

# Hierarchical chromatin domains from Hi-C data

Bartek Wilczyński

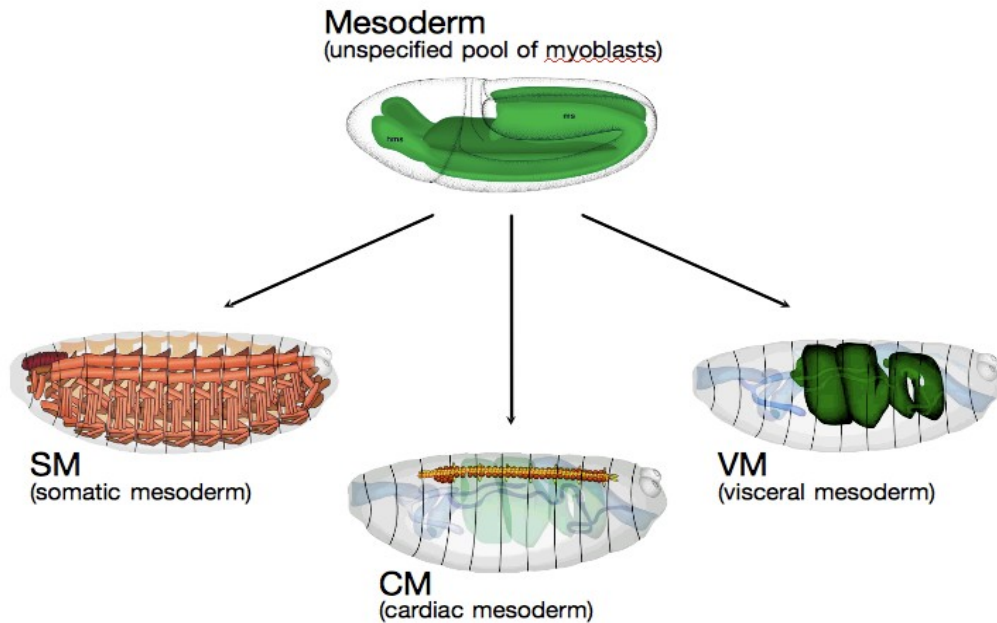
[bartek@mimuw.edu.pl](mailto:bartek@mimuw.edu.pl)

<http://regulomics.mimuw.edu.pl/>

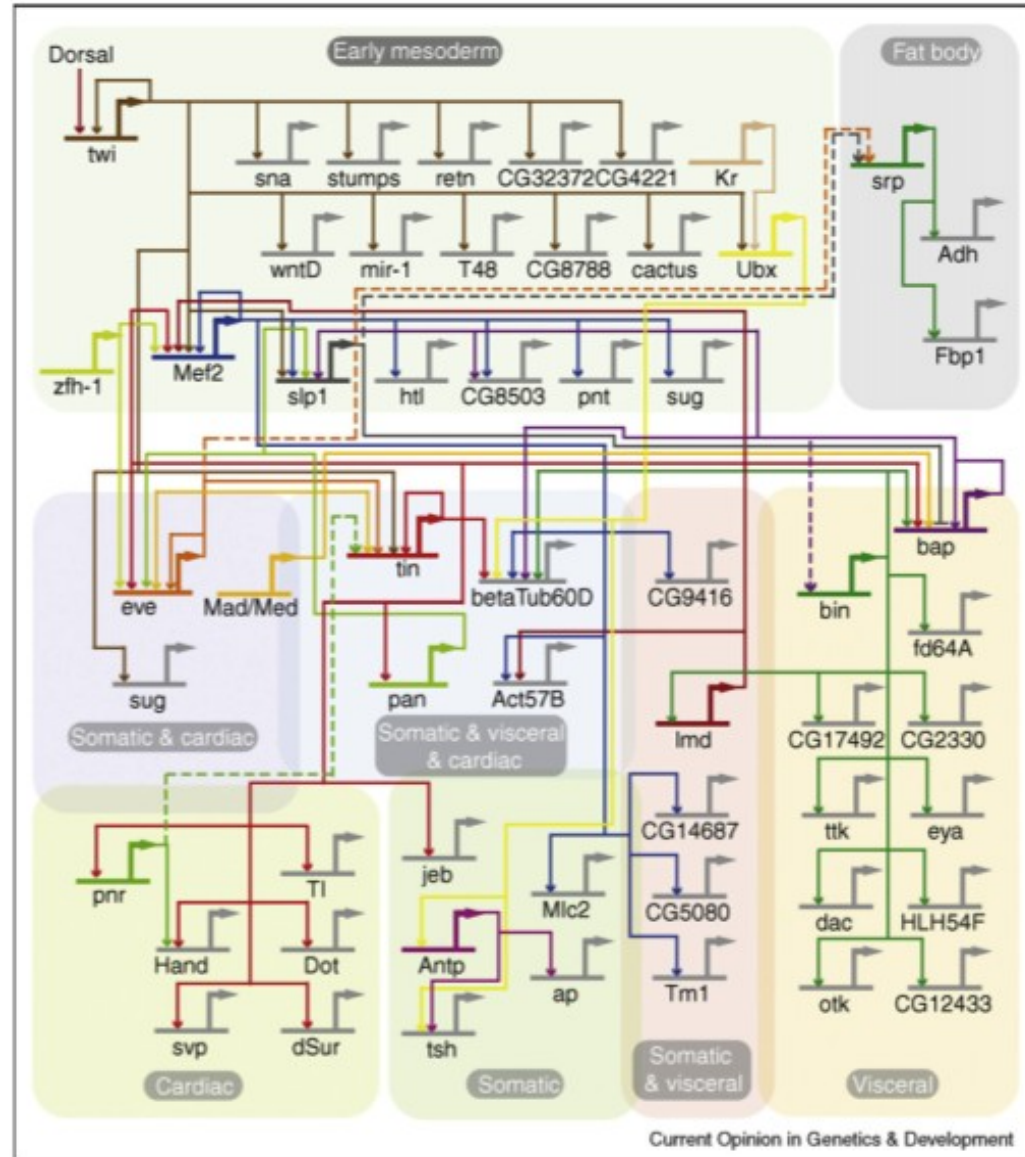
University of Warsaw  
Faculty of Mathematics,  
Informatics and Mechanics

Workshop on regulatory genomics & epigenomics

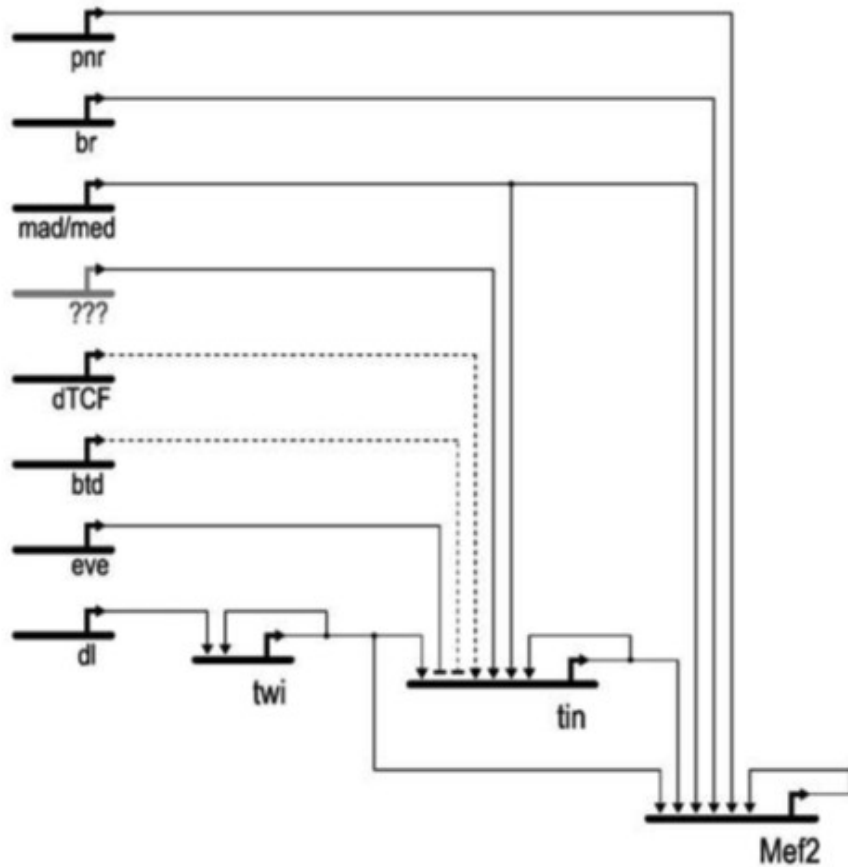
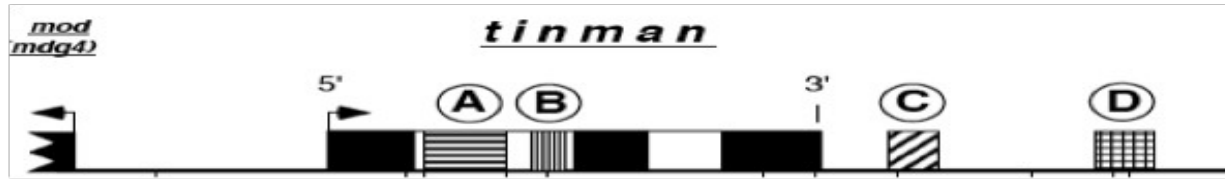
# Modelling transcriptional networks



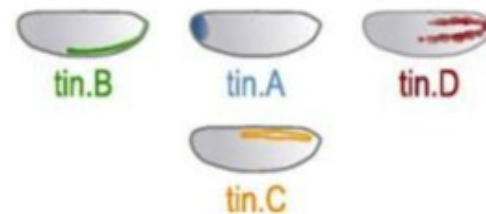
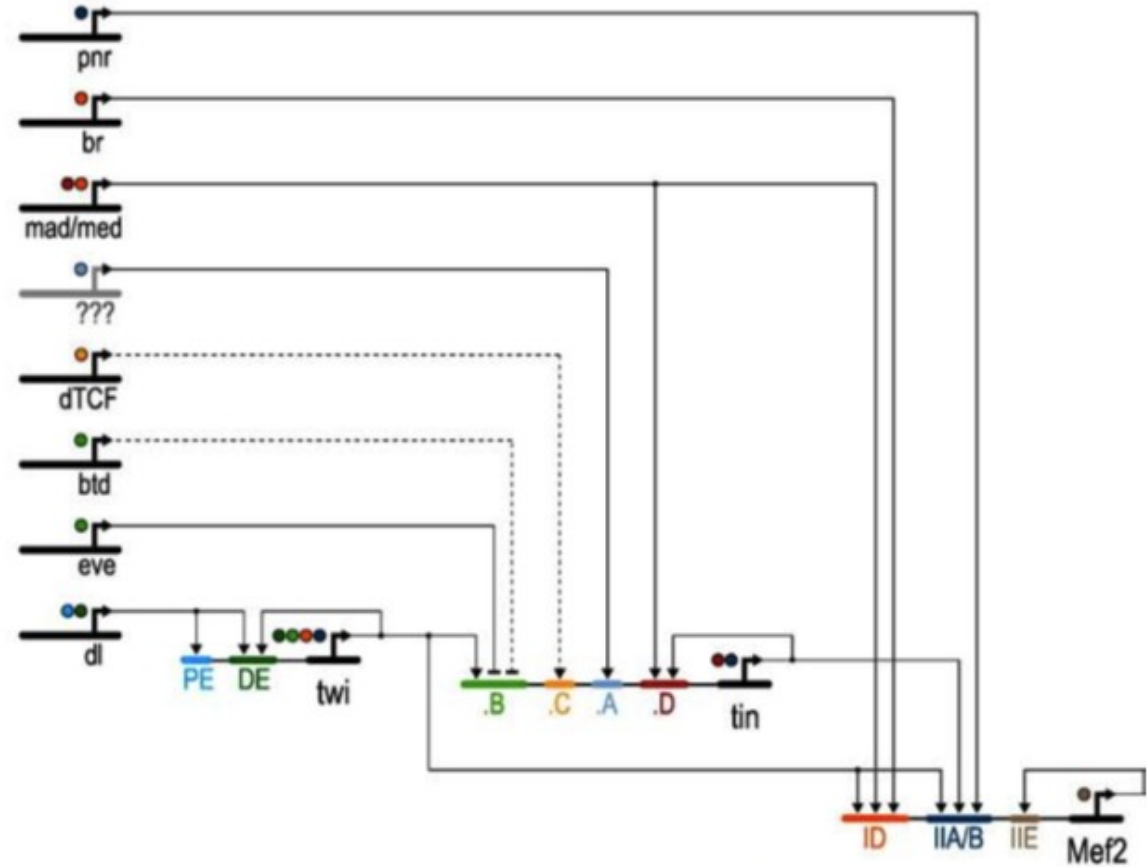
- Transcriptional networks as wiring diagrams
- Complexity on two different levels: many genes with many inputs



