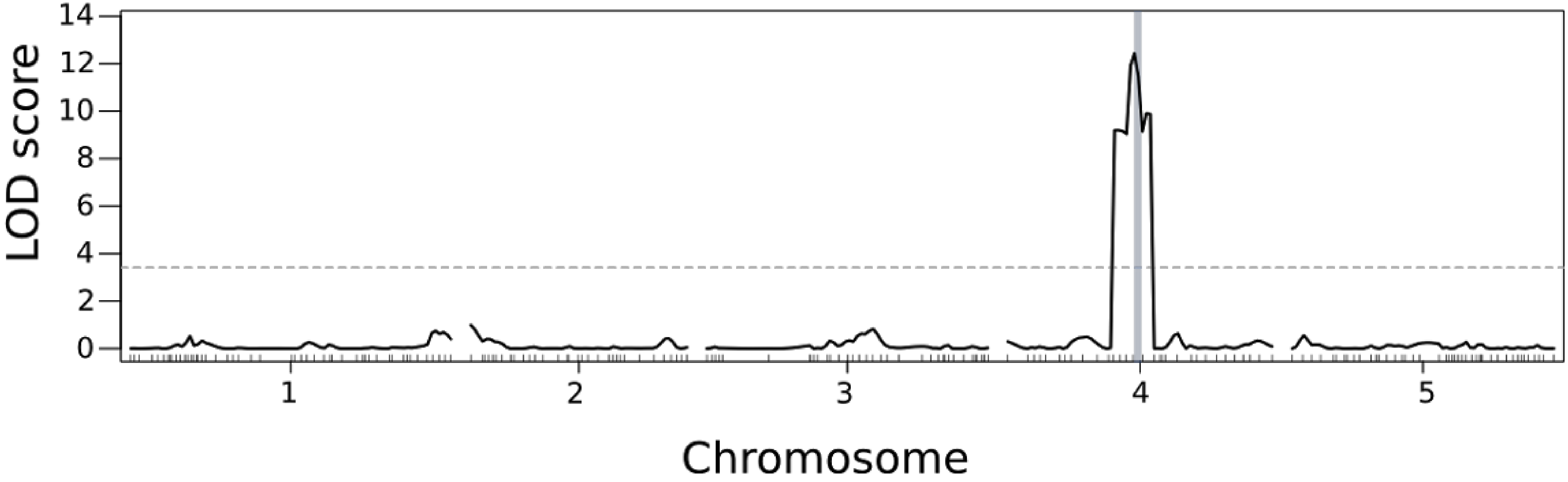


GWAS of CHH DMRs



CMT2


$-\log_{10}(p-\text{value})$
12
10
8
6
4
2
0
0
8
16
24
0
8
0
8
16
0
8
0
8
16
Mb


LOD score
14
12
10
8
6
4
2
0
1
2
3
4
5
Chromosome

Temperature adaptation?

Temperature adaptation?

- CHH methylation (and TE silencing?) is temperature sensitive

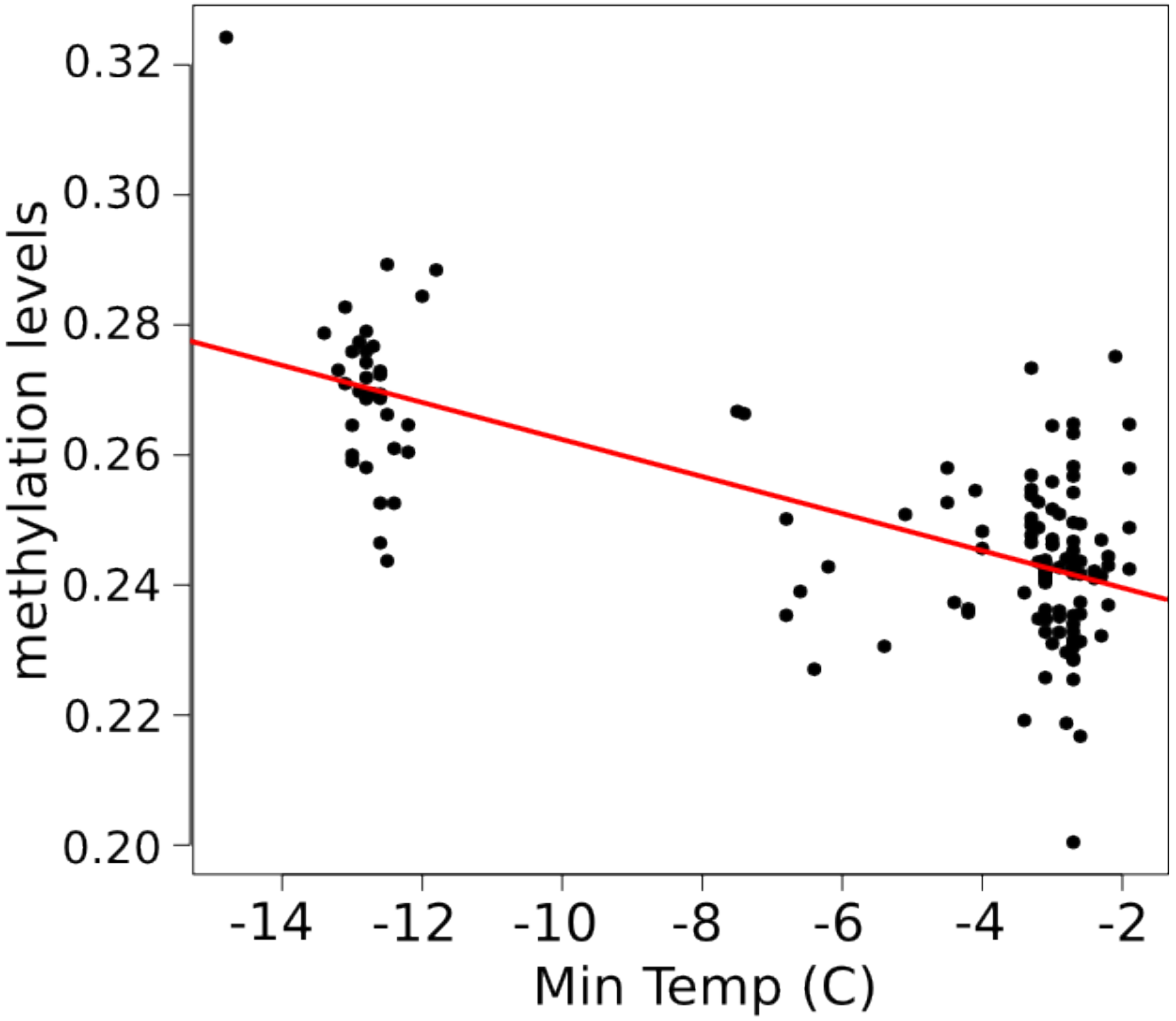
Temperature adaptation?

