TADbit

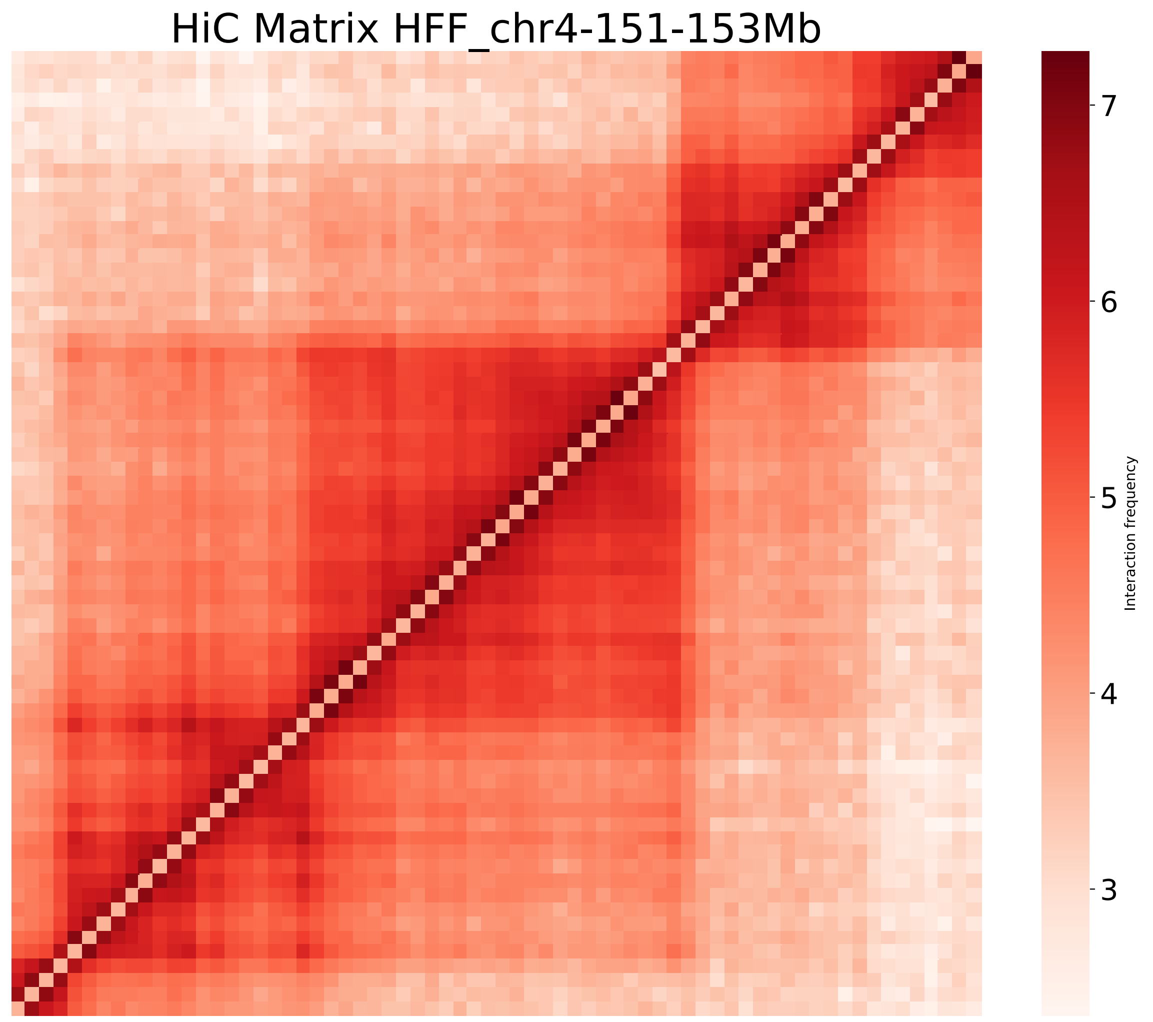
Comparing experiments

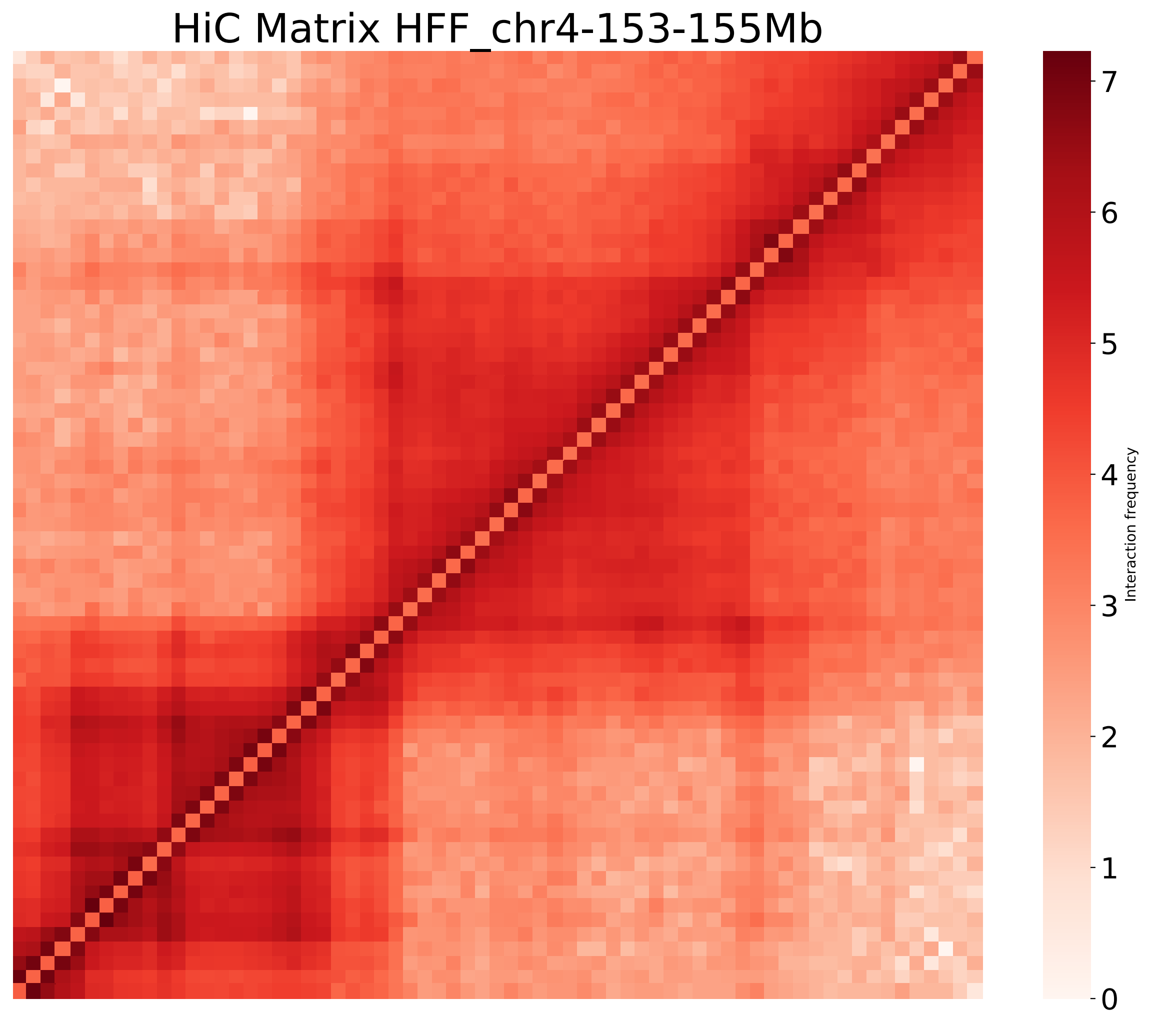
Marco Di Stefano, David Castillo & Marc A. Marti-Renom
Structural Genomics Group (CNAG-CRG)

cnag CRG Centre for Genomic Regulation

Comparing Hi-C experiments

- Assess reproducibility (replicates)
- Identify structural differences (different conditions)


