

# 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

Local chromatin  
conformation

Discussion

Acknowledgements

## Introduction

# Unravelling higher order chromatin organisation through statistical analysis



INSTITUTE OF GENETICS

& MOLECULAR MEDICINE

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?

# 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

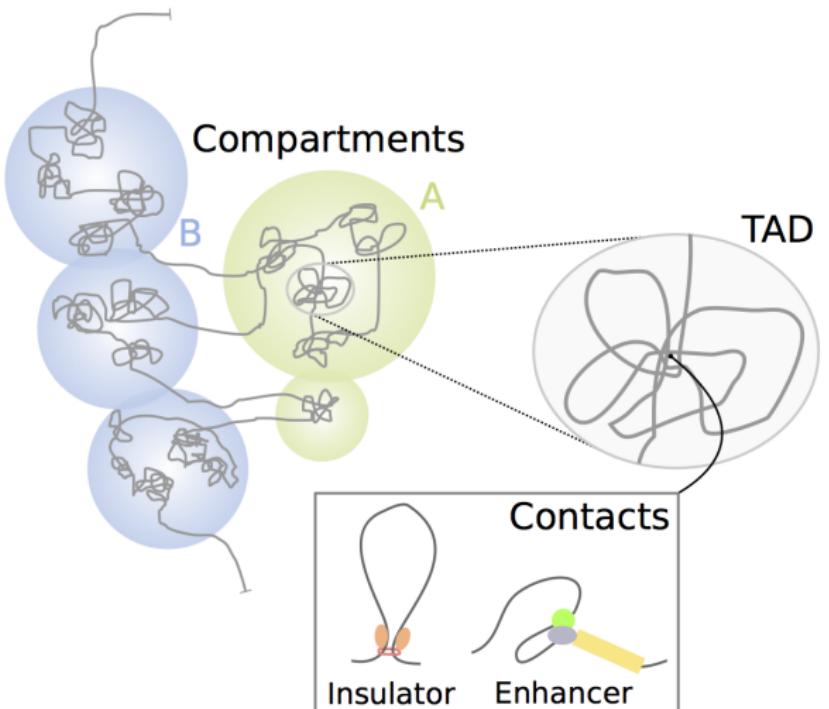
Integrative  
modelling

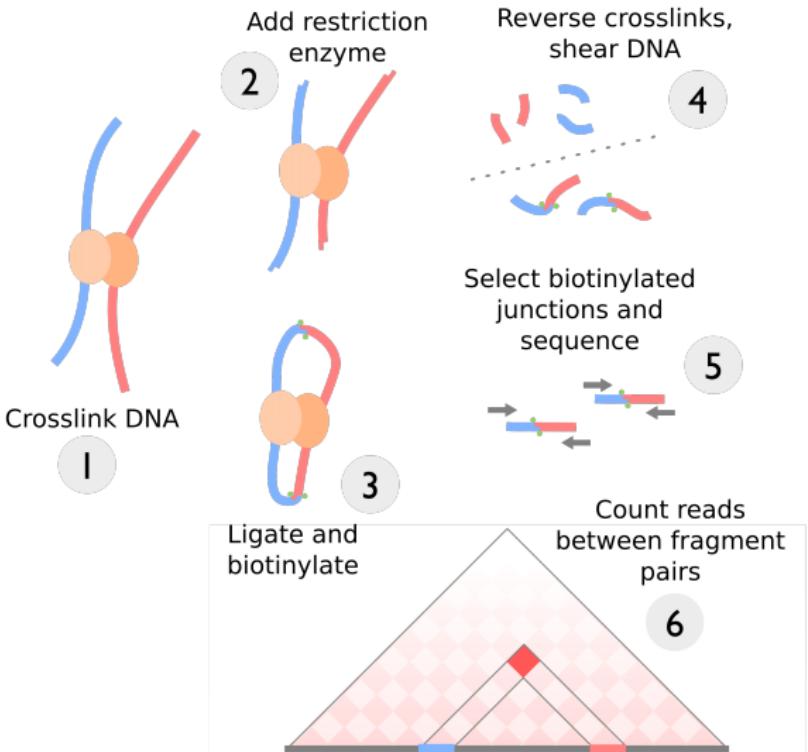
Boundaries

Local chromatin  
conformation

Discussion

Acknowledgements





Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Local chromatin  
conformation