# Genes *integrate* action of multiple enhancers

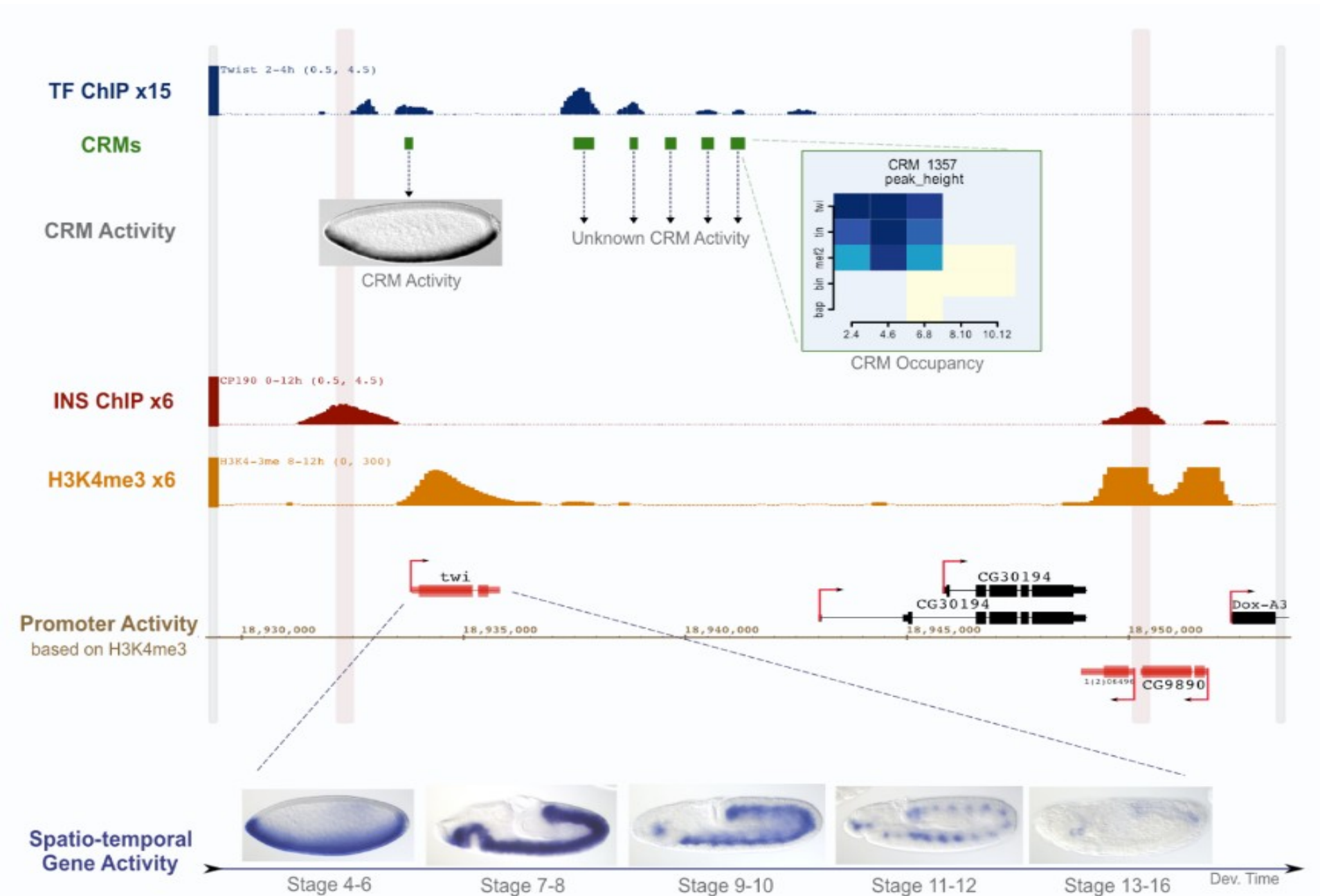


A

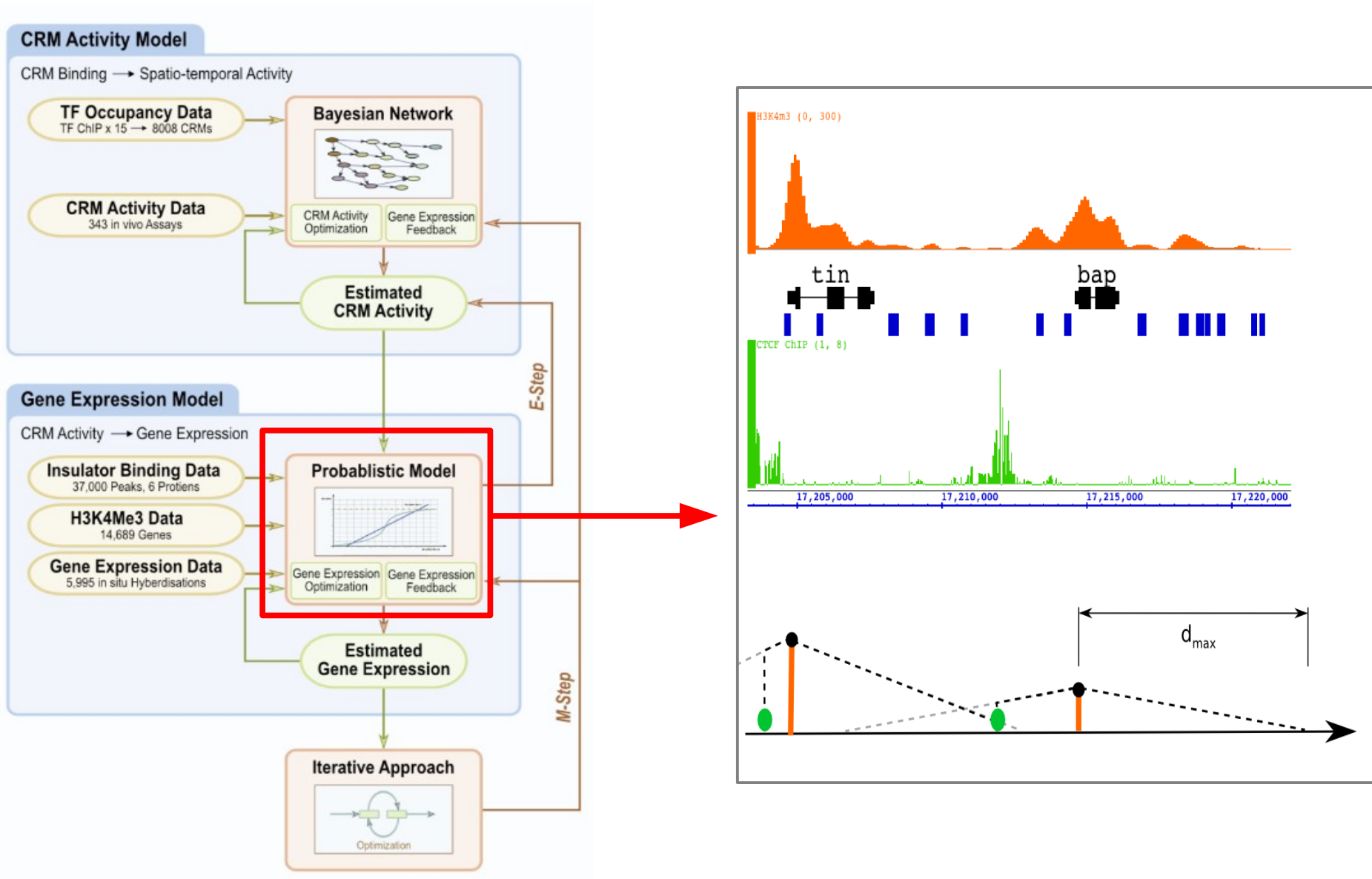


B

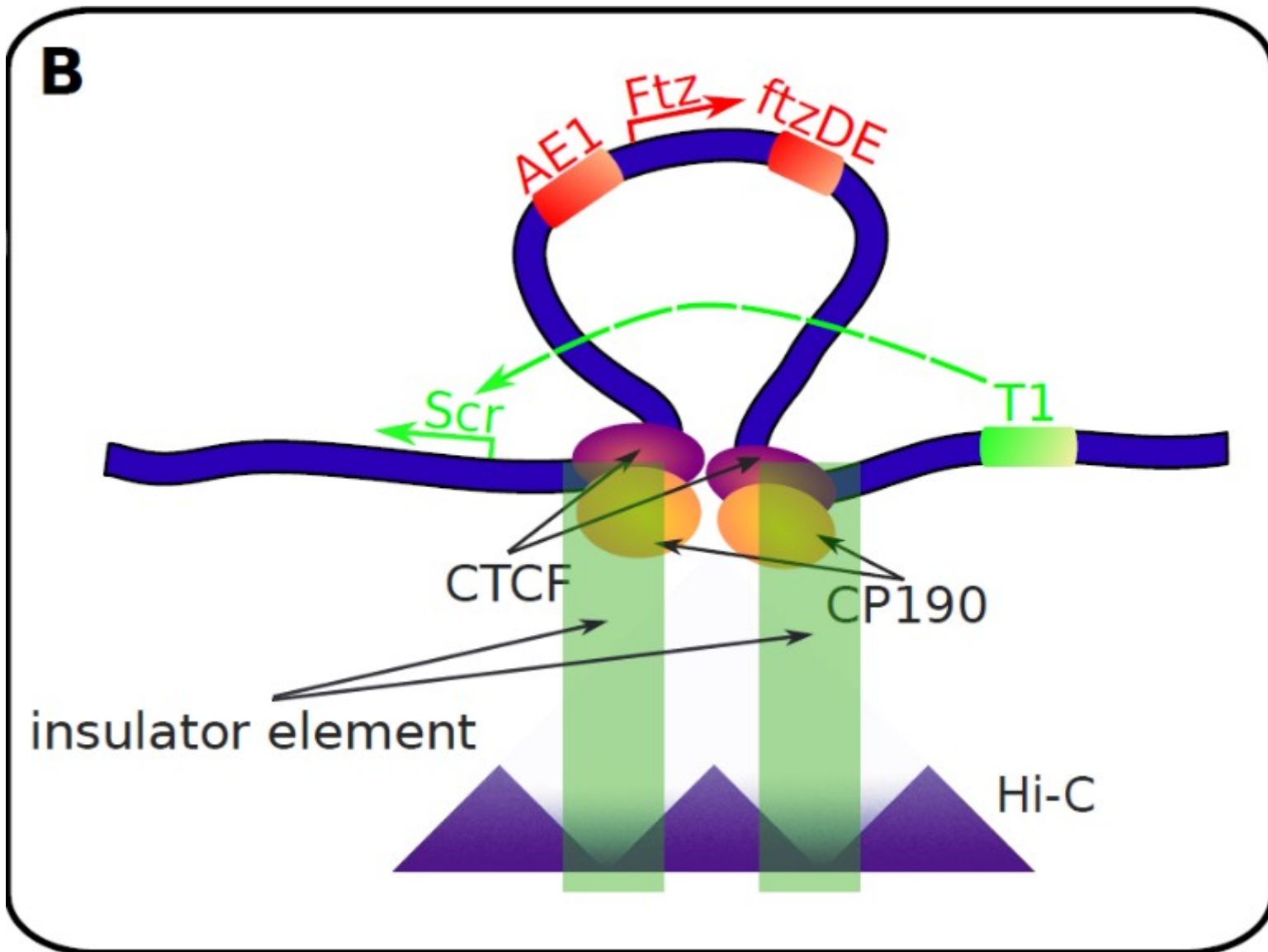
# Multiple layers of data in transcriptional regulation



# Model structure and optimization

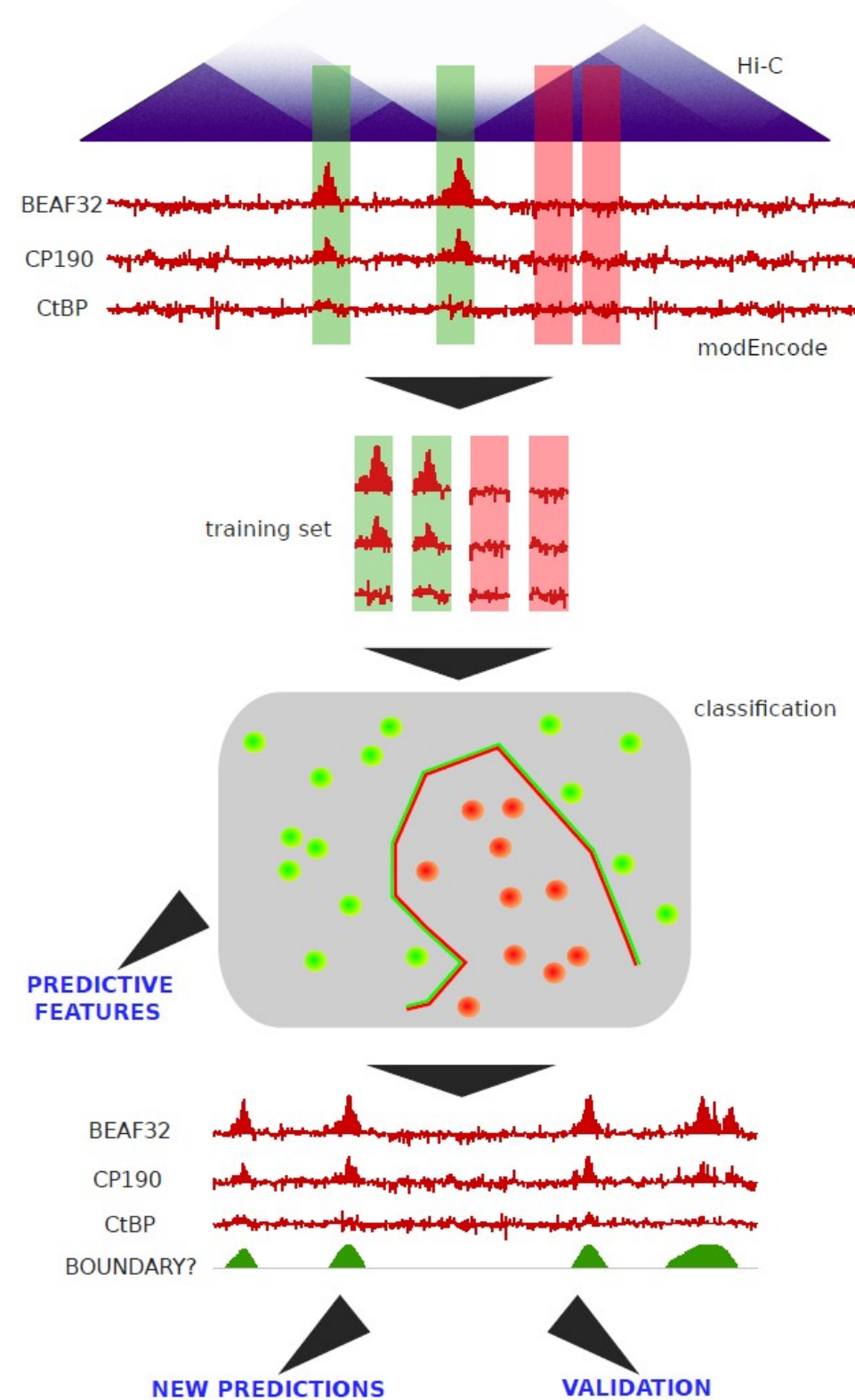


# Insulator elements

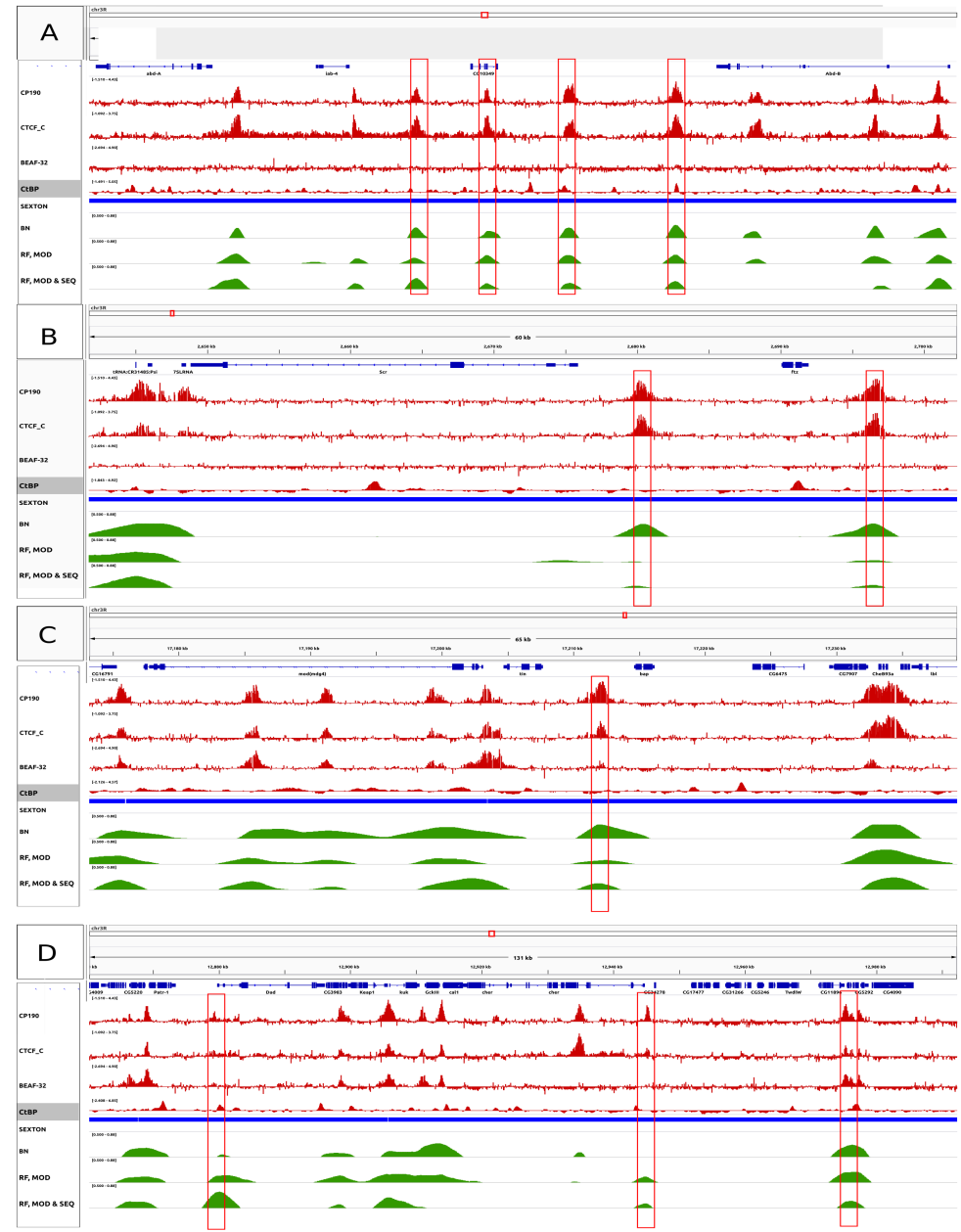
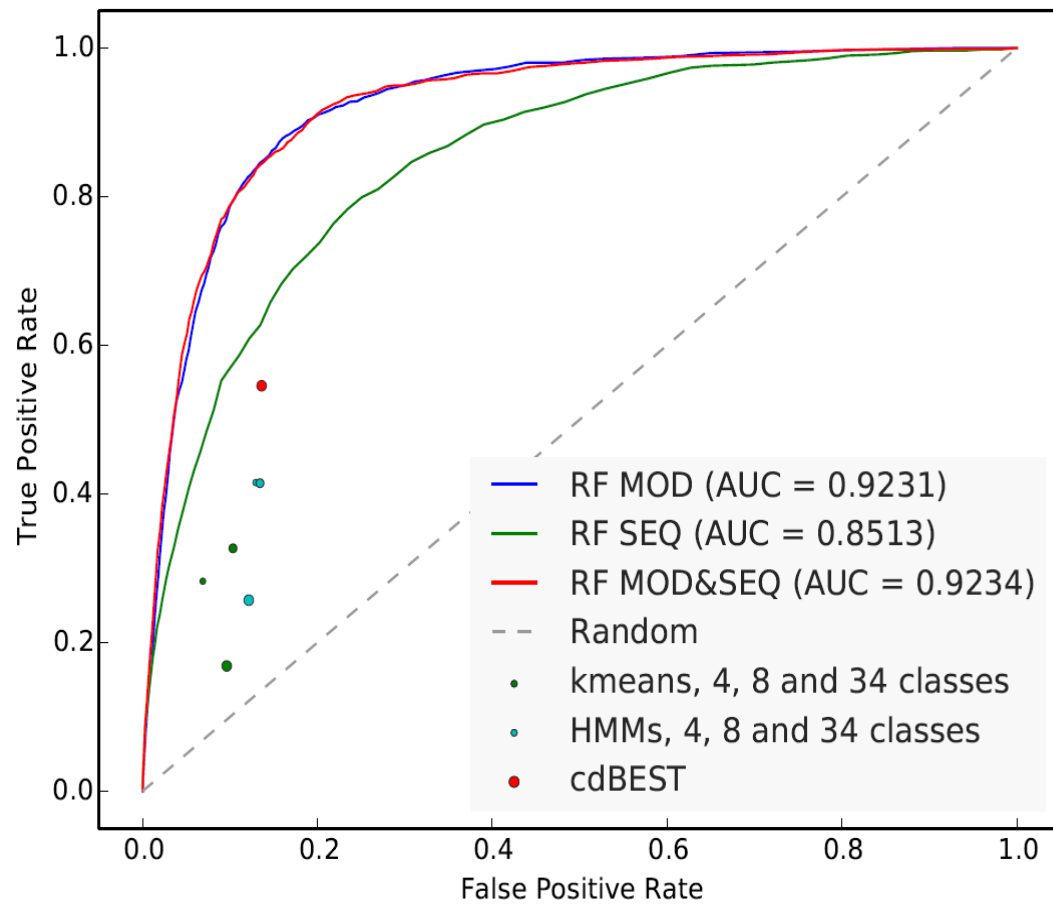


# Supervised learning

- Starting with  $\sim 1k$  boundaries determined by scaling factor of Hi-C data
- Training a Random Forest classifier to discern boundaries from „non-boundaries” by their Chip-Seq signal from modEncode

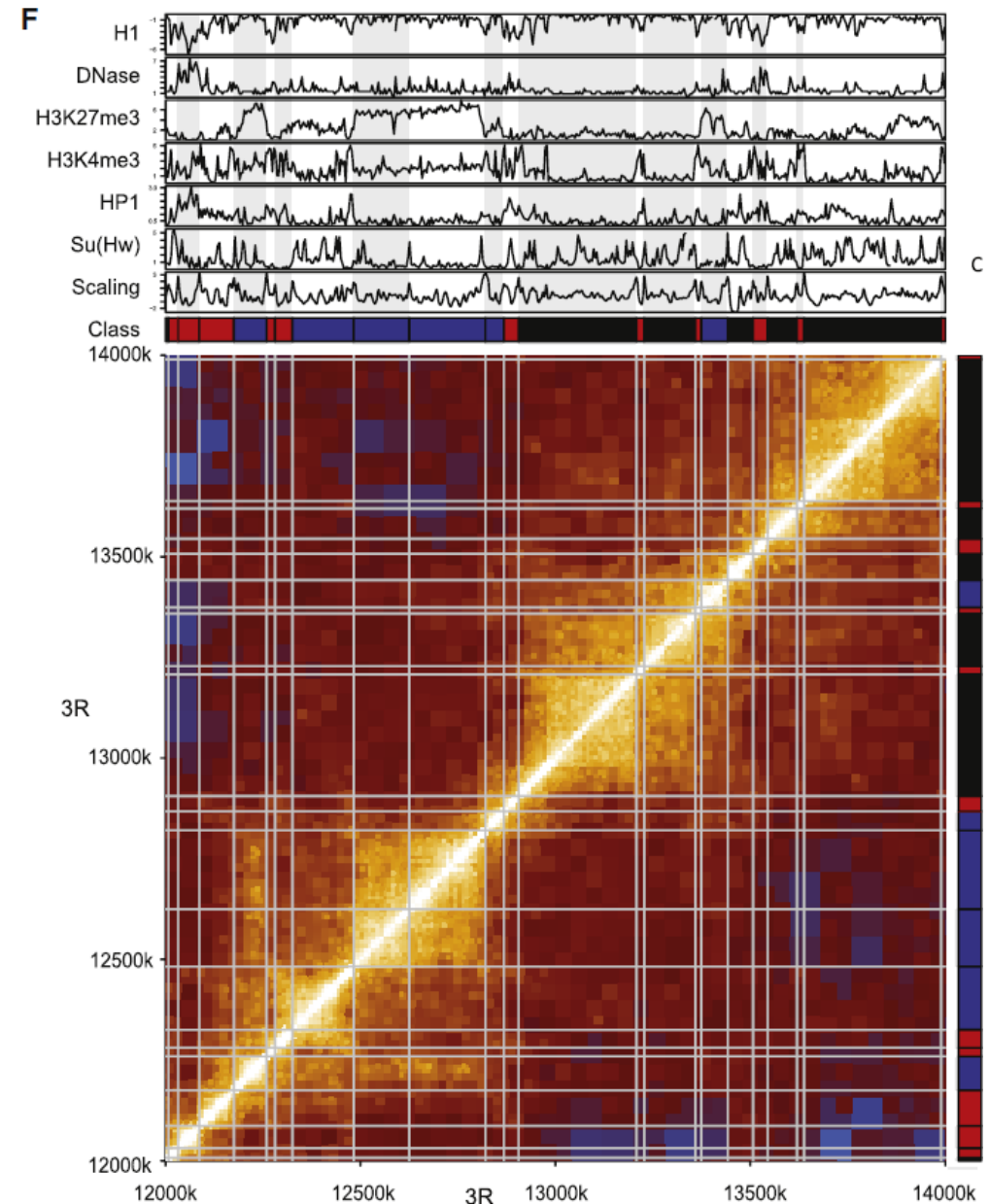
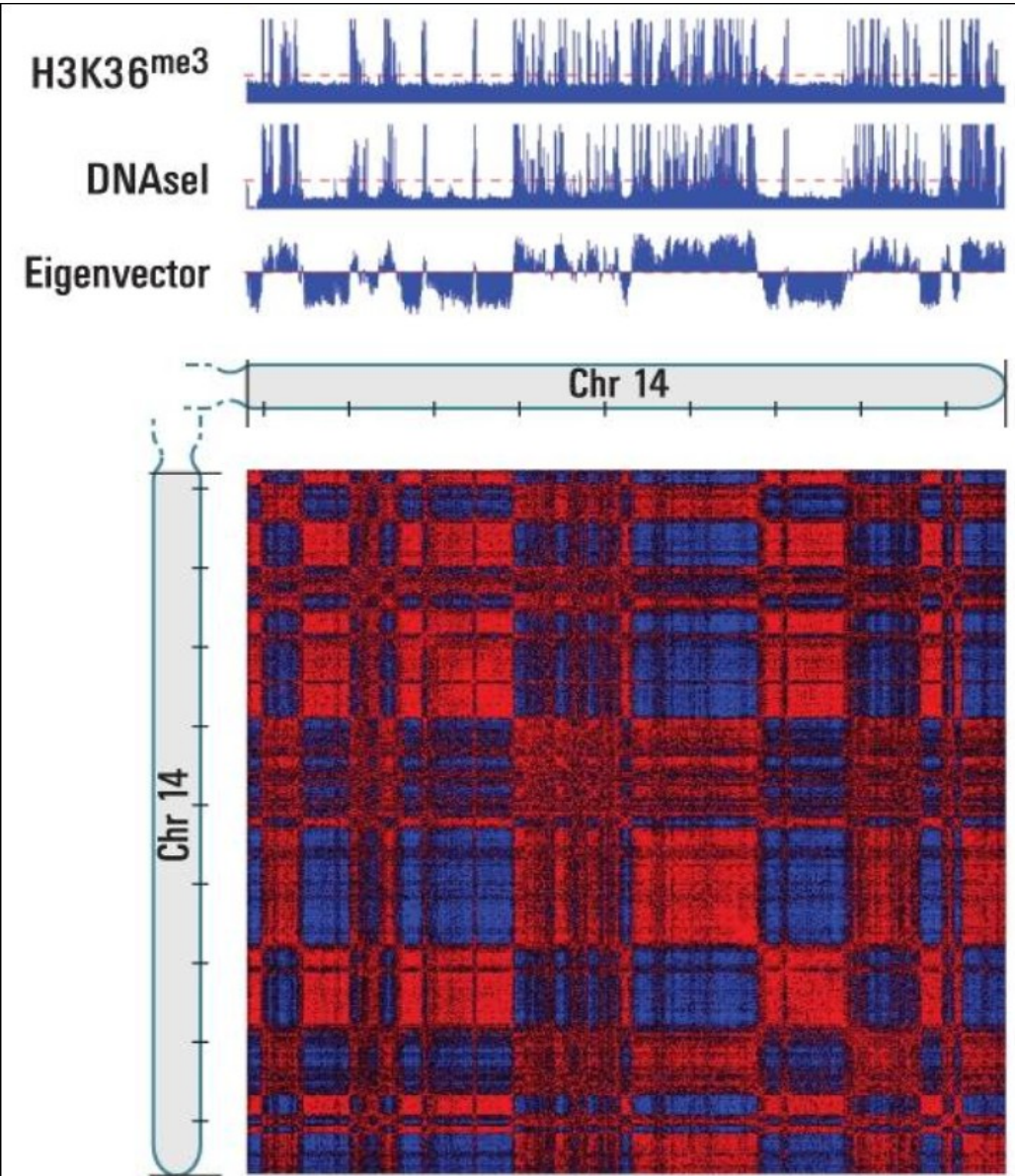


# Results for insulators





# A/B compartments vs Topologically Associating Domains

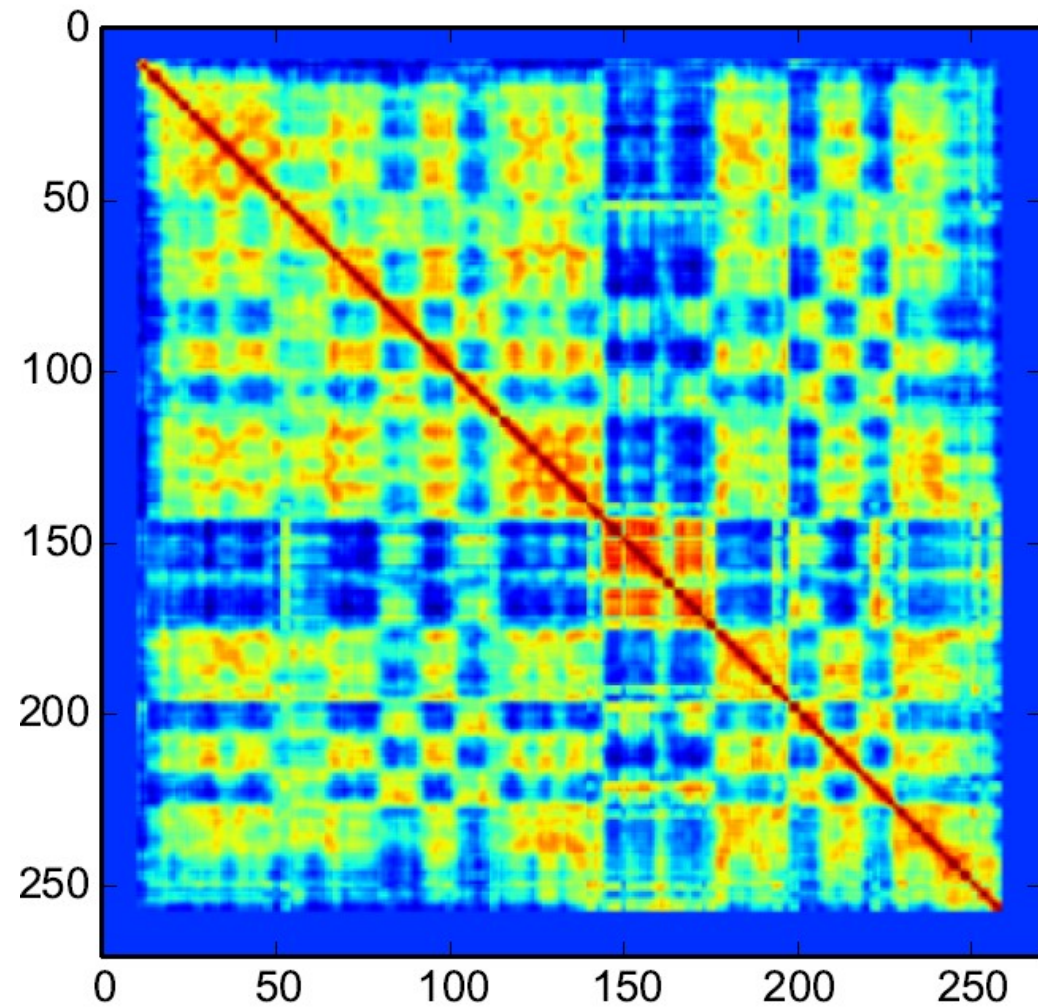
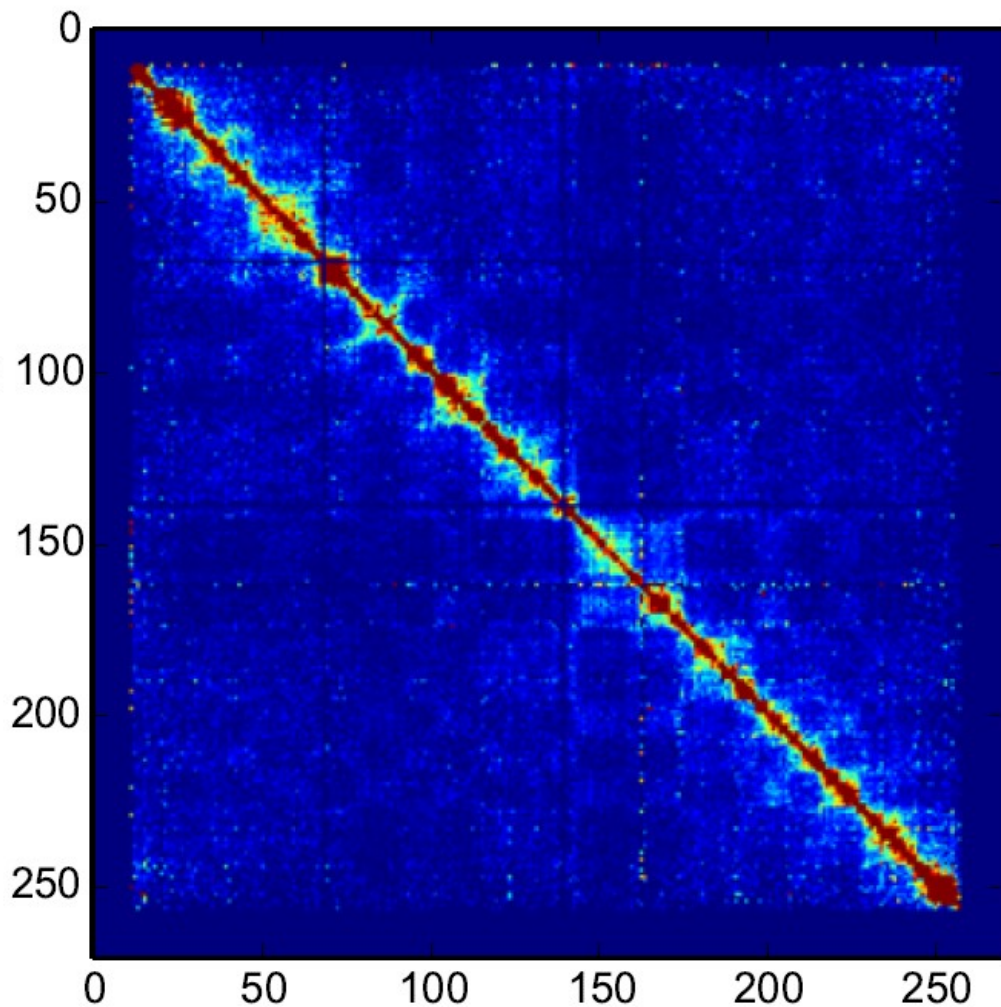


Lieberman-Aiden et al, Nature, 2009

Sexton et al, Cell, 2012

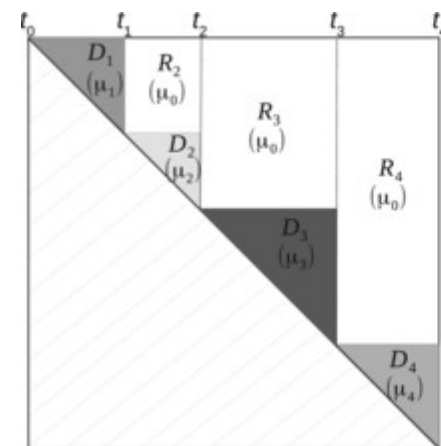
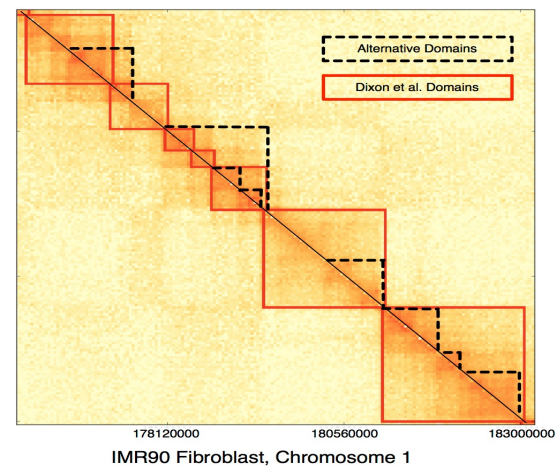
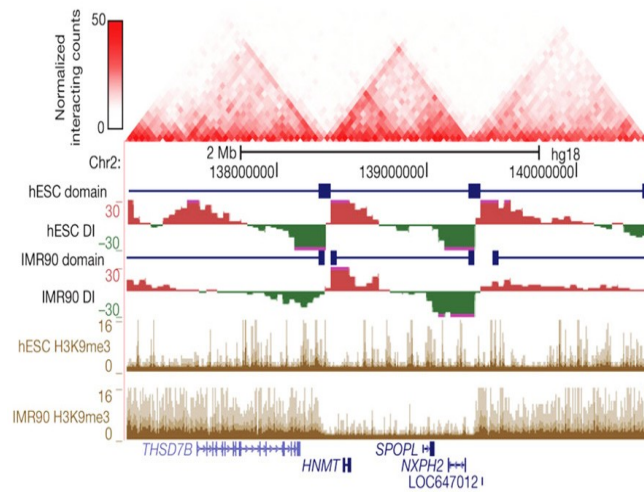
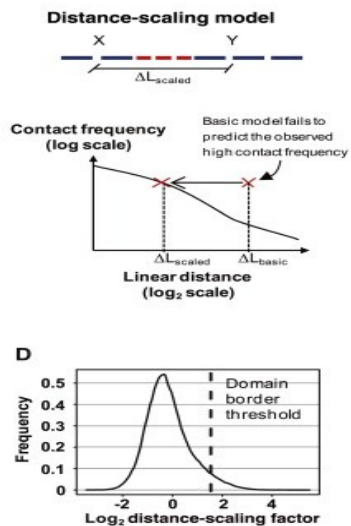
Depending on processing of the data, TADs or compartments are more visible.

While compartments tend to be larger than TADs, they do overlap in terms of scale and some TADs span more than one compartment.

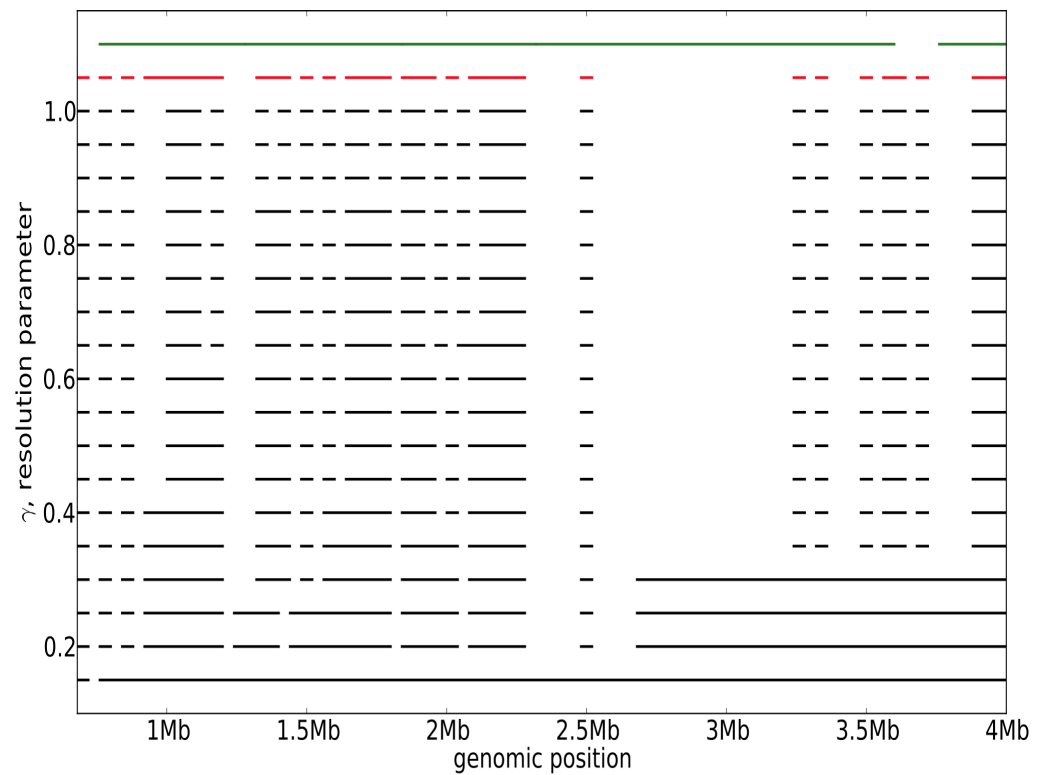
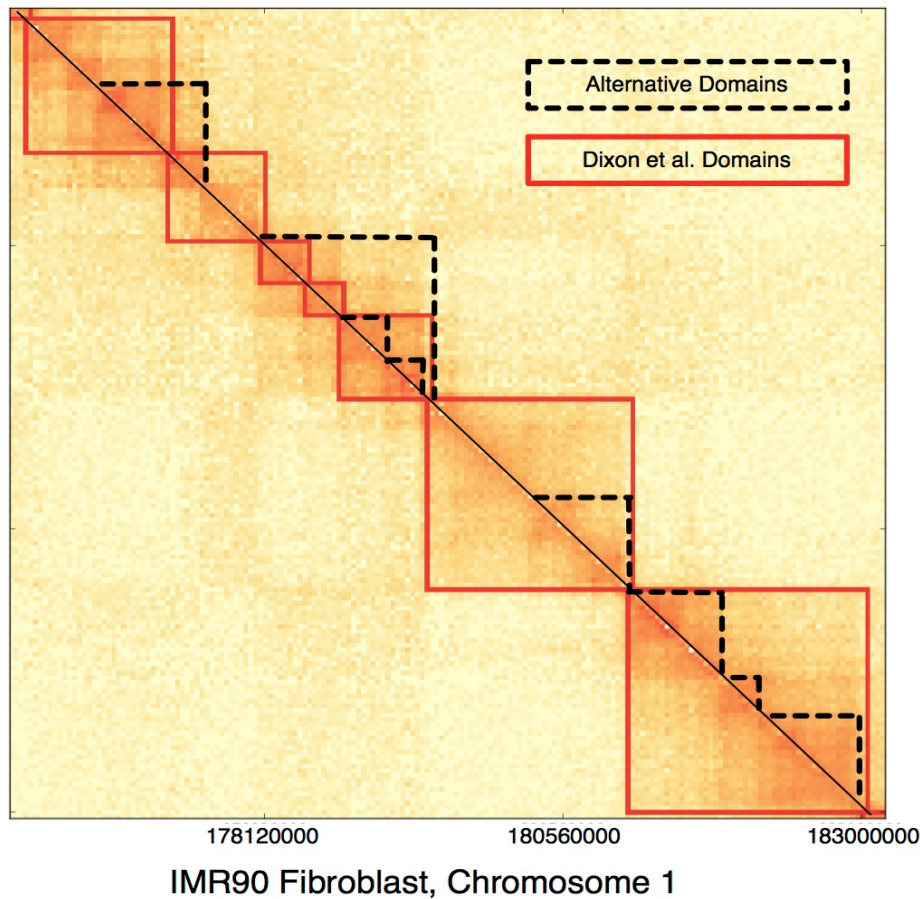


# Methods for delineating TADs

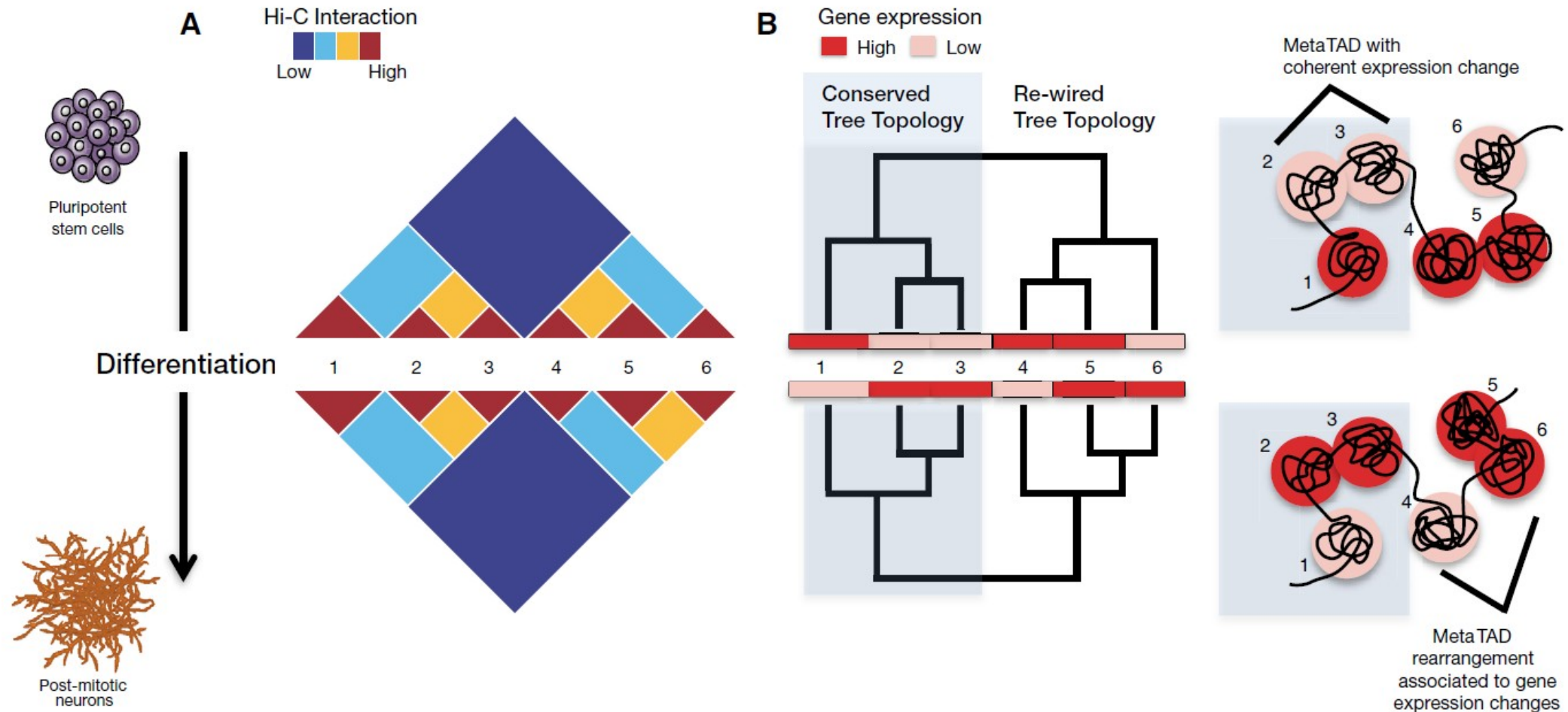
- Scaling factor (Sexton et al. 2012)
- Directionality Index (Dixon et al. 2012)
- Armatus (Filippova et al 2013)
- Statistical potential (Levy-Leduc et al 2014)
- More and more coming...



# ARMATUS - shifting parameters leads to different segmentations

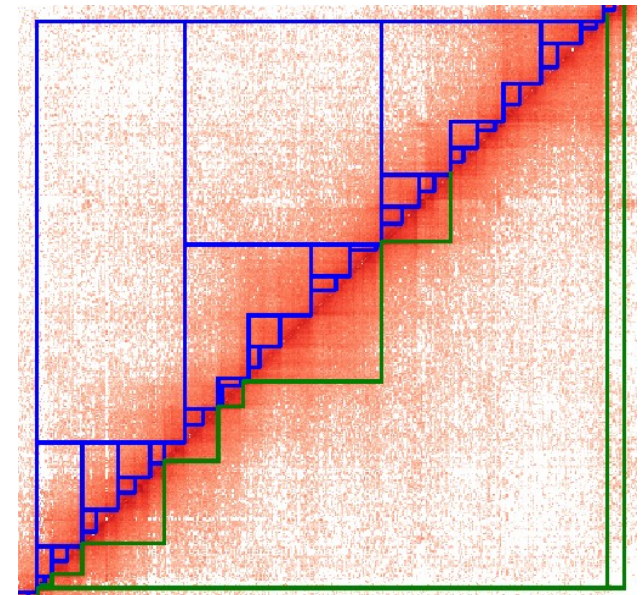
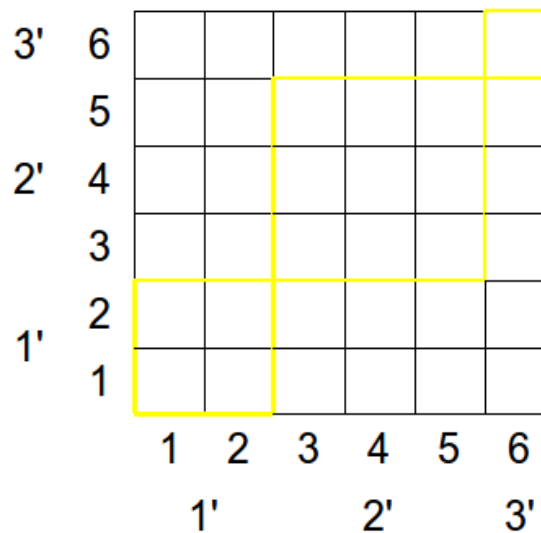
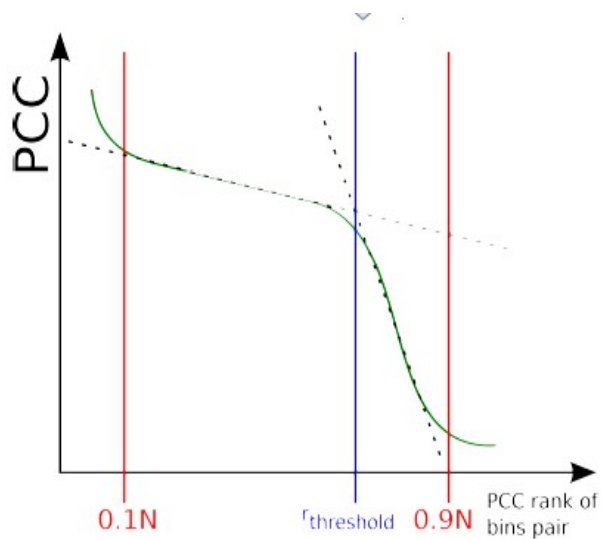
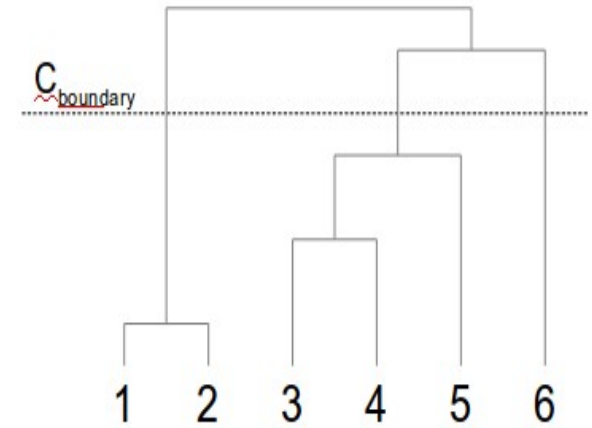
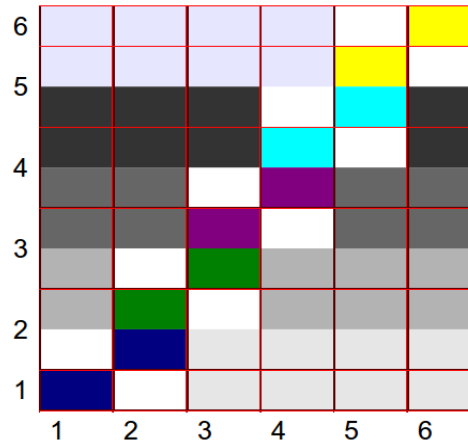
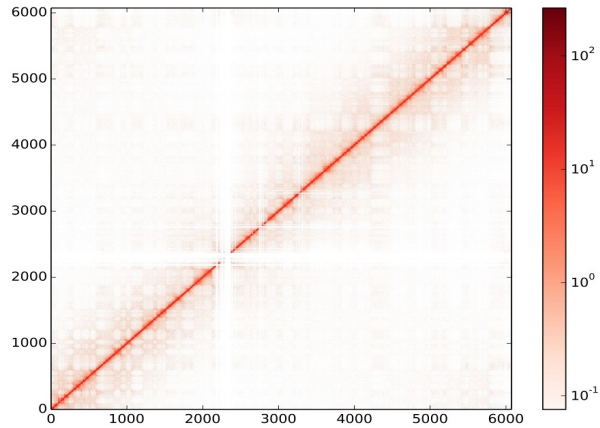


# Aggregation of TADs by their contact frequency

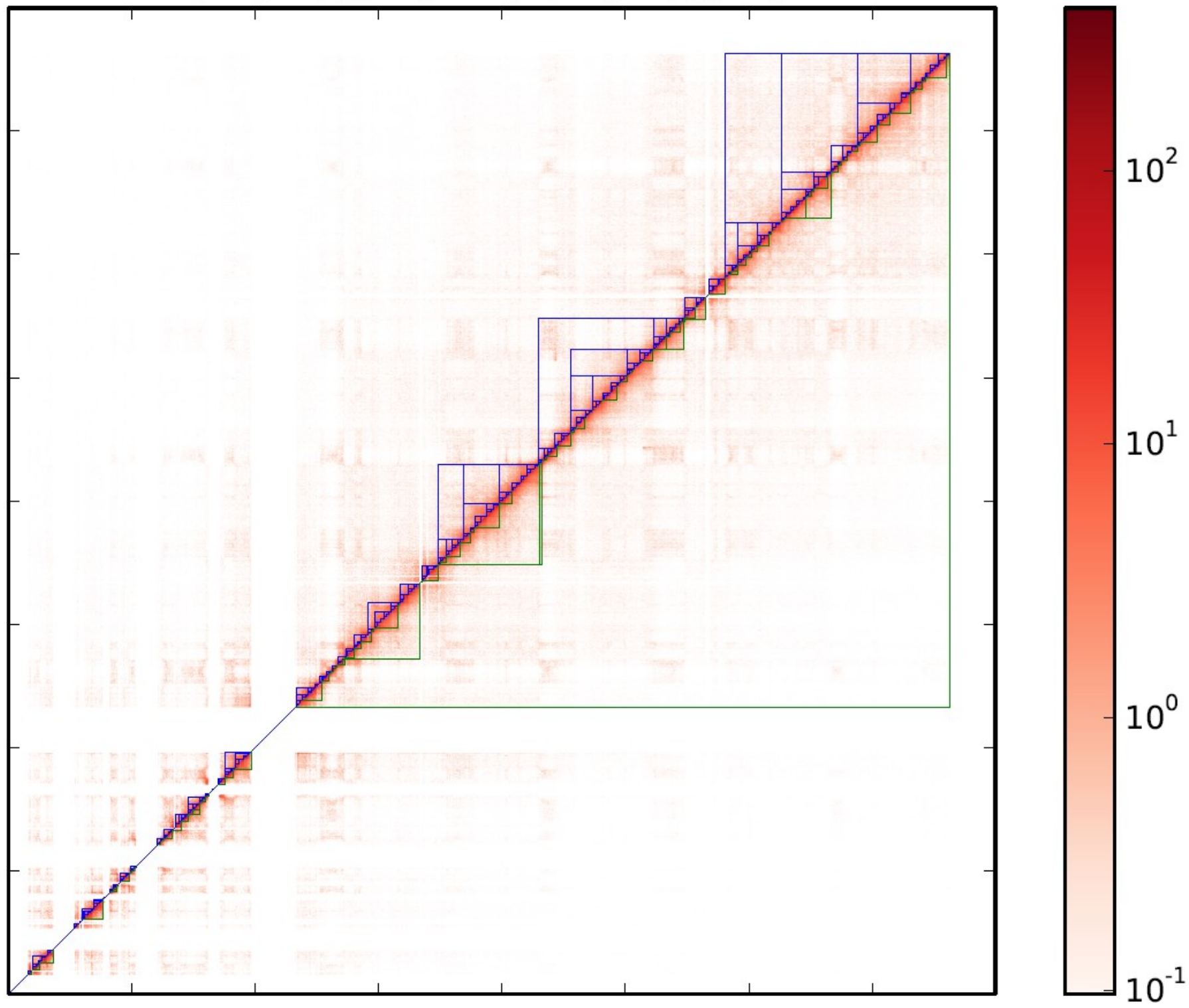


# SHiERPA in 6 pictures

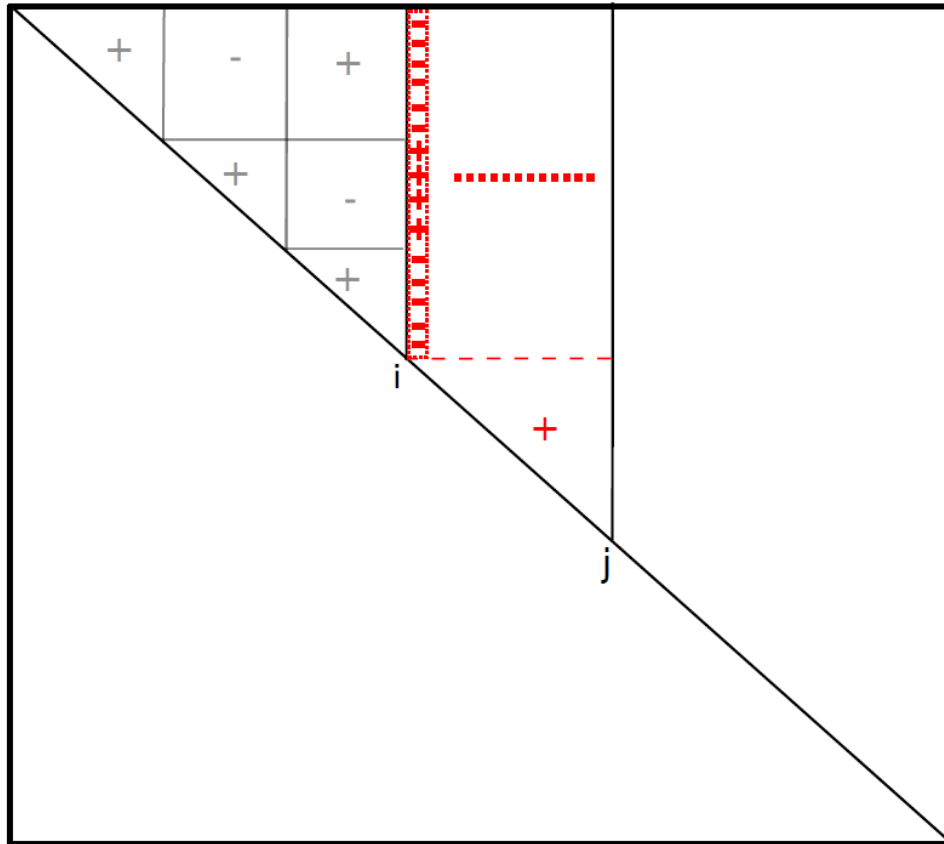
(Simple HiErarchical Profile Aggregation)



chromosome 7 vs chromosome 7



# Top-down approach for hierarchical „compartment” delineation - OPPA

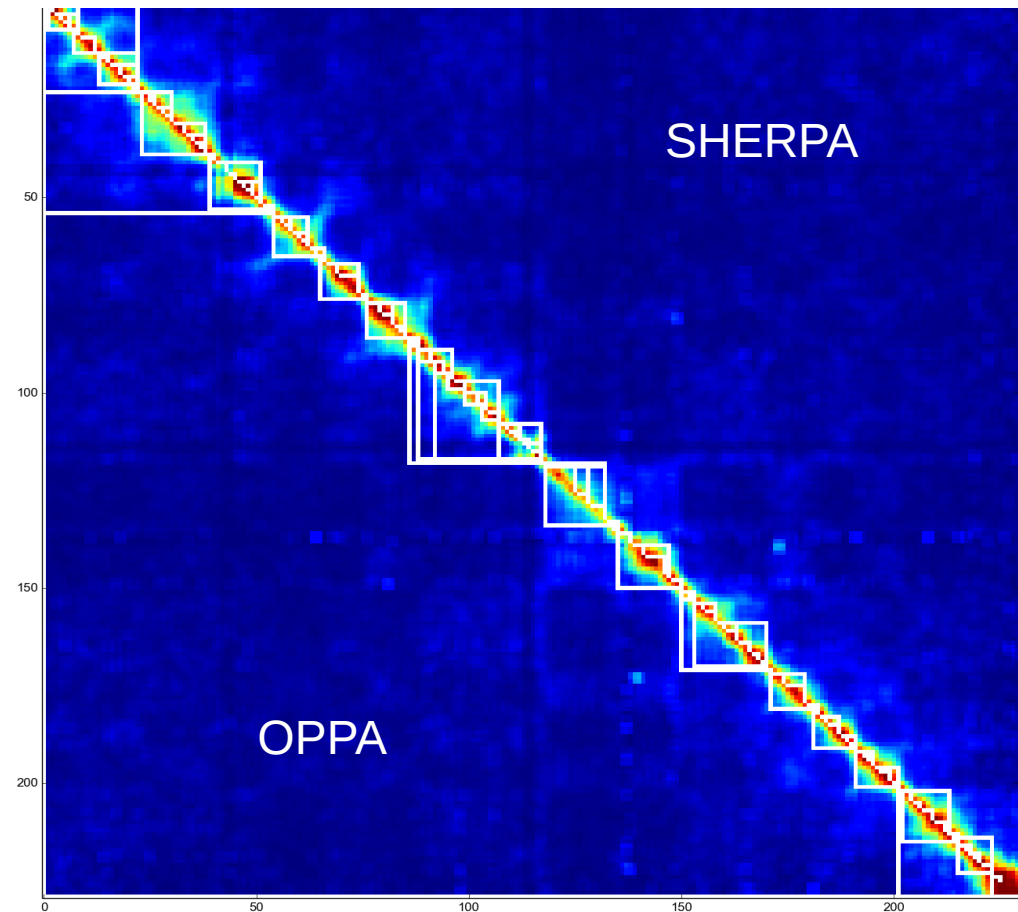


- Start with the correlation matrix for a given chromosome
- Using dynamic programming find the optimal segmentation into „opposite” compartments
- The dynamic algorithm will find the global optimum, provided that the data is consistent with the compartment model
- For every compartment run the method recursively, on re-normalized sub-matrix
- The process naturally stops when all vectors in the compartment have positive correlation



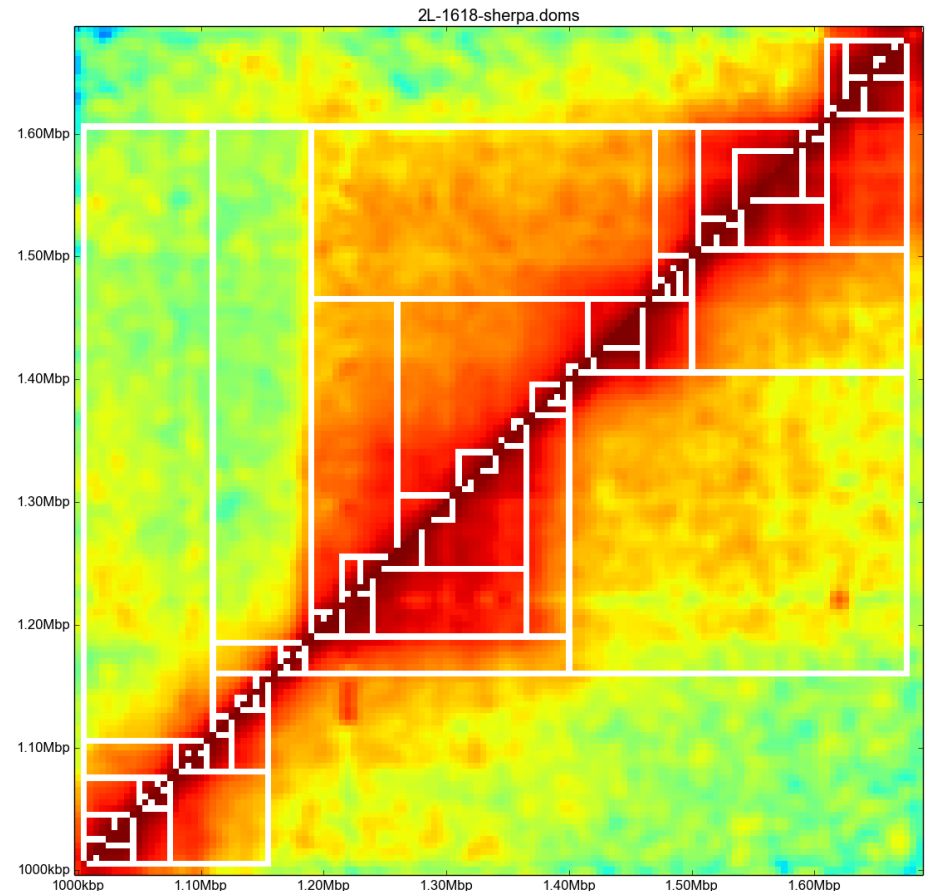
# Comparing SHERPA to OPPA

- Both methods have natural means of stopping the segmentation
  - SHERPA stops aggregation when there is no gain over simpler model
  - OPPA stops recursion, when all profiles are positively correlated
- They tend to stop at the same level, however sometimes their scales do overlap



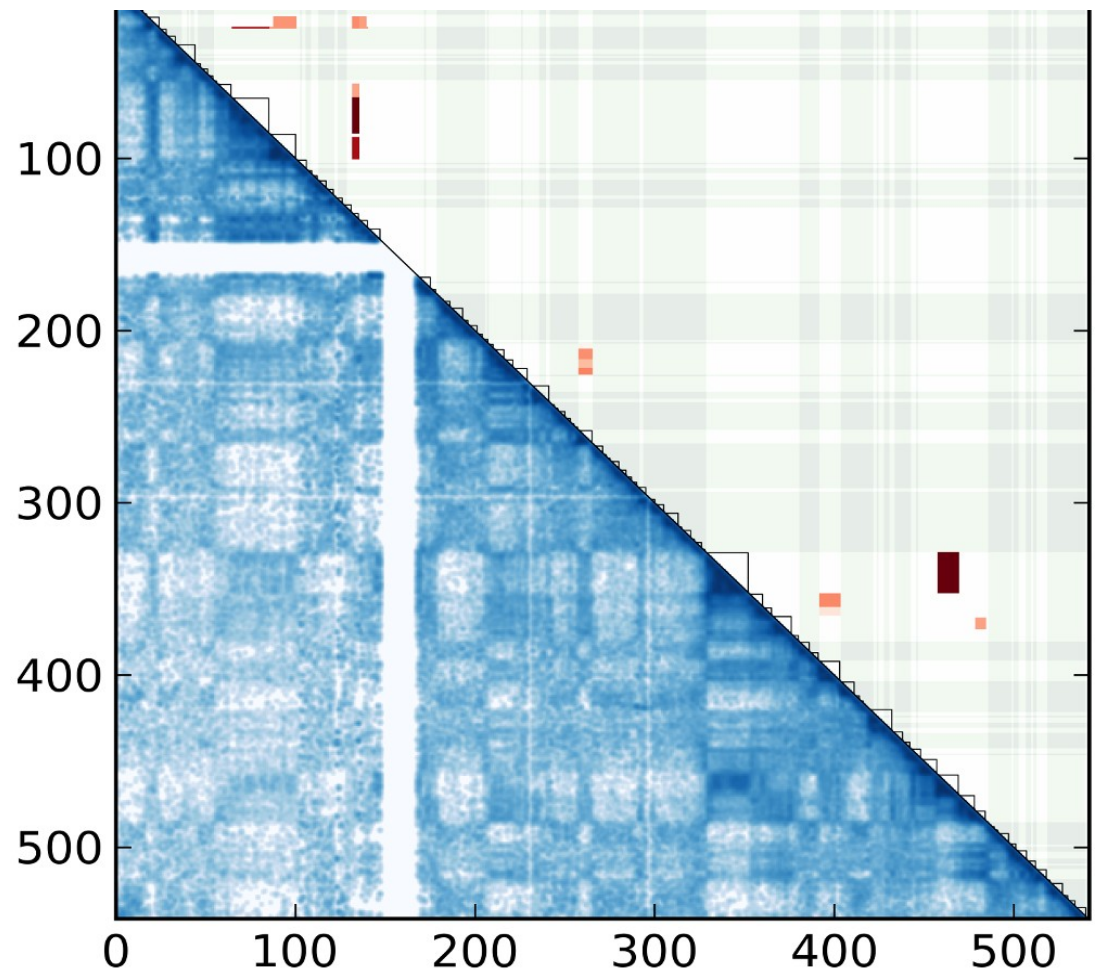
# Identification of domain hierarchy rearrangements between cell types

- Identification of hierarchy allows for identification of differences in neighboring domain association between cell types
- This is very frequent at the sub-tad level as opposed to the rather infrequent neighboring „super-TAD” rearrangements



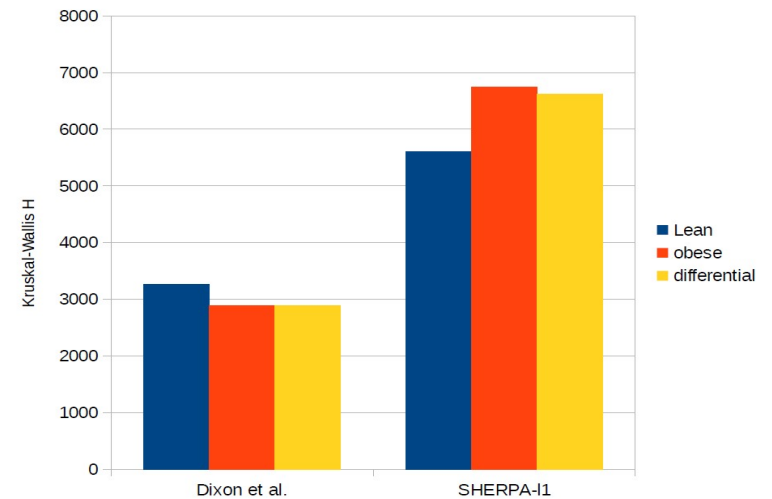
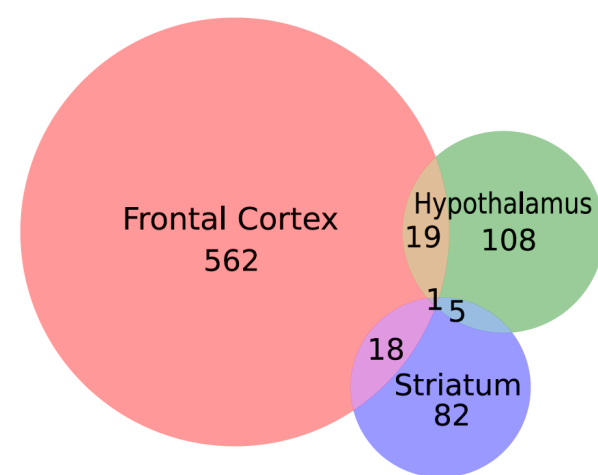
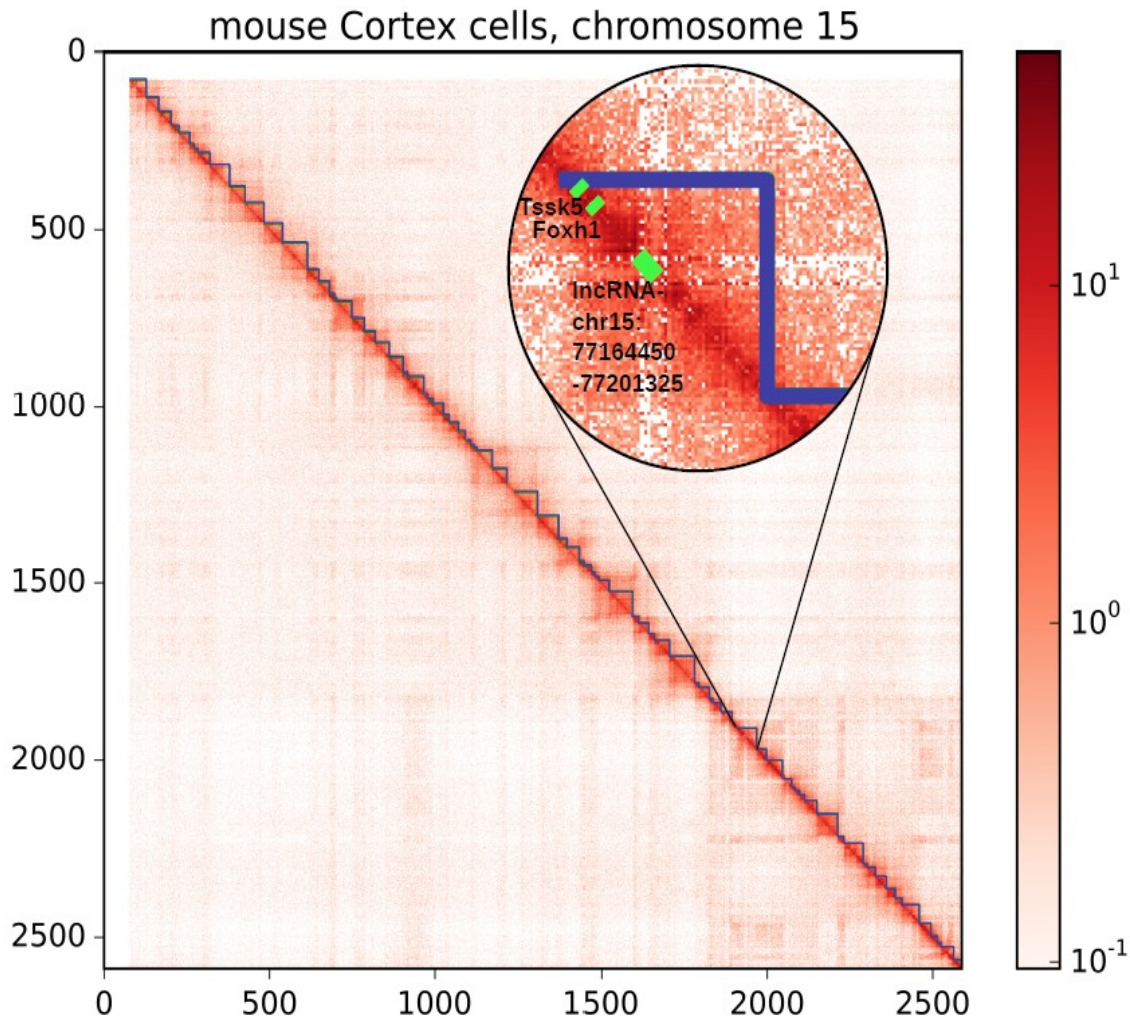
# Long range-contacts identification

- Having a hierarchy of domains greatly helps in identification of enriched domain-domain contacts
- Using the domain structure the problem of long range contact identification is reduced to statistical hypothesis testing

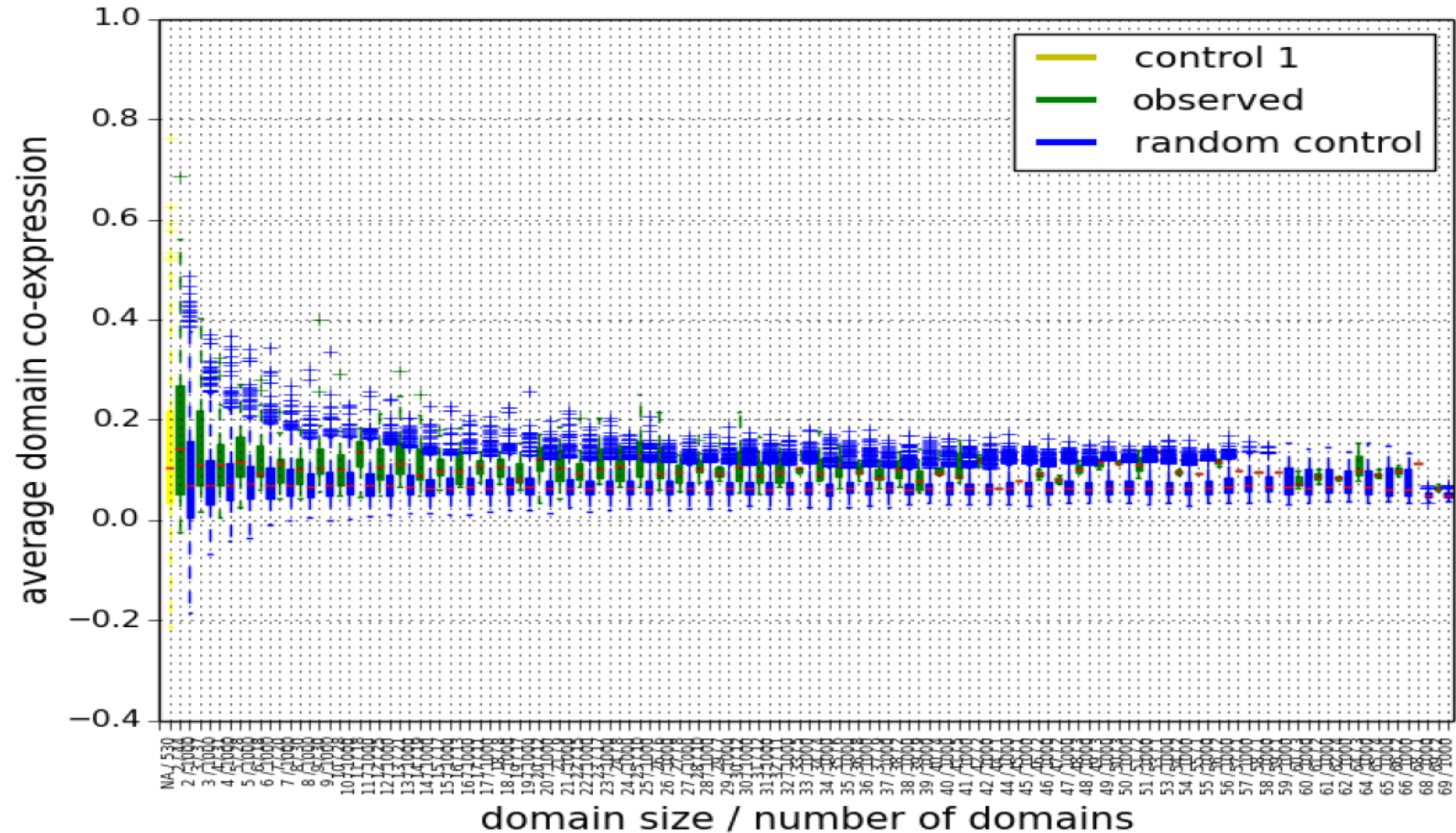


Cell-type specific contacts on human chromosome 17

# Comparing distribution of gene expression changes in chromatin domains upon stress



# Analysing co-expression of genes in chromatin domains



# Summary

- The chromosomes are organized at different scales into compartments and domains
- Both types of structures seem to overlap and show hierarchy within them
- Our algorithms, SHERPA and OPPA can be helpful in identifying these segmentations
- The segments we identify show agreement with gene expression
- They can be useful in downstream applications including long range contact identification and differential hi-c analysis

# 2016 Autumn School on Computational Approaches to Chromatin Organisation



Warsaw Center  
of Mathematics  
and Computer Science



10-13 Oct. 2016  
Będlewo, Poland

## Speakers:

**Minna U. Kaikkonen**, University of Eastern Finland, Kuopio

**Noam Kaplan**, Technion, Haifa

**Mario Nicodemi**, University of Naples Federico II, Napoli

**Aleksandra Pękowska**, National Institutes of Health, Bethesda, USA

**Dariusz Plewczyński**, Centre of New Technologies, Warsaw

**Ana Pombo**, Humboldt University of Berlin, Berlin

Registration deadline **April 30th 2016**

<http://regulomics.mimuw.edu.pl/wp/autumn-school-2016/>

**Fellowships available!**

- **Paweł Bednarz**
  - Julia Herman-Iżycka
  - **Ilona Grabowicz**
  - *Krzysiek Królak*
  - **Rafał Zaborowski**
  - **Irina Tuszyńska**
  - Michał Własnowolski
- 
- Eileen Furlong (EMBL)
  - Mara Dierssen (CRG)
  - Minna Kaikonen (UEF)
  - Ewa Szczurek (MIM)



Thanks !



NARODOWE CENTRUM NAUKI



EUROPEAN UNION  
EUROPEAN REGIONAL  
DEVELOPMENT FUND