HiC Matrix HFF_chr4-151-153Mb


HiC Matrix HFF_chr4-153-155Mb

Traditional metrics

Pearson and Spearman correlation of each value in the matrix

Drawbacks:

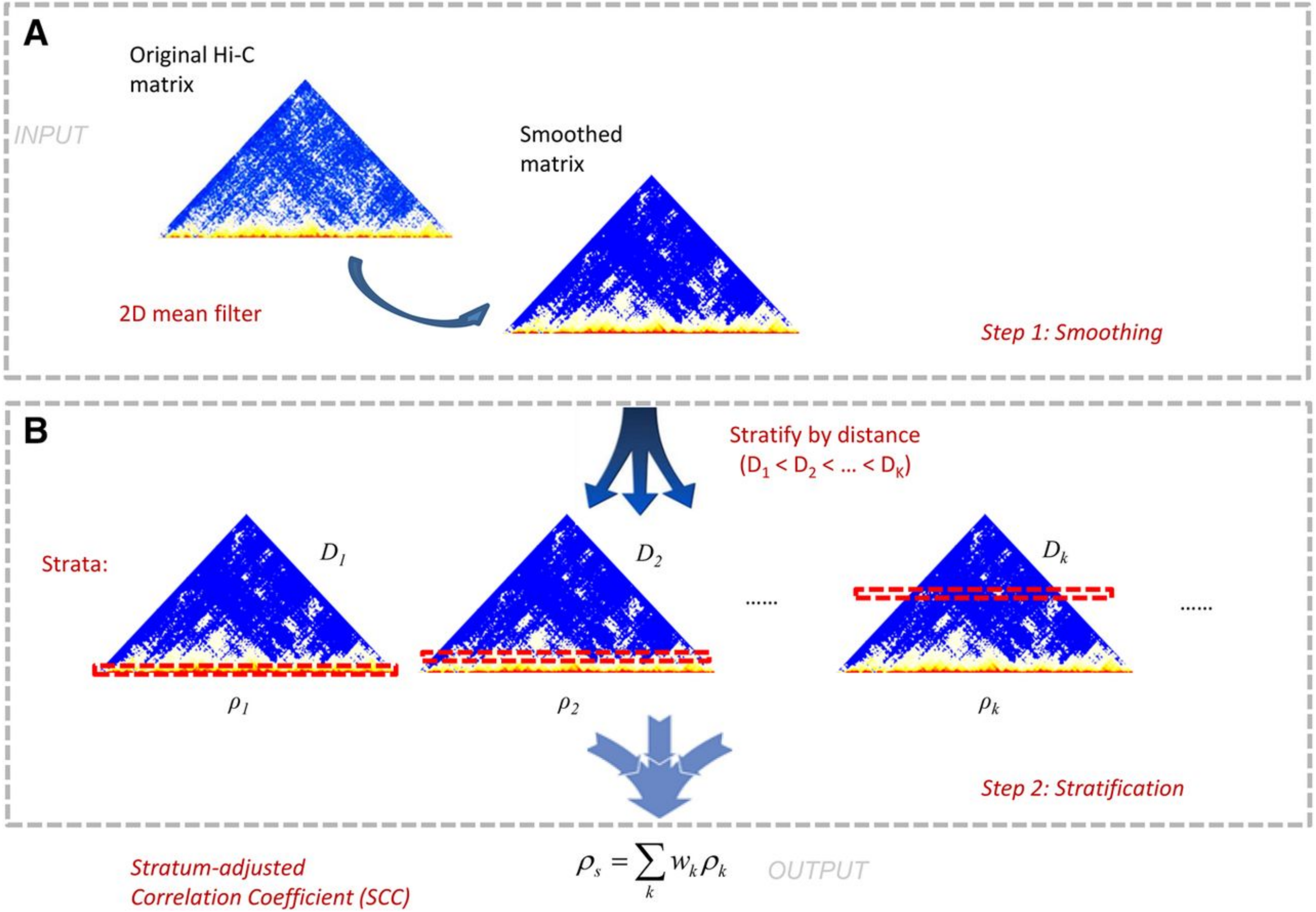
- Susceptible to outliers
- Treats all elements of the Hi-C matrix as independent measurements
- Most of the signal comes from pairs that are less than 1Mb apart. Correlation coefficient is dominated by these short-range interactions

