Characterization of *Vitis vinifera* chromatin domains via Hi-C sequencing

Aldo Tocci[1,2,3], R. Schwope[2,3], F. Marroni[2,3], M. Morgante[2,3]

[1]Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy
[2]Dip. di Scienze agroalimentari, ambientali e animali (Di4a), Udine, Italy
[3]Istituto di Genomica Applicata (IGA), Udine, Italy

SISSA — *ma per seguir virtute e canoscenza*
UNIVERSITÀ DEGLI STUDI DI UDINE
erc — European Research Council
IGA — ISTITUTO DI GENOMICA APPLICATA

Introduction

The three-dimensional structure of chromatin and the DNA function are in a strict relationship affecting each other. We use Pinot Noir grapevine (*Vitis vinifera*), an economically significant crop, to investigate this relationship. The genome of Pinot Noir contains thousands of large insertions or deletions, known as Structural Variants (SV) that can influence chromatin organization and the activity of DNA. This influence can be investigated using Hi-C. Recently, a closer inspection of Hi-C data by Principal Component Analysis (PCA) revealed the existence of discrete Structural Domains (SD) in genomes (Grob, *et al.* 2014), classified in Compact Structural Domains (CSD, related to regions of inactive chromatin) and Loose Structural Domains (LSD, related to regions of active chromatin).
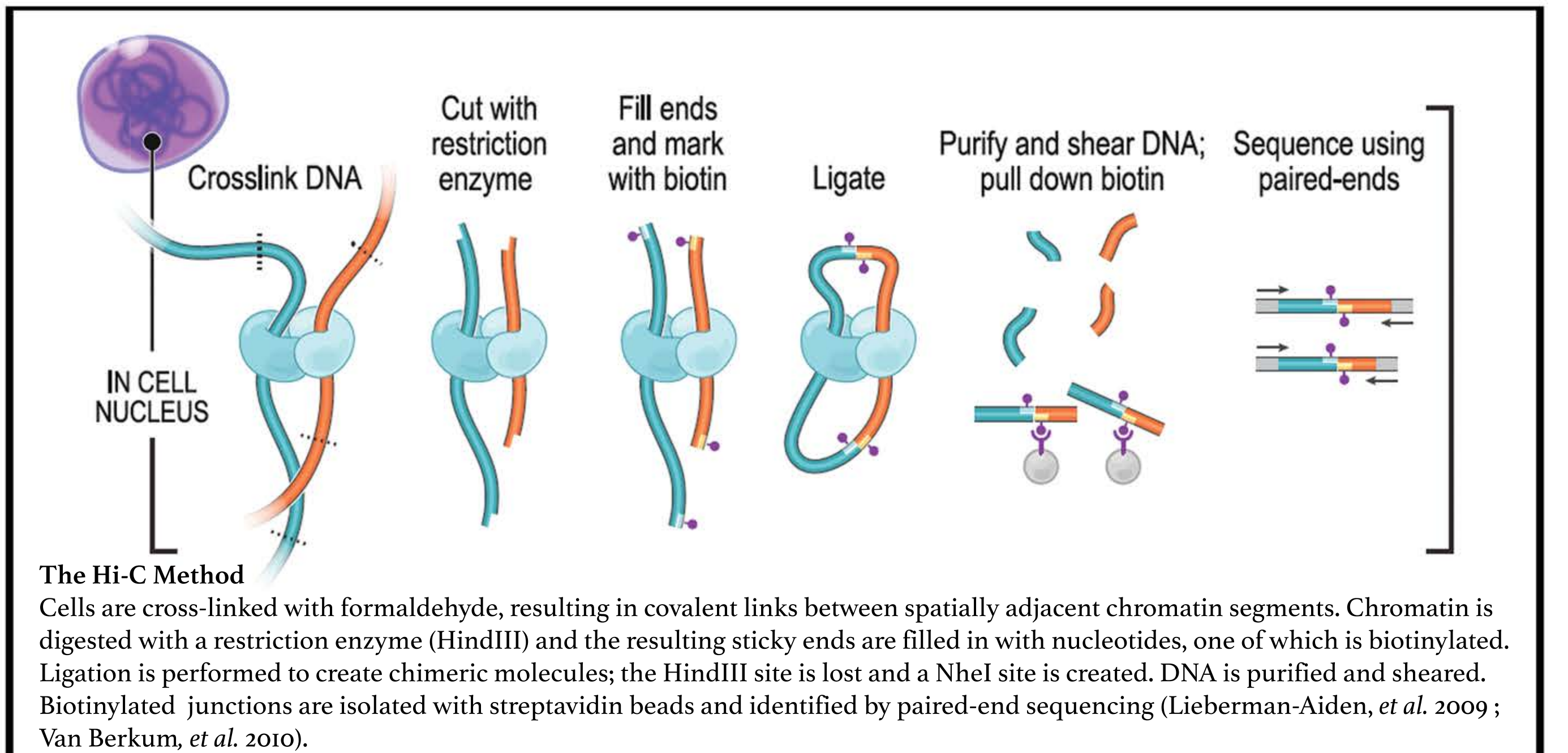
Methods