- CHH methylation (and TE silencing?) is temperature sensitive
- *CMT2* plays a major role in CHH methylation variation in Sweden

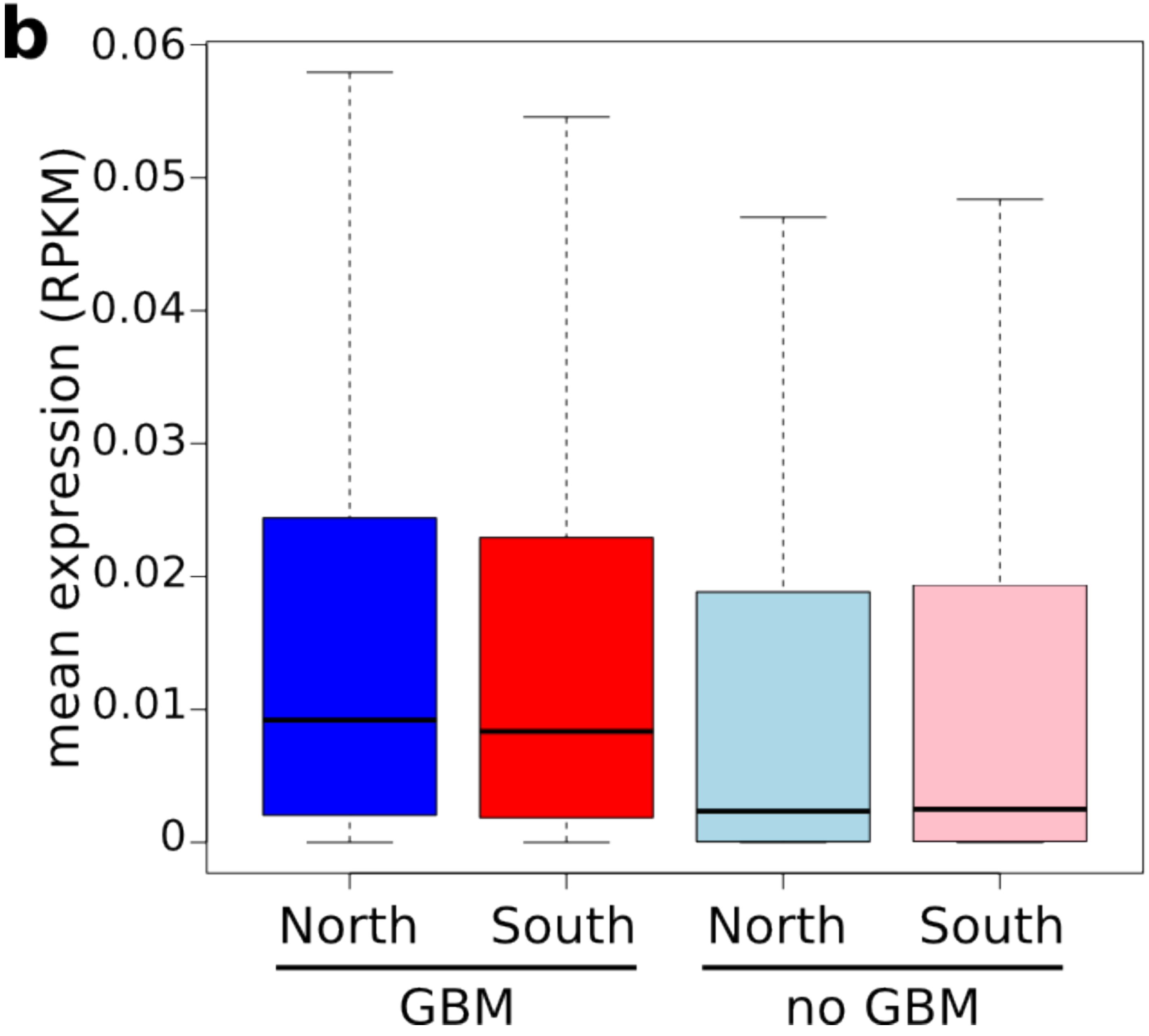
Temperature adaptation?

- CHH methylation (and TE silencing?) is temperature sensitive
- *CMT2* plays a major role in CHH methylation variation in Sweden
- *CMT2* shows up in a global scan of SNPs associated with "temperature seasonality" (Shen et al. 2014)

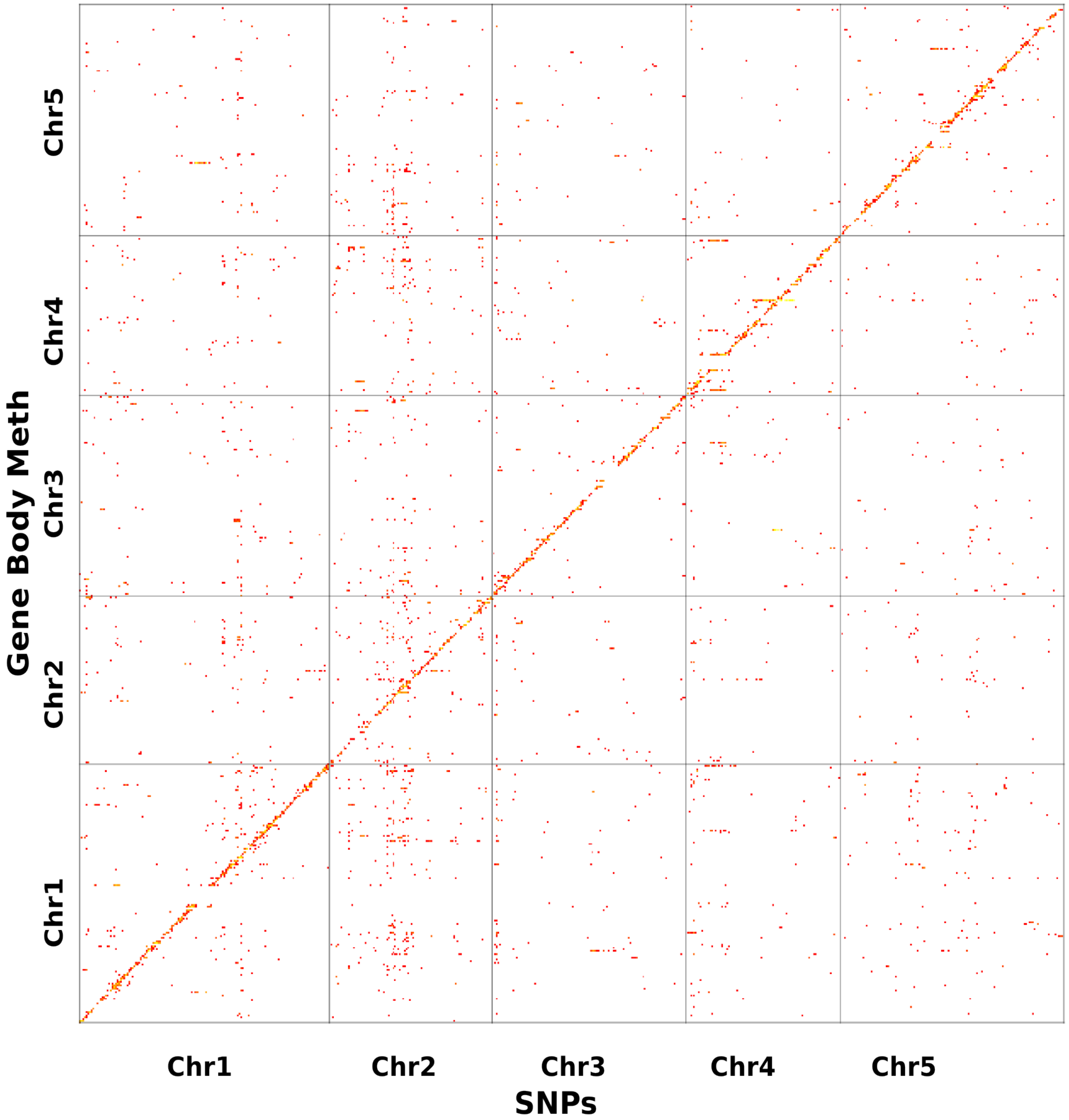
Northern lines have more gene-body methylation