Tailored metrics

- Based on correlation
  - HICRep (https://pubmed.ncbi.nlm.nih.gov/28855260/)
- Based on decomposition
  - HiC-spector (https://academic.oup.com/bioinformatics/article/33/14/2199/3078603)

HICRep

Stratum-adjusted Correlation Coefficient (SCC)



$$\rho_s = \sum_k w_k \rho_k$$

HiC-spector

Reproducibility score Q

Laplacian

$$\mathscr{L} = D - W$$

$$D_{ii} = \sum_j W_{ij} \qquad \begin{pmatrix} D_{11} & 0 & 0 & 0 & \dots \\ 0 & D_{22} & 0 & 0 & \dots \\ 0 & 0 & D_{33} & 0 & \dots \\ 0 & 0 & 0 & D_{44} & \dots \end{pmatrix}$$

Normalized Laplacian

$$l = D^{-1/2} L D^{-1/2}$$

↓ n first eigenvectors

$$\{v_0, v_1, \dots, v_{n-1}\}$$

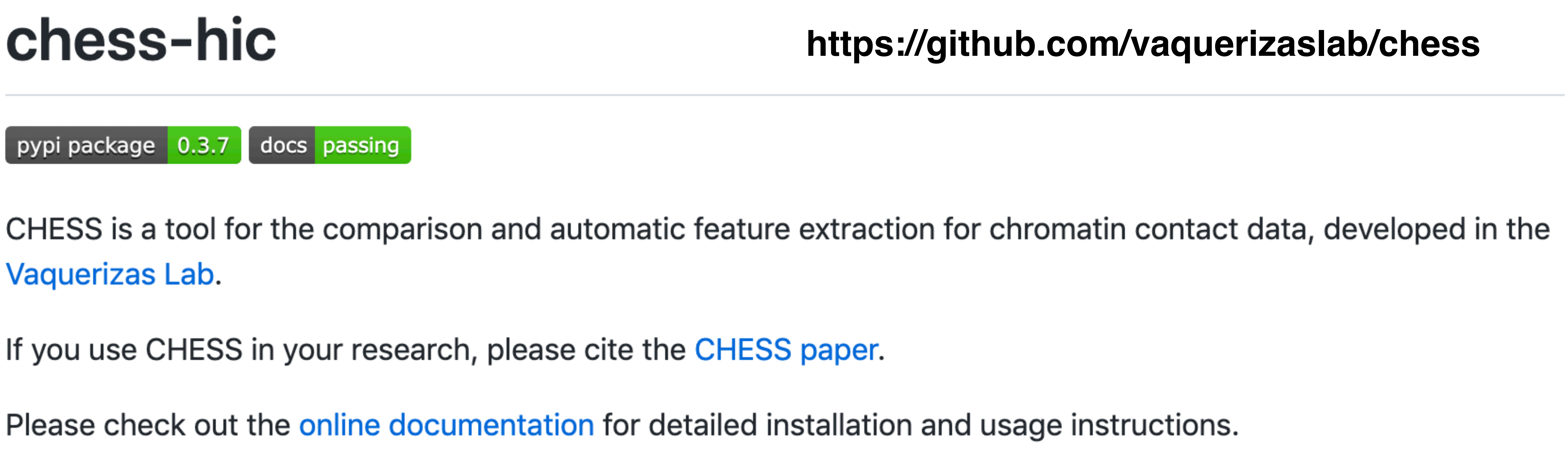
$$S_d(A, B) = \sum_{i=0}^{n-1} \| v_i^A - v_i^B \|$$

Q *is* $S_d(A, B)$ *linearly rescaled*

Reproducibility metrics

- Do not provide a systematic approach to identify structural changes

- Visual inspection is needed to pinpoint the differences https://github.com/vaquerizaslab/chess

