Random loops model can explain the appearance of Topologically Associating Domains (TADs)

Ofir Shukron

September 15, 2014

Introduction

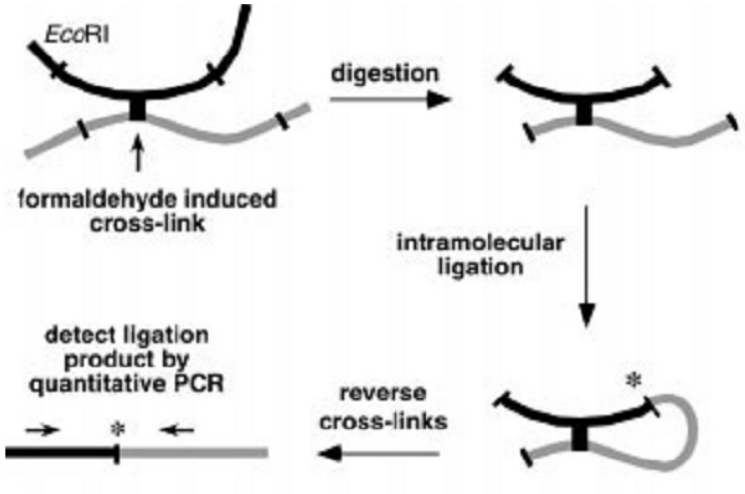
1. The spatio-temporal organization of the chromatin has significant implication of cellular activities, such as gene expression and regulation.
2. however, the spatial organization of the chromatin is not entirely known.
3. DNA looping has shown to be a mechanism for long range gene regulation.
4. here we show that using random polymer looping model and fitting it to the experimental chromatin looping data we can explain the appearance of conserved structures in chromatin called TADs.

Agenda

Chromosome Conformation Capture Experiments

A set of methods to simultaneously record millions of looping events occurring within the genome (specific or unspecific).
The general steps are:

1. intact nuclei are extracted from millions of cells
2. Formaldehyde induces protein-DNA and protein-protein cross-links
3. restriction enzymes digest the cross-linked DNA
4. cross-linked DNA is purified, diluted and ligated
5. cross-links are reversed
6. PCR to amplify ligation junctions
7. histogram of segment encounter is produced



The experimental data

- Two replicate of the CC experiments were conducted by Nora et. al 2012.
- we focus on a 920,432 bp subset of the data, around the X inactivation center of the X chromosome in mouse embryonic stem cells.
- the region harbors the Xist enhancer and Tisx promoter.
- we have the segments' encounter frequencies from the two experimental replicates.

Topologically Associating Domains (TADs)

Conserved structures of chromosome interactions on the Mb scale, with higher inter than intra-segment interactions It is believed that the TAD forms a 'regulatory unit' for regulating gene expression, as can be seen by the correlation of gene expression located on the same TAD

