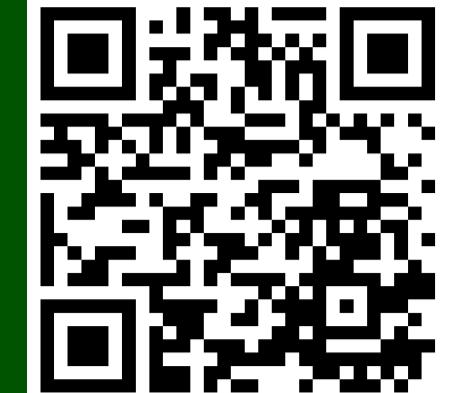


Tharvesh Moideen Liyakat Ali, Jonas Paulsen and Philippe Collas

Department of Molecular Medicine, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway

Contact: jonas.paulsen@medisin.uio.no, tmali@medisin.uio.no

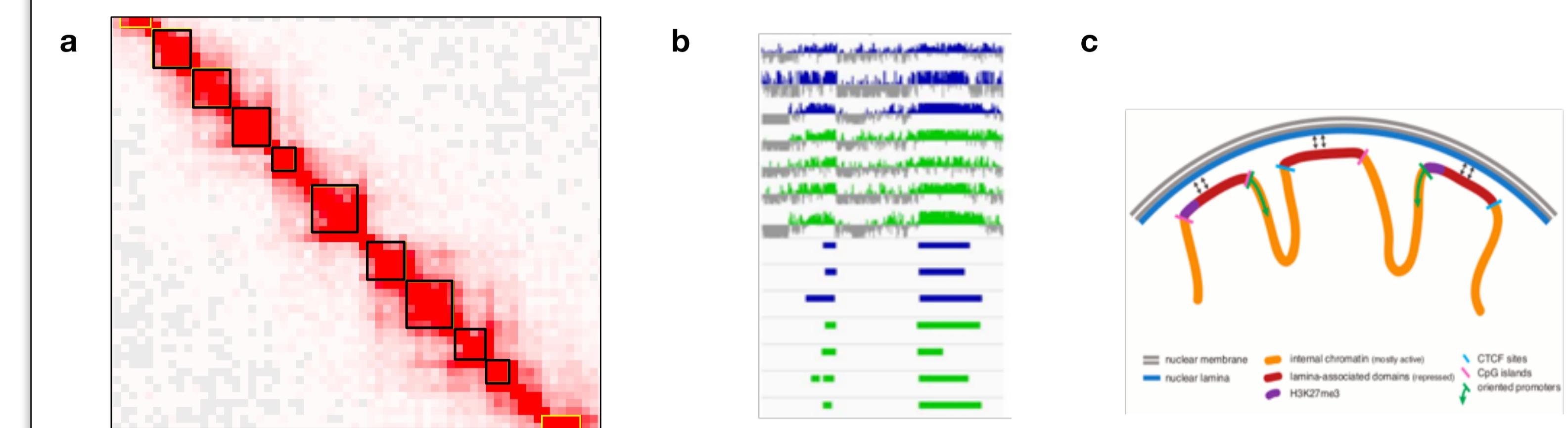


Summary:

- Chrom3D is a computational modelling platform specifically designed to model genome 3D structure
- We have developed a pipeline to combine Hi-C and lamin ChIP-Seq datasets to model TAD placement in the 3D nucleus
- In our protocol, we use the HiC contact matrix to define Topologically Associating Domains (TADs), and compute significant TAD-TAD interactions using a statistical test
- Lamin ChIP-Seq data is used to define Lamin Associating Domains (LADs)
- The resulting 3D models can be used for placing genomic data in 3D space, as well as for generating statistical distributions of radial and pairwise distances between genomic loci

Defining Topological Associating Domains (TADs) and Lamin Associating Domains (LADs):