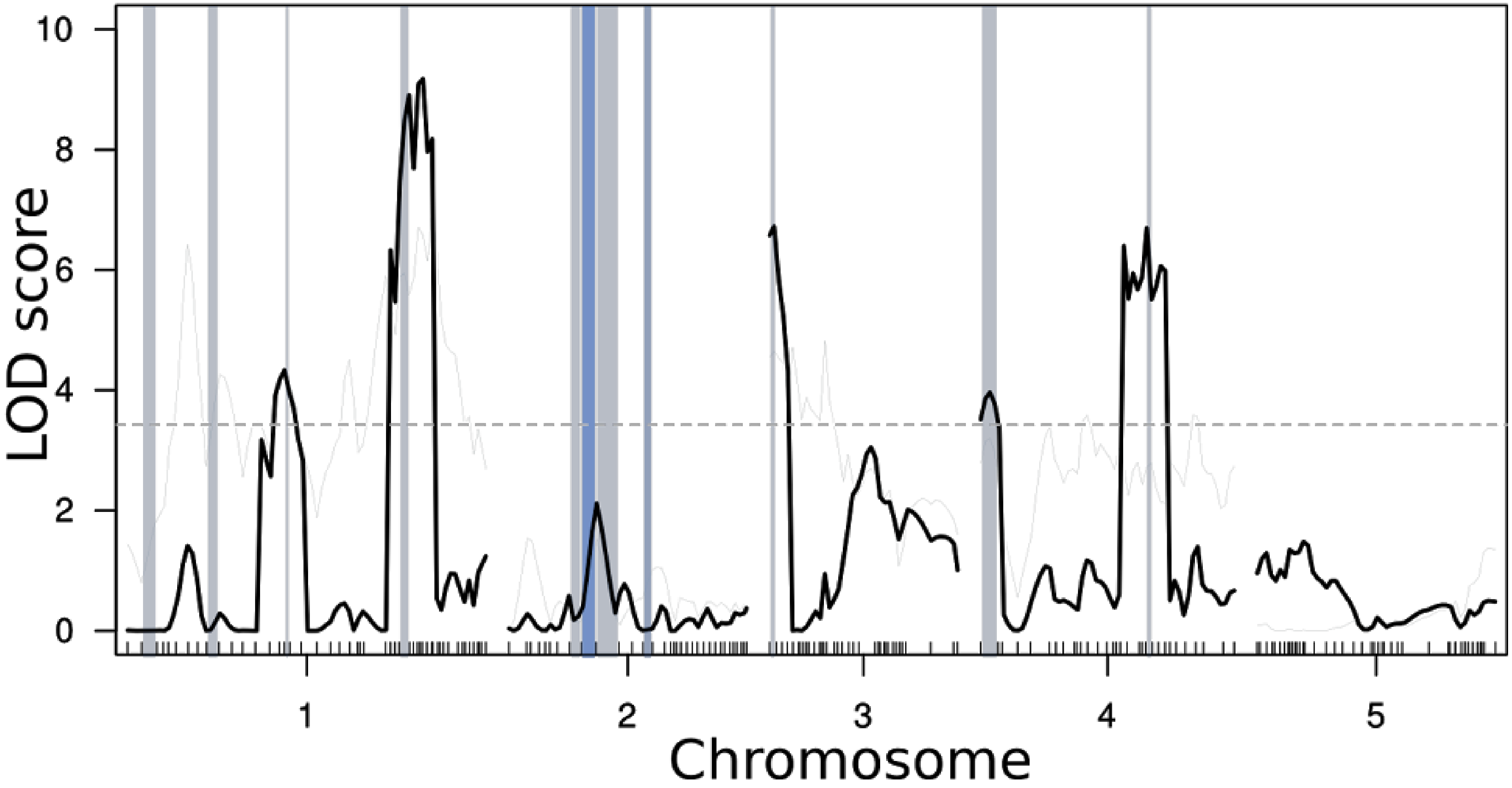
Northern lines have higher gene expression



GWAS of gene-body methylation



At least some are real…



Selection responsible for pattern in gbM?

Selection responsible for pattern in gbM?

- Non-reference allele found in north only, and associated with increased methylation

Selection responsible for pattern in gbM?

- Non-reference allele found in north only, and associated with increased methylation
- 9-fold increase in “selective sweep” regions

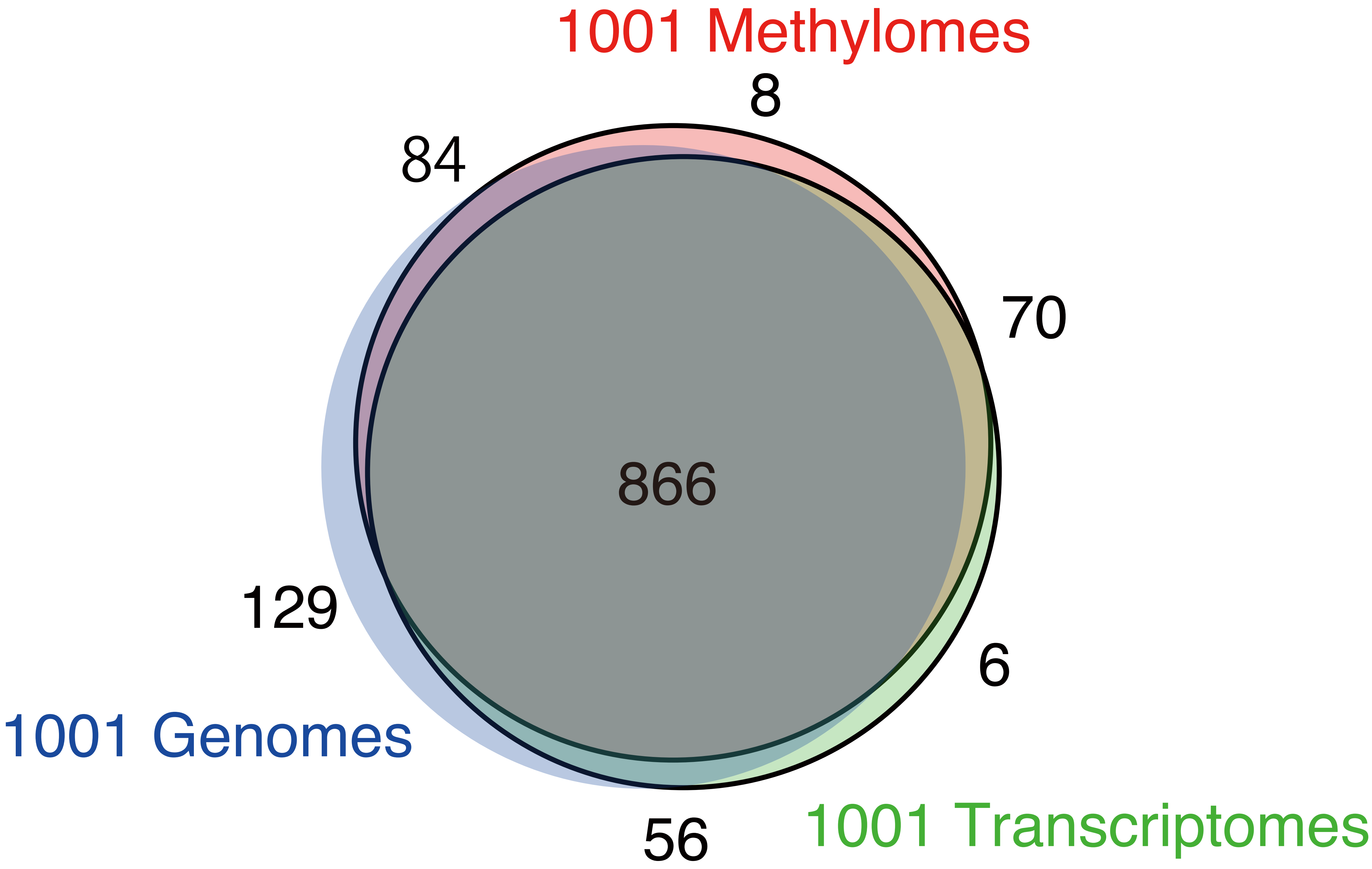
Selection responsible for pattern in gbM?