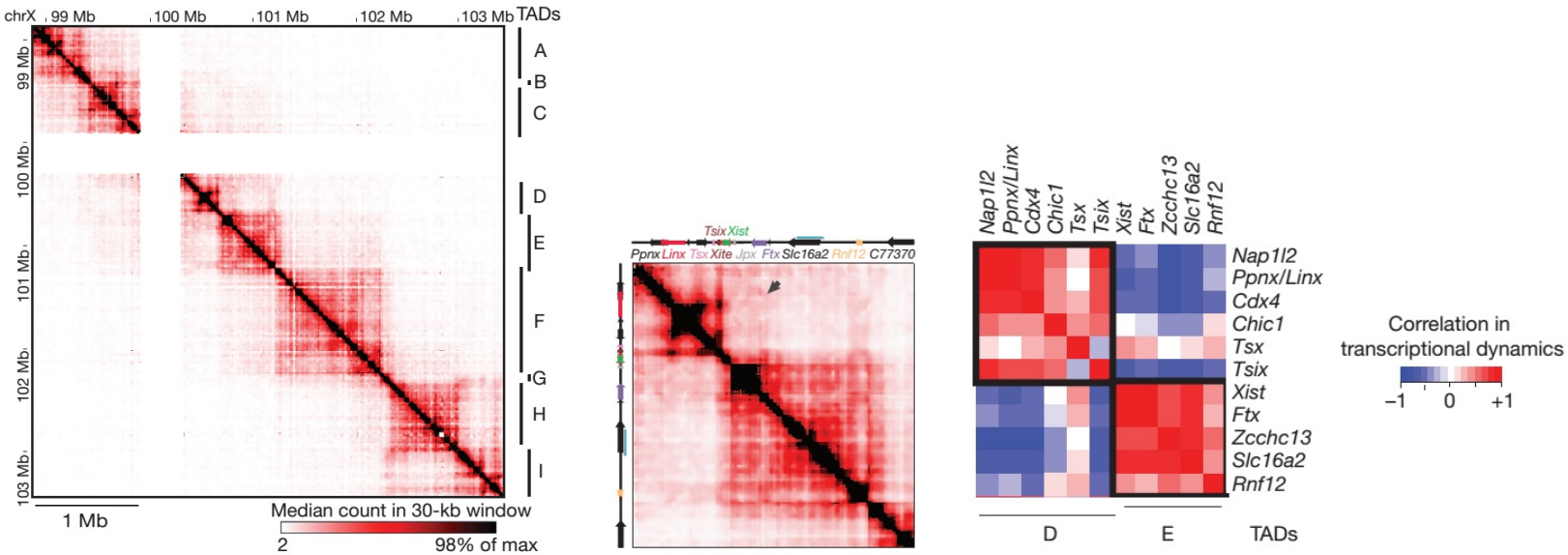
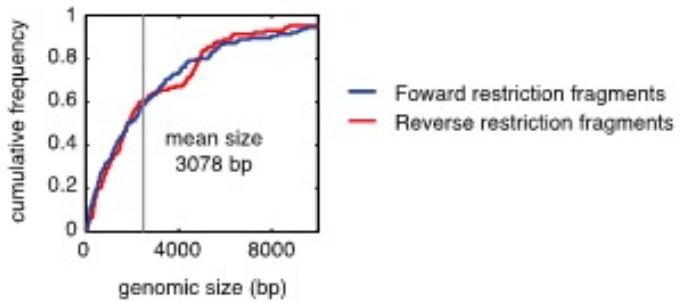


Figure: The 4.5 Mb region (left), enlargement of TAD D and E (center). Displayed median count in a 30kb window every 6kb, gene expression correlation (right)

From restriction segments to beads

- To coarse-grain the data, we choose a bead-size of 3000 bp, corresponding to the mean segment length resulted from the digestion of EcoRIII enzyme.



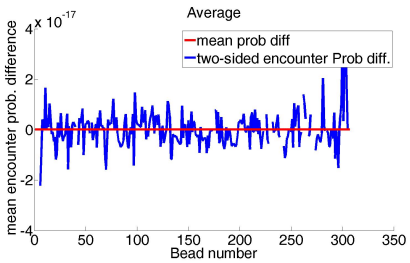
- the genomic section was evenly partitioned by 3000 bp beads. Each segment receives a start and end index according to the beads it covers.
- for example,

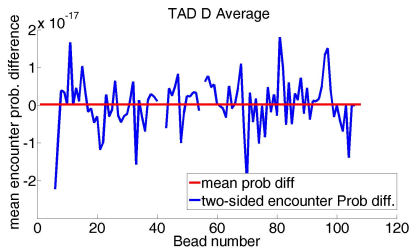
bp range	start ind	end ind
500-3500	1	2
4000-4500	2	2
5000-12001	2	4

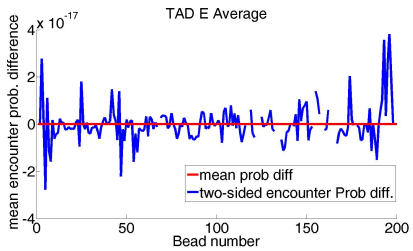
Bead encounter frequecy

TAD D and E

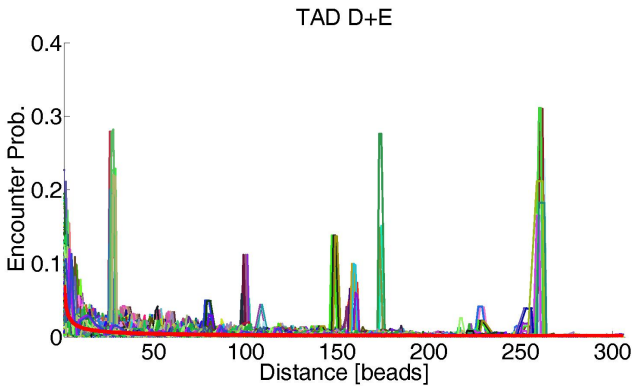
- We work with the average of the two experimental replicates
- a total length of 920,432 bp - resulting in 307 beads (TAD D 107 beads, TAD E 200 beads)
- We calculate the 'one-sided' encounter probability vs. distance (bead units) for each bead
- the mean encounter probability difference, shows that the data is left-right symmetric

