Can we better understand HOT regions based on 3D genome organization?

Koon-Kiu Yan (Gerstein lab)
6/23/2016

Motivation

- HOT regions are heavily clustered with transcription factor binding sites. The high accessibility should be related to the 3D structure of genome

- HOT regions have been identified in worm, fly and human (e.g. Araya et al., Boyle et al. Nature 2014). There are various Hi-C data performed in human, worm and fly.

Hi-C data in worm and fly

LETTER

doi:10.1038/nature14450

Condensin-driven remodelling of X chromosome topology during dosage compensation

worm embryo

Emily Crane[1]†*, Qian Bian[1]*, Rachel Patton McCord[2]*, Bryan R. Lajoie[2]*, Bayly S. Wheeler[1], Edward J. Ralston[1], Satoru Uzawa[1], Job Dekker[2] & Barbara J. Meyer[1]

Cell

fly cell lines: s2, Kc167, DmBG3-c2, OSC

Molecular Cell
Resource

Kc167

Three-Dimensional Folding and Functional Organization Principles of the *Drosophila* Genome

Tom Sexton,[1,4] Eitan Yaffe,[2,4] Ephraim Kenigsberg,[2] Frédéric Bantignies,[1] Benjamin Leblanc,[1] Michael Hoichman,[2] Hugues Parrinello,[3] Amos Tanay,[2,*] and Giacomo Cavalli[1,*]
[1]Institut de Génétique Humaine, UPR 1142, CNRS, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, France
[2]Department of Computer Science and Applied Mathematics and Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel
[3]Montpellier GenomiX IBiSA, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, France

Gene Density, Transcription, and Insulators Contribute to the Partition of the *Drosophila* Genome into Physical Domains

Chunhui Hou,[1,3] Li Li,[2,3] Zhaohui S. Qin,[2,*] and Victor G. Corces[1,*]
[1]Department of Biology
[2]Department of Biostatistics and Bioinformatics
Emory University, Atlanta, GA 30322, USA

Research

Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains

Sergey V. Ulianov,[1,2,12] Ekaterina E. Khrameeva,[3,4,12] Alexey A. Gavrilov,[1] Ilya M. Flyamer,[1,2] Pavel Kos,[5] Elena A. Mikhaleva,[6] Aleksey A. Penin,[4,7,8] Maria D. Logacheva,[4,8,9] Maxim V. Imakaev,[10] Alexander Chertovich,[5] Mikhail S. Gelfand,[4,11] Yuri Y. Shevelyov,[6] and Sergey V. Razin[1,2]

[1]Institute of Gene Biology, RAS, 119334 Moscow, Russia; [2]Department of Molecular Biology, Lomonosov Moscow State University, 119991 Moscow, Russia; [3]Skolkovo Institute of Science and Technology, 143026 Skolkovo, Russia; [4]Institute for Information Transmission Problems (Kharkevich Institute), RAS, 127051 Moscow, Russia; [5]Physics Department, Lomonosov Moscow State