- **Hi-C:** a derivative method of Chromosome Conformation Capture that explores the three-dimensional architecture of whole genomes by coupling proximity-based ligation with next generation sequencing.
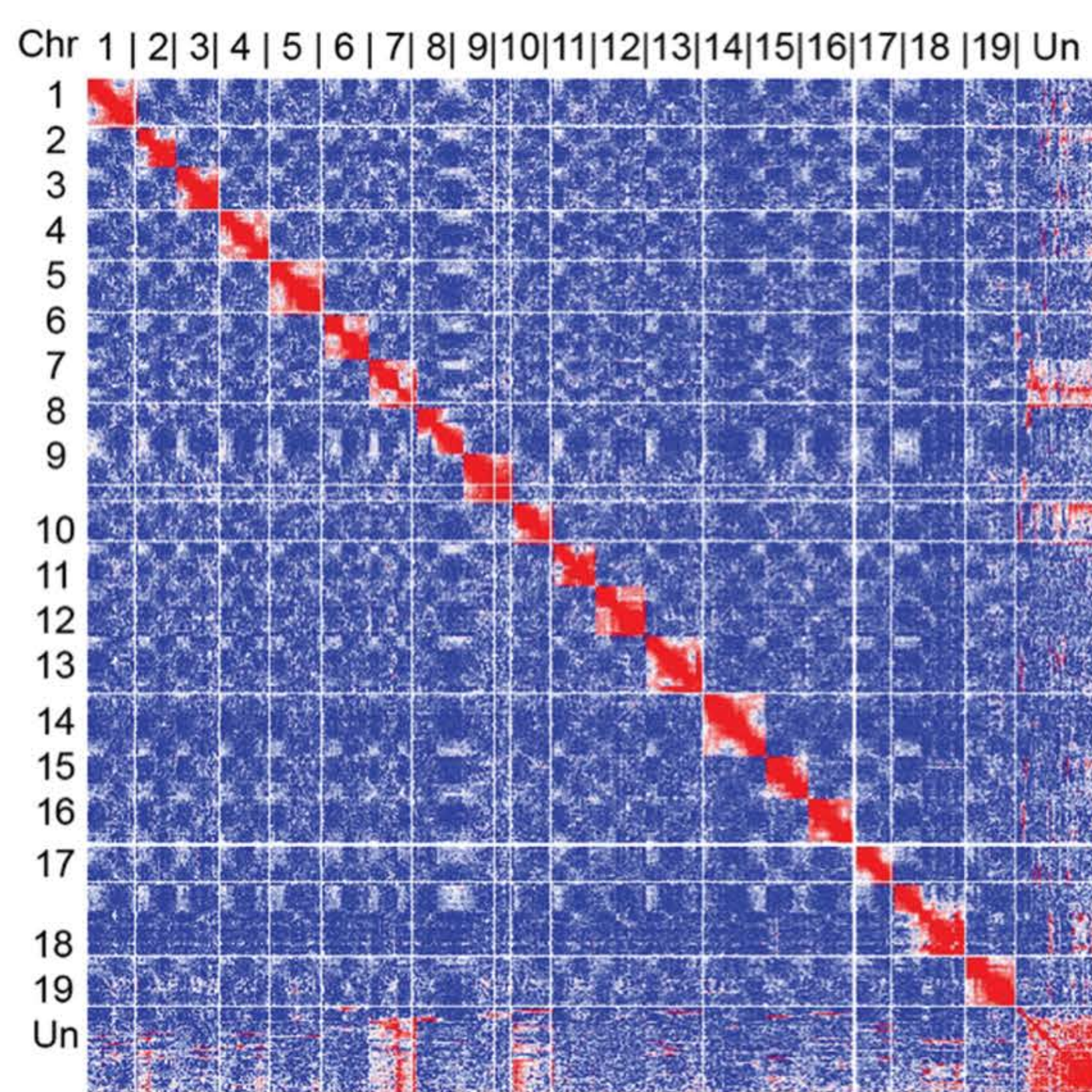
- **Alignment:** BWA-mem.