- Non-reference allele found in north only, and associated with increased methylation
- 9-fold increase in "selective sweep" regions
- 5-fold increase near temperature-associated SNPs

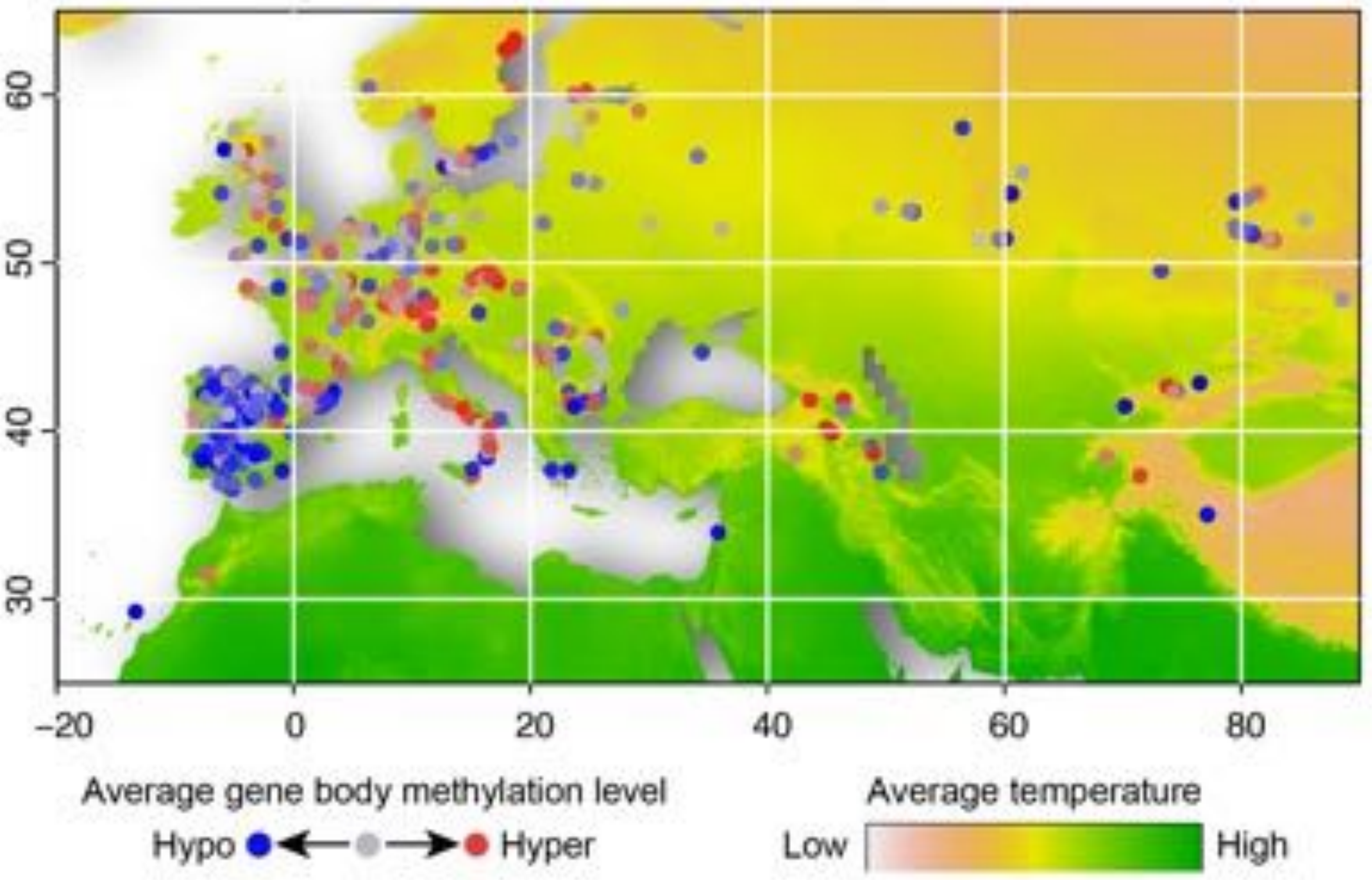
Selection responsible for pattern in gbM?

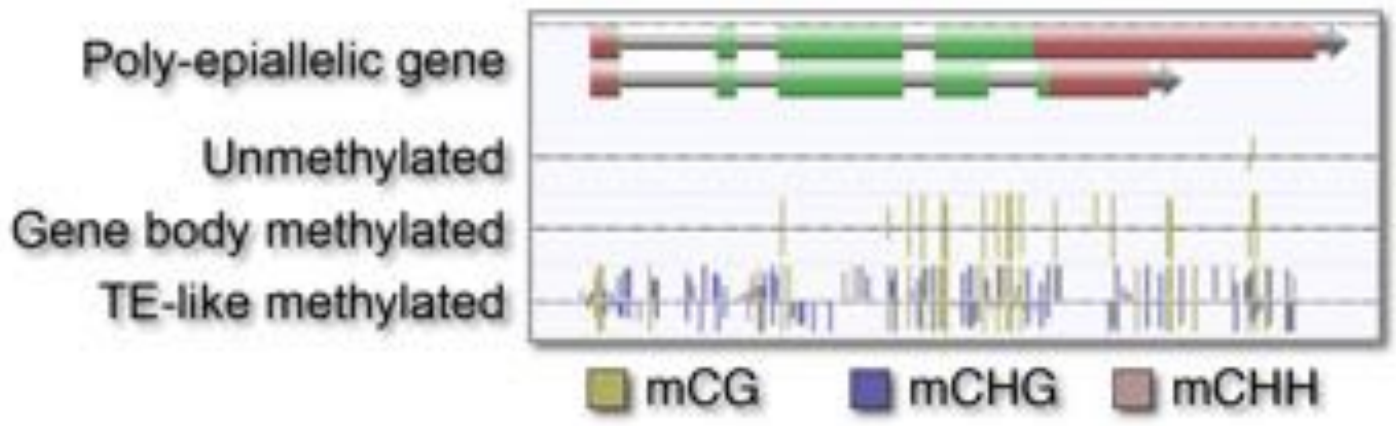
- Non-reference allele found in north only, and associated with increased methylation
- 9-fold increase in “selective sweep” regions
- 5-fold increase near temperature-associated SNPs
- Over-representation near genes involved in gene expression and/or chromatin accessibility, but not DNA methylation