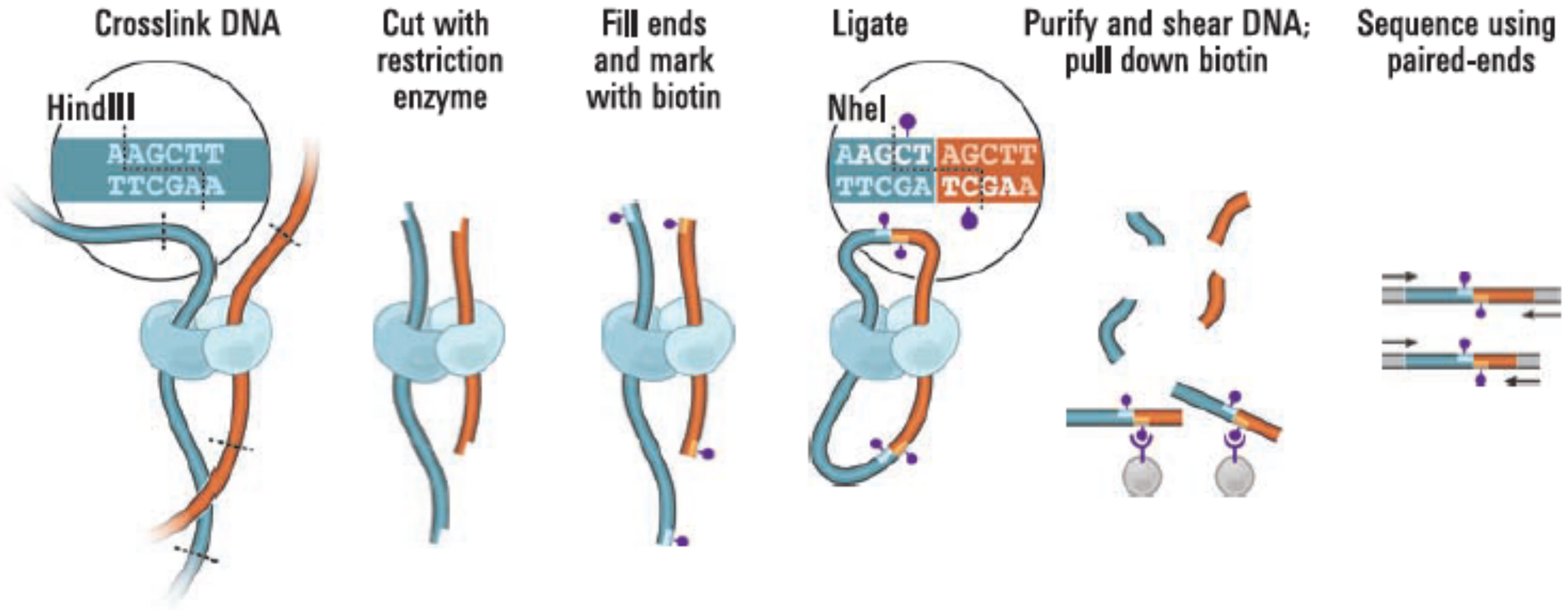
fly embryo

Chromosome conformation capture (3C) and Hi-C

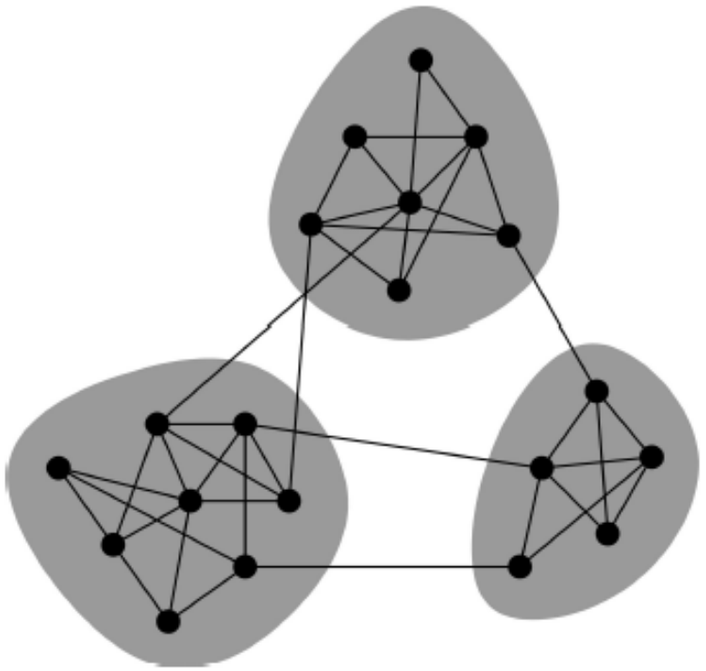
Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome

Erez Lieberman-Aiden,[1,2,3,4*] Nynke L. van Berkum,[5*] Louise Williams,[1] Maxim Imakaev,[2] Tobias Ragoczy,[6,7] Agnes Telling,[6,7] Ido Amit,[1] Bryan R. Lajoie,[5] Peter J. Sabo,[8] Michael O. Dorschner,[8] Richard Sandstrom,[8] Bradley Bernstein,[1,9] M. A. Bender,[10] Mark Groudine,[6,7] Andreas Gnirke,[1] John Stamatoyannopoulos,[8] Leonid A. Mirny,[2,11] Eric S. Lander,[1,12,13]† Job Dekker[5]†

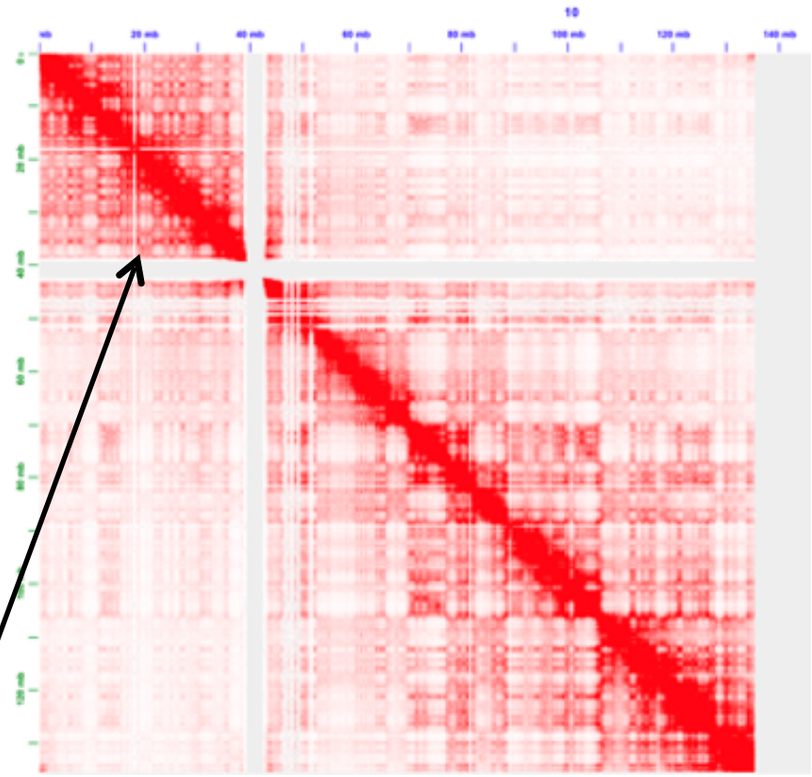
SCIENCE VOL 326 9 OCTOBER 2009



A network-based approach to find Topologically Associating Domains (TADs)



network	contact map
node	chromosome bin
edge	Hi-C contact
# of connections	coverage
module	domain



DNA picture adapted from Weinreb et al. Bioinformatics 2015

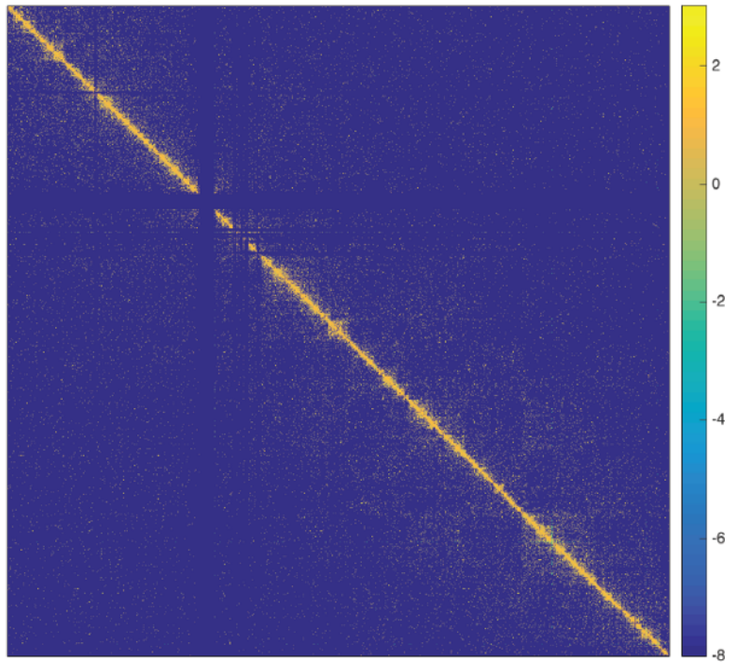
Modularity maximization

$$Q = \frac{1}{2m} \sum_{i,j} \left(W_{ij} - \frac{k_i k_j}{2m} \right) \delta_{\sigma_i \sigma_j}$$

