

# Unravelling higher order chromatin organisation through statistical analysis

Ben Moore

19<sup>th</sup> November 2015



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

# Introduction

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis

Broad aim: investigate the relationship between structure and function of the genome

Some specific questions:

1. How does higher order chromatin structure compare across human cell types?
2. Can we predict higher order chromatin structure from locus-level features?
3. How do the characteristics of boundaries demarcating higher order domains vary between cell types and domain classes?



Introduction

Reanalysis of Hi-C datasets

Integrative modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

# What's known about genome structure

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

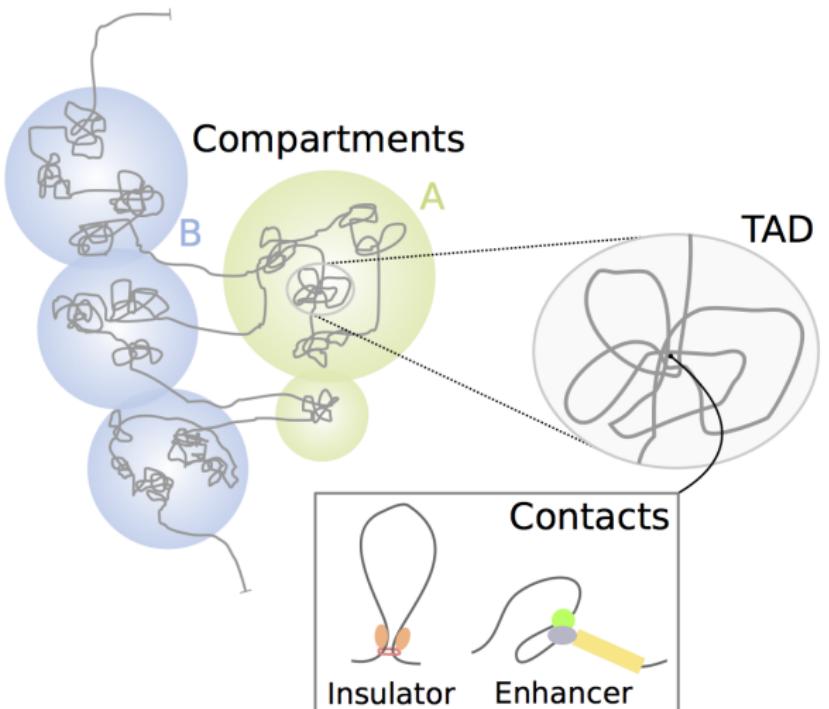
Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