1001 Methylomes

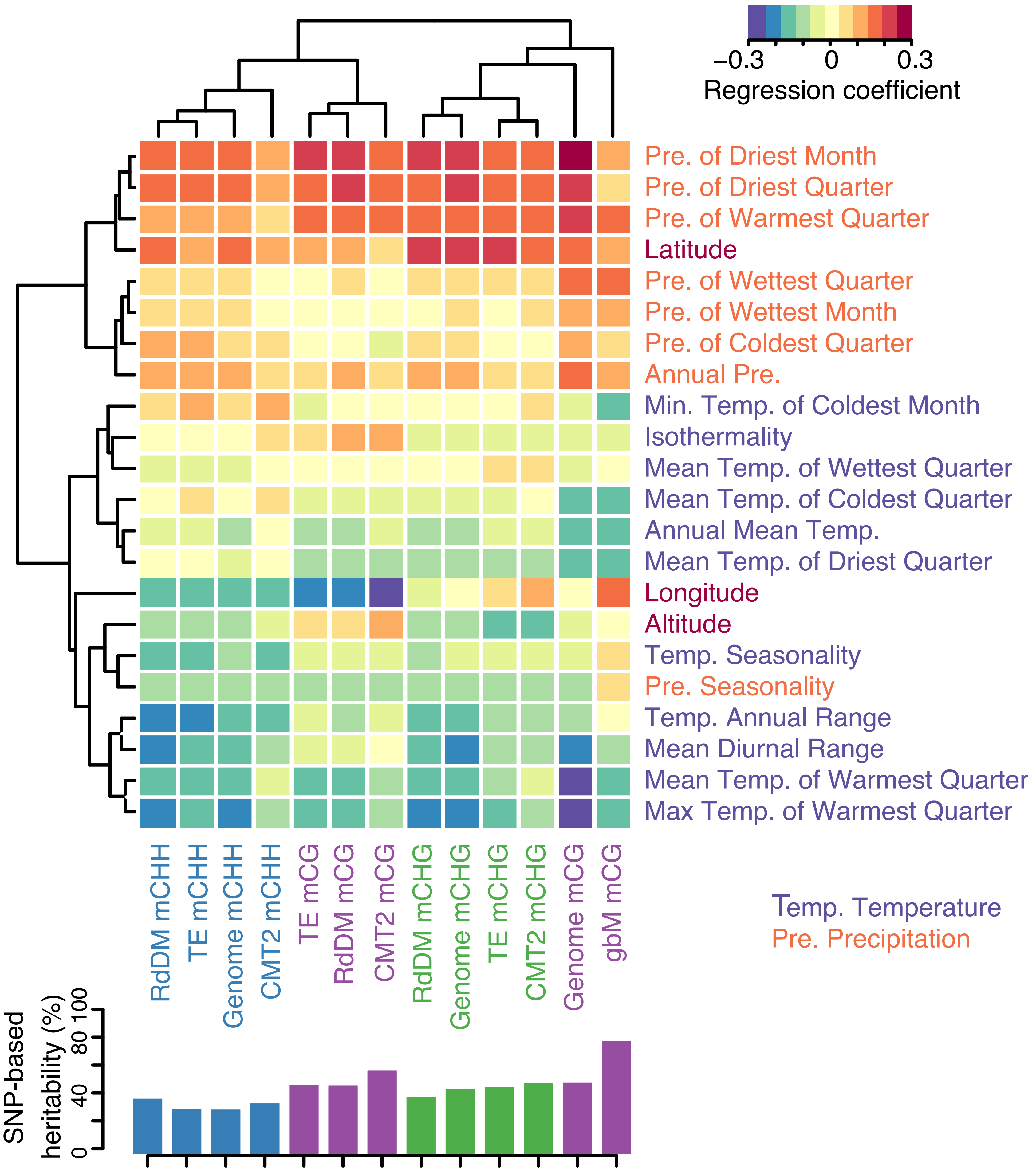


Climate correlations

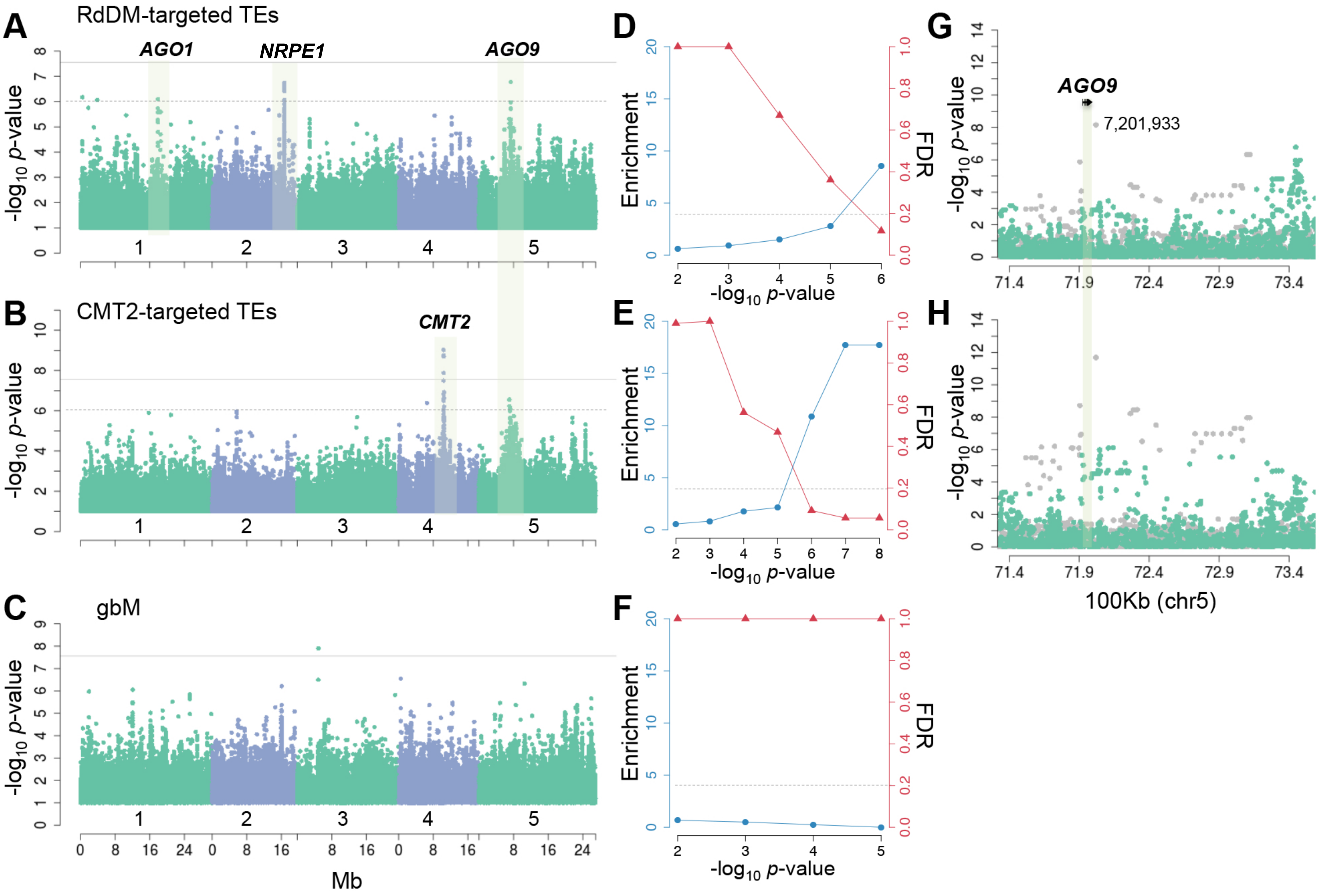




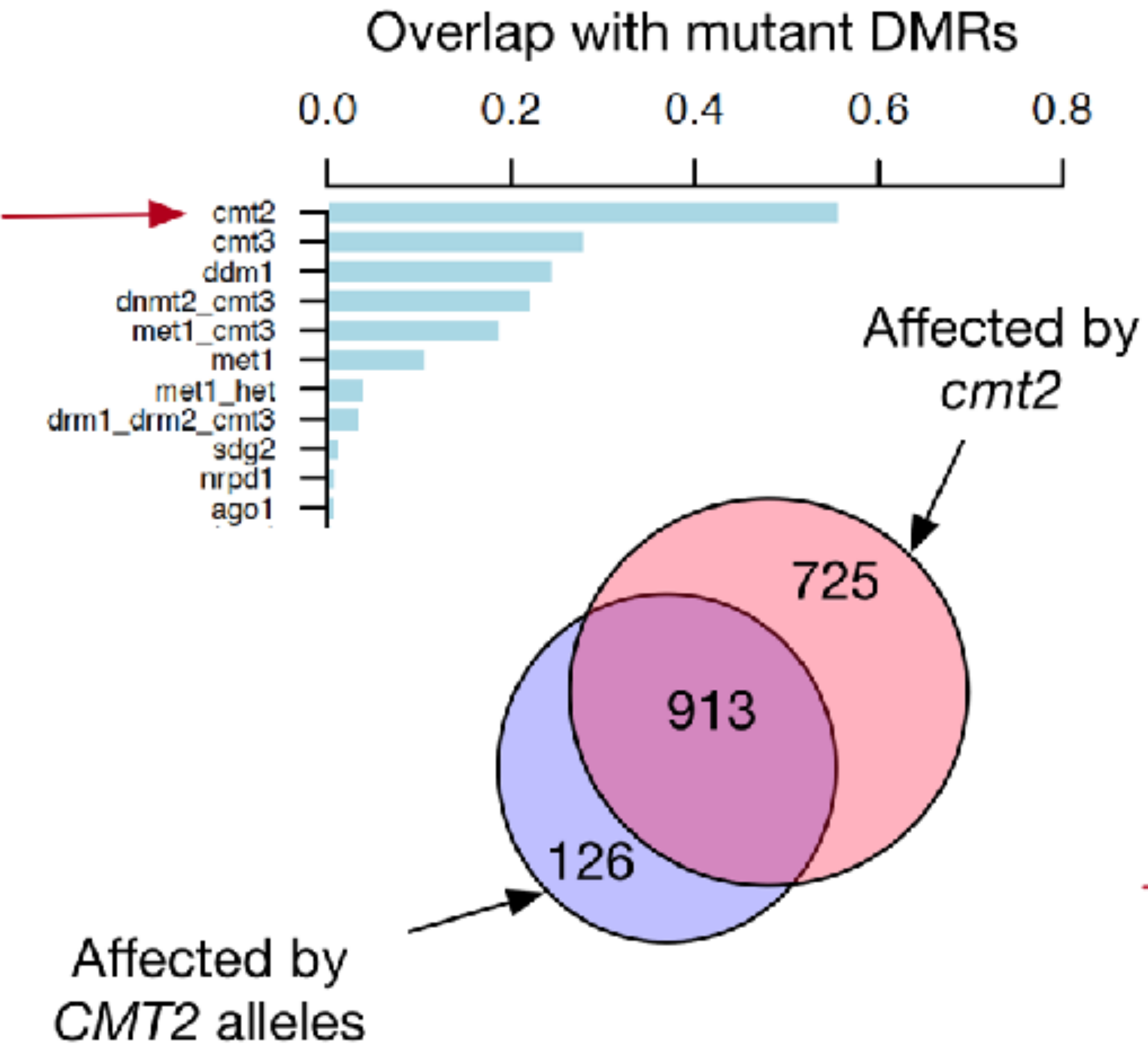
Correlations between types of methylation