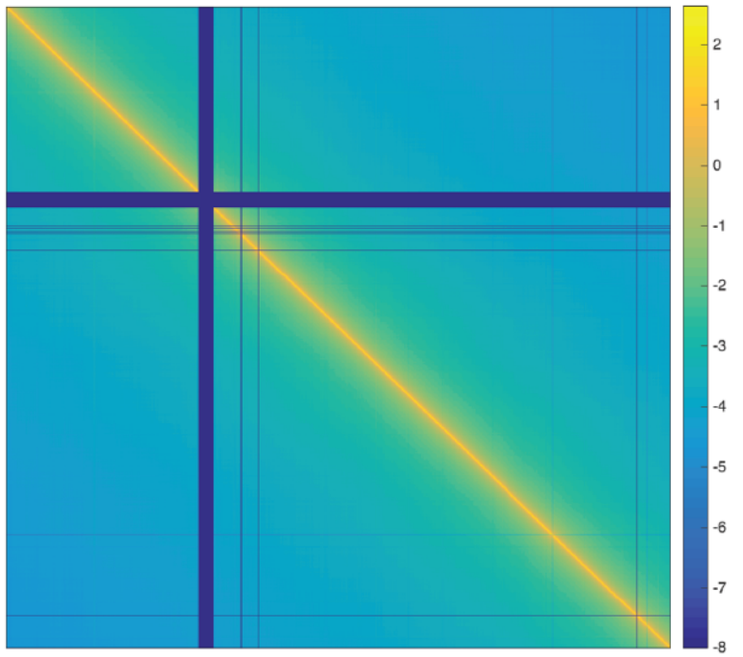
TADs have apparent hierarchical organization