- **Contact maps and PCA software:** "HOMER" (Heinz, *et al.* 2010).

- **Additional analysis software:** "HiC-Pro" (Servant, *et al.* 2015).

Crosslink DNA · **Cut with restriction enzyme** · **Fill ends and mark with biotin** · **Ligate** · **Purify and shear DNA; pull down biotin** · **Sequence using paired-ends**

IN CELL NUCLEUS

The Hi-C Method
Cells are cross-linked with formaldehyde, resulting in covalent links between spatially adjacent chromatin segments. Chromatin is digested with a restriction enzyme (HindIII) and the resulting sticky ends are filled in with nucleotides, one of which is biotinylated. Ligation is performed to create chimeric molecules; the HindIII site is lost and a NheI site is created. DNA is purified and sheared. Biotinylated junctions are isolated with streptavidin beads and identified by paired-end sequencing (Lieberman-Aiden, *et al.* 2009 ; Van Berkum, *et al.* 2010).

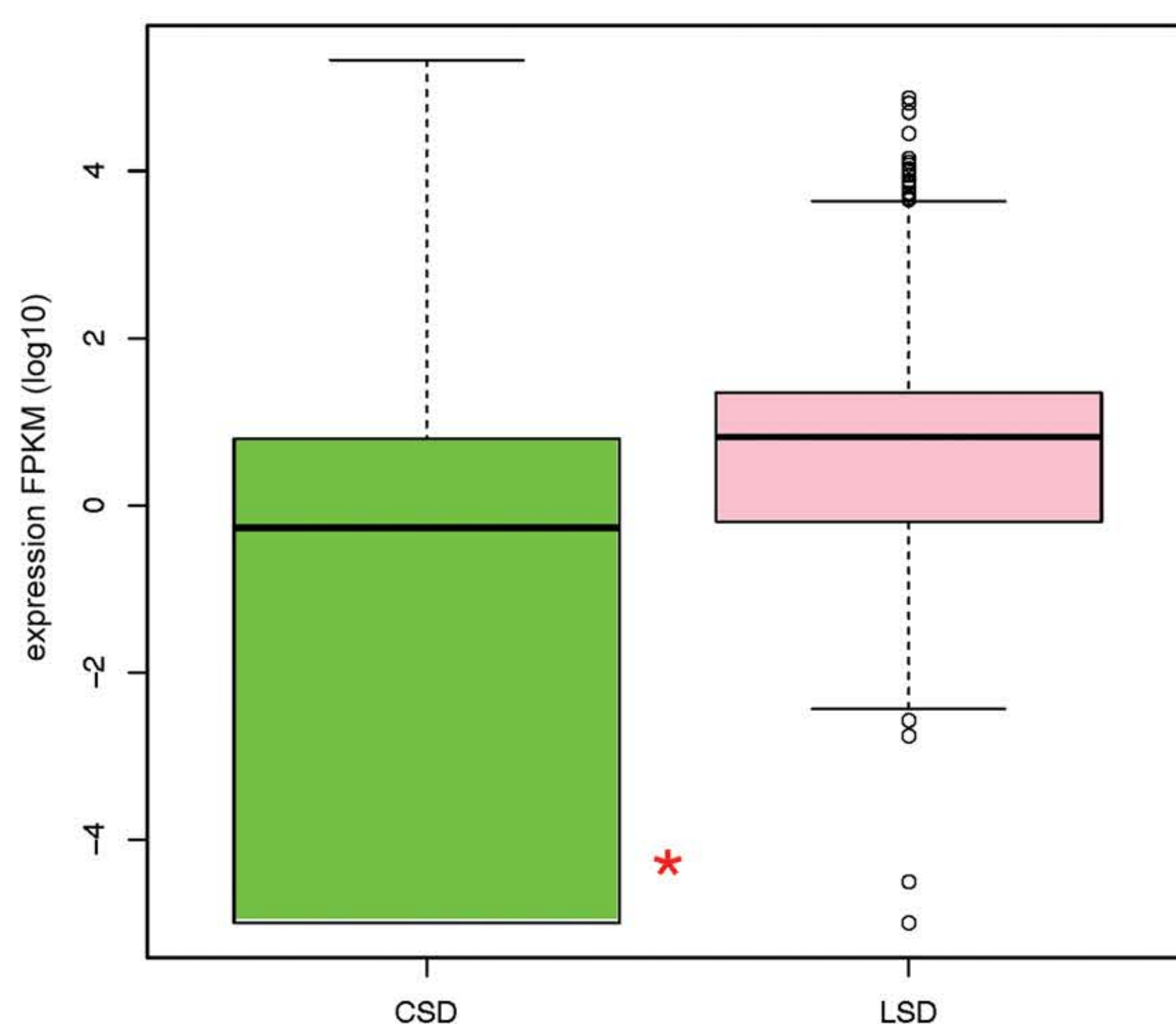## Results

Vitis vinifera Genomic Contact Map

The contact map reports the frequency of interaction for the 19 chromosomes of *Vitis vinifera*, plus the "chromosome Unknown", a set of scaffolds that could not be associated to any chromosome during the assembly. The strong interaction chrUn-chr7 and chrUn-chr10 suggest the potential to correct misassembly issues. (Color scale: blue to red for increasing interaction frequency).