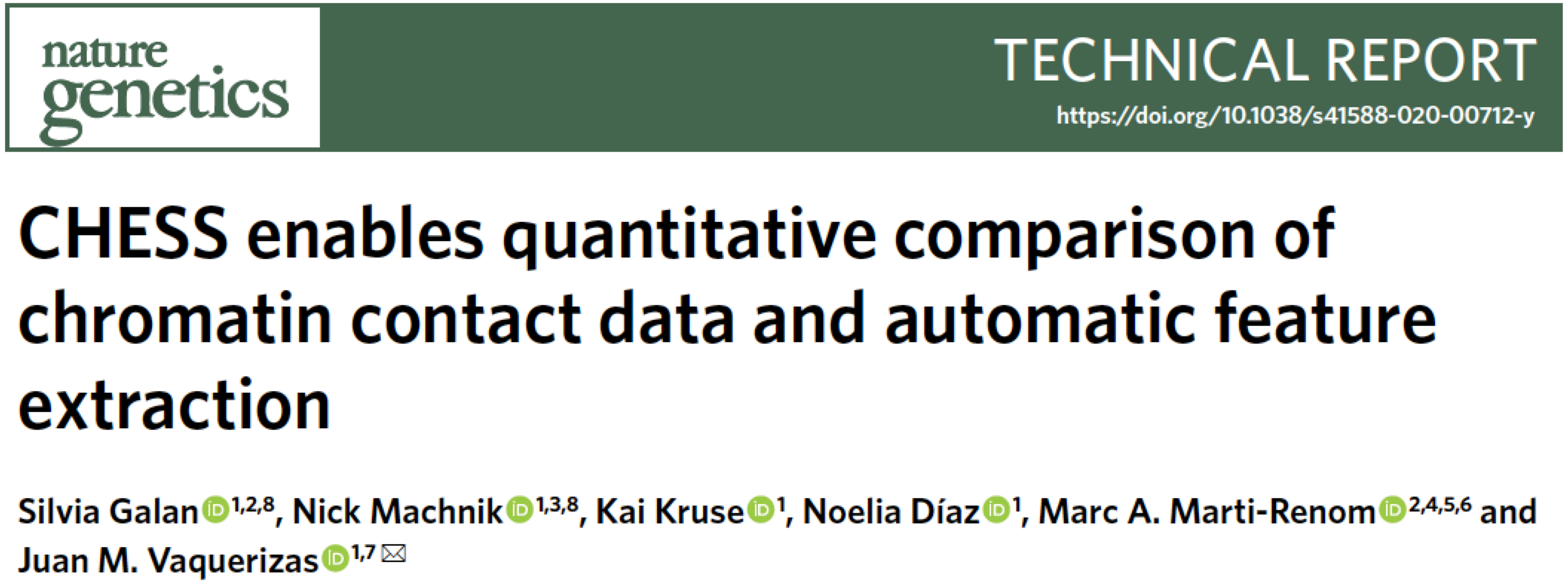


chess-hic

CHESS is a tool for the comparison and automatic feature extraction for chromatin contact data, developed in the Vaquerizas Lab.

If you use CHESS in your research, please cite the CHESS paper.

Please check out the online documentation for detailed installation and usage instructions.



nature genetics

TECHNICAL REPORT

https://doi.org/10.1038/s41588-020-00712-y

CHESS enables quantitative comparison of chromatin contact data and automatic feature extraction

Silvia Galan[1,2,8], Nick Machnik[1,3,8], Kai Kruse[1], Noelia Díaz[1], Marc A. Marti-Renom[2,4,5,6] and Juan M. Vaquerizas[1,7] ✉