Workflow

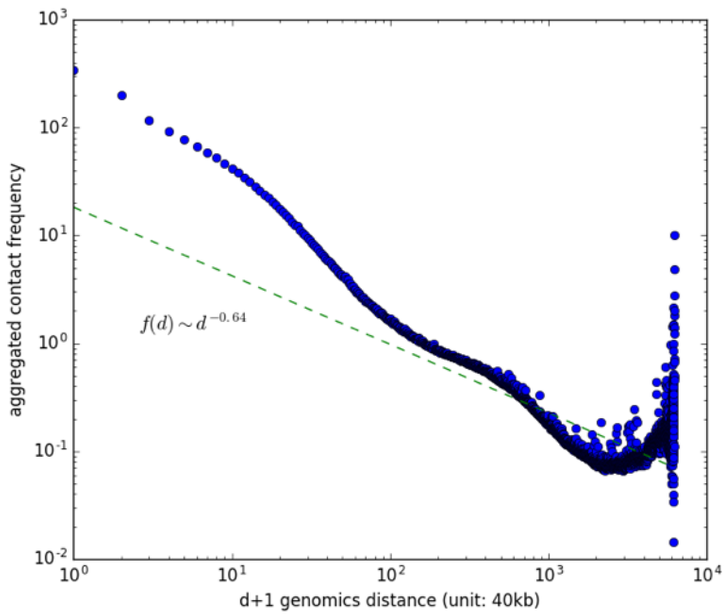
input: contact map W



null model E



$$E_{ij} = c_i^* c_j^* f(|i-j|)$$



Choose a particular resolution γ
Optimize Q over all possible partitions

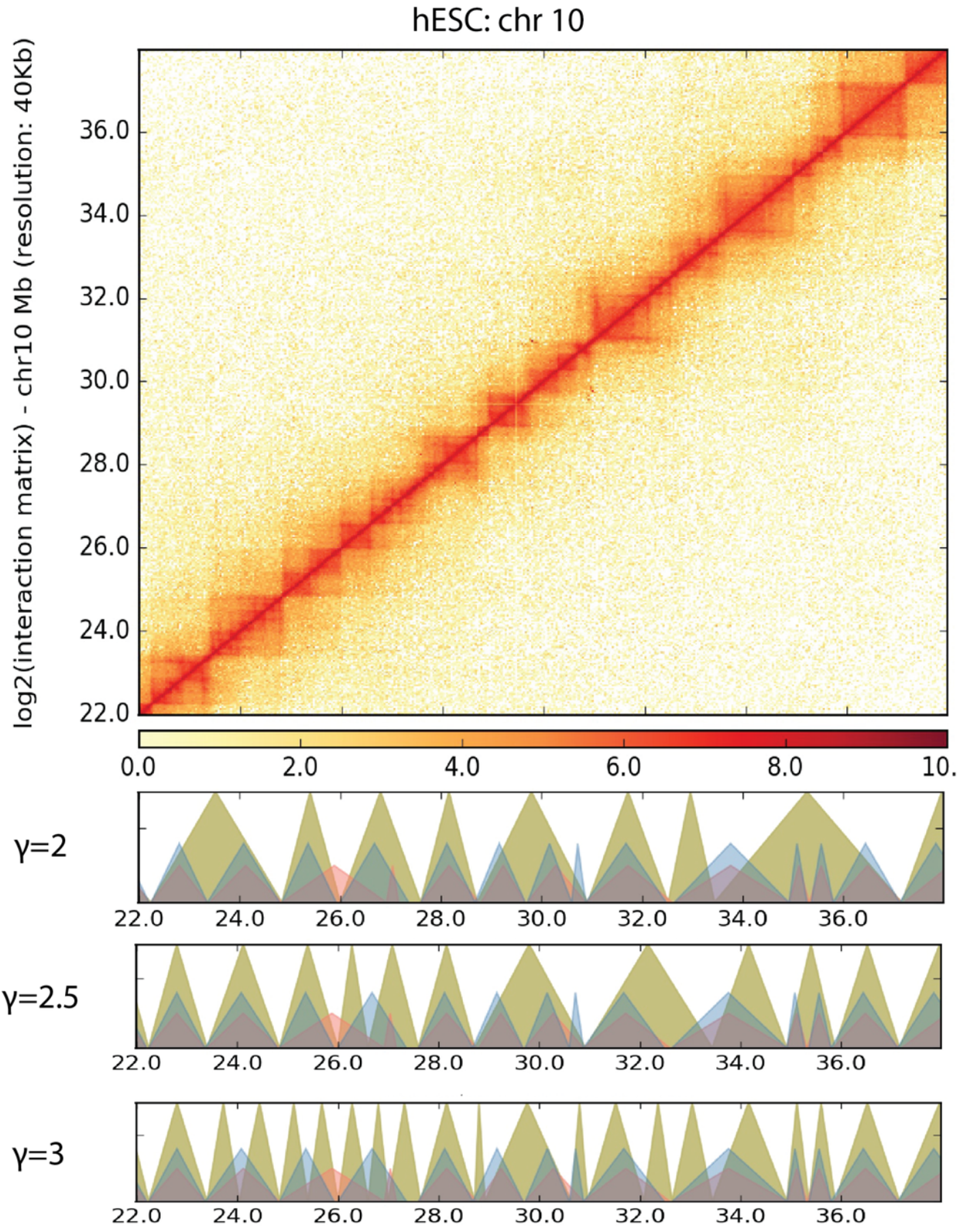
$$Q = \frac{1}{2N} \sum_{ij} (W_{ij} - \gamma E_{ij}) \delta_{\sigma_i \sigma_j}$$