### PCA identifies Structural Domains in Vitis chromosomes
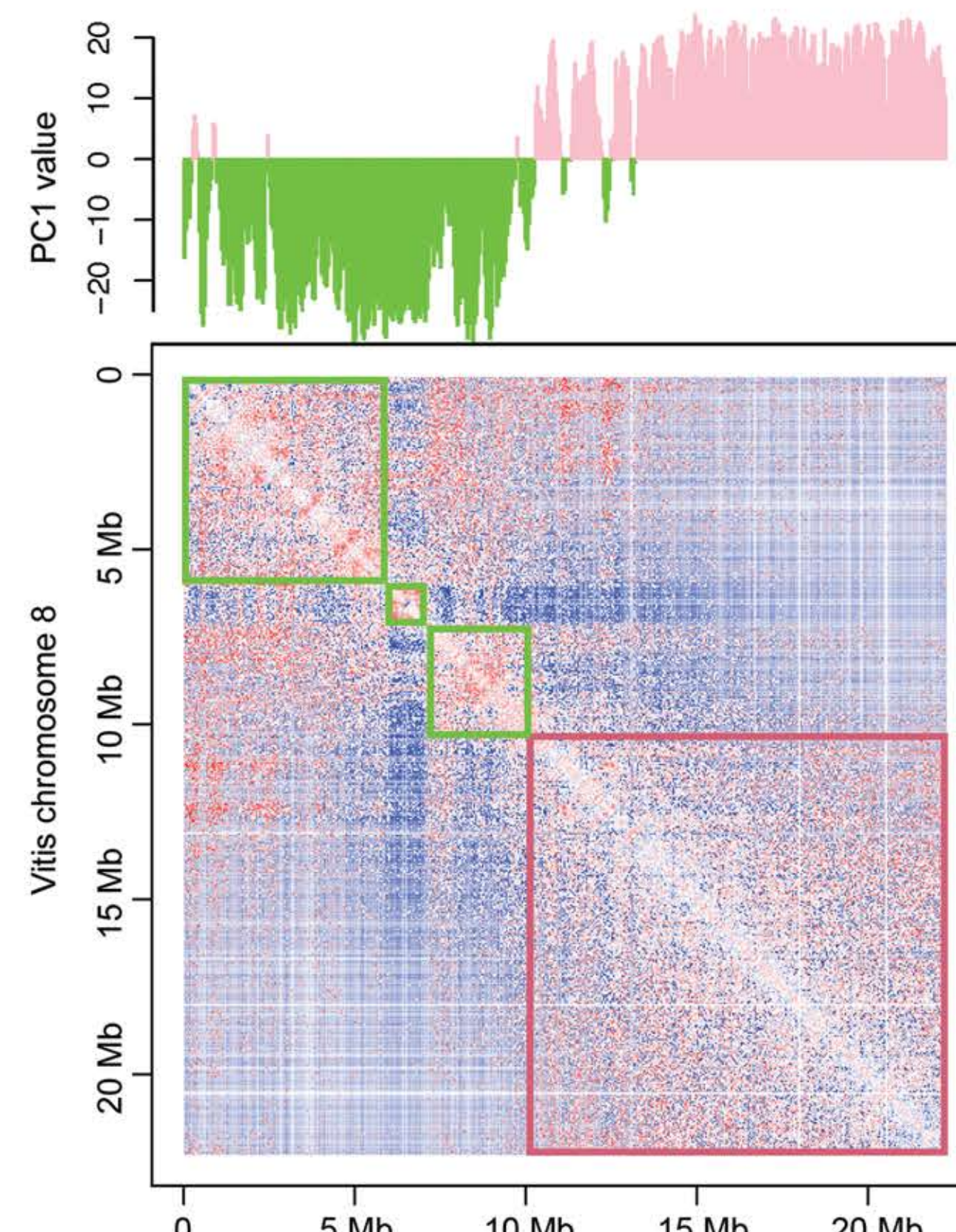
Contact map of *Vitis vinifera* chromosome 8: different contact patterns can be distinguished, corresponding to different SDs. PCA provides a numeric classification of SDs in CSD and LSD (in green and pink, respectively) depending on the sign of Principal Component 1.

Vitis vinifera SV distribution in SDs

Deletions in Vitis Structural Domains

CSD: 0.683 / 3397 · LSD: 0.457 / 2064

Insertions in Vitis Structural Domains

CSD: 1.344 / 6682 · LSD: 0.837 / 3781

The two barplots report the average number of deletions (upper plot) and insertions (lower plot) per 50 Kb window in *Vitis vinifera* genome in CSD or LSD (green bar and pink bar, respectively). The total number of observed SVs is also reported. (Red star= significant chisquare test).