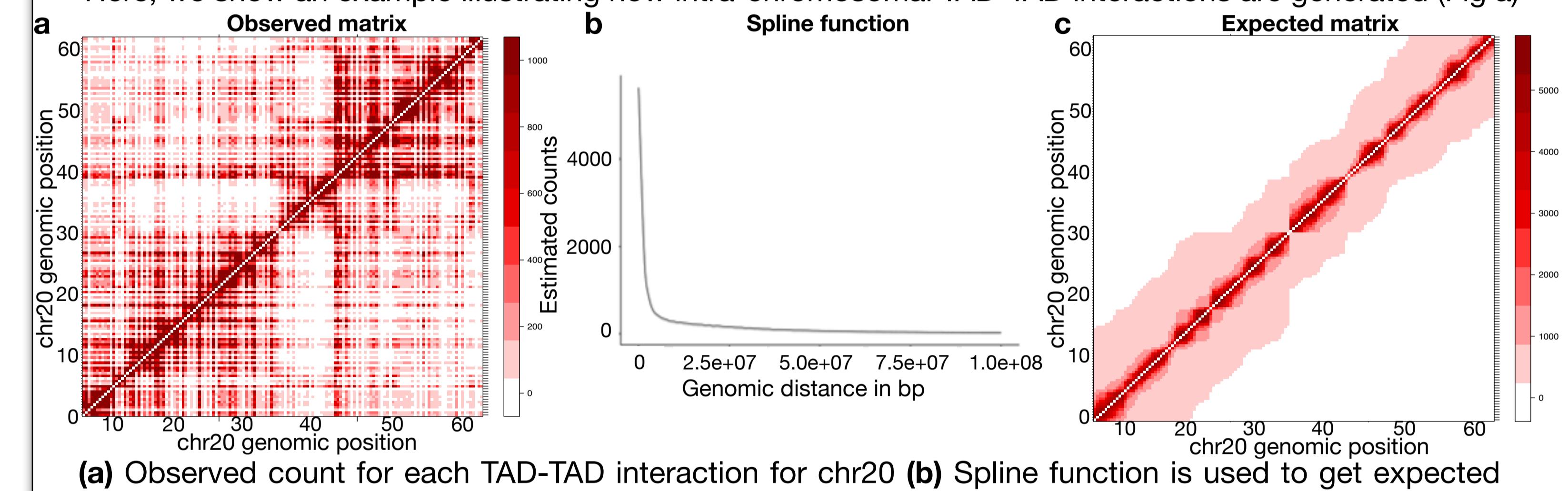
- In this example we used TADs defined using the Arrowhead algorithm (Rao et al) with IMR90 cell line (GSE63525) (Fig a). TAD callers such as TopDom or Armatus can also be used to define TADs in this pipeline
- We used Enriched Domain Detector (EDD) algorithm to define LADs using lamin ChIP-Seq data (Fig b). These domains are associated with lamins at the nuclear periphery as illustrated (Fig c).