consensus boundaries based on the boundary scores

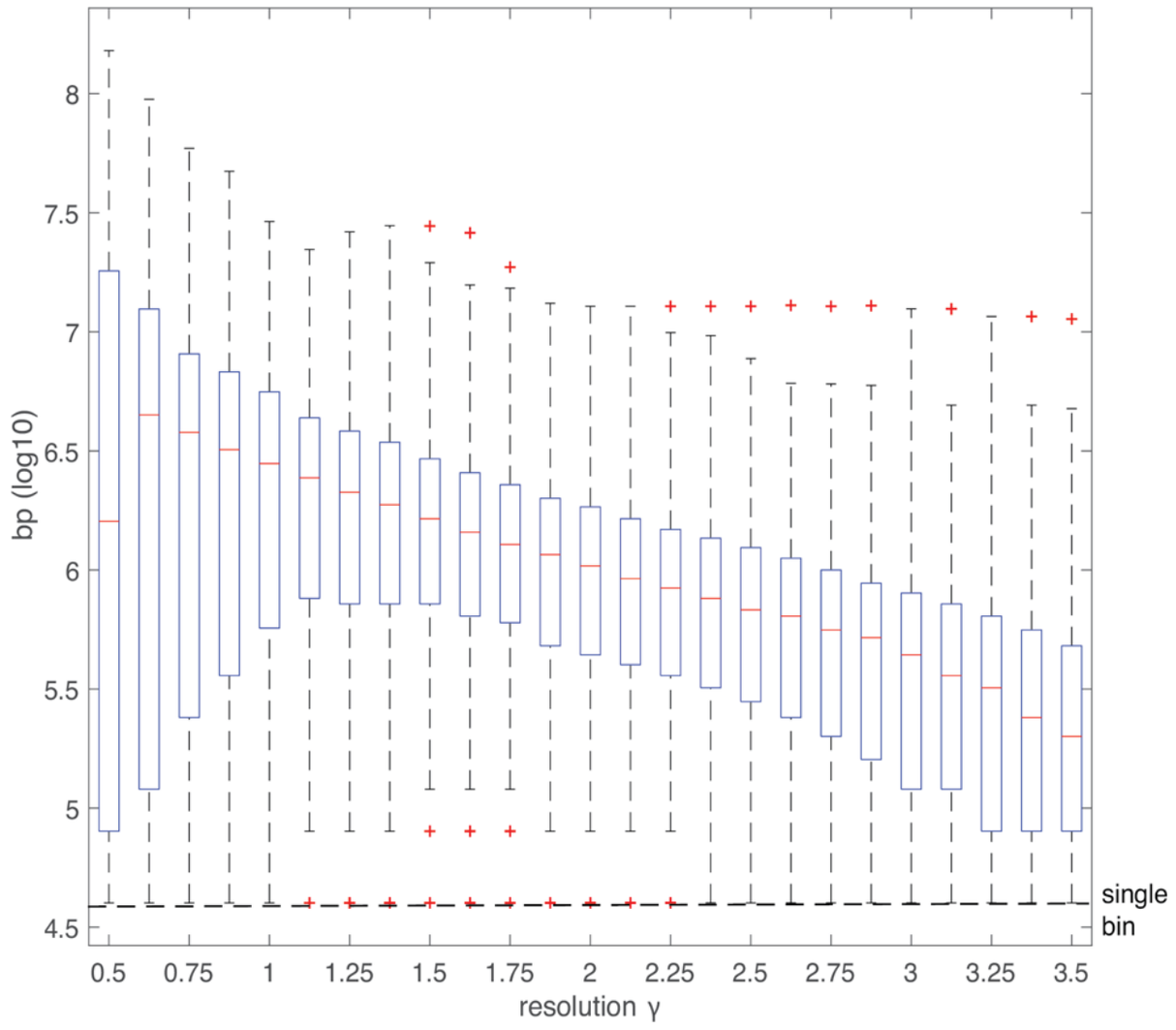
consensus domains

multiple trials to define boundary scores for all adjacent bins