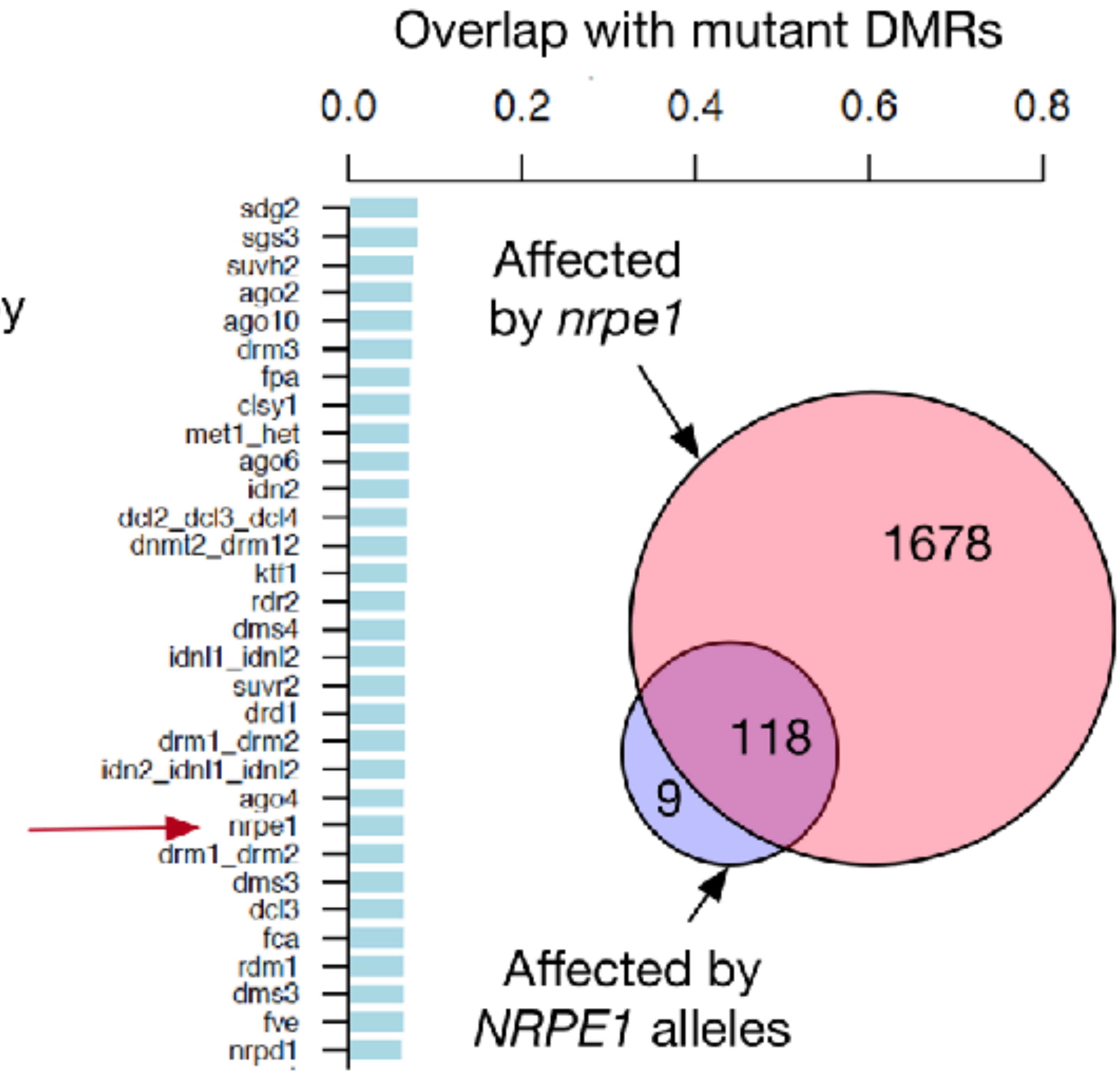


Genotype or phenotype?