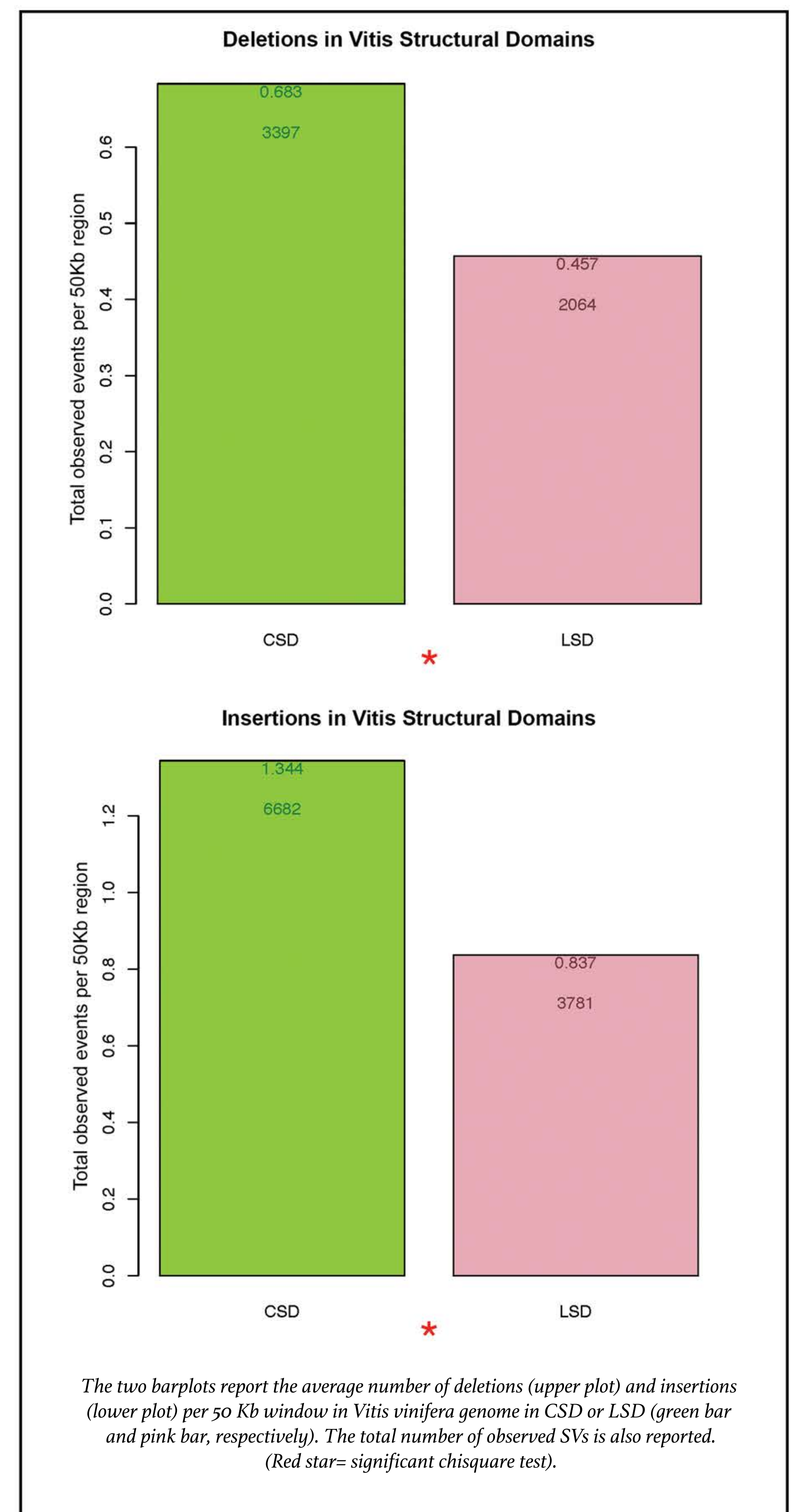### Expression levels in Vitis Structural Domains

The boxplot reports the average FPKM (log_{10}) as a measure of the expression levels in different SDs for windows of 50Kb in *Vitis vinifera* genome. LSDs show significant higher expression levels (pink box) when compared to CSDs (green box). (Red star= significant Wilcoxon test).