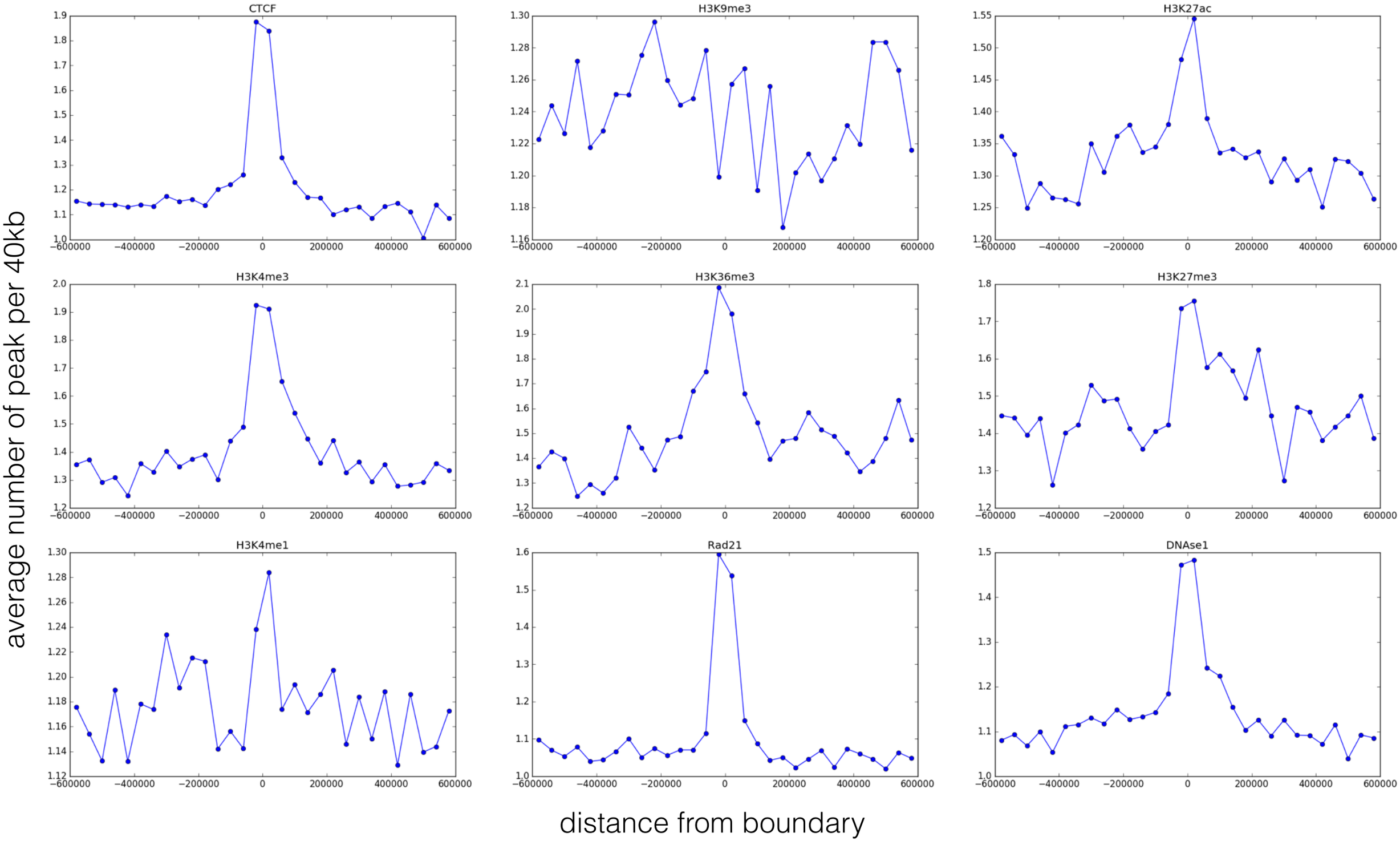
TADs in different resolutions



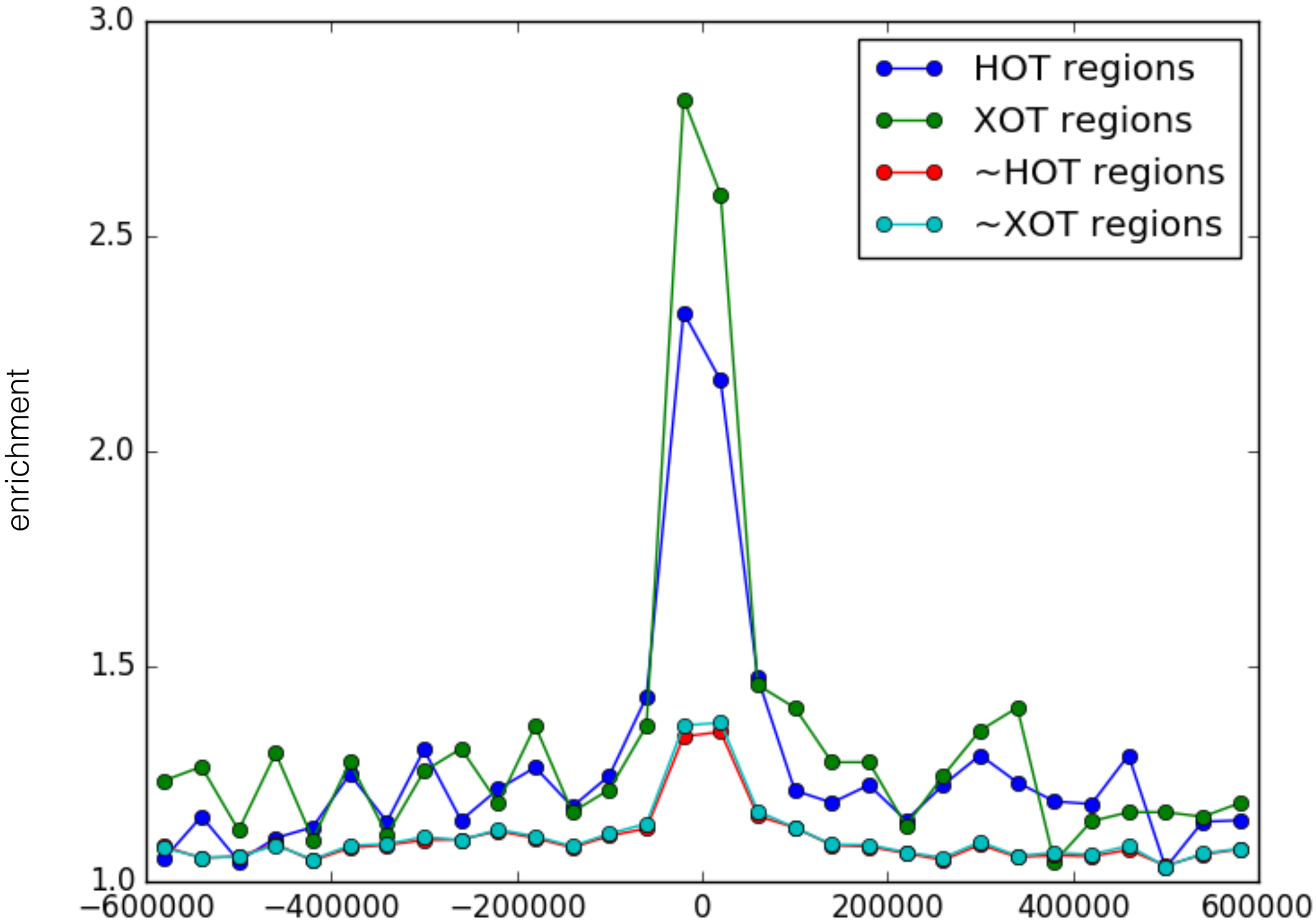
smaller TADs but are detected
as the resolution increases