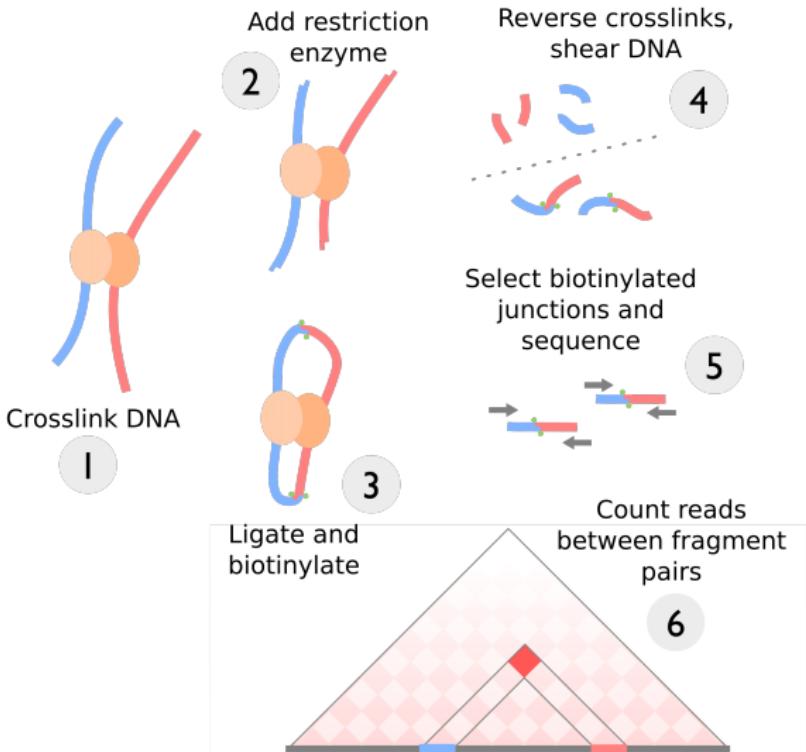


# Hi-C: a genome-wide chromosome conformation assay

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE



Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

# Chromosome compartments from Hi-C data

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

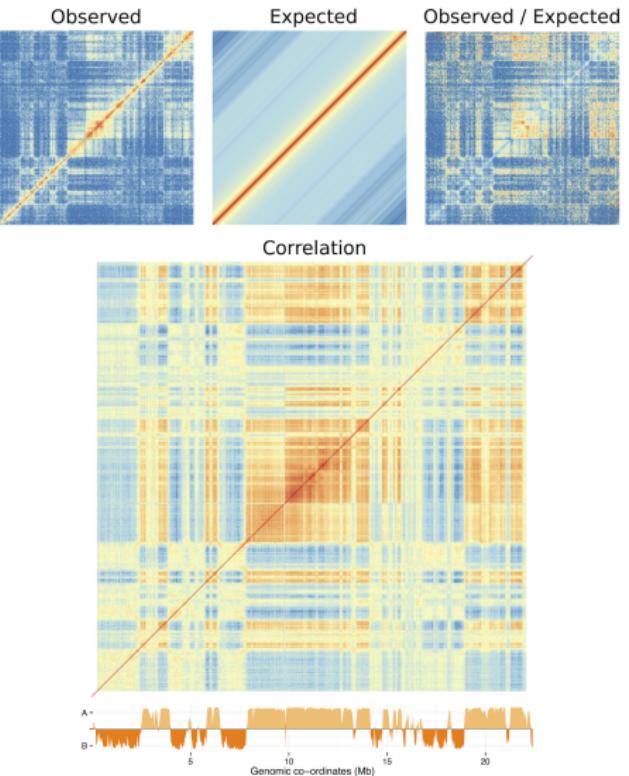
Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements



Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

# Strategy

- ▶ Integrate publicly available Hi-C data
- ▶ Uniformly reprocess each dataset
- ▶ Call compartments, TADs from reprocessed data
- ▶ Integrate cell-matched ENCODE epigenomic data

Then:

1. Compare/contrast cell types after reprocessing
2. Attempt predictive models of compartments and TADs from epigenomic features
3. Analyse boundary composition in terms of epigenomic features

Related collaborative work:

4. Investigate concept of "metaTADs"
5. Analyse conformation changes at specific locus of interest

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

## Results 1: Reanalysis of Hi-C data

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

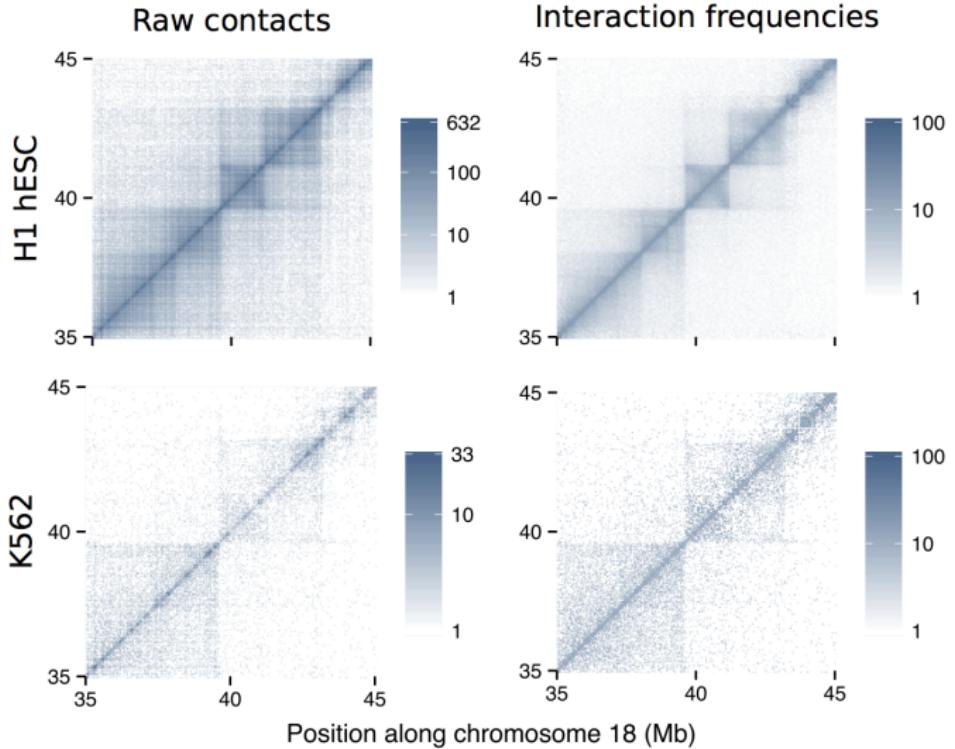
Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements



# Reanalysis of Hi-C datasets

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

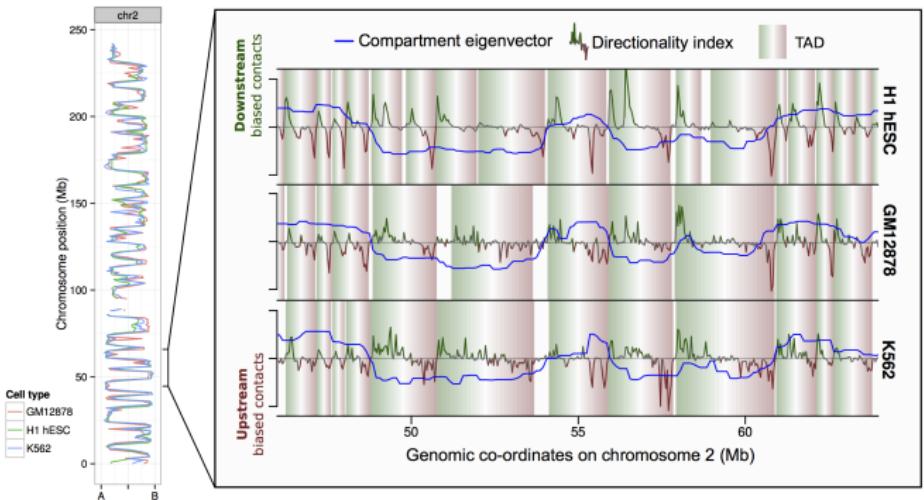
Collaborations

Next project

Thesis plan

Acknowledgements

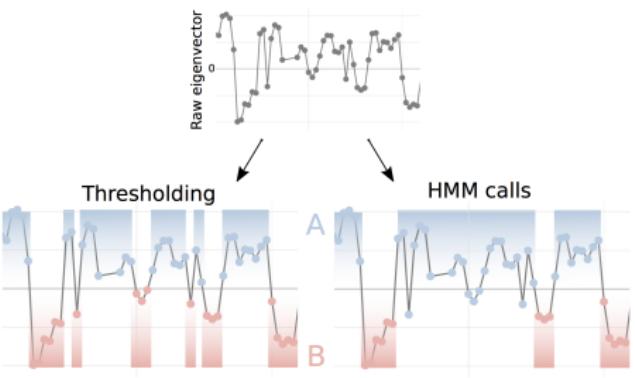
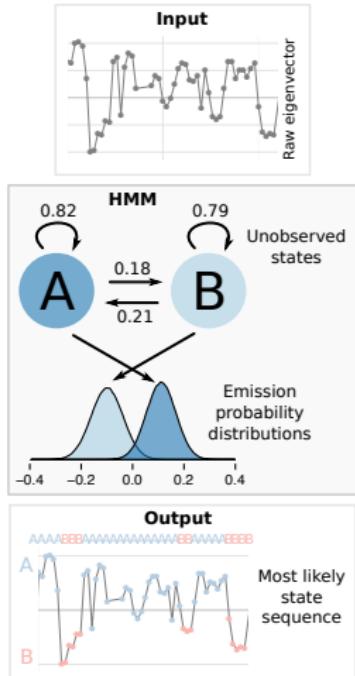
Despite this, reprocessed Hi-C data is well-correlated:



Justifies going forward with between cell-line analysis

# Improved compartment calling algorithm

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



igmm

INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

# Regions of variable compartment structure

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

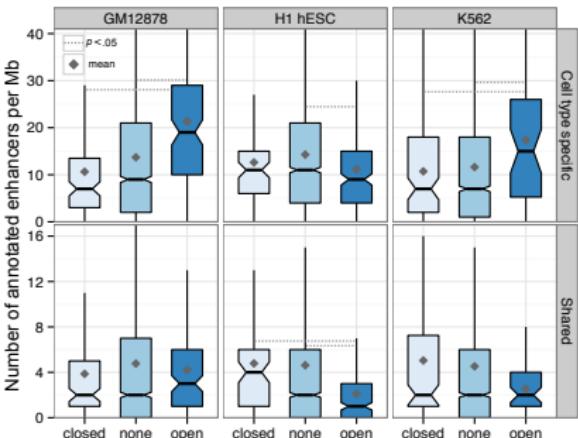
Boundaries

Collaborations

Next project

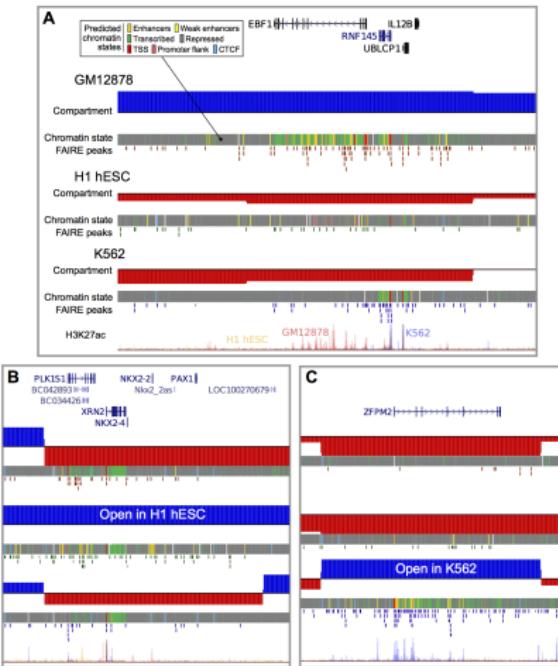
Thesis plan

Acknowledgements



# Regions of variable compartment structure

Some harbour genes with known cell type specific function



igmm

INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

## Results 2: Integrative modelling as a tool to explore biological systems

# Integrative modelling

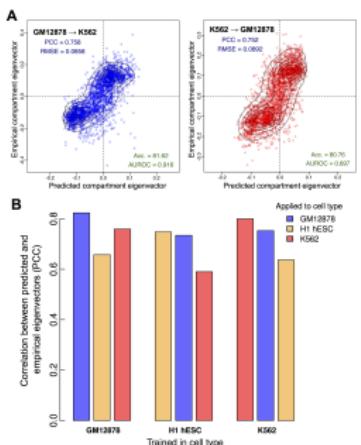
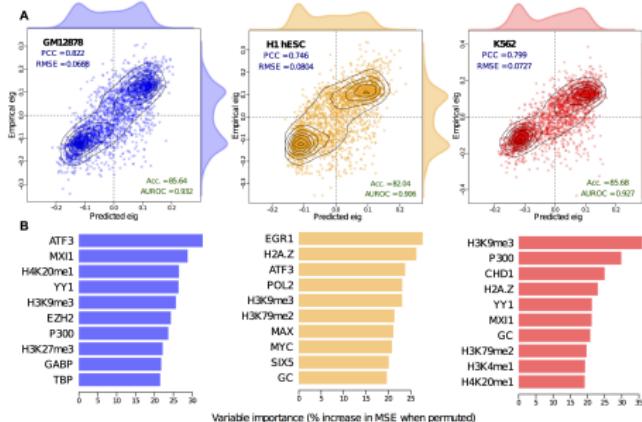
p67

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis

igmm

INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Chromosome compartment profiles can be accurately predicted and models generalise across cell types:



Introduction

Reanalysis of Hi-C datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

# Integrative modelling

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Model dissection a focus of much of this chapter:

- ▶ Variable importance
  - ▶ Varies between cell type models but multicollinearity in feature set
  - ▶ In some cases interpretable (p70), many of the usual suspects of chromatin conformation
- ▶ Comparing machine learning methods
- ▶ Regularised models

Introduction

Reanalysis of Hi-C datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

# Integrative modelling

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis

Most results in these two chapters were published:



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Research

Highly accessed

Open Access

## Integrative modeling reveals the principles of multi-scale chromatin boundary formation in human nuclear organization

Benjamin L Moore, Stuart Aitken and Colin A Semple\*

\* Corresponding author: Colin A Semple [colin.semple@igmm.ed.ac.uk](mailto:colin.semple@igmm.ed.ac.uk)

▼ Author Affiliations

MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, UK

For all author emails, please [log on](#).

Genome Biology 2015, **16**:110 doi:10.1186/s13059-015-0661-x

Introduction

Reanalysis of Hi-C datasets

Integrative modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

## Results 3: Chromatin domain boundaries

# Boundaries

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

TAD boundaries have been shown to be enriched for various epigenomic marks.

- ▶ Which of these boundary enrichments are statistically significant relative to a null model?
- ▶ Are the same features enriched at chromatin compartment boundaries?
- ▶ Are these enrichments shared across cell types?