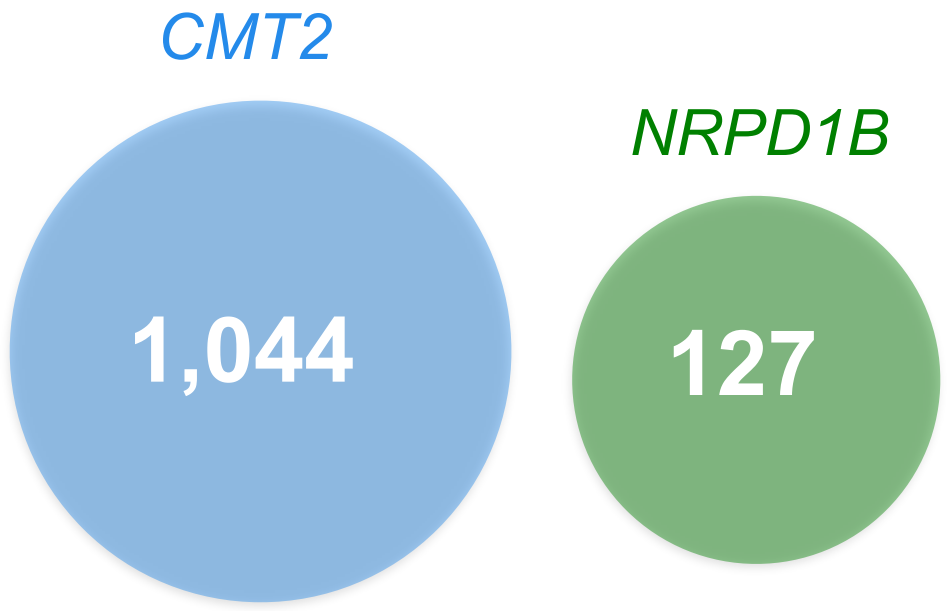


Comparison with mutant phenotypes



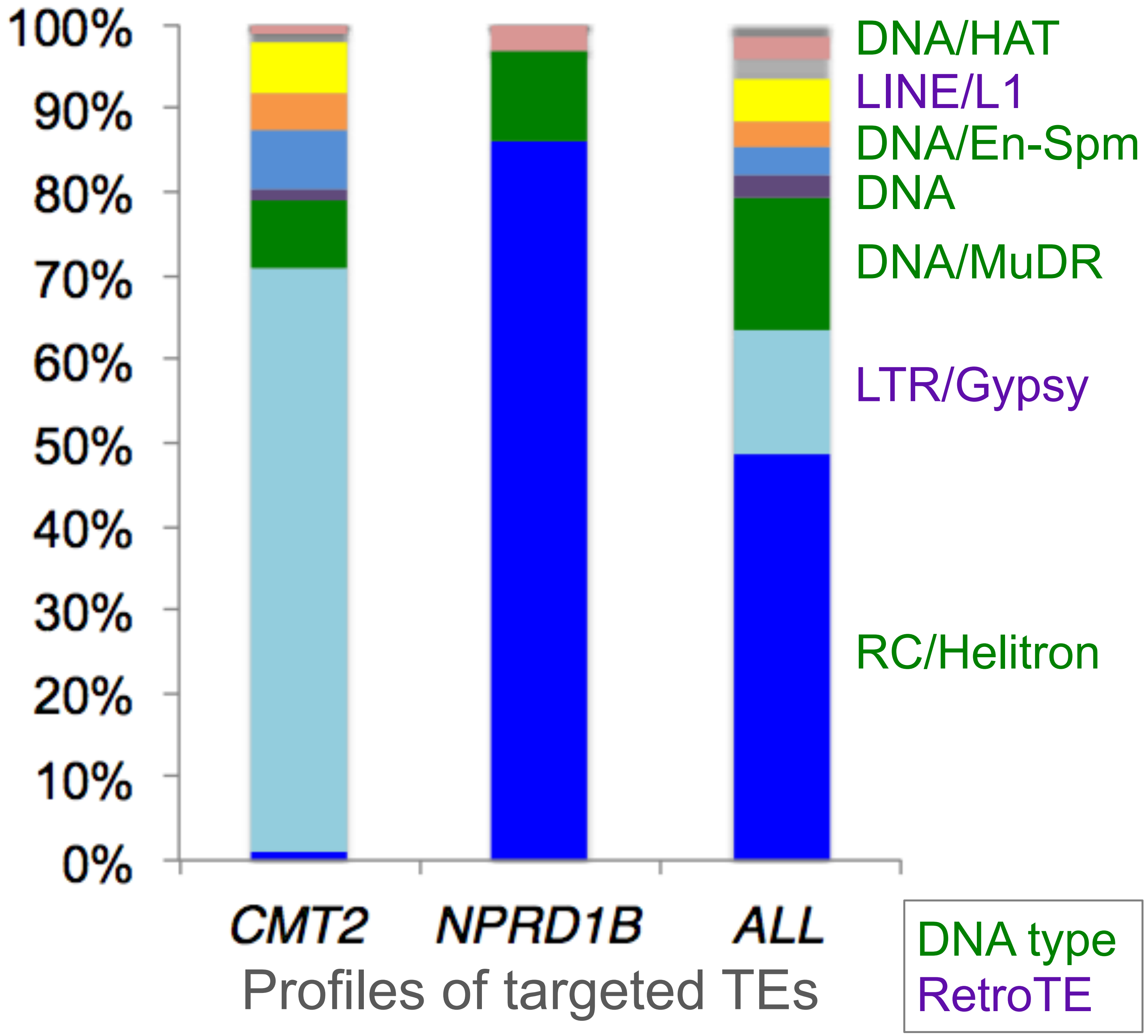


CMT2 & NRPE1 target different TEs



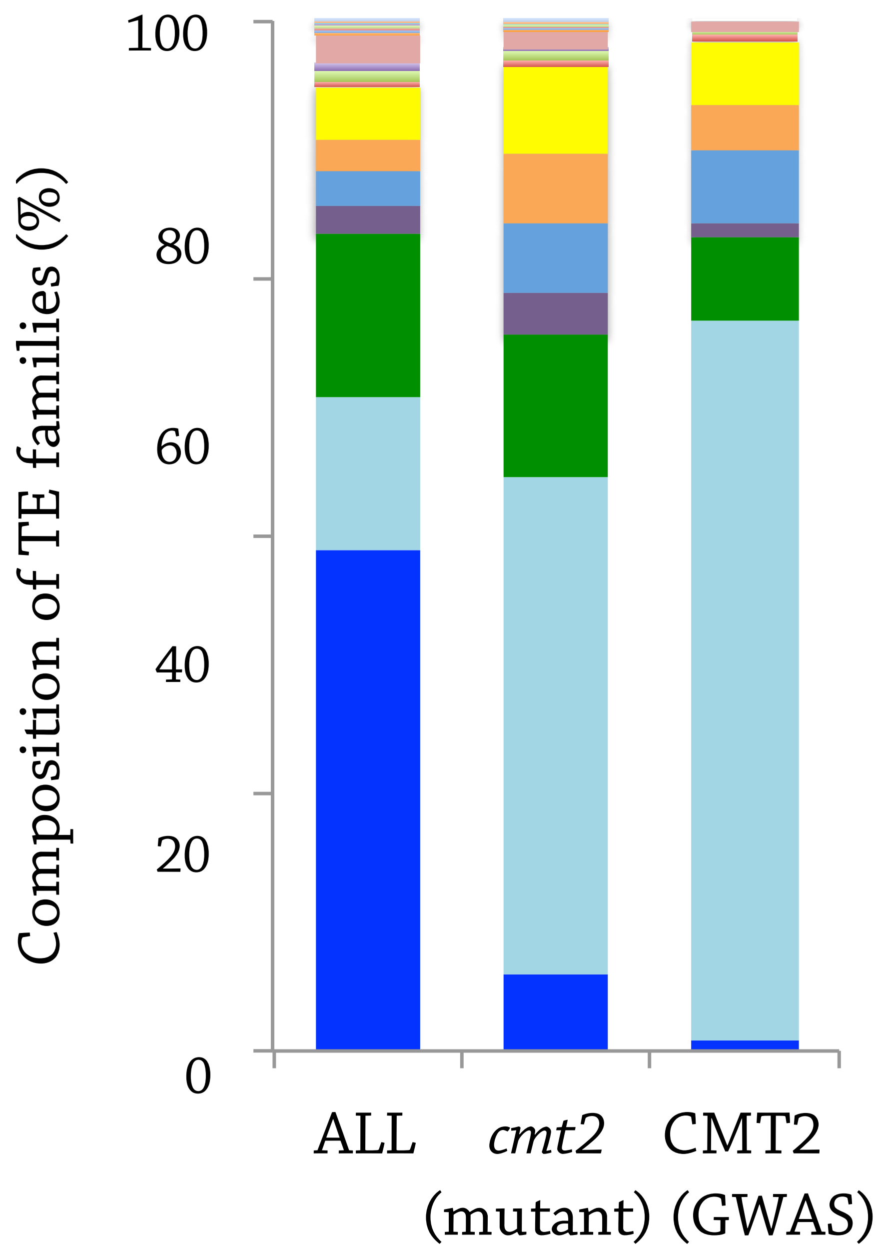
CMT2
MAF5%, p<1e-06   MAF5%, p<1e-06

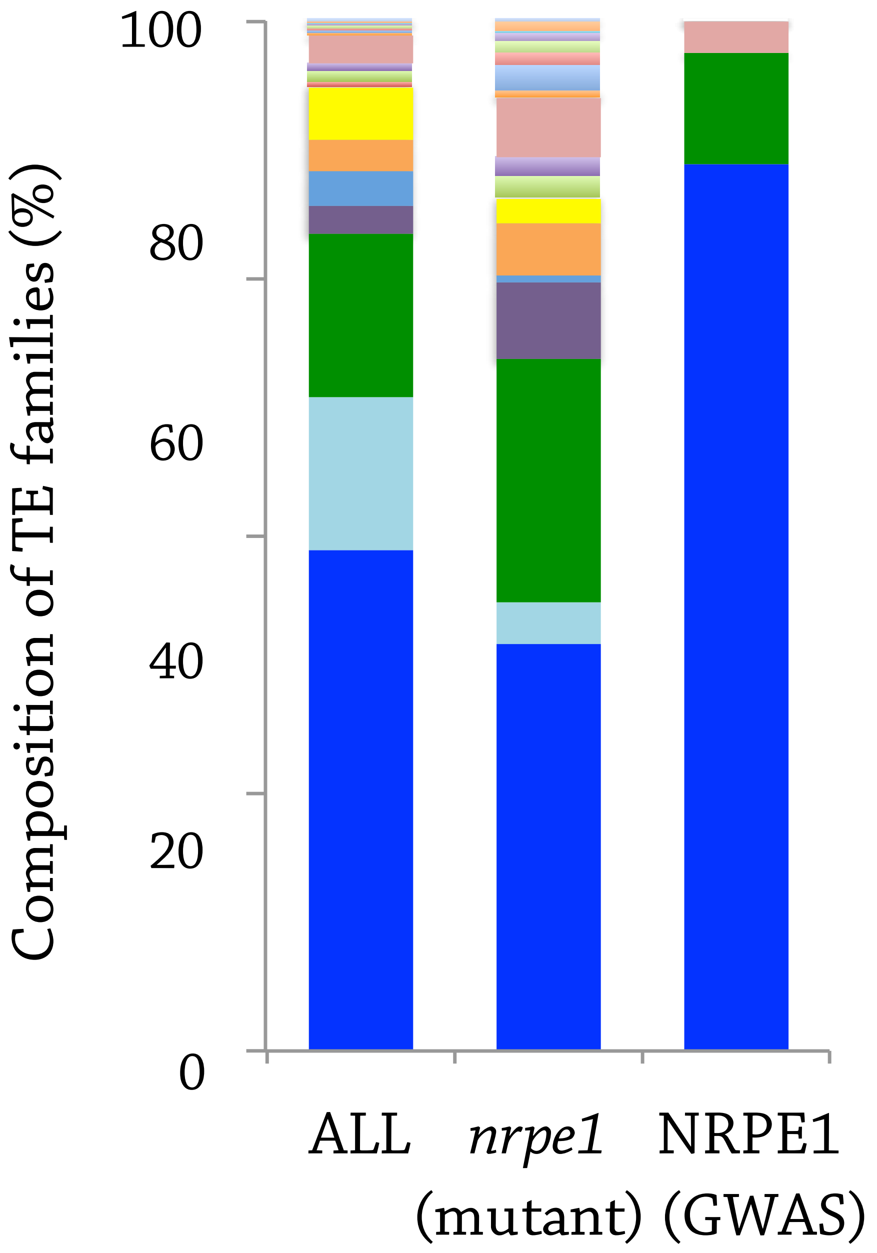
Number of targeted TEs



Profiles of targeted TEs

Natural alleles have higher specificity?





Summary

- The methylation pattern reflect temperature *of origin* — suggesting that plants remember their past environment
- Methylation variation has strong genetic basis — suggesting that the memory may simply be genetic

But…

- Why is pattern correlated with the environment?
- Why such an unusual genetic architecture?
- Both suggest selection — but what’s the phenotype?