Enrichment of chromatin marks near TAD boundaries

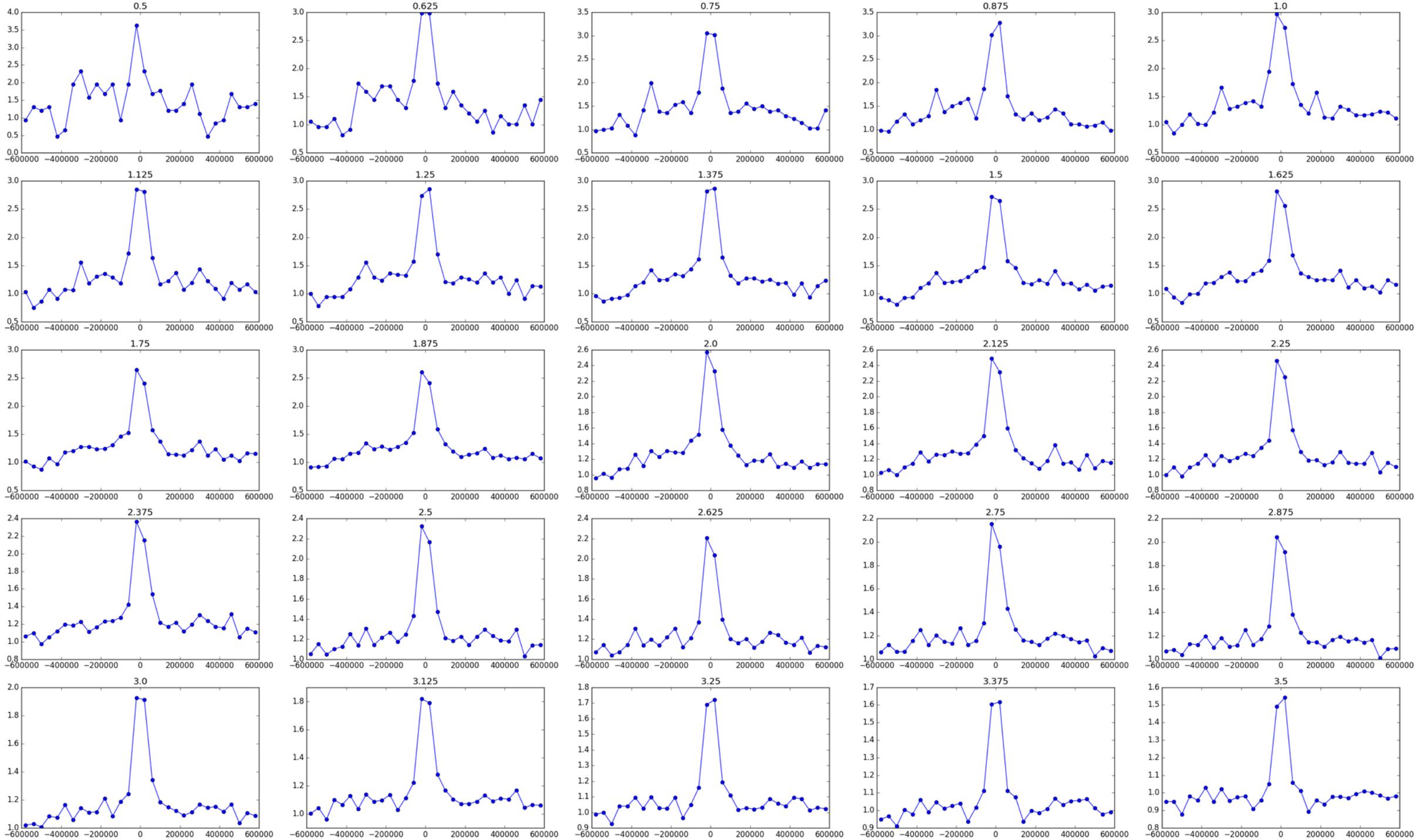


Enrichment of HOT regions near TAD boundaries

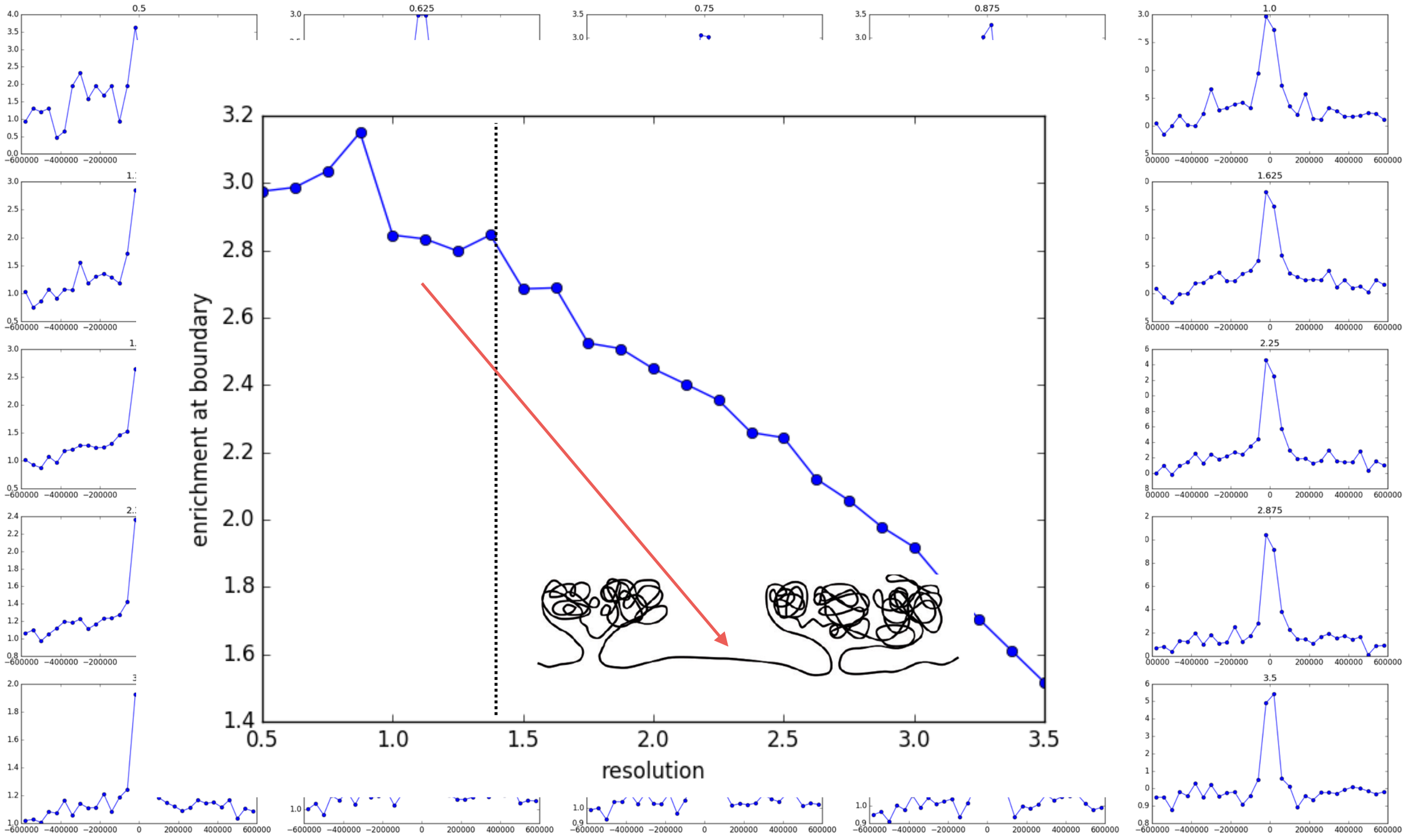


HOT regions:
http://www.stanford.edu/~claraya/metrn/data/hot/

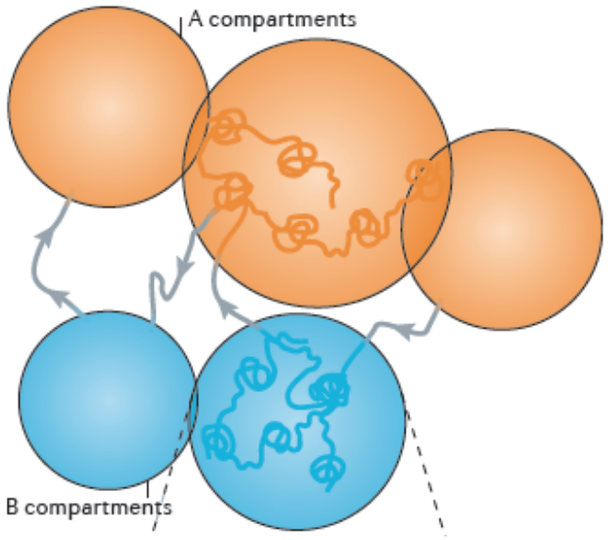
HOT regions in different resolutions



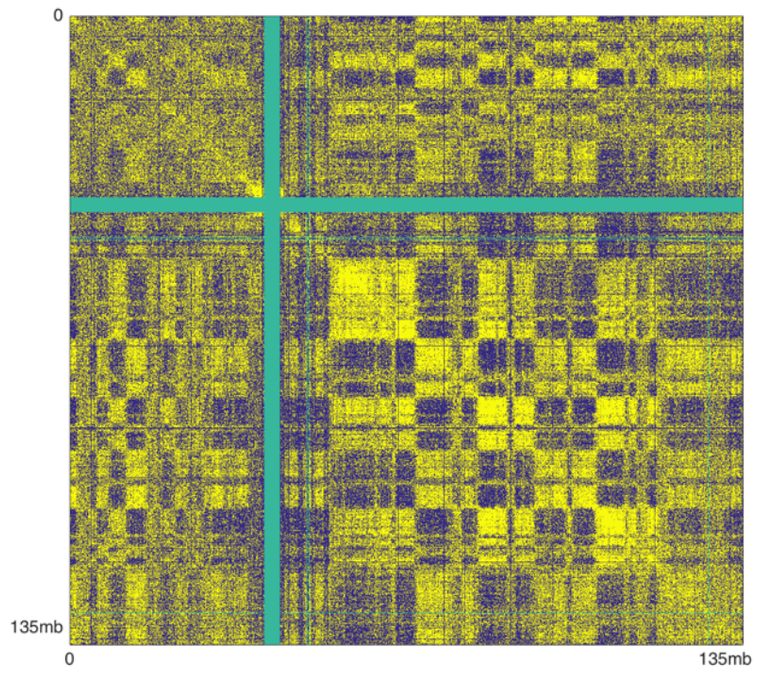
HOT regions in different resolutions

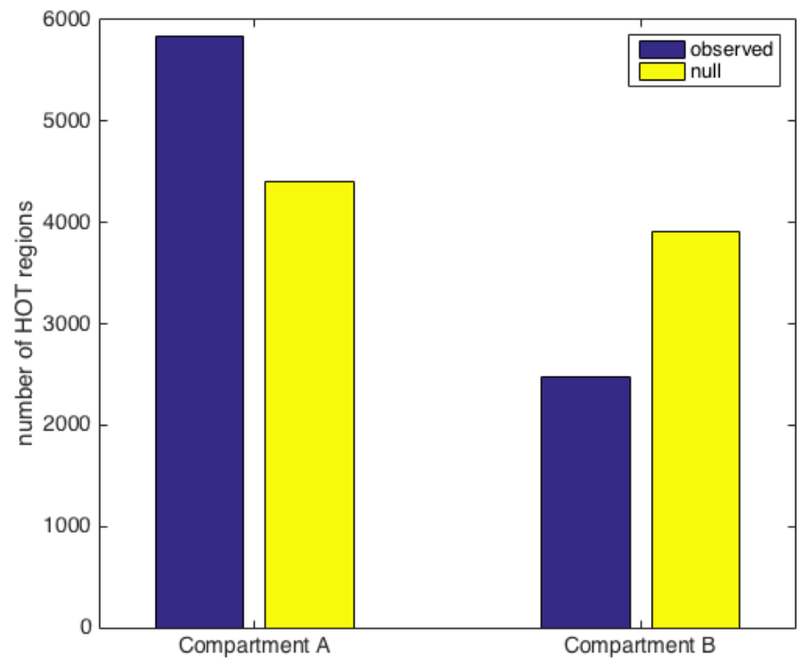


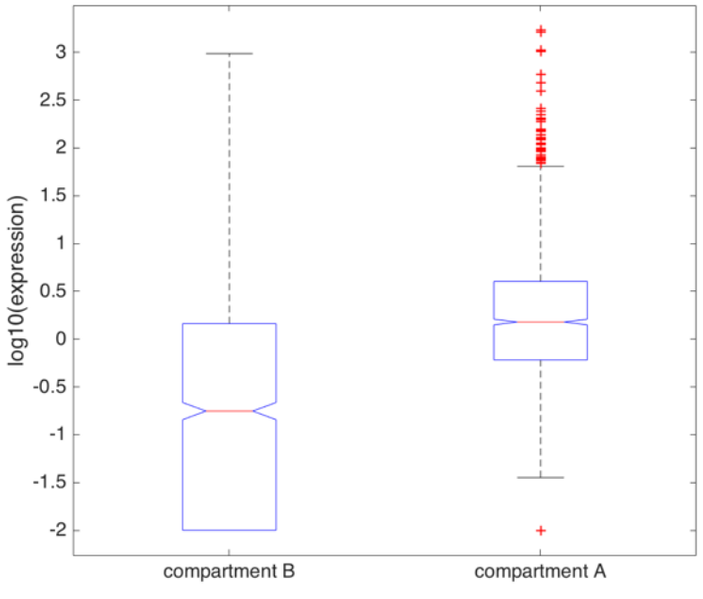
HOT regions are enriched in the Compartment A