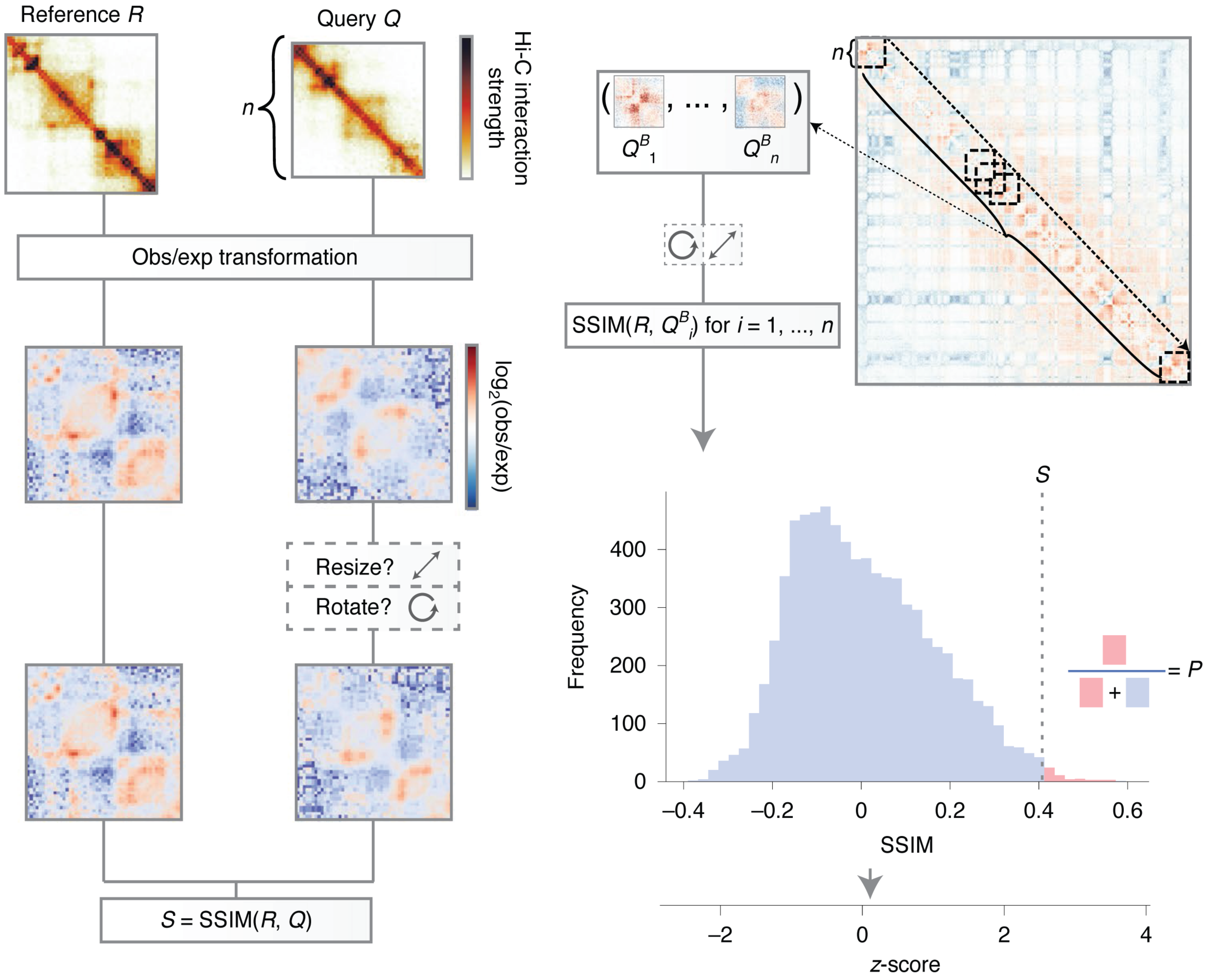
CHESS

- Systematic and quantitative comparison of 3C regions
- Identify and classify specific similarities or differences and features:
  - feature-free approach: not specifically looking for tads or loops.

Structural similarity index (SSIM)

- Widely used in image analysis
- SSIM score is a single value combining:
  - brightness: mean of the signal intensity
  - contrast: variance in signal
  - structure differences: correlation between signal values of two matrices

Identify regions that are significantly different between two conditions



Identify features that change between two conditions