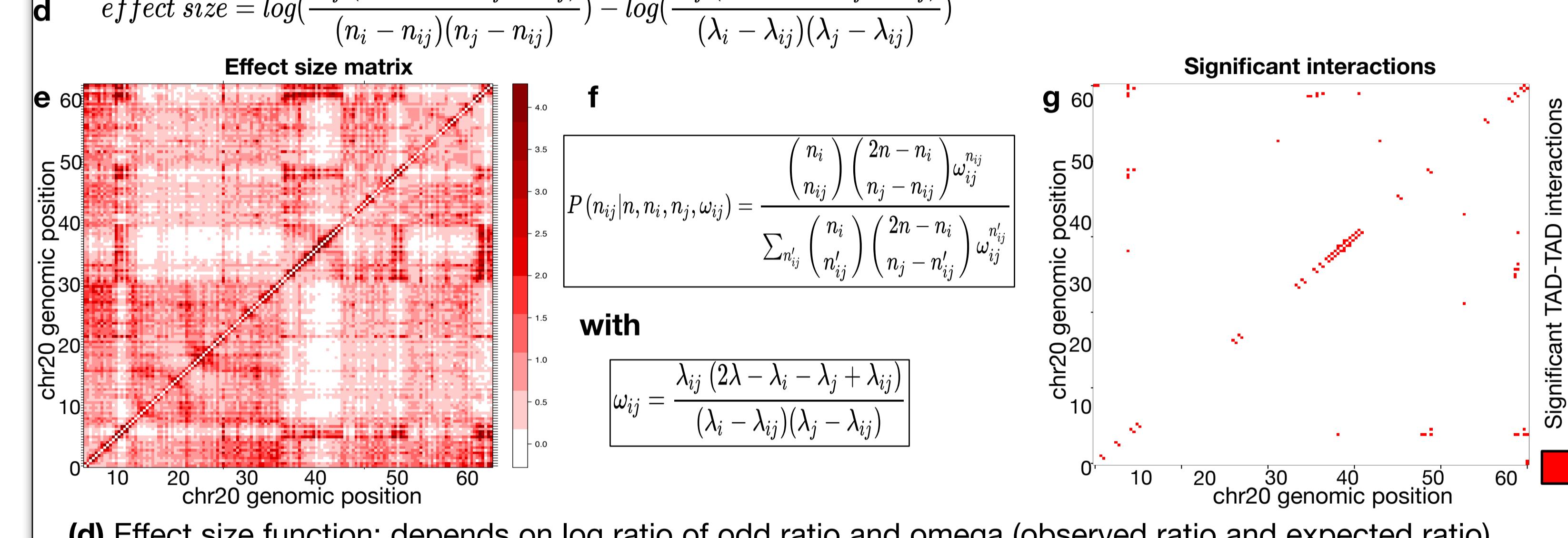
(a) Juicebox view of an example HiC matrix; black squares represent TADs defined using the Arrowhead algorithm (b) IGV snapshot of peaks and LADs defined using the EDD algorithm (c) Illustration of domains interacting with lamin proteins at the nuclear periphery

Defining TAD-TAD interactions:

- TAD-TAD interactions were defined using the Non Central Hypergeometric Test (**NCHG**) and significant interactions were selected based on *p*-value and effect size
- Here, we show an example illustrating how intra-chromosomal TAD-TAD interactions are generated (Fig a)