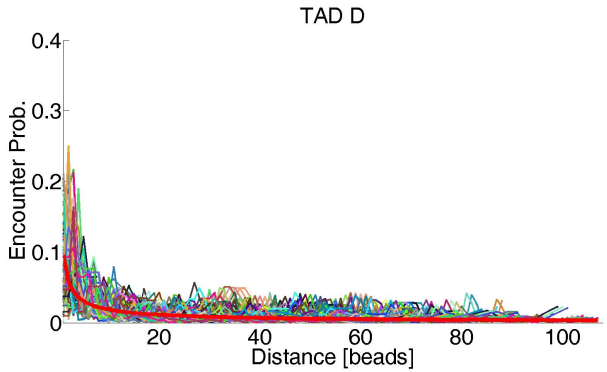


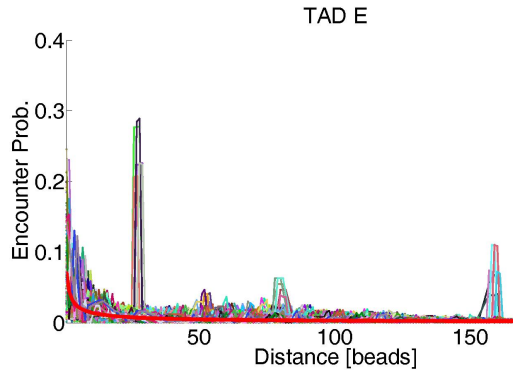




- TAD E has several strong specific interactions. TAD D has almost no specific interactions. Strong inter-TAD specific interactions







Peaks of the encounter data

- About half of the peaks in the encounter data result from specific interactions between TADs
- The other half comes from specific internal interactions of TAD E.
- To get an impression, a manual marking of the peaks shows

Bead numbers	Encountered beads	TAD
23-26	280-290	$D \leftrightarrow E$
49-53	148-155	$D \leftrightarrow E$
56-59	80-90	$D \leftrightarrow D$
115-117	165-170	$E \leftrightarrow E$
161-162	187 190	$E \leftrightarrow E$
182-184	260-264	$E \leftrightarrow E$
185-186	253-255	$E \leftrightarrow E$
234-236	184-189	$E \leftrightarrow E$
234-236	4-11	$E \leftrightarrow D$
243	88	$E \leftrightarrow D$
264	89-90	$E \leftrightarrow D$
274-277	113-120	$E \leftrightarrow E$

The enconter probability

For the case of TAD D, TAD E, and the two together, we estimate the bead encounter probability, p, and fit it with a function of the form

$$p_n(d) = \alpha d^{-\beta}$$

where, d is the distance in bead units, $\alpha = \frac{1}{\sum_{j=1}^{d_{max}} j^{-\beta}}$ and β is a parameter to be estimated.

Theoretical model

The Rouse model

We start with the classical and most simple model, the Rouse chain.

- A Rouse chain describes polymer dynamics as a stochastic motion of a collection of microscopic "beads" connected by harmonic springs
- the 3D motion of bead n in the chain of N beads

$$\frac{dR_n}{dt} = -\frac{3D}{b^2}(2R_n(t) - R_{n+1}(t) - R_{n-1}(t)) + f_n(t)$$

- R_n- the position of bead n
  b- the standard deviation of the distance between adjacent beads
  D- the diffusion constant
  f_n- white Gaussian noise
- From the theory, $Pr(\|R_n - R_m\| < \epsilon) \sim |n-m|^{-1.5}$

Simulation with simple rouse chain

- we first check whether a simple model can produce the TADs.
- we examine the results of simulating chain of 64 beads

Loops corresponding to the peaks of the encounter data

Dynamic Loop Model

- some beads in the same TAD have affinity toward one another
- affine beads located within a distance less than ϵ (the encounter distance) are connected
- the rate of disconnection between beads is k_{off}

Dynamic loops model with beads affinity

Enlarging the encounter distance

3C experiment with stiff connectors

Next, we simulate 64 bead chains with stiff connectors.
Stiff connectors are Rouse spring that stay fixed.

Future perspective

- A model with variable encounter distance
-