Characterizing cis-regulatory elements in grapevine

Rachel Schwope

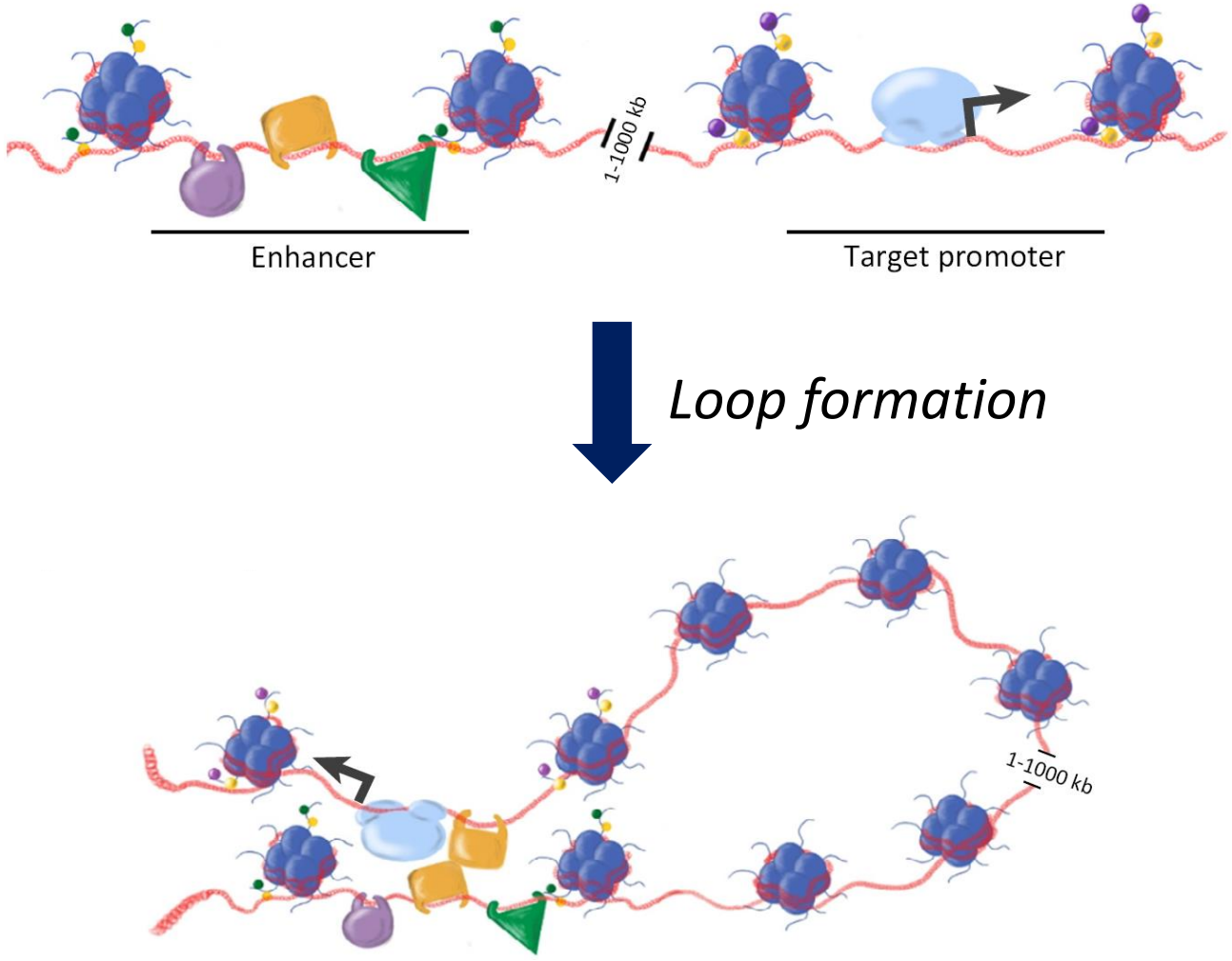
June 20, 2018

IGA

UNIVERSITÀ DEGLI STUDI DI UDINE

Cis-regulatory elements (CREs) affect gene function

- Regions of 300-600 bp where transcription factors bind
- Non-promoter elements
- Can increase or decrease gene expression of genes up to 1 Mb away