Discussion

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

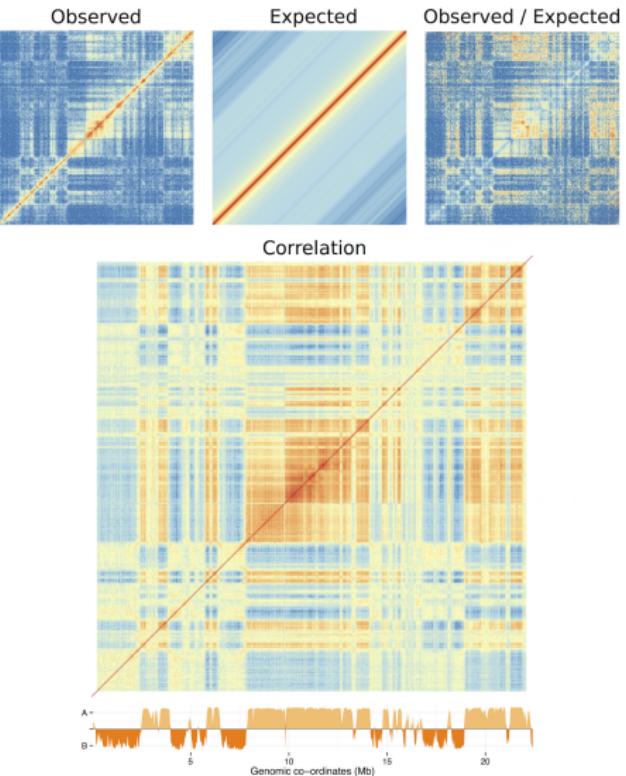
Integrative  
modelling

Boundaries

Local chromatin  
conformation

Discussion

Acknowledgements



Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Local chromatin  
conformation