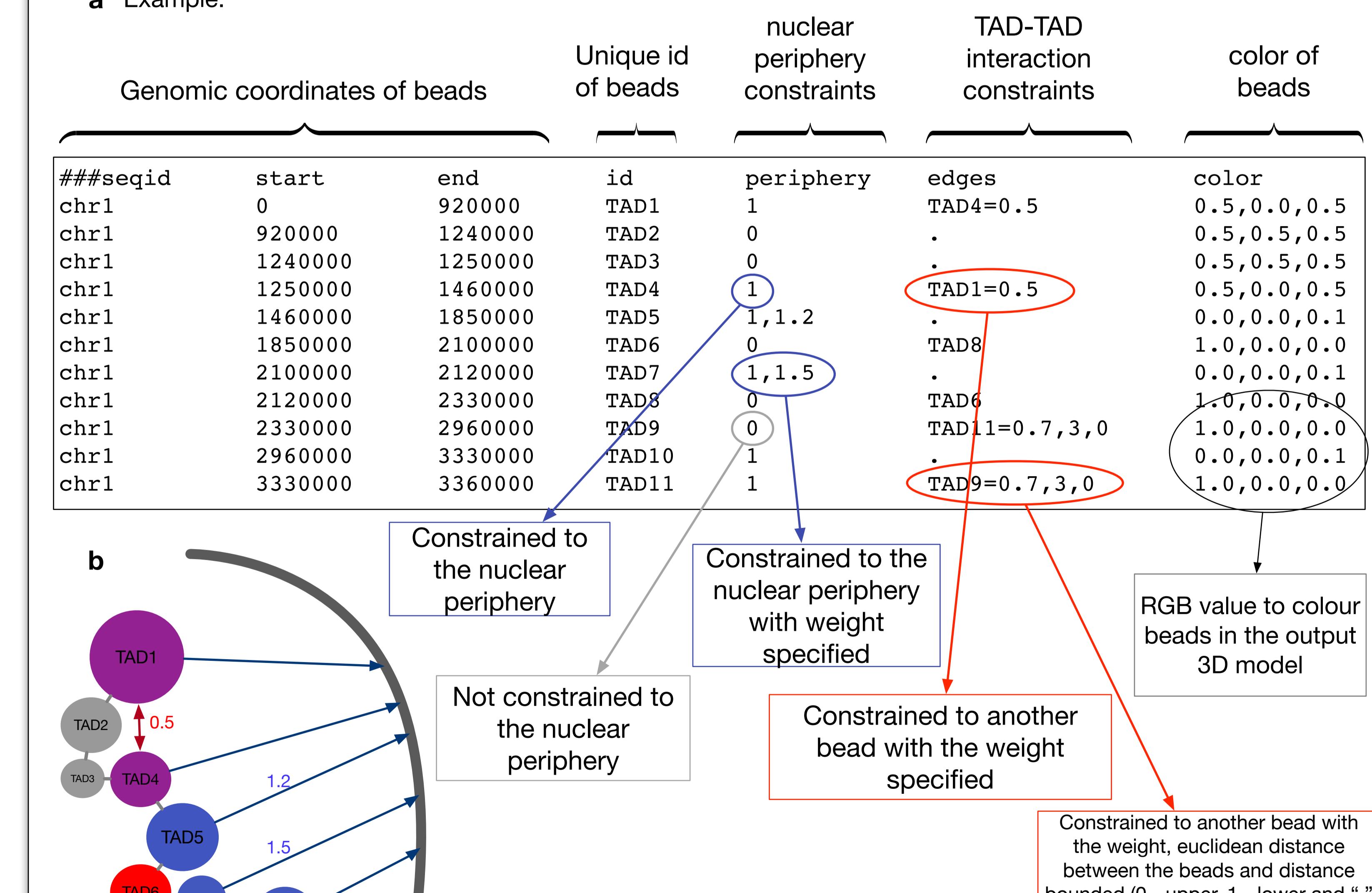


(a) Observed count for each TAD-TAD interaction for chr20 (b) Spline function is used to get expected counts for each genomic distance to get (c) Expected TAD-TAD interaction counts



(d) Effect size function; depends on log ratio of odd ratio and omega (observed ratio and expected ratio) to get (e) Effect size matrix of chr20. Then, (f) Non central hypergeometric probability function is used to define (g) significant TAD-TAD interactions based on *p*-value and effect size

Model Setup File used as input in Chrom3D (GTrack):