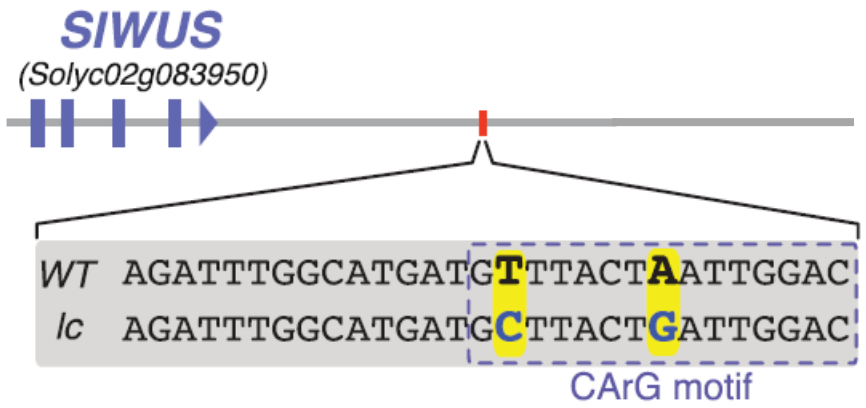


Adapted from Weber et al., *Trends in Plant Science* 2016

Small changes in one CRE can cause dramatic effects

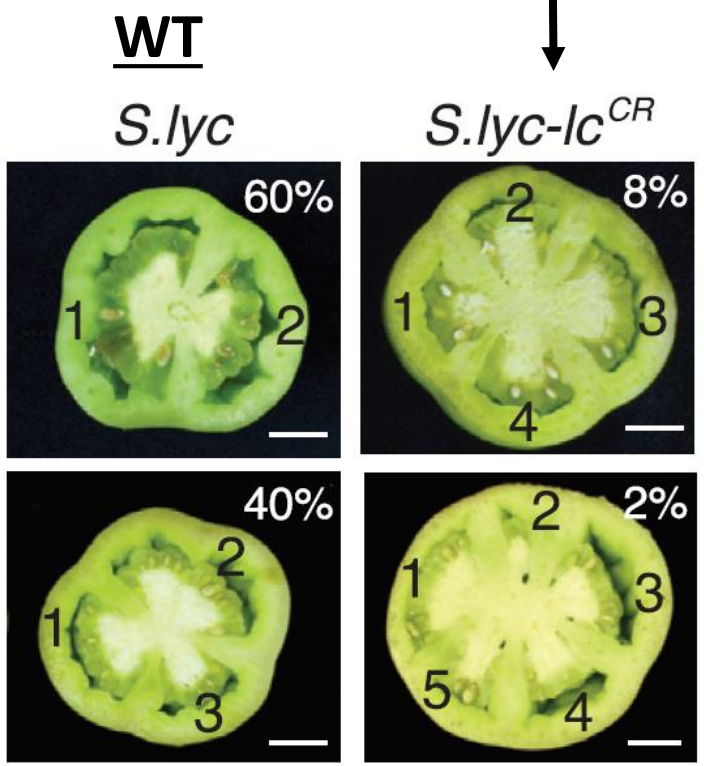
Example: disruption of *lc* repressive CRE in tomato

SIWUS gene promotes growth



Repressor site

Mutation of repressor site = increased number of tomato locules



Adapted from Rodríguez-Leal et al., 2017, Cell 171

Grapevine CRE project goals

http://botanyprofessor.blogspot.it

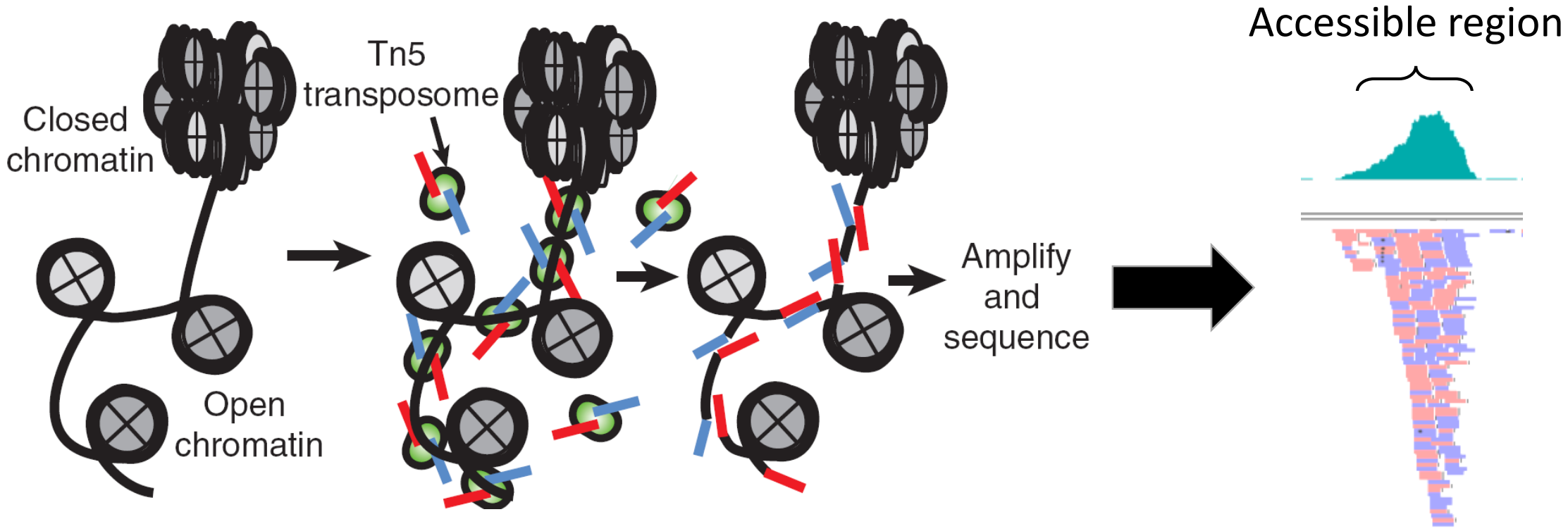
Next Generation Sequencing

Bioinformatics analysis

1. Identify cis-regulatory elements across genome in grapevine leaf
2. Understand their characteristics and effects on expression