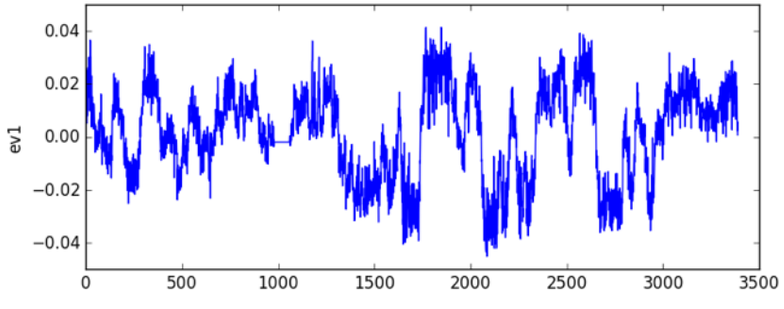


$$C_{ij} = cor(W_{ij}/E_{ij})$$









Summary and Possible threads

- Based on hES cells, the location of HOT regions is related to 3D genome organization.
- Possible threads to follow:
  - Identify TADs in worm and fly; we expect similar observations.
  - Make further use of our ChIP-Seq data:
    - architectural proteins for domain formation:
      - CTCF, YY1, Rad21 in human
      - fly: Zw5, dCTCF, Su(Hw)… worm?
    - use the binding of specific TFs to predict domains/ boundaries formation