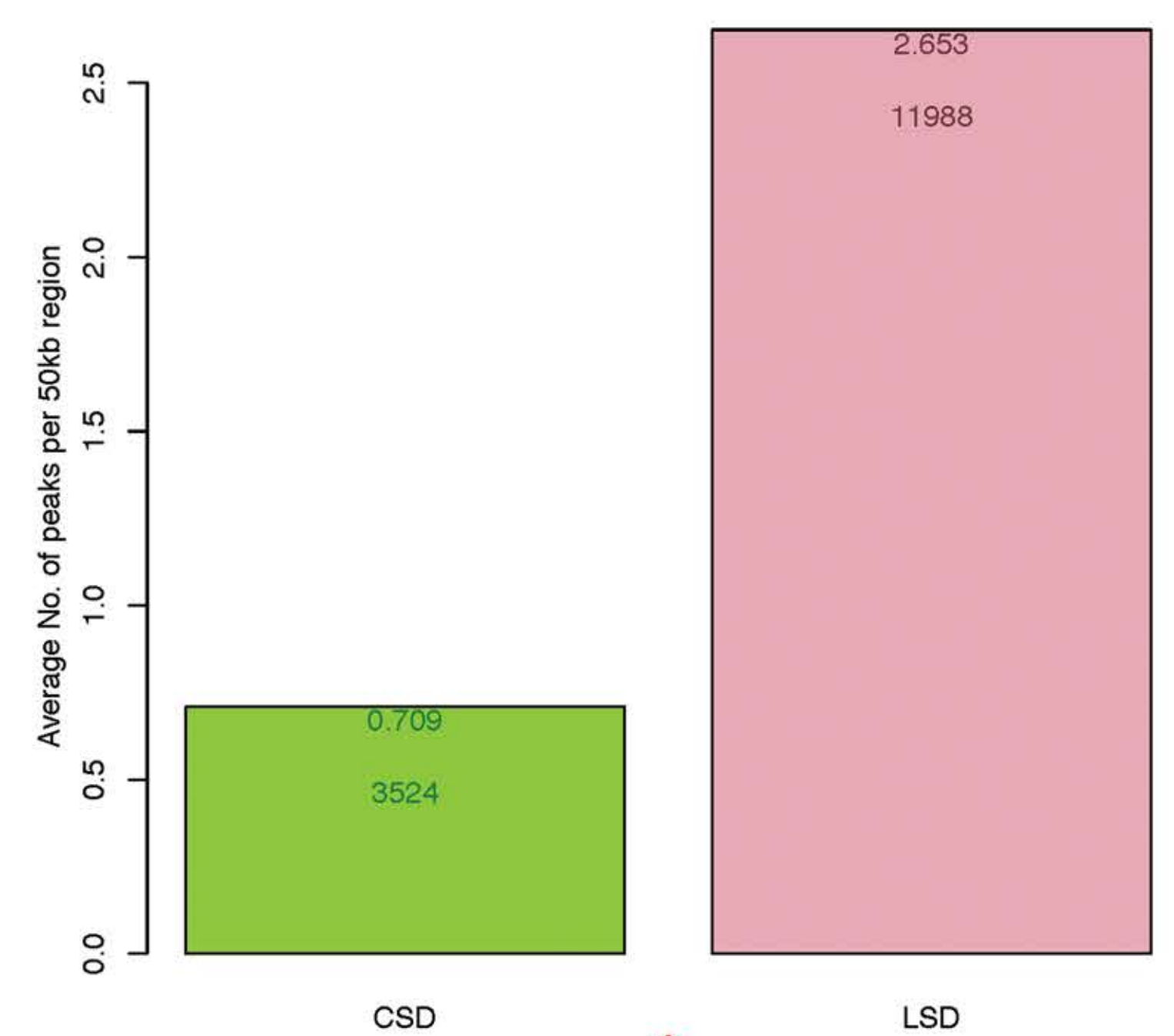### CG methylation in Vitis Structural Domains
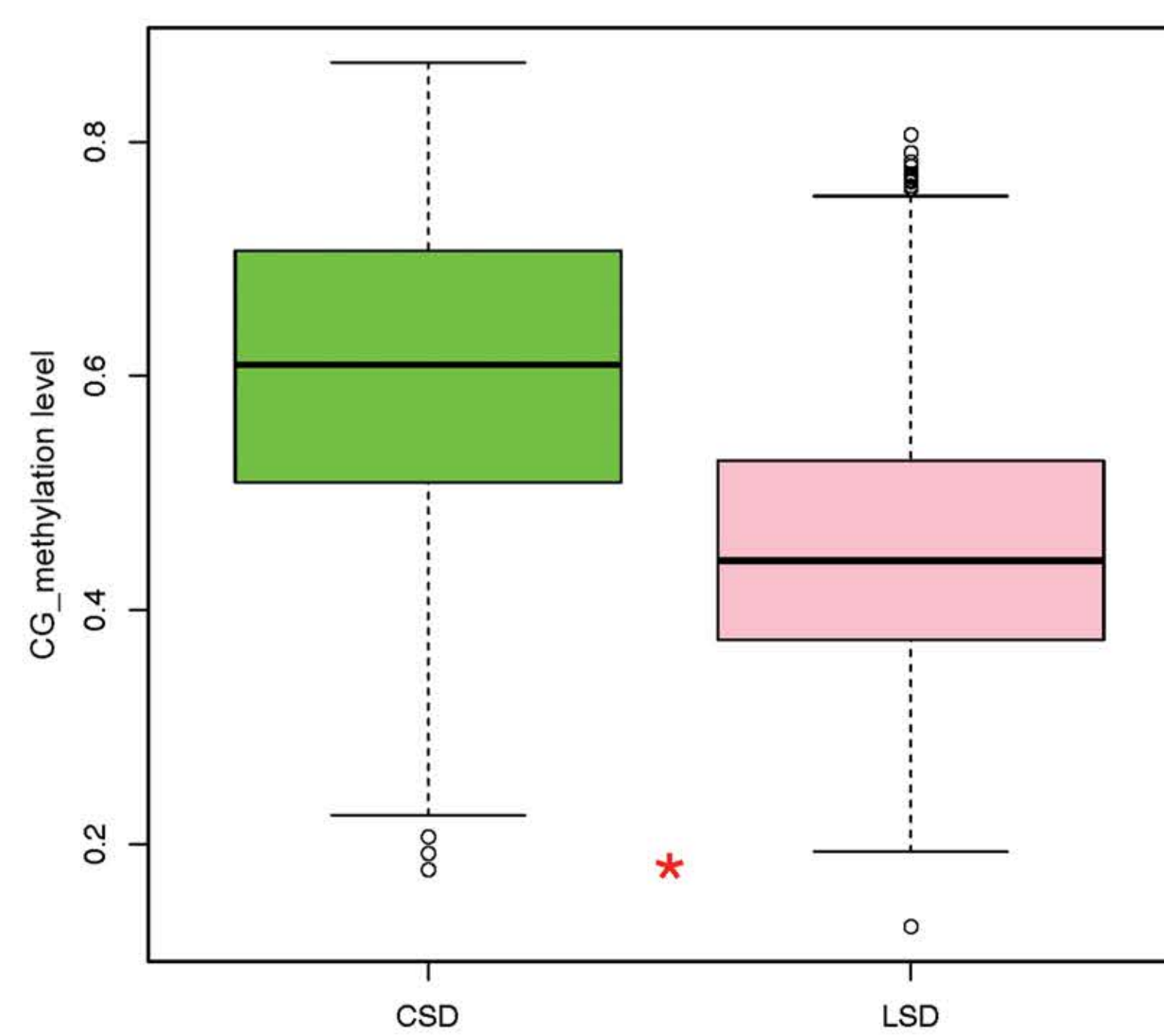
The boxplot reports the average CG methylation level in different SDs for windows of 50Kb in *Vitis vinifera* genome. CG methylation is higher in CSD (green box). (Red star= significant Wilcoxon test).

H3K4me3 Peaks in Vitis Structural Domains

CSD: 0.709 / 3524 · LSD: 2.653 / 11988

The barplot reports the average number of H3K4me3 peaks found in CSD or LSD (green bar and pink bar, respectively) in 50 Kb windows of *Vitis vinifera* genome. The total number of observed events is also reported. (Red star= significant chisquare test).

Conclusions

We used Hi-C to build the contact map for *Vitis vinifera*. Even this first result can give a hint of the power of this method to improve the existing *Vitis vinifera* genome assembly. We then identified the SDs through PCA and integrated such information with existing genomic feature data. We show that CG DNA methylation is found at higher levels in CSD, while higher gene expression occurs in LSD regions. Moreover we found a high level of SVs in CSD regions, while LSD regions are characterized by higher presence of chromatin activation markers like H3k4me3.

Future Perspectives

Hi-C data can be furhter employed for:
a. The resolution of haplotypes at a molecular level;
b. Allele specific analysis of grapevine genome to investigate the effect of SVs on chromosomal interactions.
c. The improvement of genome assembly.

Acknowledgements

Contacts

Aldo Tocci, Istituto di Genomica Applicata (IGA)
via J. Linussio, 51-Udine 33100 Italy
Phone: 0432 629761
e-mail: atocci@appliedgenomics.org