Identifying the transcription factors mediating enhancer–target gene regulation in the human genome

Yih-Chii Hwang | yihhwang@upenn.edu
Pavel P. Kuksa | pkuksa@upenn.edu

University of Pennsylvania
Genomics and Computational Biology Program
Perelman School of Medicine

American Society of Human Genetics | October 9th 2015, Baltimore
Regulatory interactions and mediating protein complexes in the Hi-C data?

Can we find DNA regulatory interactions and the DNA binding transcription factor complexes mediating these interactions?
Protein complexes mediate enhancer–promoter interactions

- Long-range regulation of beta-globin gene by locus control region

[Cell (2014), Deng et al.]
Who are the players for long-range regulations?

Considering the whole-genome...

What is the global landscape of **enhancer elements** and their **target genes**?

Which **transcription factors** mediate these interactions?
Restriction enzyme cutting suggests physically-interacting regions (PIRs)

★ Hi-C reads properties:
1. pile up around the cleavage/ligation sites
2. suggest the relative position of the physically interacting region from the cleavage/ligation sites
Two converging cleavage/ligation sites enclose a physically-interacting region

Determine the positions of the PIRs relative to the cleavage/ligation sites by:
1. Hi-C read mapping positions
2. distances from the paired-end reads to their nearest restriction sites
3. strand orientations

The PIR is identified with two converging and consistently cleavage/ligation sites
PIRs correlate with open chromatin regions

- 84% of open chromatin regions overlap with PIRs
- PIRs have an average length of 994 nucleotides, spanning 1–2 restriction fragments on average

Hi-C data from *Cell* (2014), Rao et al.
Open Chromatin data from [ENCODE Open Chromatin Track]
Promoter-touching PIRs are likely to be regulatory elements.

- We identified **11,848** enhancer–promoter interactions, with **8,552** enhancers and **4,612** promoters.

Significant interactions identified by Fit-Hi-C [Genome Res. (2014), Ay et al.]
Histone modification and ChIP-seq peaks data from [ENCODE]
Transcription factor binding sites are enriched in promoter-touching enhancers

Promoter-touching PIRs with enhancer-associated histone marks are enriched for TF bindings

PWM-based ChIP-seq peaks computational identified motifs (on Enhancers)

*: P values <= 0.05, binomial test with bonferroni correction

Clusters of TF binding motifs enriched in promoter-interacting enhancers

TFs for enhancer: TAL1, GATA1, CTCF, AP1, YY1, ETS1, SP1, etc.
TF–TF interactions are involved in enhancer–promoter regulations.

Among the enriched 30 TFs, we discovered 99 out of 115 (86%) known PPIs mediating the enhancer–promoter interactions.

TF complexes linking enhancer–promoter interactions

TFs for enhancers: SP1, YY1, AP1, RUNX1, etc.

TFs for promoters: SP1, YY1, AP1, RUNX1, TBP, etc.
Summary

• We developed a methodology to utilize Hi-C readout and discovered physically-interacting regions (PIRs)
• The PIRs cover chromatin-accessible regions, and promoter-interacting PIRs are enriched for epigenomic signals associated with enhancer activity
• We identified candidate transcription factor complexes that specifically regulate enhancer–promoter interactions
• Hi-C bioinformatics can shed light on long-range regulation and the transcription factors mediating regulatory interactions
## Acknowledgments

### Advisors
- **Li-San Wang**
- **Brian D. Gregory**

### Thesis committee
- Doris Wagner
- Gerard Schellenberg
- Nancy Zhang
- Uwe Ohler (MDC in Berlin-Buch)

### Lab members
- Pavel P. Kuksa
- Fanny Leung
- Matthew Willmann
- Xiang Yu
- Ian Silverman
- Nate Berkowitz
- Lee Vandivier
- Alex Amlie-Wolf
- Shawn Foley
- Lucy Shan
- Steve Anderson
- Otto Valladares
- Weixin (Jacky) Wang
- Mugdha Khaladkar
- Mitchell Tang

### Bioengineering
- Jennifer Phillips-Cremins

---

**Grant**
- NIGMS R01-GM099962
- NIA U24-AG041689