Discussion

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

Local chromatin  
conformation

Discussion

Acknowledgements

## Results 1: Reanalysis of Hi-C data

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

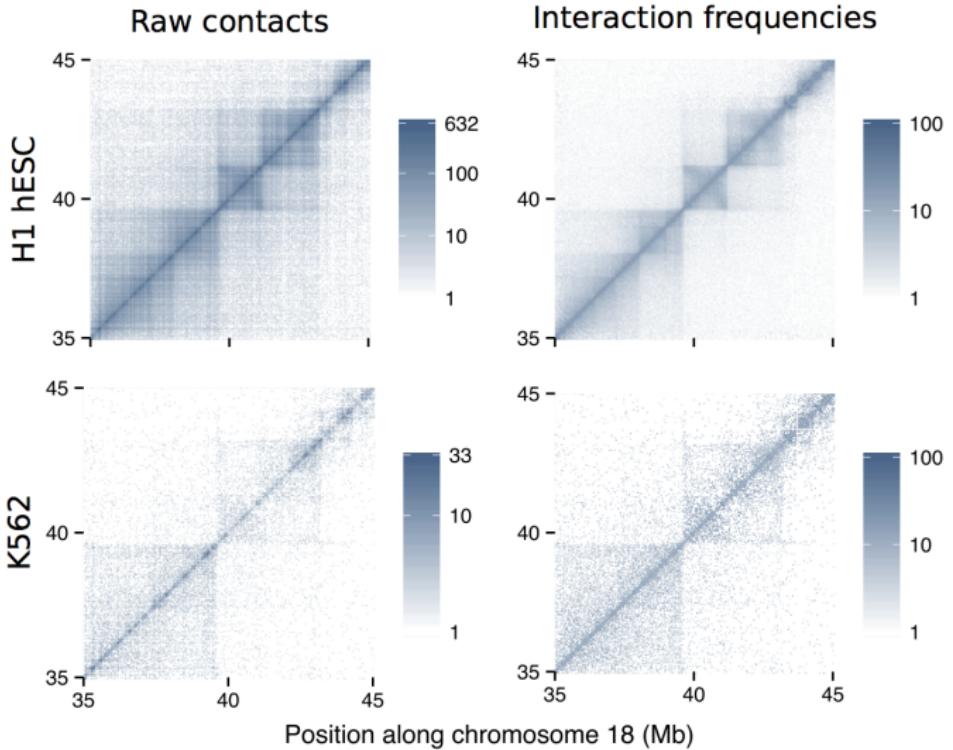
Local chromatin  
conformation

Discussion

Acknowledgements

# Reanalysis of Hi-C datasets

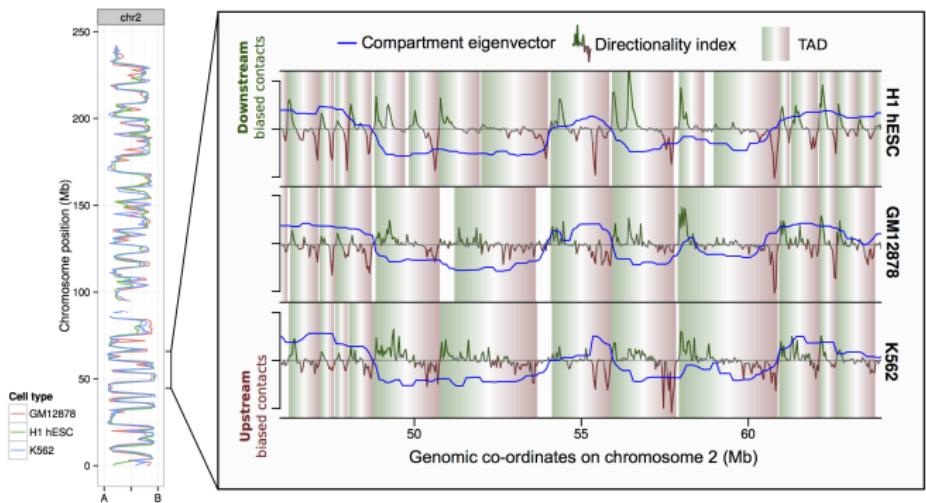
Very different sequencing depths between the input datasets:



## Reanalysis of Hi-C datasets

## Unravelling higher order chromatin organisation through statistical analysis

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



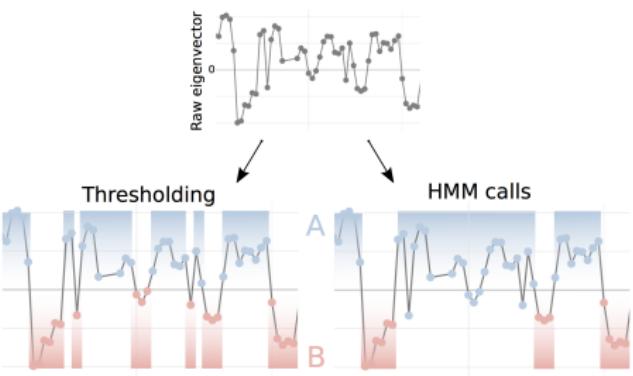
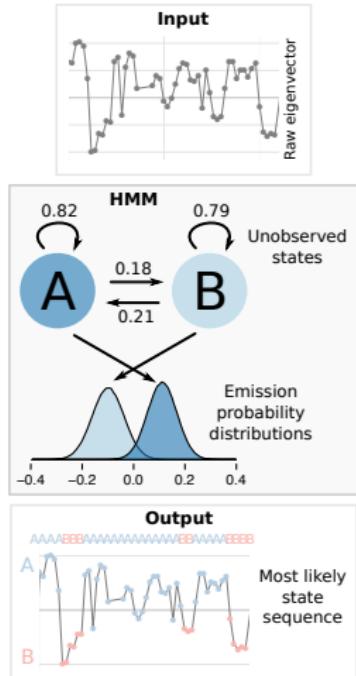
Justifies going forward with between cell-line analysis

# Improved compartment calling algorithm

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE



Introduction

Reanalysis of Hi-C datasets

Integrative modelling

Boundaries

Local chromatin conformation

Discussion

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

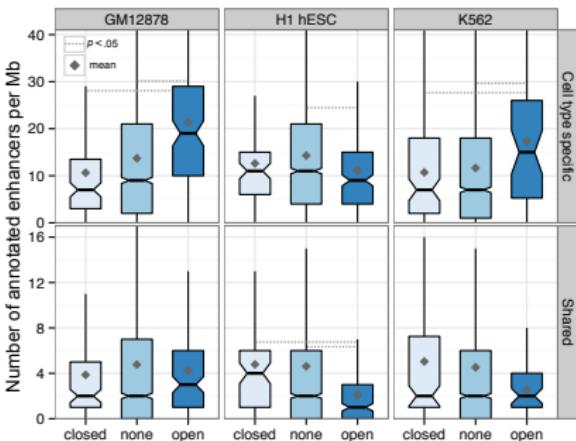
Integrative  
modelling

Boundaries

Local chromatin  
conformation

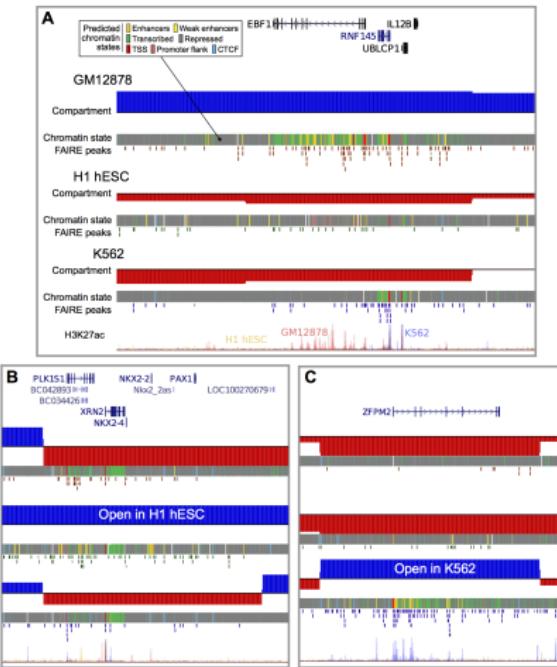
Discussion

Acknowledgements



# Regions of variable compartment structure

Some harbour genes with known cell type specific function



Introduction

Reanalysis of Hi-C datasets

Integrative modelling

Boundaries

Local chromatin conformation

Discussion

Acknowledgements

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Local chromatin  
conformation

Discussion

Acknowledgements

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

# Integrative modelling

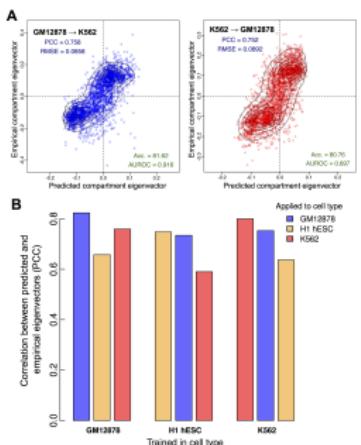
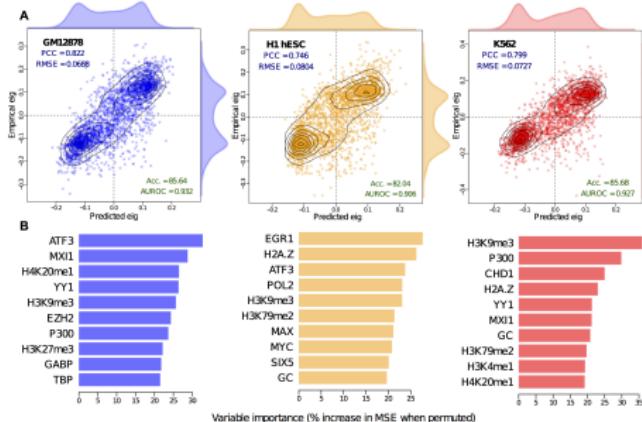
p67

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

Local chromatin  
conformation

Discussion

Acknowledgements

# Integrative modelling

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



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

Local chromatin conformation

Discussion

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

Local chromatin conformation

Discussion

Acknowledgements

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Local chromatin  
conformation

Discussion

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

Local chromatin conformation

Discussion

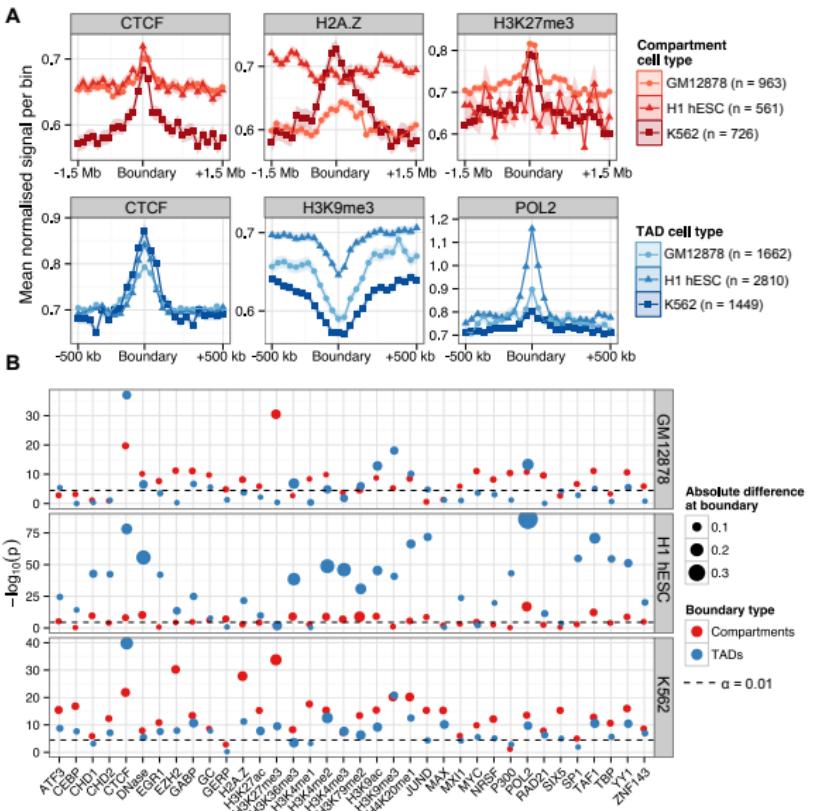
Acknowledgements

# 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

Local chromatin  
conformation

Discussion

Acknowledgements

# Predicting TAD boundaries

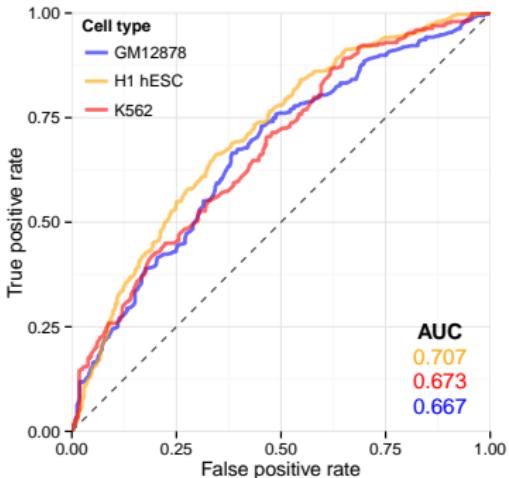
p96

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

Local chromatin  
conformation

Discussion

Acknowledgements

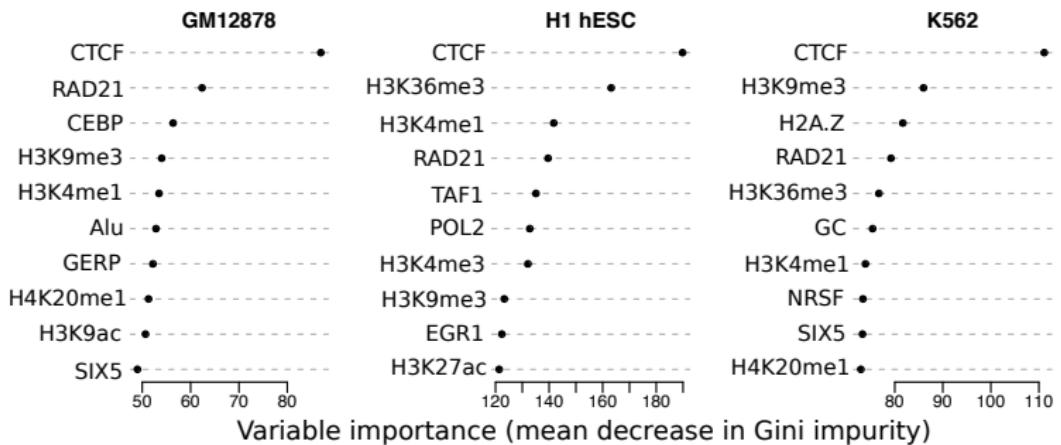
# Predicting TAD boundaries

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

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



Introduction

Reanalysis of Hi-C datasets

Integrative modelling

Boundaries

Local chromatin conformation

Discussion

Acknowledgements

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Local chromatin  
conformation

Discussion

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

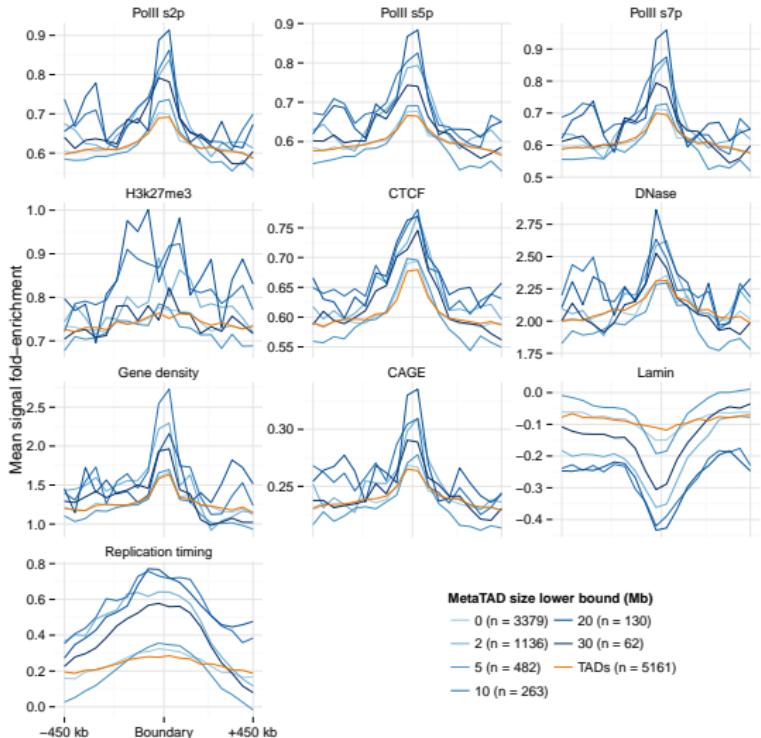
Integrative  
modelling

Boundaries

Local chromatin  
conformation

Discussion

Acknowledgements



## MetaTADs and LADs

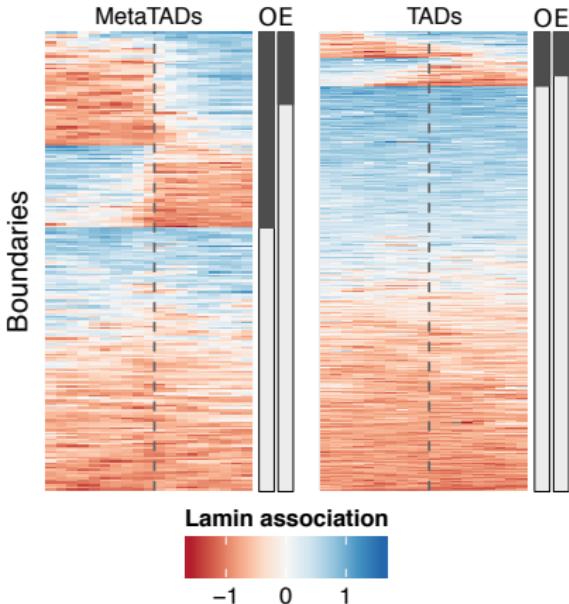
p100

# Unravelling higher order chromatin organisation through statistical analysis

igmm

INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Are lamin associated domains being captured as metaTADs?



## Introduction

## Reanalysis of Hi-C datasets

## Integrative modelling

## Boundaries

## Discussion

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Local chromatin  
conformation

Discussion

Acknowledgements

## Results 4: Local chromatin conformation

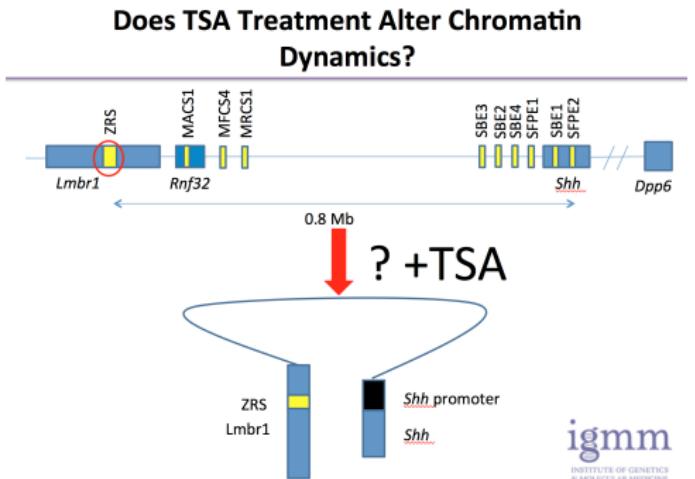
# Local chromatin conformation

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Collaboration w/ Adam Douglas (Bob Hill's group) in  
analysing his 3C-seq (4C) data. Their experimental question:



Introduction

Reanalysis of Hi-C datasets

Integrative modelling

Boundaries

Local chromatin conformation

Discussion

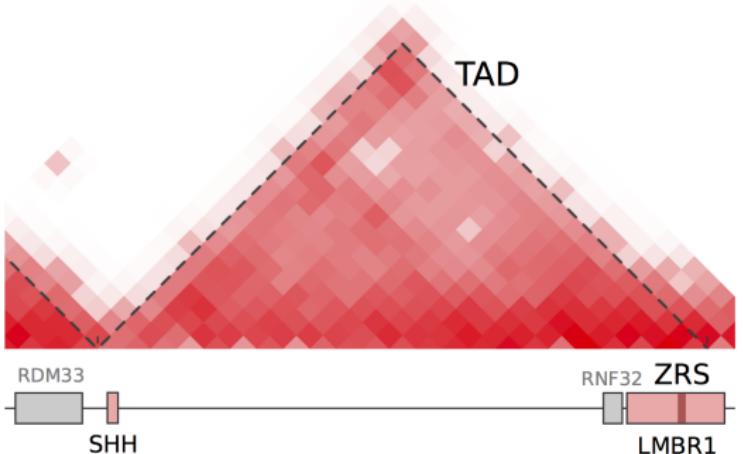
Acknowledgements

My contribution: analysis of 4C, 5C data and inference of  
3D chromatin trajectory

## Shh locus

# Unravelling higher order chromatin organisation through statistical analysis

## ZRS-Shh is a model system for long range *cis*-regulation through an enhancer–promoter contact



The Hill lab has developed an *Shh*-inducible system through TSA treatment, though exact mechanism unclear

# 4C results

p112

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Introduction

Reanalysis of Hi-C  
datasets

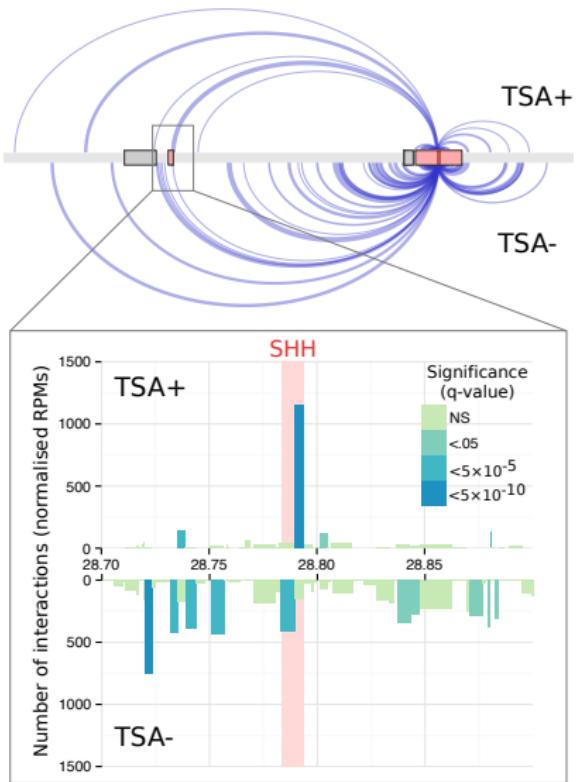
Integrative  
modelling

Boundaries

Local chromatin  
conformation

Discussion

Acknowledgements



Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Local chromatin  
conformation

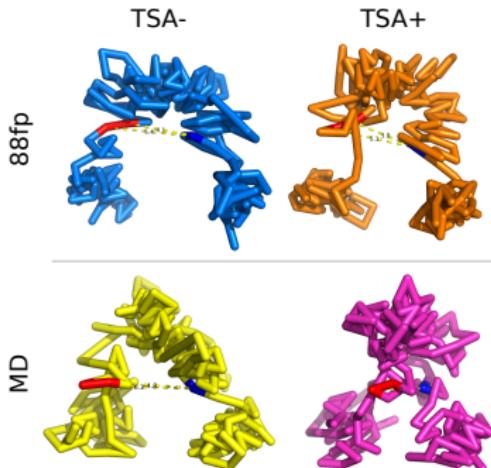
Discussion

Acknowledgements

## 5C results

p118

Used 5C data over the same region, passed to 3D  
reconstruction algorithm and measured distances



Contrary to expectations, little change in inducible cell line  
and large shift in control mandibular cell line (non-*shh*  
expressing)

# 5C results

p121

Unravelling higher  
order chromatin  
organisation  
through statistical  
analysis



INSTITUTE OF GENETICS  
& MOLECULAR MEDICINE

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

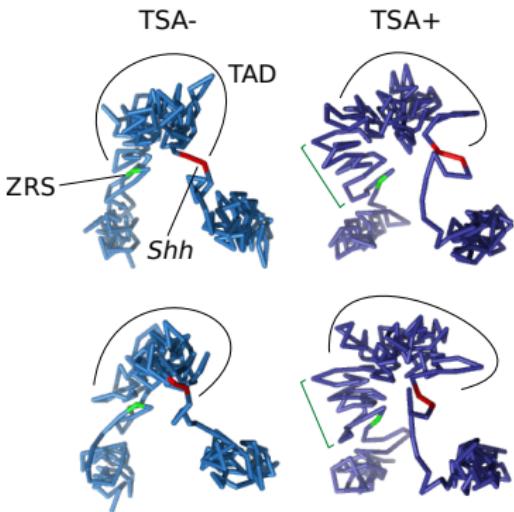
Boundaries

Local chromatin  
conformation

Discussion

Acknowledgements

Reconstructing TAD lends some credibility to the structures



HDAC causing some distension around ZRS in 14fp?

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Local chromatin  
conformation

Discussion

Acknowledgements

## Discussion and Summary

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Local chromatin  
conformation

Discussion

Acknowledgements

# Discussion

p123

I suggest some extensions and discuss some open questions:

- ▶ Could a multi-tiered model predict TADs, MetaTADs and compartments from epigenomic data?
- ▶ Are boundaries important or a side-effect of domains?  
(Likely something inbetween)

Ultimately our results agree with a model of genome organisation where large constitutive structures (compartments, LADs) anchor the genome but more local, cell-type specific interactions can then fine-tune transcriptional events at a local level

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Local chromatin  
conformation

Discussion

Acknowledgements

# Summary

Results presented in this thesis:

- ▶ Chromosome compartments can be accurately predicted from epigenomic features alone
- ▶ This link between locus-level epigenomics and higher order structure is particularly evident at boundary regions
- ▶ Further evidence for combinatorial importance of CTCF, YY1, RAD21 in genome organisation
- ▶ MetaTADs appear a useful and biologically-relevant novel organisational strata
- ▶ TSA inducible *Shh* expression intimately linked with chromosome conformation, through details still unclear

More generally:

- ▶ Data recycling: most of this thesis done with publicly-available data from different sources
- ▶ Machine learning offers useful tools to dissect complex biological systems

Introduction

Reanalysis of Hi-C  
datasets

Integrative  
modelling

Boundaries

Local chromatin  
conformation

Discussion

Acknowledgements

Thanks to supervisors:  
Colin Semple and Stuart Aitken

Collaborators:  
Hill lab (University of Edinburgh)  
Pombo lab (Max Delbrück center)