Using ATAC-seq to find regulatory elements in grape

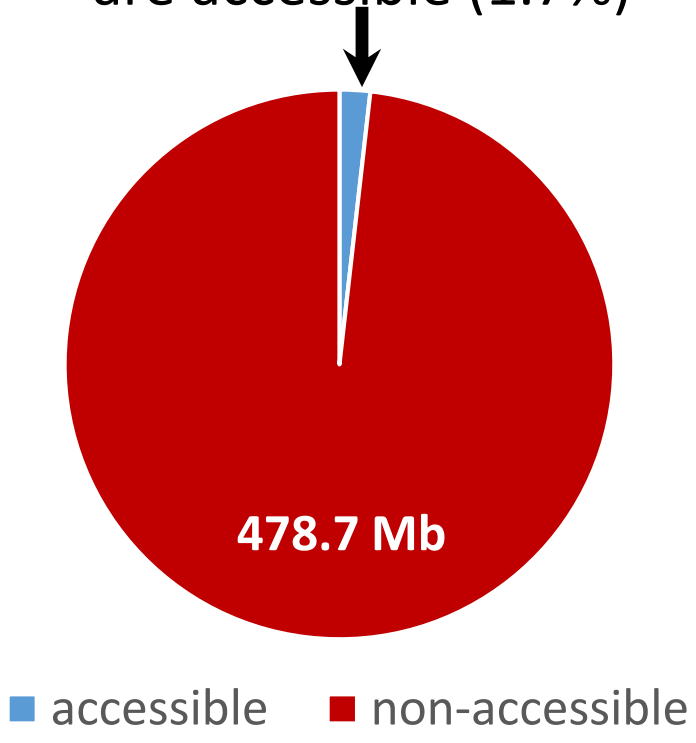
- Conveniently, TF binding sites are associated with **open chromatin**



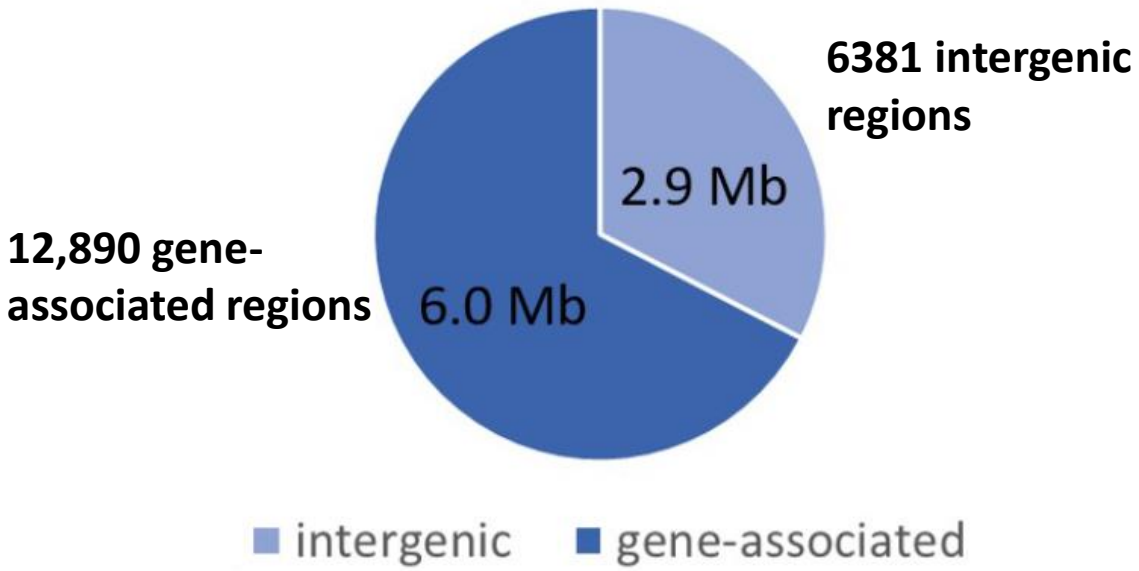
Adapted from Buenrostro et al., *Nature* 2013

What does ATAC-seq tell us about accessible loci?

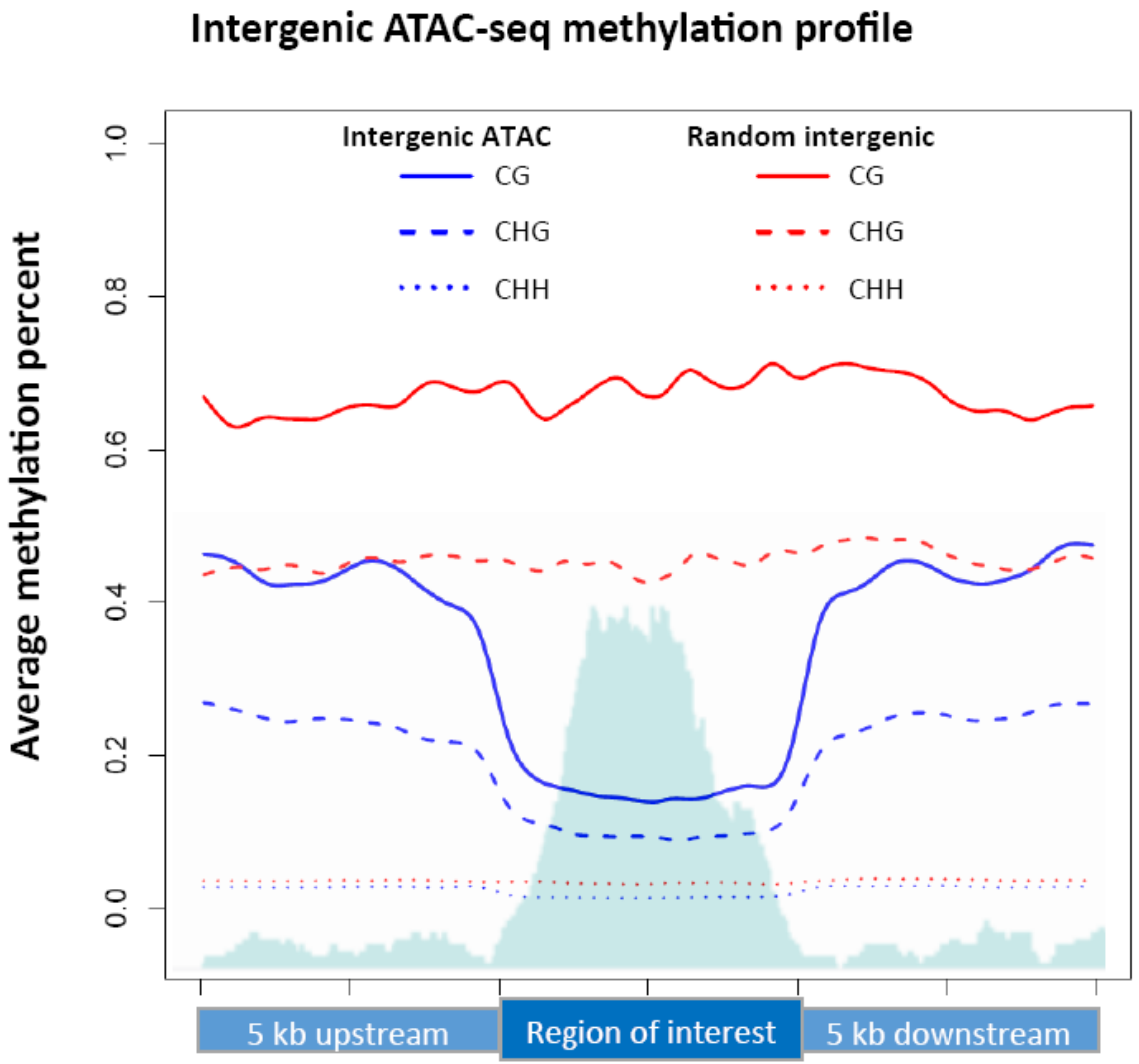
8.3 Mb out of 487 Mb are accessible (1.7%)



About two-thirds of accessible bases are gene-associated



Accessible regions lack cytosine methylation

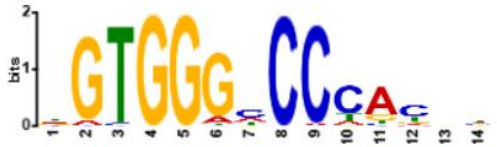


Control regions: 500 bp intergenic, non-TE windows

Bisulfite sequencing: Mirko Celii

Accessible regions have TF binding sites and are likely CREs

TCP



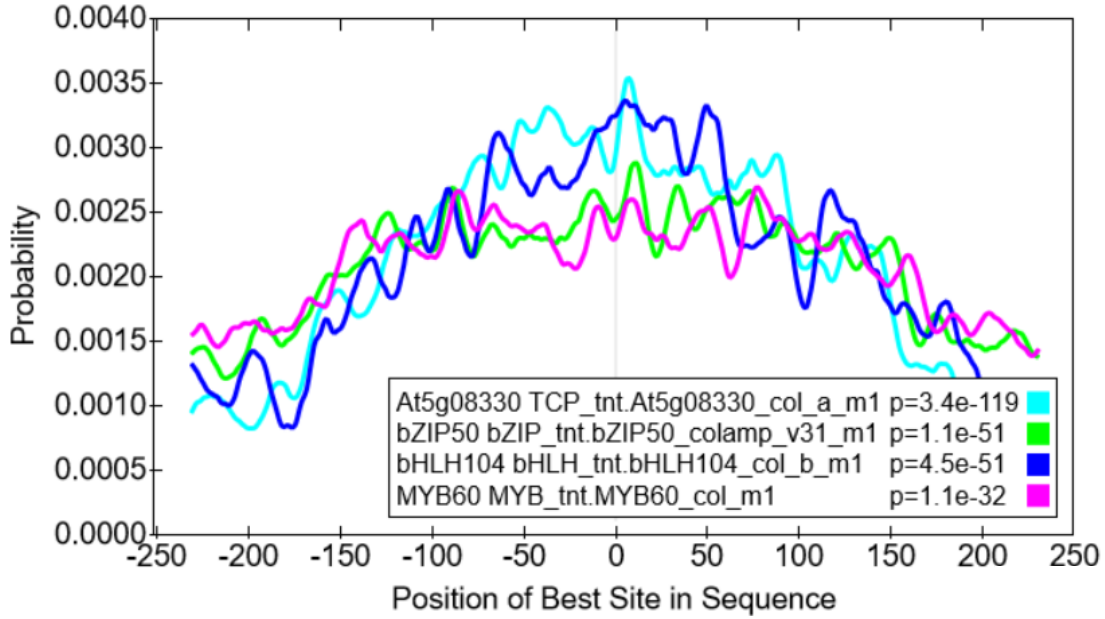
bZIP50

bHLH104

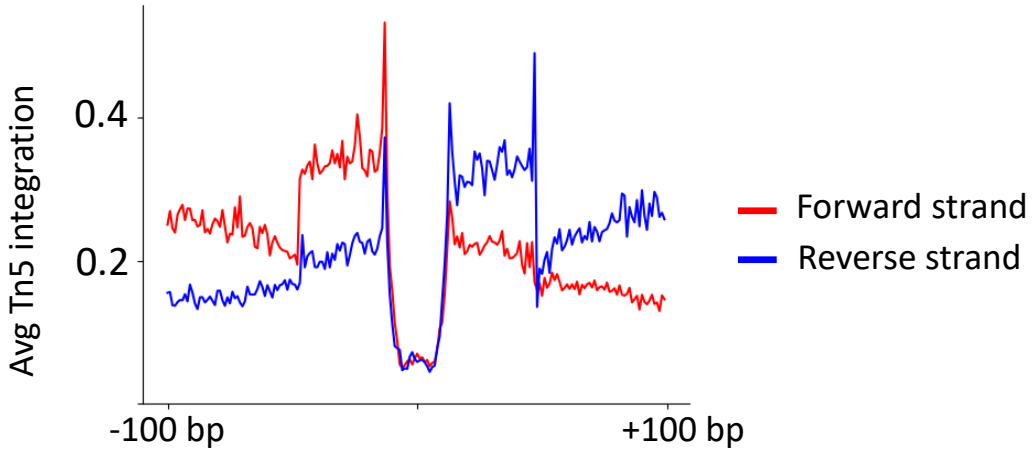
Myb60

Output from MEME-Suite v 4.12.0

Motifs are centrally enriched in intergenic ATAC-seq peaks

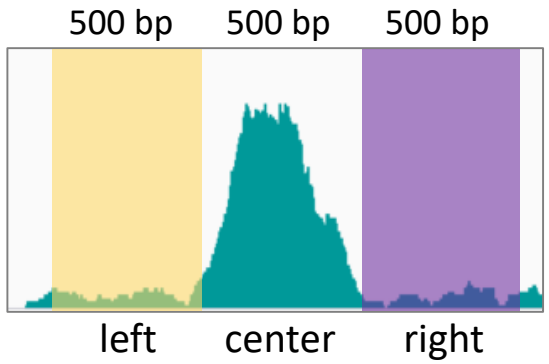


Transposase integration footprints in ATAC-seq regions suggest motifs are likely protein-bound

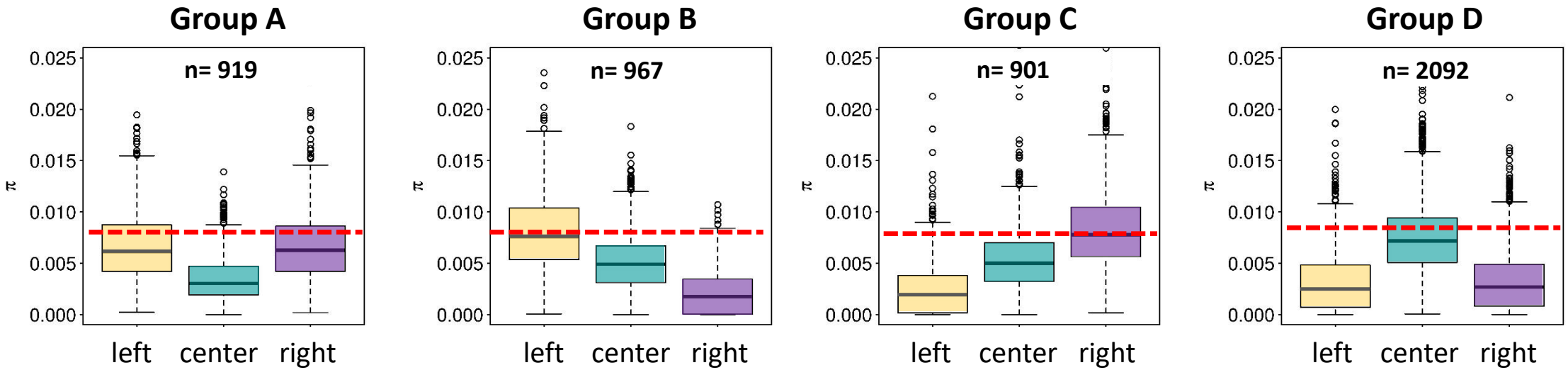


Output from pyDNase v 0.2.4

Four nucleotide diversity trends observed at peaks



Nucleotide diversity (π)



Genome π average = 0.0078

Nucleotide diversity: Gabriele Magris

Putative CREs frequently near TF genes

Significantly enriched Molecular Function GO terms for each group:

	GO.ID	Term	p-value
Group A	GO:0003677	DNA binding	4.37E-16
	GO:0003700	transcription factor activity	4.37E-16

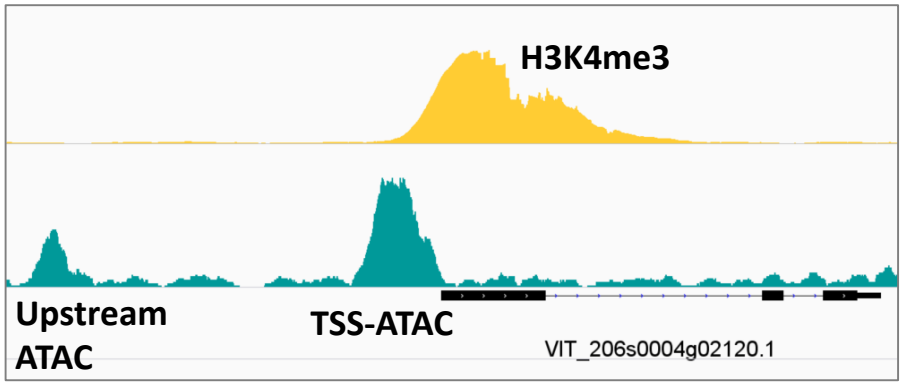
	GO.ID	Term	p-value
Group B	GO:0003700	transcription factor activity	5.52E-08
	GO:0003677	DNA binding	1.14E-07

	GO.ID	Term	p-value
Group C	GO:0003677	DNA binding	5.28E-11
	GO:0003700	transcription factor activity	1.06E-10

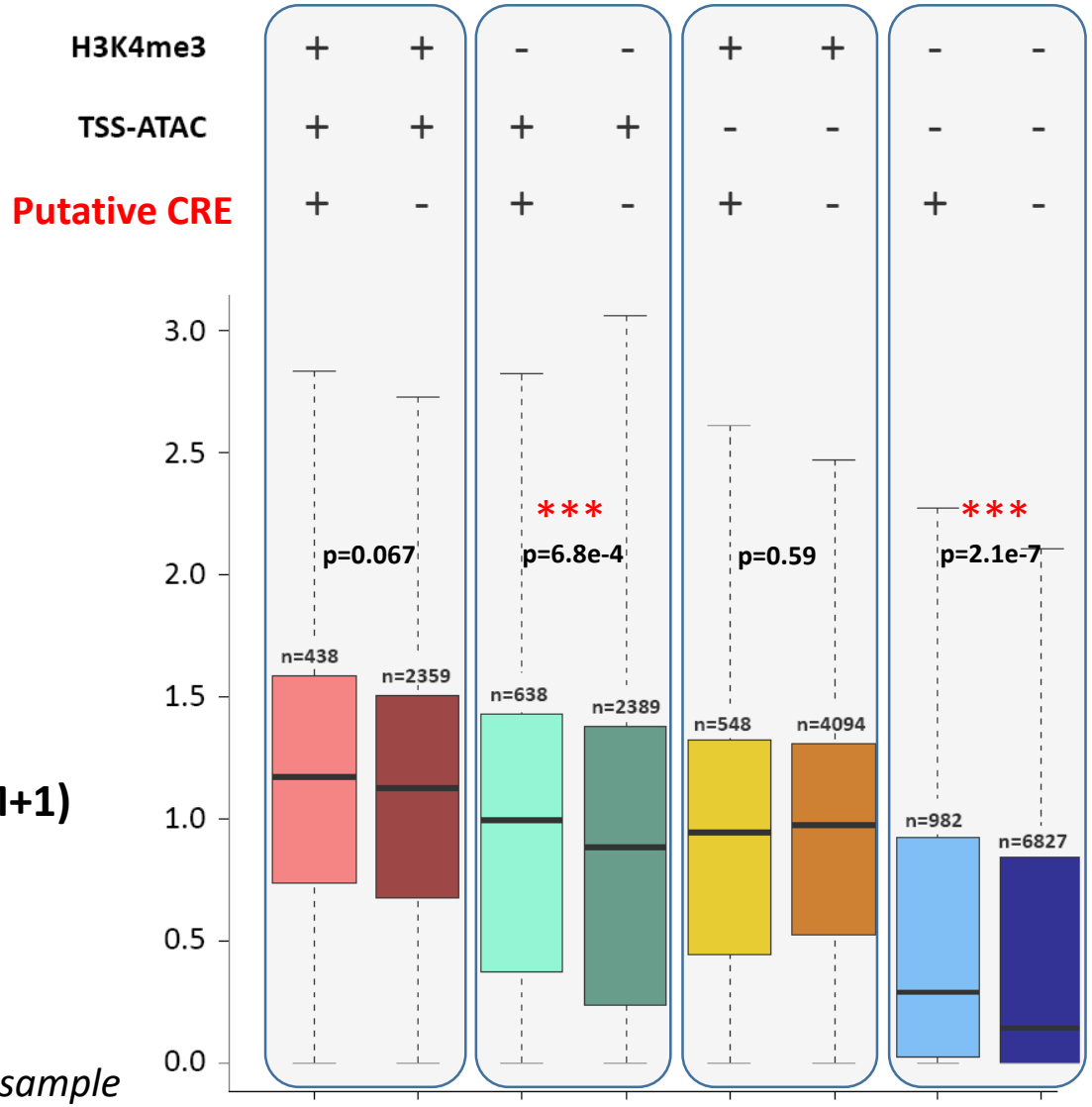
	GO.ID	Term	p-value
Group D	GO:0003700	transcription factor activity	1.99E-23
	GO:0003677	DNA binding	2.16E-20

Effects of local environment on expression

Example gene environment:



Putative CRE



Log10(FPKM+1)

p-values: *unpaired two-sample Wilcoxon test*

RNA-seq: Eleonora Paparelli

Expression is a product of multiple variables

- Linear model shows **TSS environment** highly indicative of FPKM
  - Nuclear compartment and gene-body methylation are also major contributors

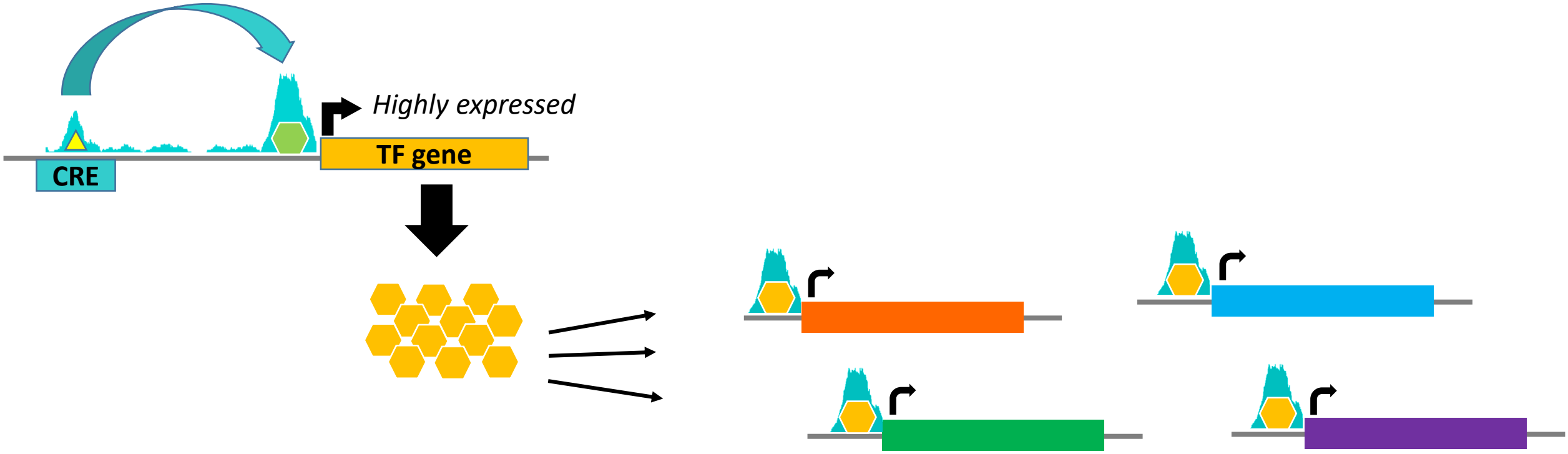
	Correlation direction	Standard Error	t-value	p-value
H3K4me3 FE	**+**	0.0017	32.95	4.13E-229
TSS ATAC-seq FE	**+**	0.0031	29.81	3.28E-189
Putative CRE FE	**+**	0.0043	8.93	4.81E-19
Promoter 5mC	-	0.0004	-10.28	1.07E-24
'A' Compartment (Hi-C)	+	0.0003	30.42	1.10E-196
Gene body 5-mCpG	+	0.0236	29.74	2.31E-188
Promoter TE density	+	0.0001	4.87	1.14E-06

Variance accounted for by model: ***28.29%***
p-value of model ***< 2.2e-16***

Linear model (R version 3.3): Gabriele Magris

Enhancer activity propagated through increased TF production?

- Of approximately 32,000 genes, only ~4700 of them have a nearest putative CRE
- What about the other genes?



Acknowledgements

Principal Investigator: Michele Morgante

IGA and University of Udine
Mirko Celii
Emanuele de Paoli
Gabriele Magris
Fabio Marroni
Eleonora Paparelli
Sara Pinosio
Aldo Tocci
Michele Vidotto

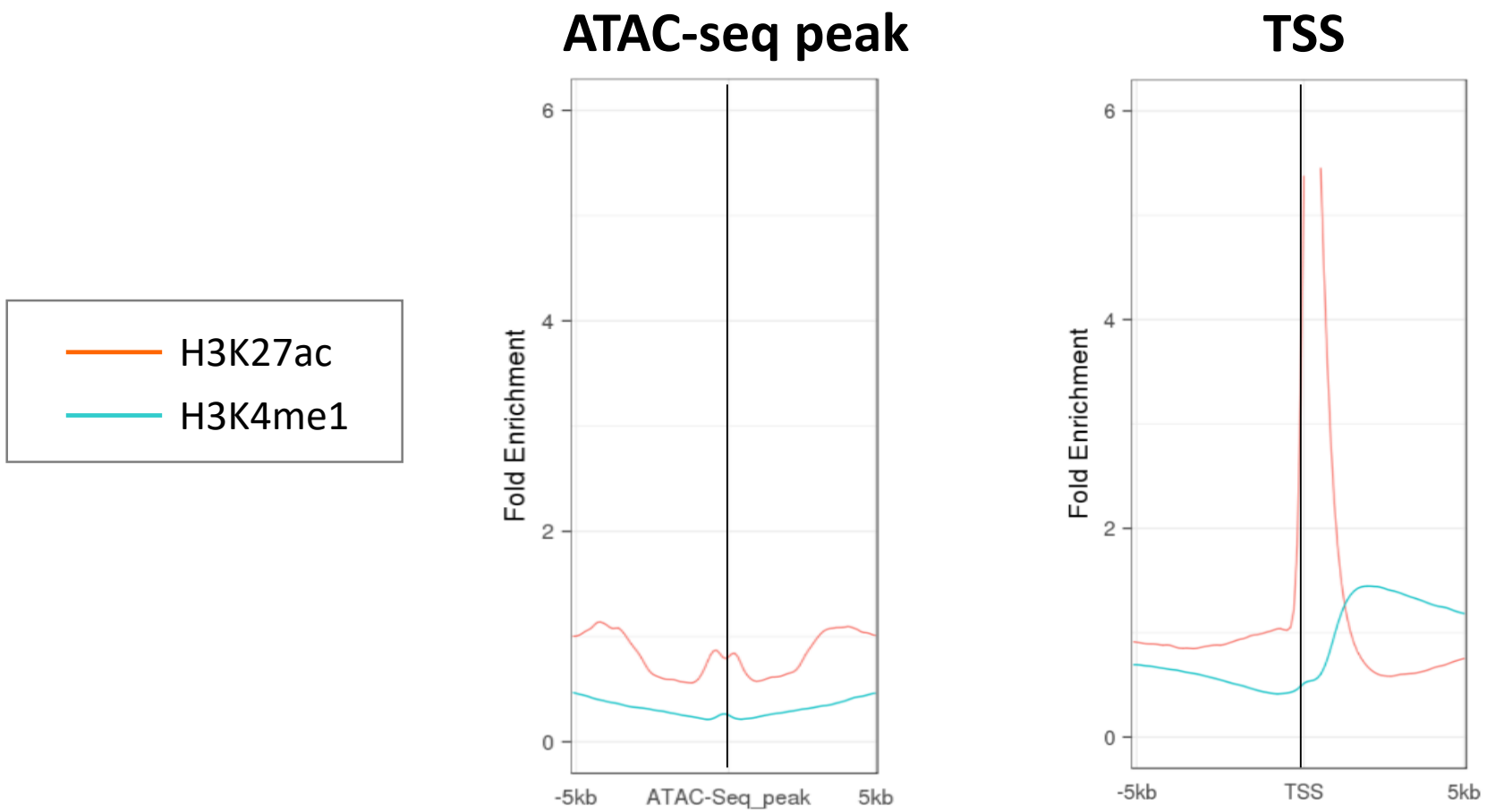
IGA Technology Services
Irena Jurman
Giacomo Prete

Lab and Administration
Nicoletta Felice
Giusi Zaina

Supported by the ERC project NOVABREED - Novel variation in plant breeding and the plant pan-genomes (Grant agreement no.: 294780)

Are regions identifiable by histone markers?

- Neither H3K27ac nor H3K4me1 are suitable markers for enhancer identification



Random control regions aggregate footprint