Introduction

Reanalysis of Hi-C datasets

Integrative modelling

Boundaries

Collaborations

Next project

Thesis plan

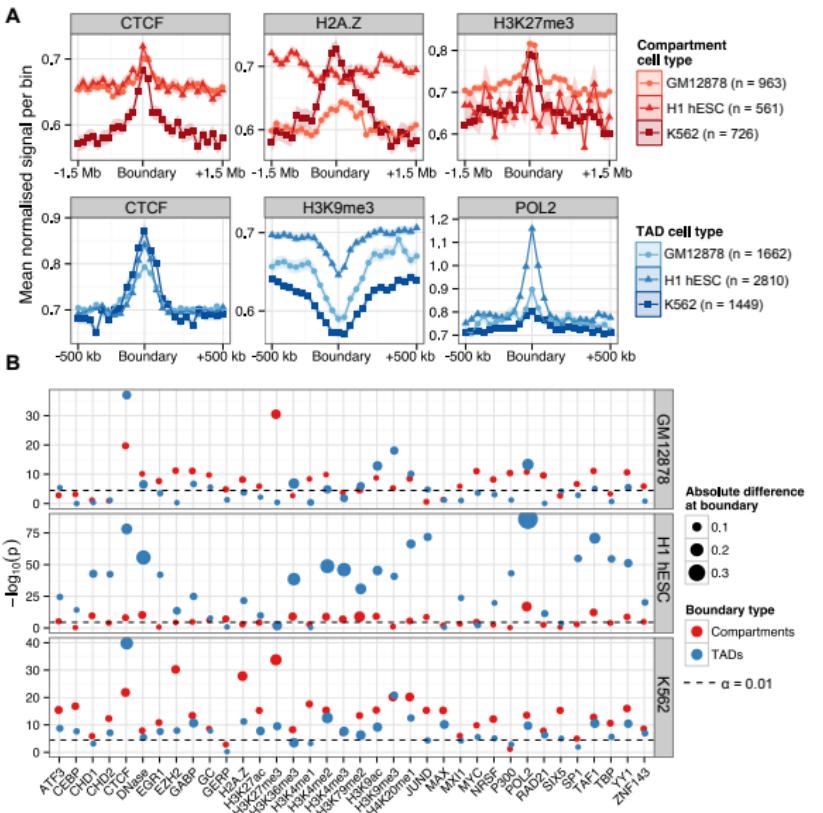
Acknowledgements

# Boundary enrichment analysis

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE



Introduction

Reanalysis of Hi-C datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

# Predicting TAD boundaries

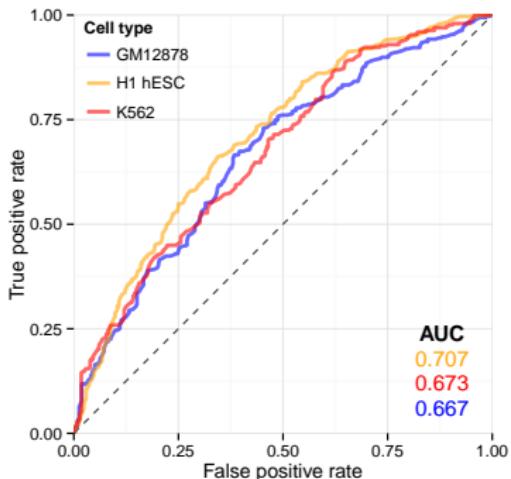
p96

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Given these enrichments exist, can we predict TAD  
boundaries in lieu of Hi-C data using epigenomic features?



Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

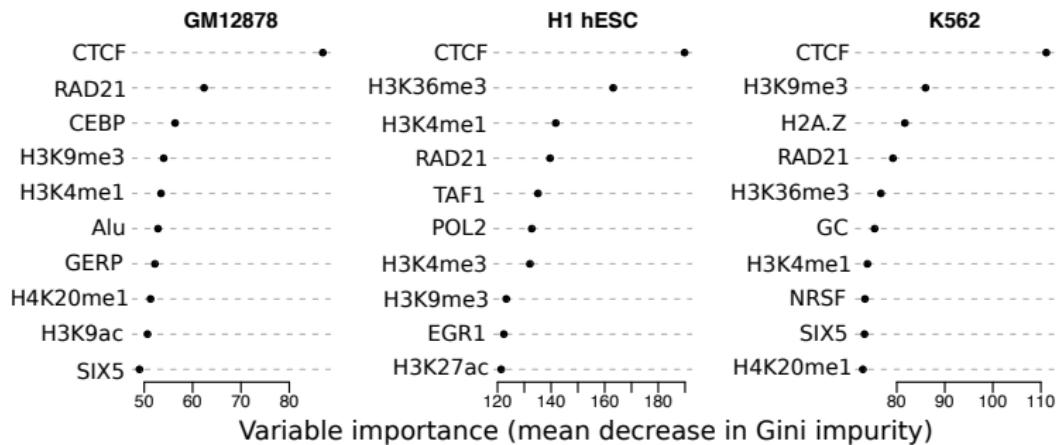
Thesis plan

Acknowledgements

# Predicting TAD boundaries

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis

In this case, more consistent and clear results from variable importance:



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

# MetaTADs

Collaborator's finding:

*Neighbour joining of interacting TADs appears to capture a novel layer of chromatin organisation: metaTADs. This attractive concept that offers a framework for fractal 3D structure and links TADs with compartments.*

My contributions:

- ▶ Analysis of MetaTAD boundary enrichments as per TADs and compartments
- ▶ Analysis of the co-incidence of MetaTAD boundaries and LAD boundaries

# MetaTAD boundary enrichments

p100

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

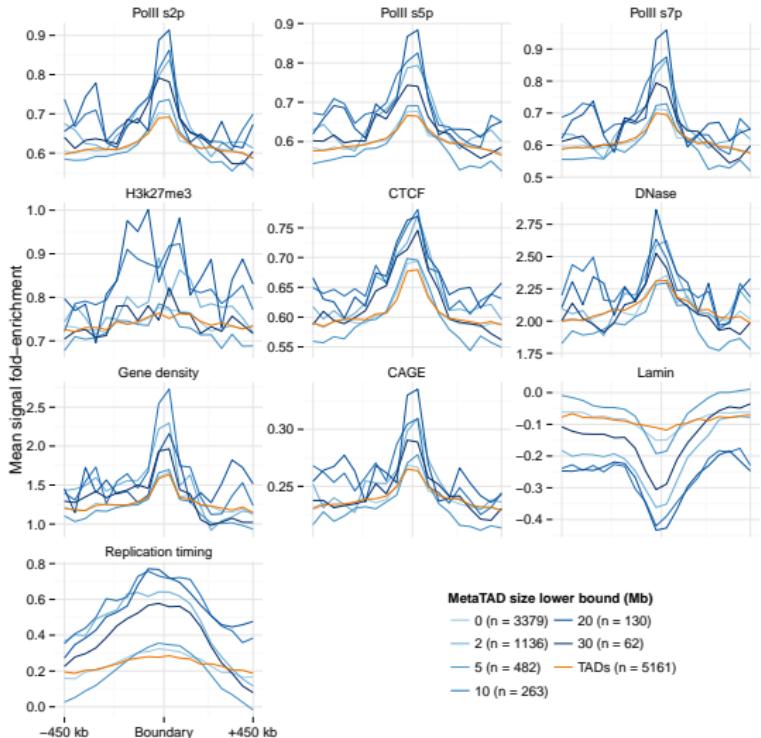
Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements



# Side project

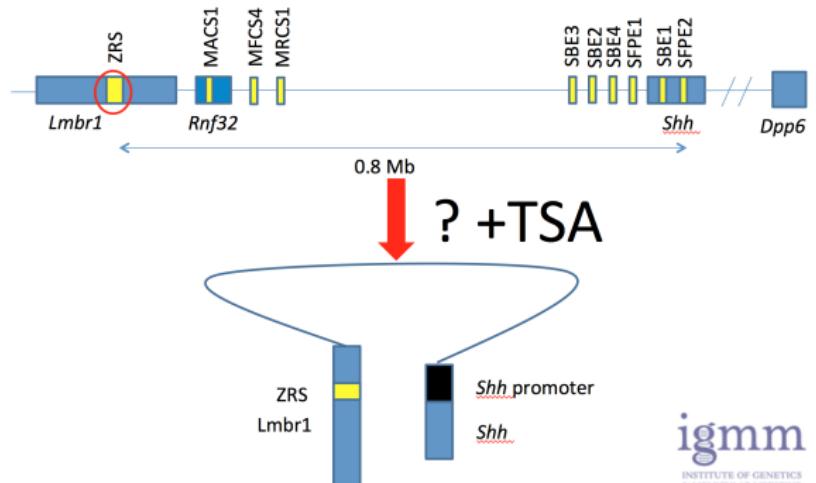
Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Collaborating with Adam Douglas (Bob Hill's group) in analysing his 3C-seq (4C) data. Their hypothesis:

## Does TSA Treatment Alter Chromatin Dynamics?



Introduction

Reanalysis of Hi-C datasets

Integrative modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

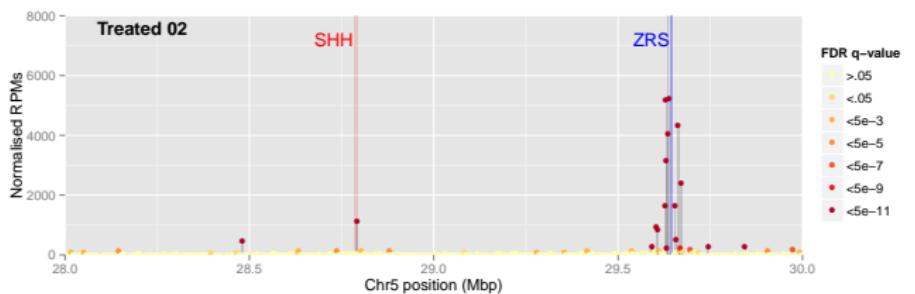
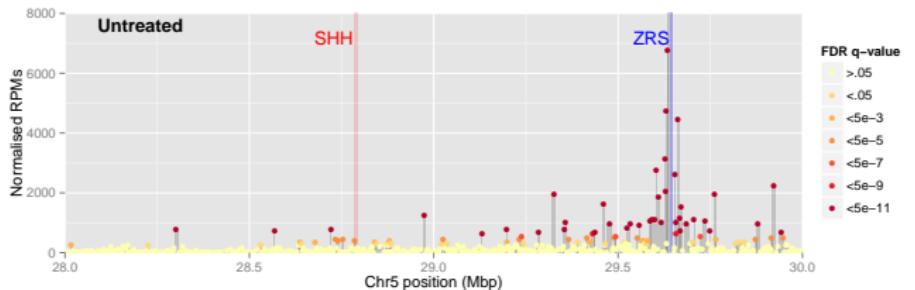
# Side project

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

## Initial results:



Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Collaborations

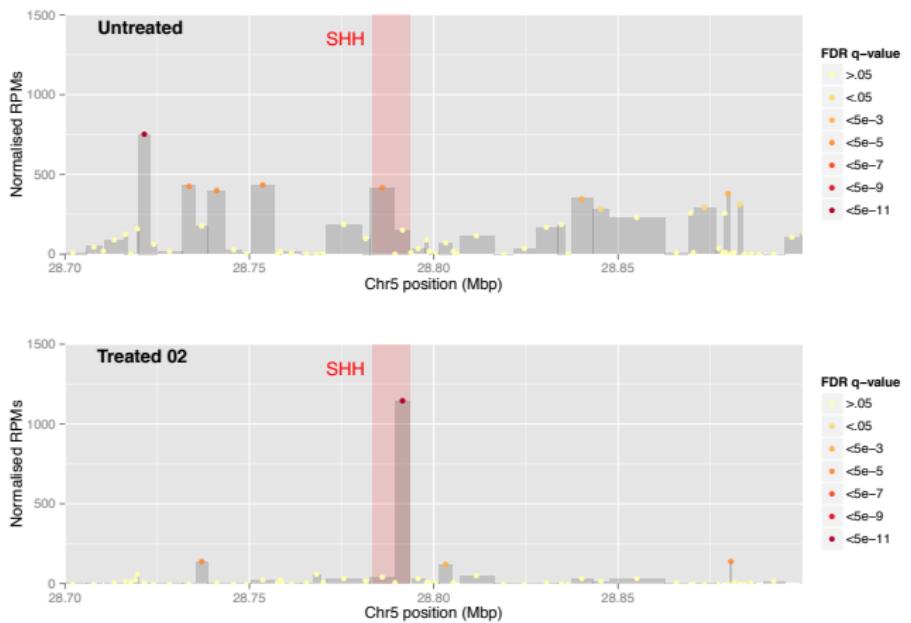
Next project

Thesis plan

Acknowledgements

# Side project

Zoomed:



Incoming: repeats, 5C / Capture-C (?), FAIRE-seq . . .

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis

igmm

INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

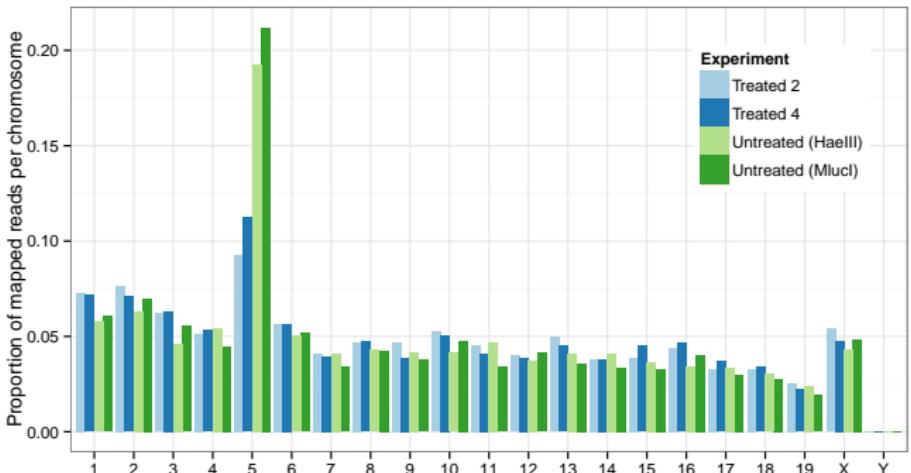
Thesis plan

Acknowledgements

# Side project

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis

All looks good except . . .



Introduction

Reanalysis of Hi-C datasets

Integrative modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

# Next project

So far haven't looked at any expression data, but CAGE available for each cell line.

Initial idea:

- ▶ Investigate TADs as “regulons”—new paper reports 20% domains act as “discrete regulatory units” with relatively homogenous epigenetic states.<sup>1</sup>

Can I find evidence for this with my data?

---

<sup>1</sup> Le Dily *et al.* (2014) Distinct structural transitions of chromatin correlate with coordinated hormone-induced gene regulation. *Genes and Development*, **28**:2151-62

# Next project

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

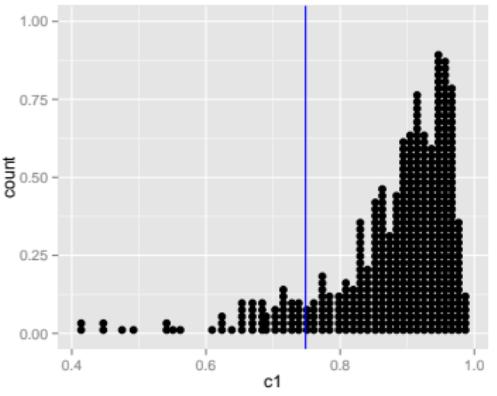
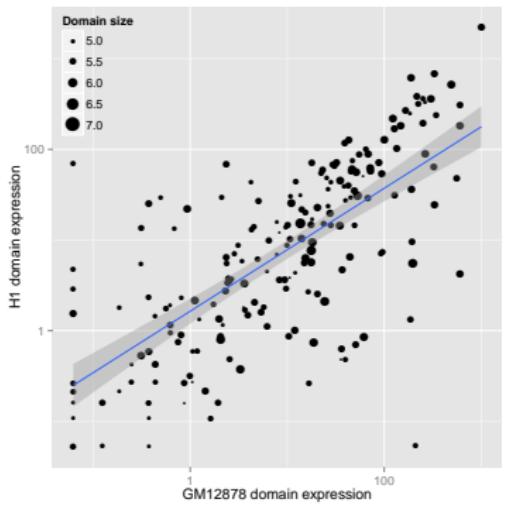
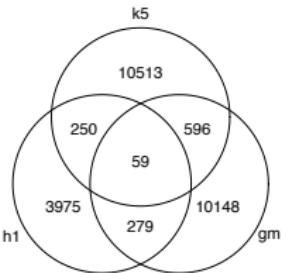
Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements



Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

# Next project

- ▶ Higher resolution Hi-C now available, better for identifying domains
- ▶ Improved methods of calling domains, where / why are some well-conserved between cell types?
- ▶ Mouse data also available

# Thesis plan p1

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

1. Introduction
2. Methods
3. **Modelling transcription and chromatin** Replicating and extending ENCODE project predicting txm output; adapt techniques to genome organisation
4. **Model dissection** How the chromatin structure models cross-apply; regularised models; variable importance
5. **Boundaries** Comparing TADs/compartments across boundaries; “super bounds”; Giemsa bands
6. ...

Introduction

Reanalysis of Hi-C datasets

Integrative modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

# Thesis plan p2

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

6. **Investigating the function of self-interacting domains** Work from future project (very early stages); possibly two chapters from the next year
7. **C-methods collaborations** Write-up Hill lab collaborations, possibly include other minor collabs
8. Discussion
9. End materials, code

Introduction

Reanalysis of Hi-C datasets

Integrative modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Collaborations

Next project

Thesis plan

Acknowledgements

Thanks to supervisors:  
Colin Semple and Stuart Aitken