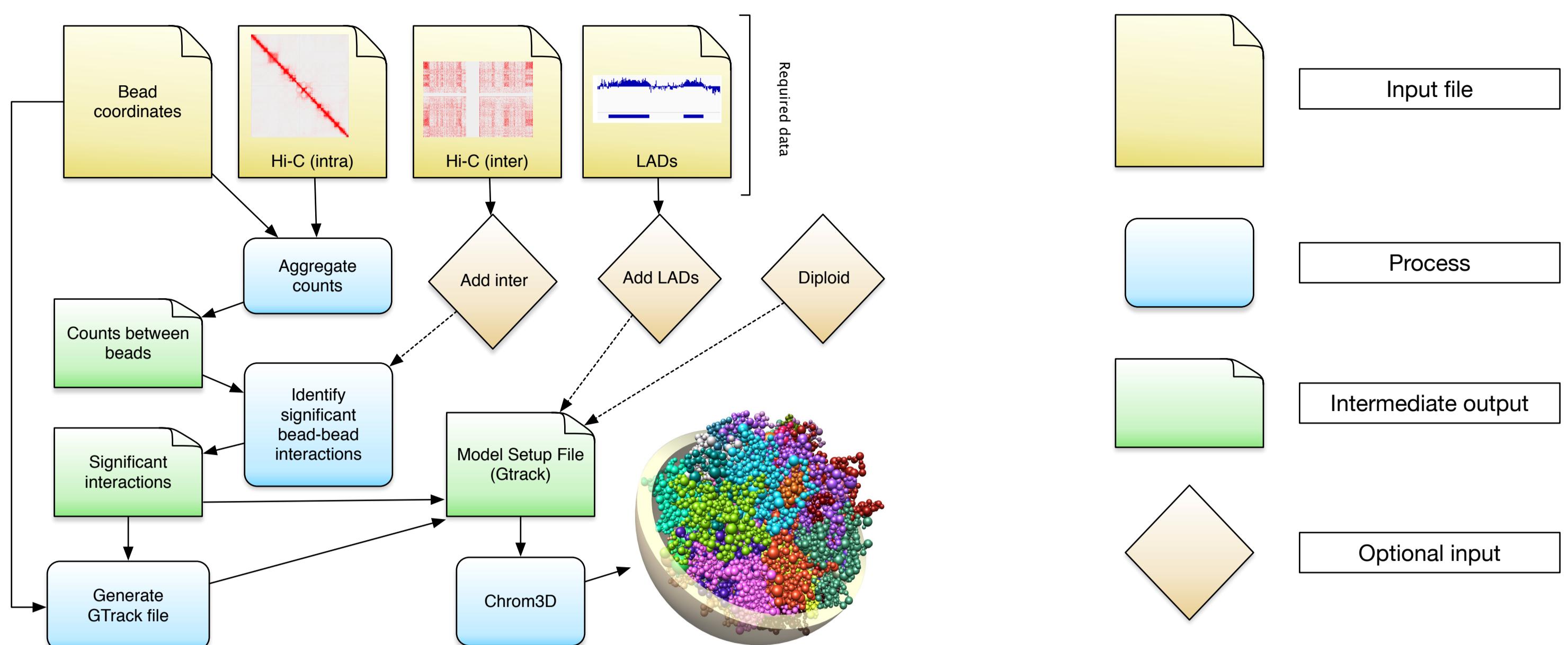


(a) Example model setup file (GTrack) (b) Graphical representation of an example GTrack file

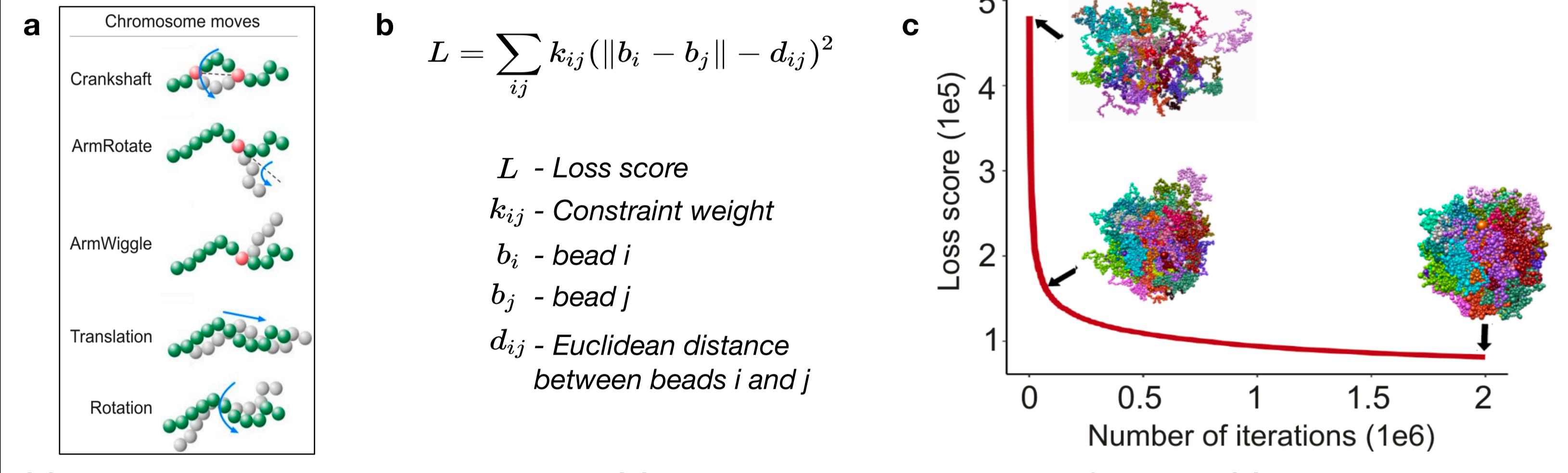
Reference:

- Paulsen et al. (2017). Chrom3D: three-dimensional genome modeling from Hi-C and nuclear lamin-genome contacts. *Genome Biology*
 Jonas Paulsen, TM Liyakat Ali, Philippe Collas. (in revision). Computational 3D genome modeling using Chrom3D
 Lund et al. (2014). Enriched Domain Detector: a program for detection of wide genomic enrichment domains robust against local variations. *Nucleic Acids Research*
 Rao et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*
 Durand et al. (2016). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell System*

Overview of the pipeline:



Monte Carlo Optimisation:



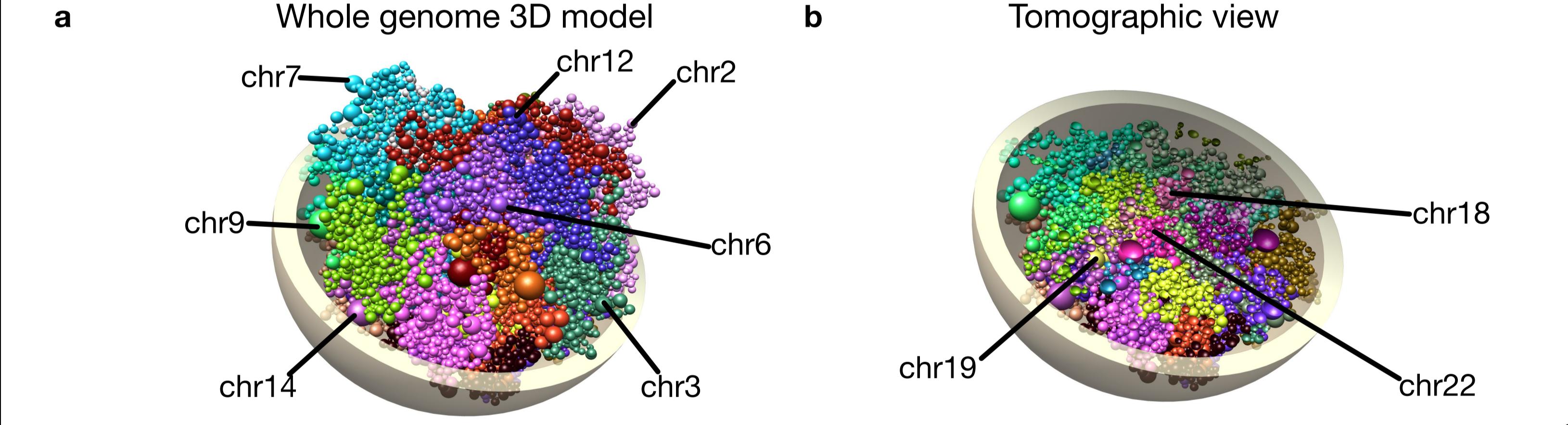
(a) Pre-defined chromosome moves and (b) Loss score function used in Chrom3D (c) Representation of minimisation of Loss score using Monte Carlo optimisation to get one of the stable whole genome 3D conformations

Simulation using Chrom3D:

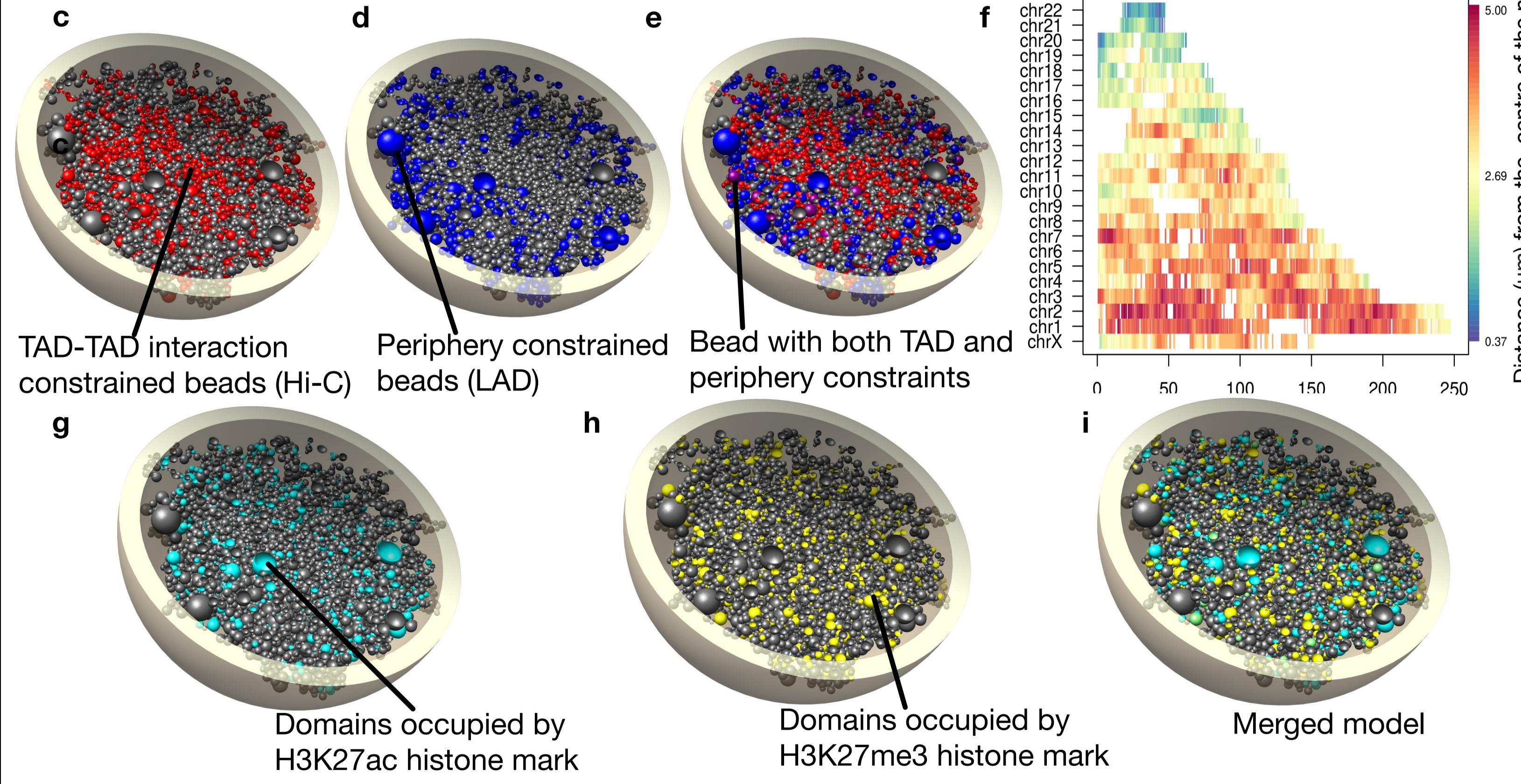
```
Chrom3D -y 0.15 -r 5.0 -l 10000 -m example_run -s 314 -n 2000000 -nucleus -o output.cmm HiC_constrained_LAD_constrained_input.gtrack
```

-y <float> - Scale total volume of the model beads relative to the volume of the nucleus. Specified as a value between 0 and 1 (default: no scaling)
 -r <float> - Radius of the nucleus
 -l <uint> - Write log to stderr
 -m <string> - Name of a model for the given run
 -s <uint> - Seed used for randomisation
 -n <uint> - Number of iterations
 -nucleus - Add constraints such that all beads are pushed towards the inside of the nucleus
 -o <string> - Output filename

Output:



Visualisation using a single 3D genome model:



Statistics using ensemble of 3D genome models:

