Bioinformatic analysis of complex, high-throughput genomic and epigenomic data in the context of $CD4^+$ T-cell differentiation and diagnosis and treatment of transplant rejection

Ryan C. Thompson
Su Lab
The Scripps Research Institute

October 24, 2019

Organ transplants are a life-saving treatment

- 36,528 transplants performed in the USA in 2018[1]

[1] organdonor.gov

Organ transplants are a life-saving treatment

- 36,528 transplants performed in the USA in 2018[1]
- 100 transplants every day!

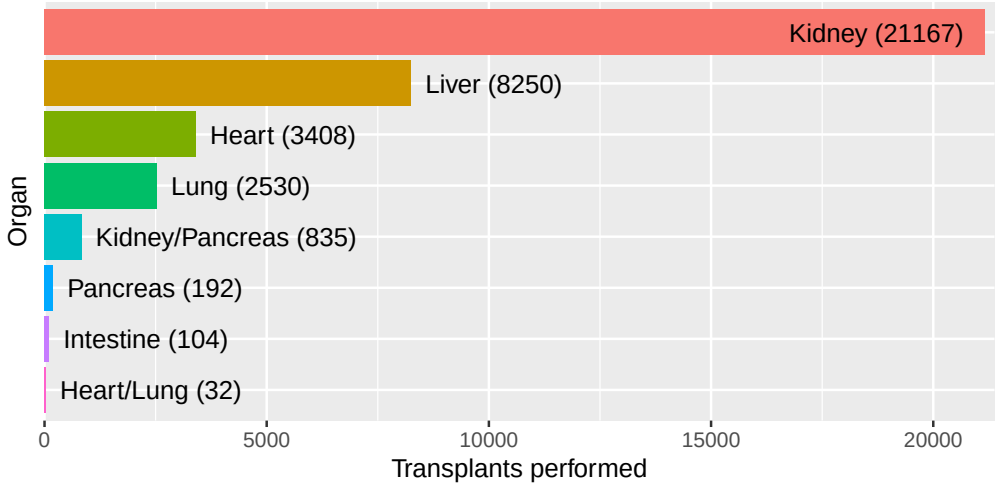
[1] organdonor.gov

Organ transplants are a life-saving treatment

- 36,528 transplants performed in the USA in 2018[1]
- 100 transplants every day!
- Over 113,000 people on the national transplant waiting list as of July 2019

[1] organdonor.gov

Organ donation statistics for the USA in 2018[2]



[2] organdonor.gov

Types of grafts

A graft is categorized based on the relationship between donor and recipient:

Types of grafts

A graft is categorized based on the relationship between donor and recipient:

- **Autograft:** Donor and recipient are the *same individual*

Types of grafts

A graft is categorized based on the relationship between donor and recipient:

- **Autograft:** Donor and recipient are the *same individual*
- **Allograft:** Donor and recipient are *different individuals* of the *same species*

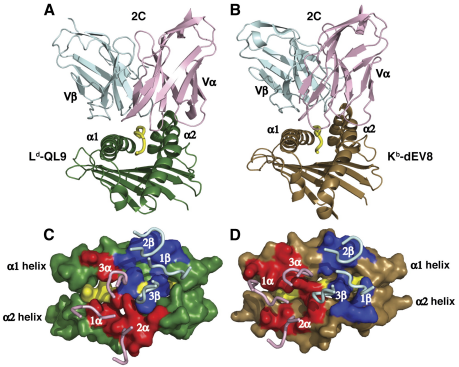
Types of grafts

A graft is categorized based on the relationship between donor and recipient:

- **Autograft:** Donor and recipient are the *same individual*
- **Allograft:** Donor and recipient are *different individuals* of the *same species*
- **Xenograft:** Donor and recipient are *different species*

Recipient T-cells reject allogenic MHCs

- TCR binds to both antigen *and* MHC surface

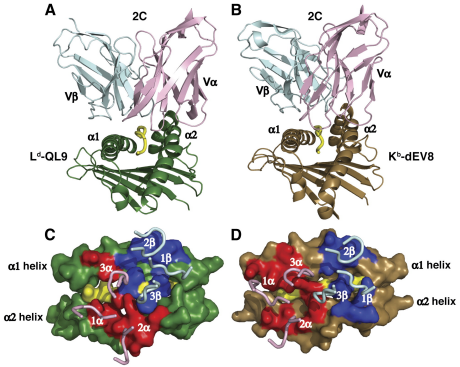


TCR binding to self (right) and allogenic (left) MHC[3]

[3] Colf, Bankovich, et al. "How a Single T Cell Receptor Recognizes Both Self and Foreign MHC". In: Cell (2007)

Recipient T-cells reject allogenic MHCs

- TCR binds to both antigen *and* MHC surface
- HLA genes encoding MHC proteins are highly polymorphic

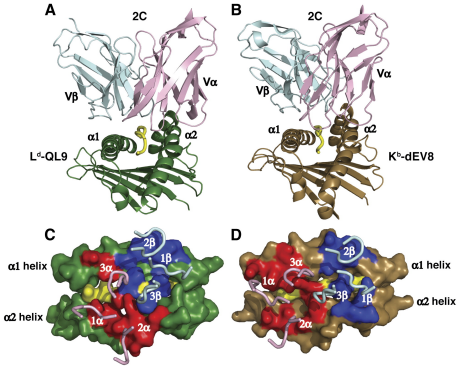


TCR binding to self (right) and allogenic (left) MHC[3]

[3] Colf, Bankovich, et al. "How a Single T Cell Receptor Recognizes Both Self and Foreign MHC". In: Cell (2007)

Recipient T-cells reject allogenic MHCs

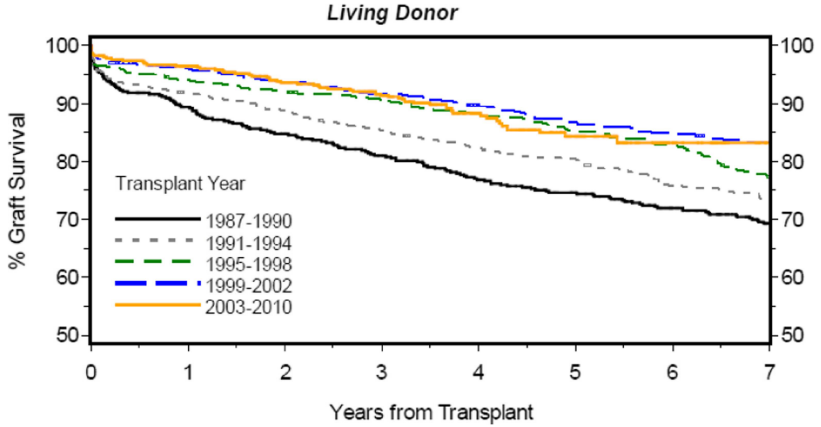
- TCR binds to both antigen *and* MHC surface
- HLA genes encoding MHC proteins are highly polymorphic
- Variants in donor MHC can trigger the same T-cell response as a foreign antigen



TCR binding to self (right) and allogenic (left) MHC[3]

[3] Colf, Bankovich, et al. "How a Single T Cell Receptor Recognizes Both Self and Foreign MHC". In: Cell (2007)

Allograft rejection is a major long-term problem



Kidney allograft survival rates in children by transplant year[4]

[4] Kim & Marks (2014)

Rejection is treated with immune suppressive drugs

- Graft recipient must take immune suppressive drugs indefinitely

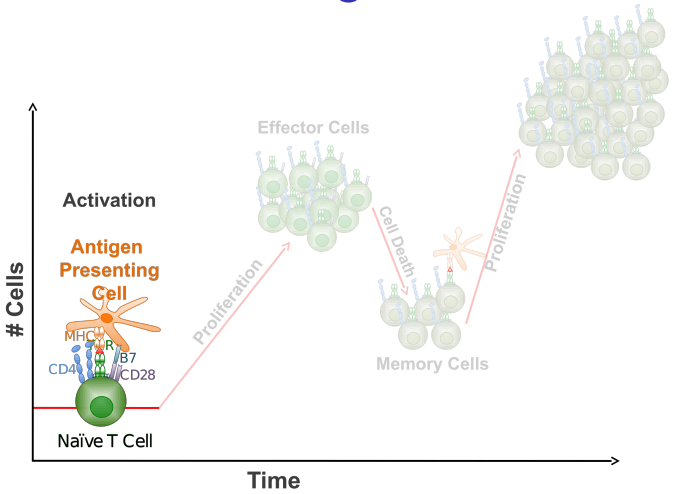
Rejection is treated with immune suppressive drugs

- Graft recipient must take immune suppressive drugs indefinitely
- Graft is monitored for rejection and dosage adjusted over time

Rejection is treated with immune suppressive drugs

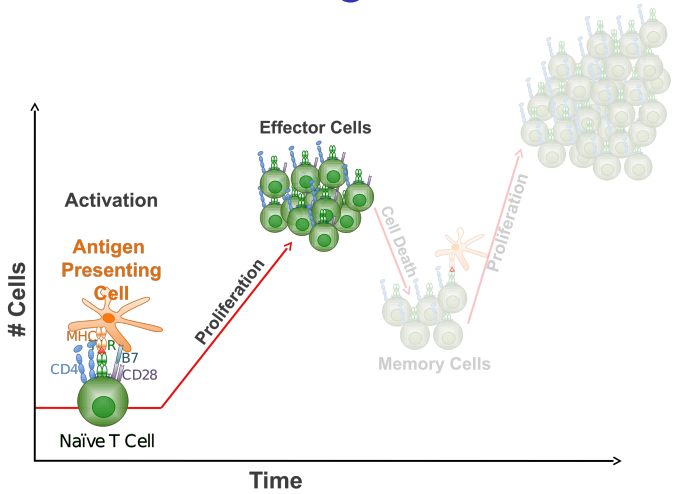
- Graft recipient must take immune suppressive drugs indefinitely
- Graft is monitored for rejection and dosage adjusted over time
- Immune suppression is a delicate balance: too much and too little are both problematic.

Memory cells: faster, stronger, and more independent



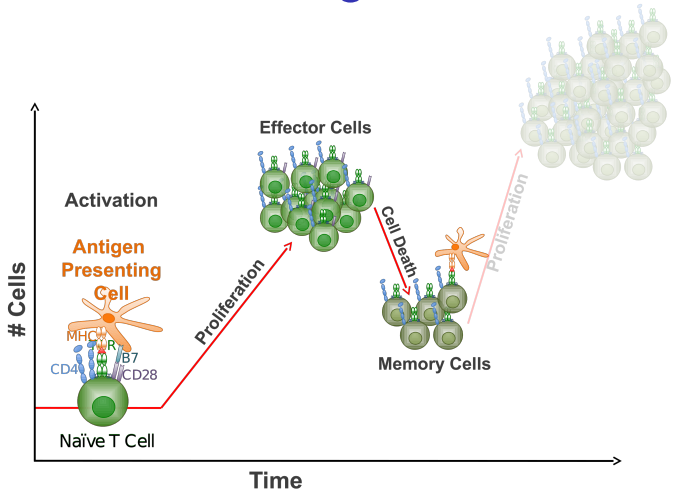
Naïve T-cell activated by APC

Memory cells: faster, stronger, and more independent



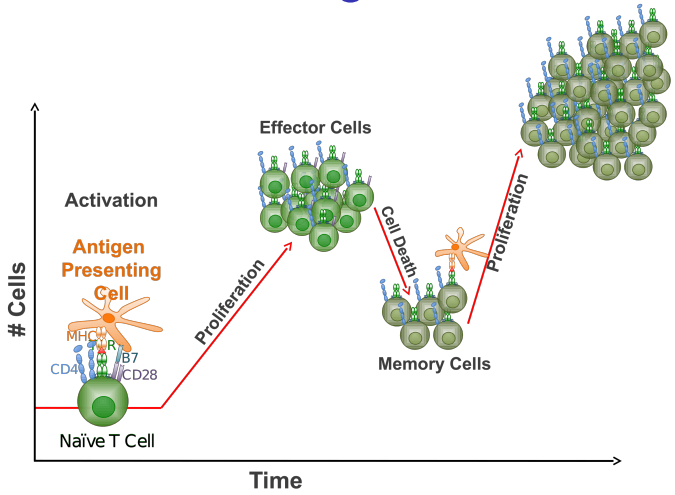
Naïve T-cell differentiates and proliferates into effector T-cells

Memory cells: faster, stronger, and more independent



Post-infection, some effectors cells remain as memory cells

Memory cells: faster, stronger, and more independent



Memory T-cells respond more strongly to activation

3 problems relating to transplant rejection

1. How are memory cells different from naïve?

2. How can we diagnose rejection noninvasively?

3. How can we evaluate effects of a rejection treatment?

3 problems relating to transplant rejection

1. How are memory cells different from naïve?

Genome-wide epigenetic analysis of H3K4 and H3K27 methylation in naïve and memory $CD4^+$ T-cell activation

2. How can we diagnose rejection noninvasively?

3. How can we evaluate effects of a rejection treatment?

3 problems relating to transplant rejection

1. How are memory cells different from naïve?

Genome-wide epigenetic analysis of H3K4 and H3K27 methylation in naïve and memory $CD4^+$ T-cell activation

2. How can we diagnose rejection noninvasively?

Improving array-based diagnostics for transplant rejection by optimizing data preprocessing

3. How can we evaluate effects of a rejection treatment?

3 problems relating to transplant rejection

1. How are memory cells different from naïve?

Genome-wide epigenetic analysis of H3K4 and H3K27 methylation in naïve and memory $CD4^+$ T-cell activation

2. How can we diagnose rejection noninvasively?

Improving array-based diagnostics for transplant rejection by optimizing data preprocessing

3. How can we evaluate effects of a rejection treatment?

Globin-blocking for more effective blood RNA-seq analysis in primate animal model for experimental graft rejection treatment

Today's focus

1. How are memory cells different from naïve?

Genome-wide epigenetic analysis of H3K4 and H3K27 methylation in naïve and memory $CD4^+$ T-cell activation

We need a better understanding of immune memory

- Cell surface markers fairly well-characterized
- But internal mechanisms poorly understood

We need a better understanding of immune memory

- Cell surface markers fairly well-characterized
- But internal mechanisms poorly understood

Hypothesis: Epigenetic regulation of gene expression through histone modification is involved in $CD4^+$ T-cell activation and memory.

Which histone marks are we looking at?

Which histone marks are we looking at?

- **H3K4me3:** "activating" mark associated with active transcription

Which histone marks are we looking at?

- **H3K4me3:** "activating" mark associated with active transcription
- **H3K4me2:** Correlated with H3K4me3, hypothesized "poised" state

Which histone marks are we looking at?

- **H3K4me3:** "activating" mark associated with active transcription
- **H3K4me2:** Correlated with H3K4me3, hypothesized "poised" state
- **H3K27me3:** "repressive" mark associated with inactive genes

Which histone marks are we looking at?

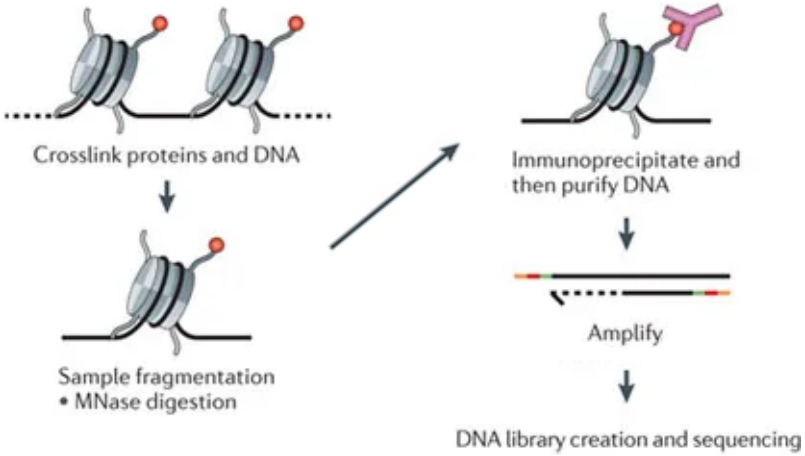
- **H3K4me3:** "activating" mark associated with active transcription
- **H3K4me2:** Correlated with H3K4me3, hypothesized "poised" state
- **H3K27me3:** "repressive" mark associated with inactive genes

Which histone marks are we looking at?

- **H3K4me3:** "activating" mark associated with active transcription
- **H3K4me2:** Correlated with H3K4me3, hypothesized "poised" state
- **H3K27me3:** "repressive" mark associated with inactive genes

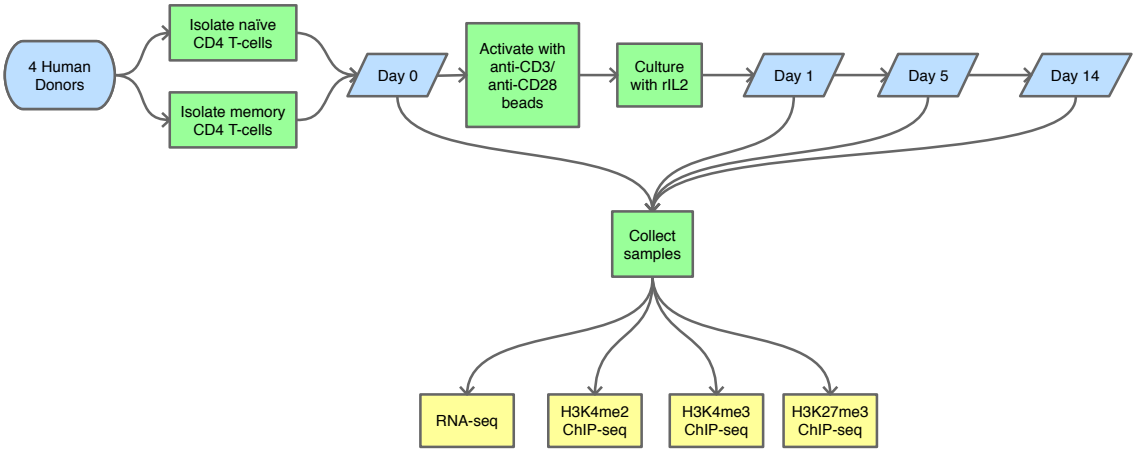
All involved in T-cell differentiation, but activation dynamics unexplored

ChIP-seq measures DNA bound to marked histones[5]



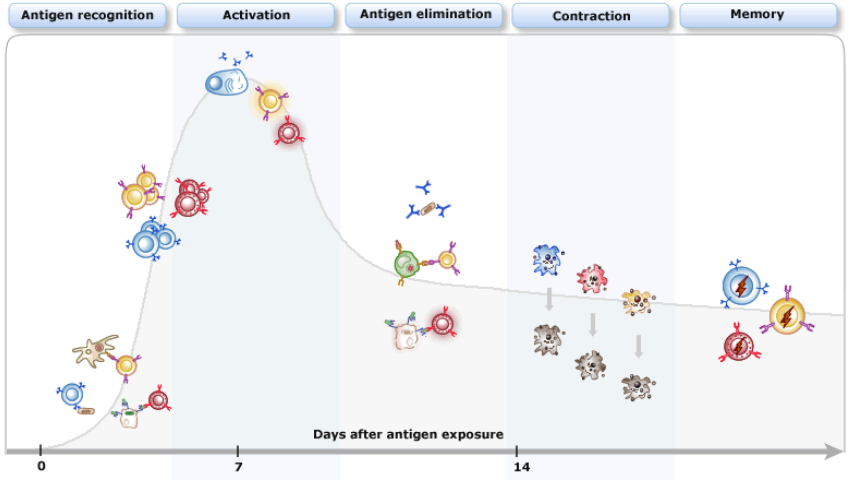
[5]Furey (2012)

Experimental design

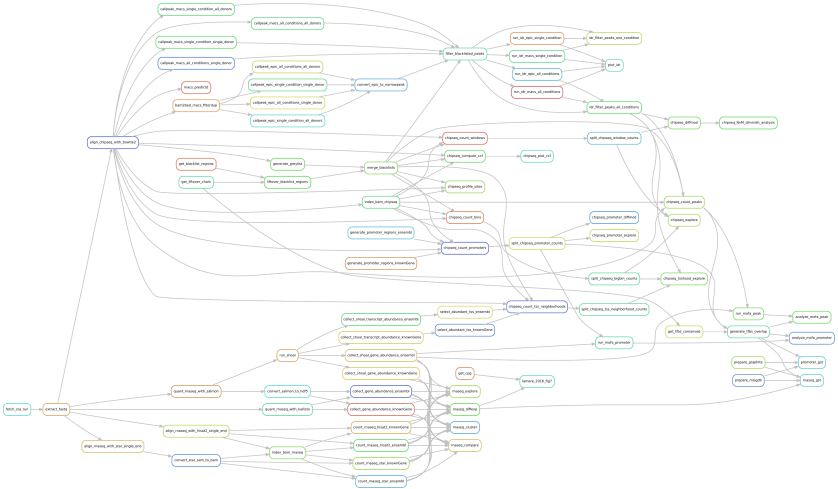


Data generated by Sarah Lamere, published in GEO as GSE73214

Time points capture phases of immune response



A few intermediate analysis steps are required



Questions to focus on

1. How do we define the "promoter region" for each gene?

Questions to focus on

1. How do we define the "promoter region" for each gene?
2. How do these histone marks behave in promoter regions?

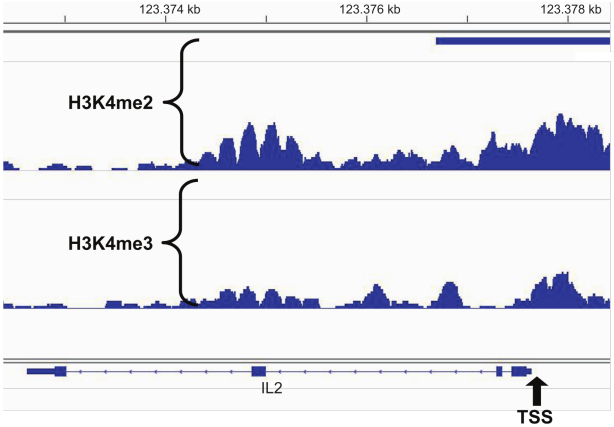
Questions to focus on

1. How do we define the "promoter region" for each gene?
2. How do these histone marks behave in promoter regions?
3. What can these histone marks tell us about T-cell activation and differentiation?

First question

How do we define the “promoter region” for each gene?

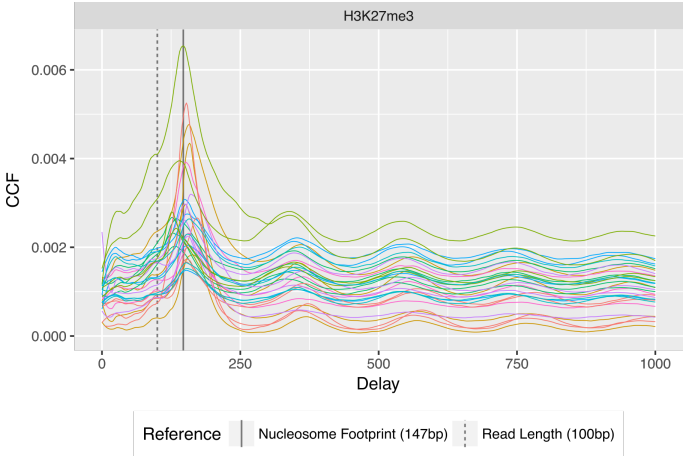
Histone modifications occur on consecutive histones



ChIP-seq coverage in IL2 gene[6]

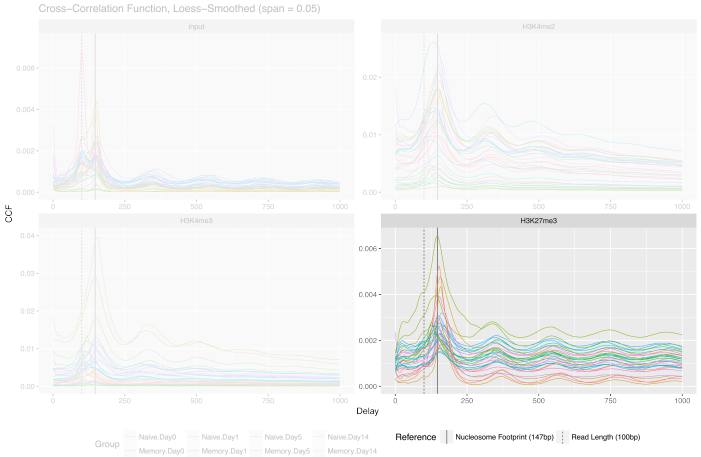
[6]Sarah LaMere. Ph.D. thesis (2015).

Histone modifications occur on consecutive histones



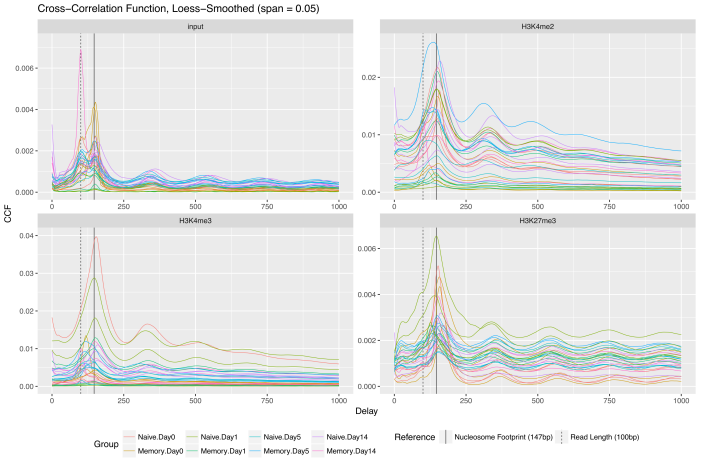
Strand cross-correlation plots show histone-sized wave pattern

Histone modifications occur on consecutive histones



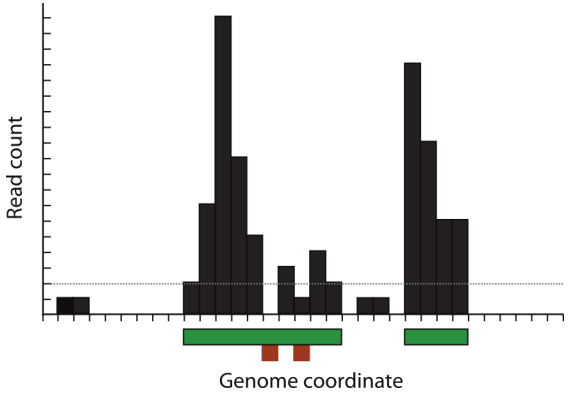
Strand cross-correlation plots show histone-sized wave pattern

Histone modifications occur on consecutive histones



Strand cross-correlation plots show histone-sized wave pattern

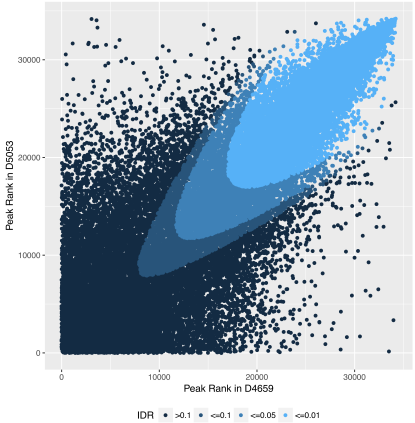
SICER identifies enriched regions across the genome



Finding "islands" of coverage with SICER[7]

[7] Zang et al. (2009)

IDR identifies *reproducible* enriched regions



Example irreproducible discovery rate[8] score consistency plot

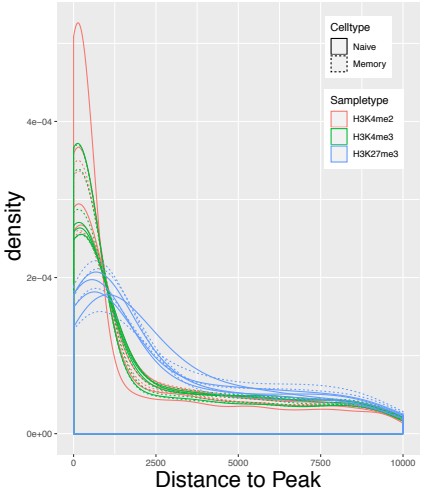
[8] Li et al. (2011)

Finding enriched regions across the genome

Histone Mark	# Peaks	Mean peak width	genome coverage	FRiP
H3K4me2	14,965	3,970	1.92%	14.2%
H3K4me3	6,163	2,946	0.588%	6.57%
H3K27me3	18,139	18,967	11.1%	22.5%

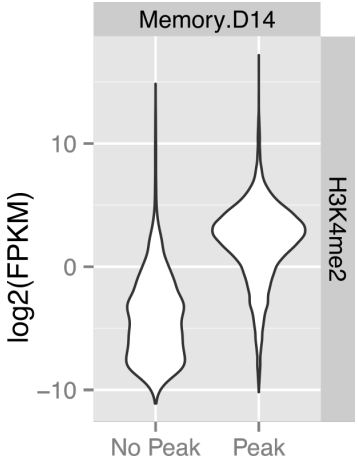
Peak-calling summary statistics

Each histone mark has an “effective promoter radius”



Enrichment of peaks near promoters

Peaks in promoters correlate with gene expression

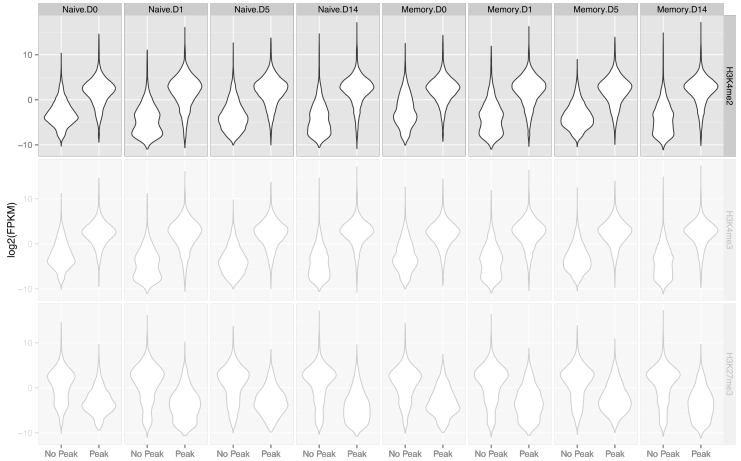


Expression distributions of genes with and without promoter peaks

Peaks in promoters correlate with gene expression

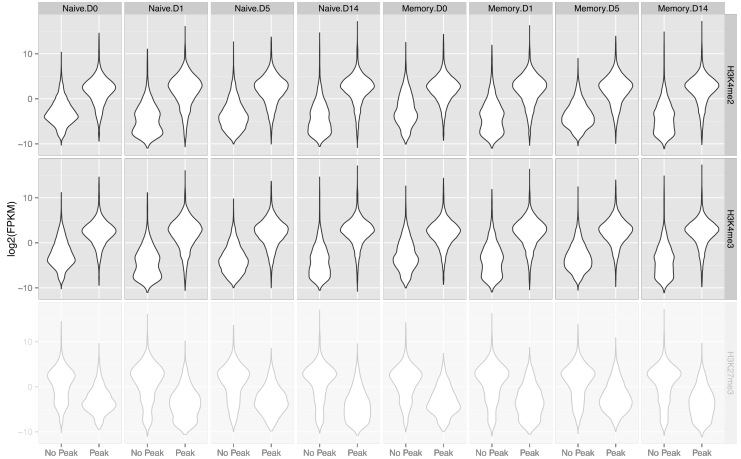
Expression distributions of genes with and without promoter peaks

Peaks in promoters correlate with gene expression



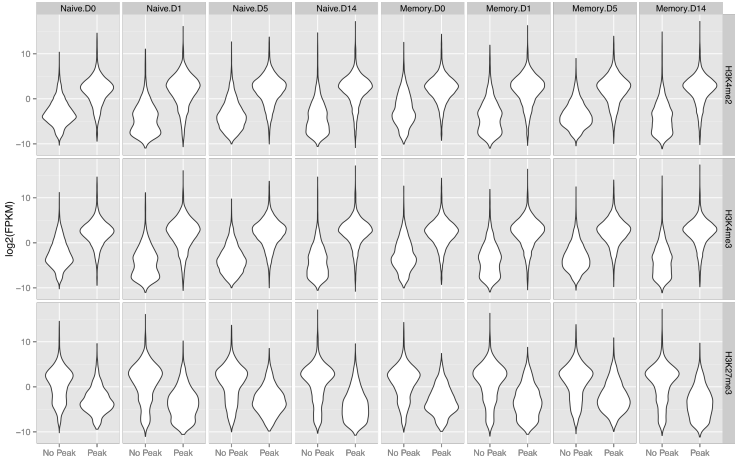
Expression distributions of genes with and without promoter peaks

Peaks in promoters correlate with gene expression



Expression distributions of genes with and without promoter peaks

Peaks in promoters correlate with gene expression



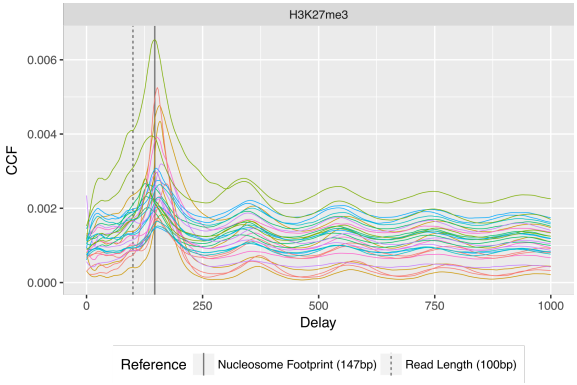
Expression distributions of genes with and without promoter peaks

First question

How do we define the “promoter region” for each gene?

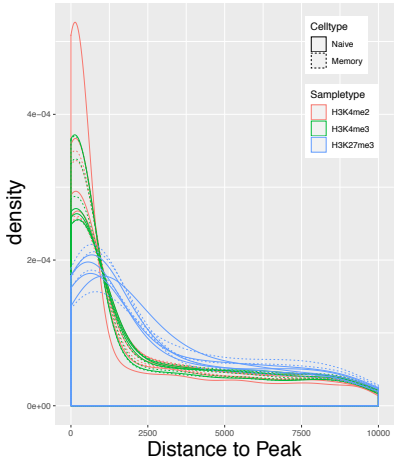
Answer: Define the promoter region empirically!

- H3K4me2, H3K4me3, and H3K27me3 occur in broad regions across the genome
- Enriched regions occur more commonly near promoters
- Each histone mark has its own "effective promoter radius"
- Presence or absence of a peak within this radius is correlated with gene expression



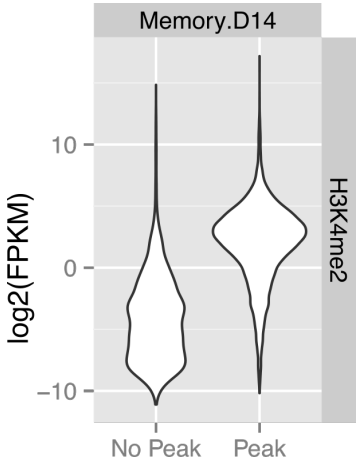
Answer: Define the promoter region empirically!

- H3K4me2, H3K4me3, and H3K27me3 occur in broad regions across the genome
- Enriched regions occur more commonly near promoters
- Each histone mark has its own "effective promoter radius"
- Presence or absence of a peak within this radius is correlated with gene expression



Answer: Define the promoter region empirically!

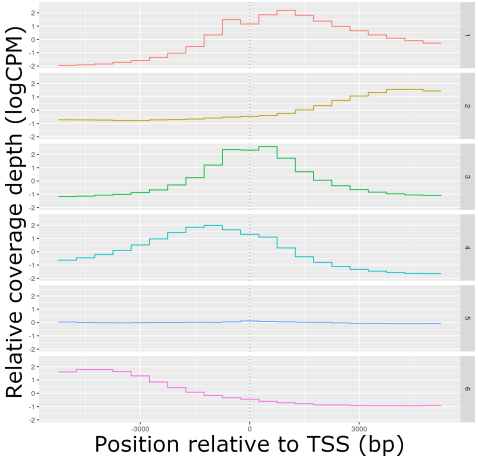
- H3K4me2, H3K4me3, and H3K27me3 occur in broad regions across the genome
- Enriched regions occur more commonly near promoters
- Each histone mark has its own "effective promoter radius"
- Presence or absence of a peak within this radius is correlated with gene expression



Next question

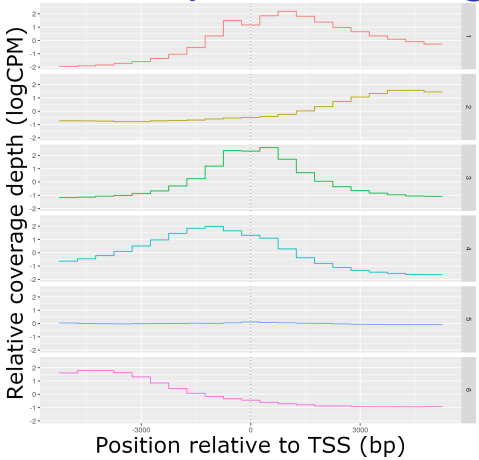
How do these histone marks behave in promoter regions?

H3K4me2 promoter neighborhood K-means clusters



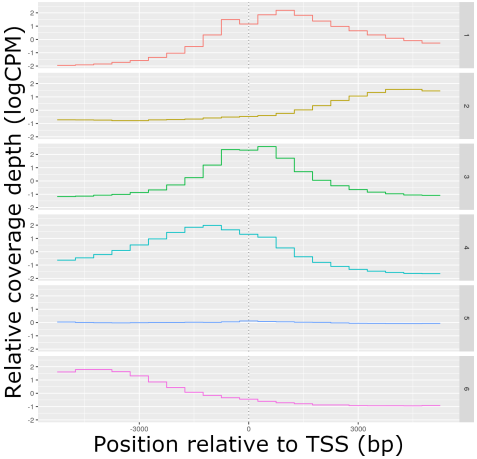
Cluster means for H3K4me2

H3K4me2 promoter neighborhood K-means clusters

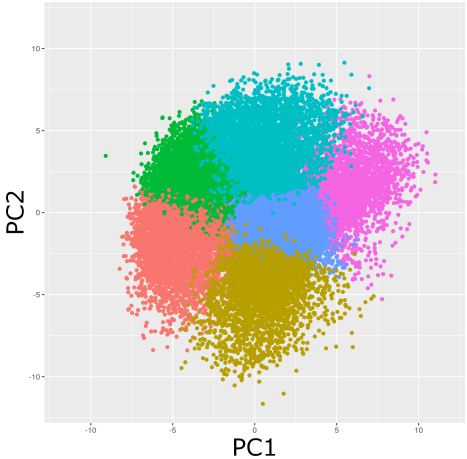


Cluster means for H3K4me2

H3K4me2 cluster PCA shows a semicircular “fan”

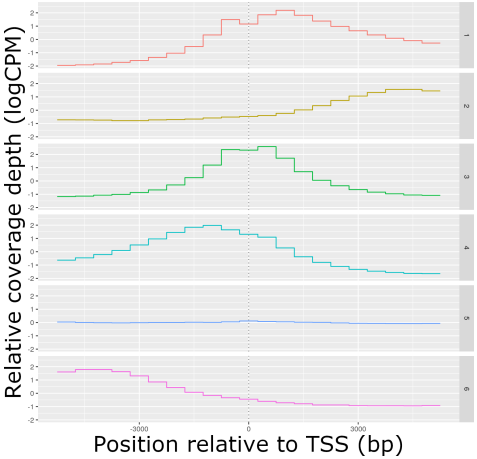


Cluster means for H3K4me2

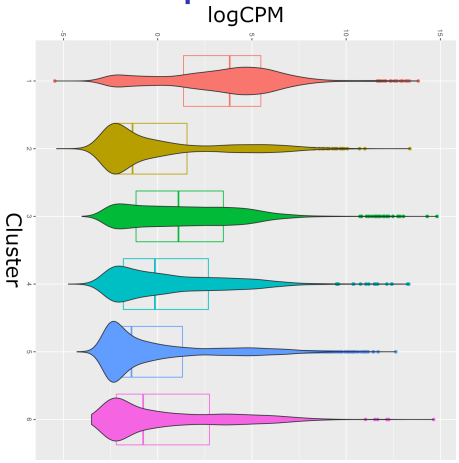


PCA plot of promoters

H3K4me2 near TSS correlates with expression

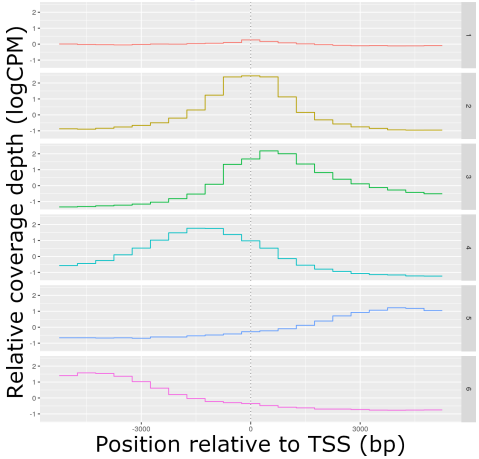


Cluster means for H3K4me2

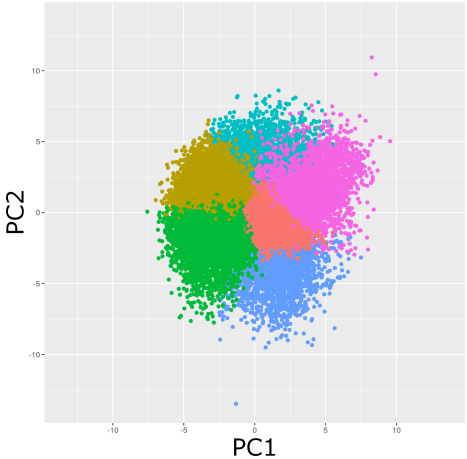


Cluster expression distributions

H3K4me3 pattern is similar to H3K4me2

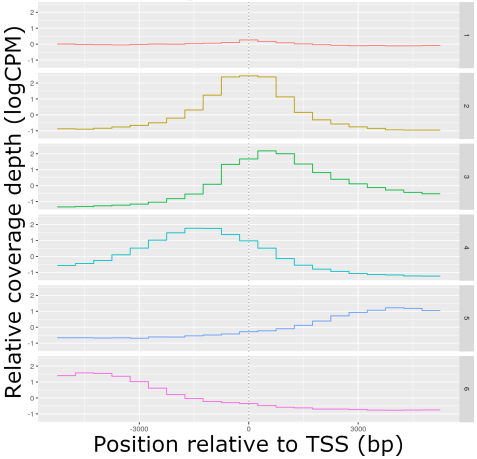


Cluster means for H3K4me3

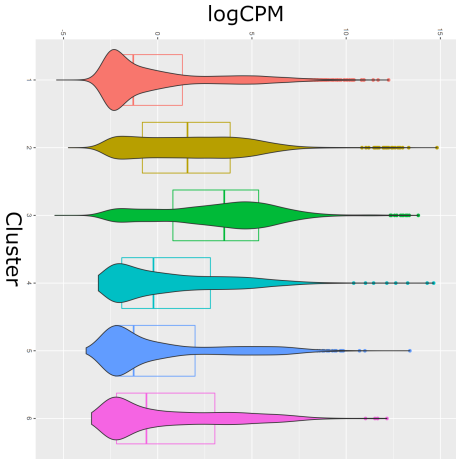


PCA plot of promoters

H3K4me3 pattern is similar to H3K4me2

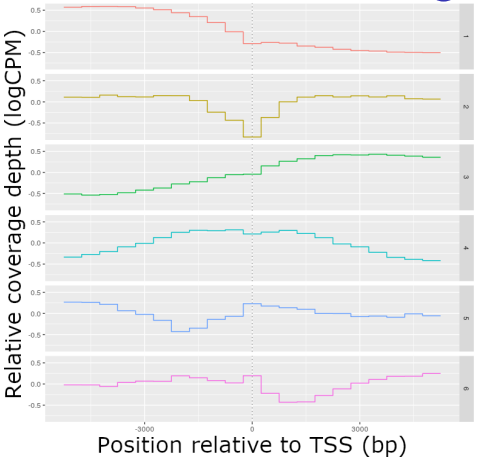


Cluster means for H3K4me3

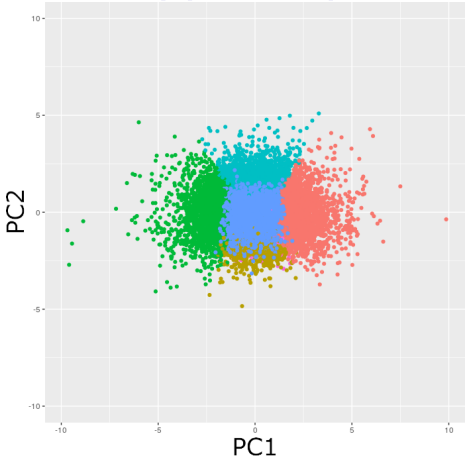


Cluster expression distributions

H3K27me3 clusters organize into 3 opposed pairs

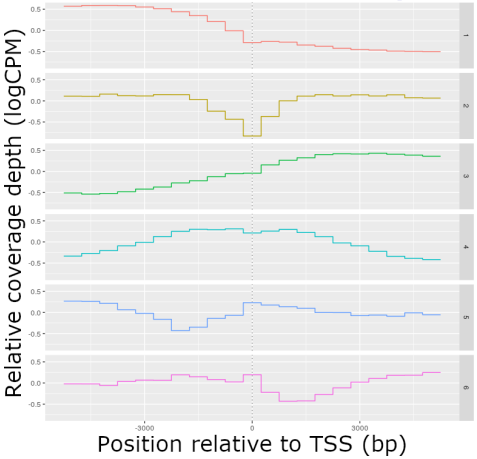


Cluster means for H3K27me3

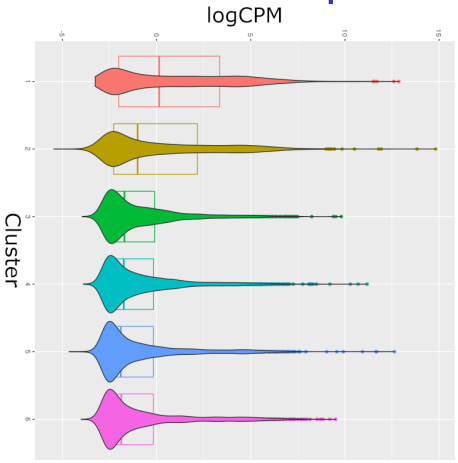


PCA plot of promoters

Specific H3K27me3 profiles show elevated expression



Cluster means for H3K27me3



Cluster expression distributions

Current question

How do these histone marks behave in promoter regions?

Answer: Presence and position both matter!

H3K4me2 & H3K4me3

- Peak closer to promoter $\Rightarrow$ higher gene expression
- Slightly asymmetric in favor of peaks downstream of TSS

Answer: Presence and position both matter!

H3K4me2 & H3K4me3

- Peak closer to promoter $\Rightarrow$ higher gene expression
- Slightly asymmetric in favor of peaks downstream of TSS

H3K27me3

- Depletion of H3K27me3 at TSS $\Rightarrow$ elevated gene expression
- Enrichment of H3K27me3 upstream of TSS $\Rightarrow$ *more* elevated expression
- Other coverage profiles: no association

Last question

What can these histone marks tell us about T-cell activation and differentiation?

Differential modification disappears by Day 14

	Number of significant promoters			Est. differentially modified promoters		
Time Point	H3K4me2	H3K4me3	H3K27me3	H3K4me2	H3K4me3	H3K27me3
Day 0	4553	927	6	9967	4149	2404
Day 1	567	278	1570	4370	2145	6598
Day 5	2313	139	490	9450	1148	4141
Day 14	0	0	0	0	0	0

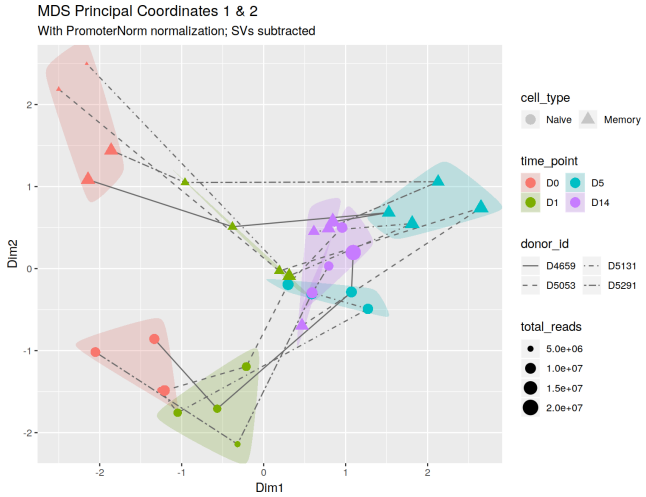
Differential modification between naïve and memory samples at each time point

Differential modification disappears by Day 14

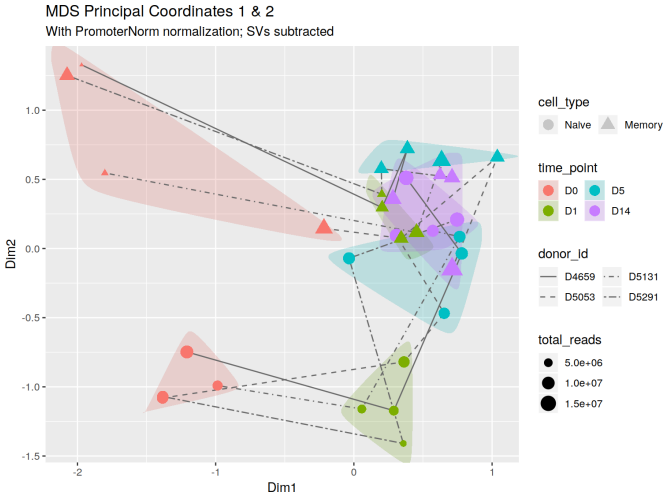
	Number of significant promoters			Est. differentially modified promoters		
Time Point	H3K4me2	H3K4me3	H3K27me3	H3K4me2	H3K4me3	H3K27me3
Day 0	4553	927	6	9967	4149	2404
Day 1	567	278	1570	4370	2145	6598
Day 5	2313	139	490	9450	1148	4141
Day 14	0	0	0	0	0	0

Differential modification between naïve and memory samples at each time point

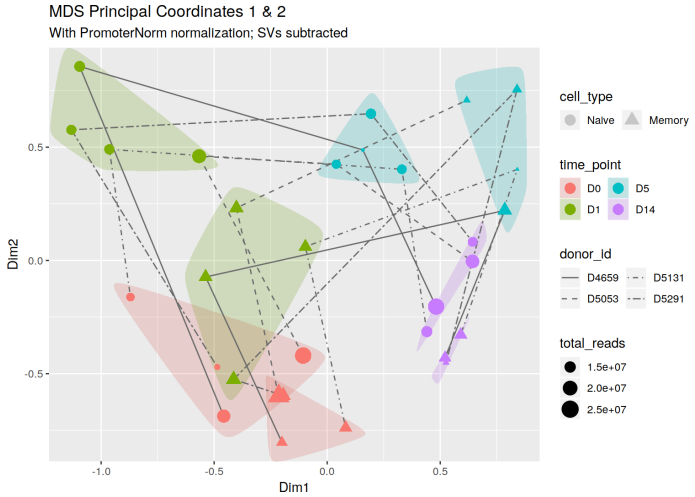
Promoter H3K4me2 levels converge at Day 14



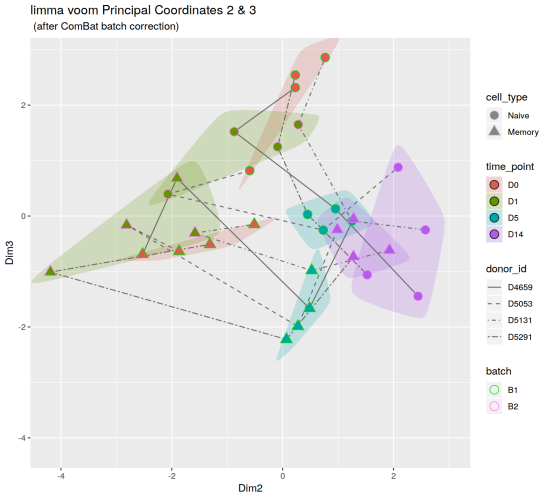
Promoter H3K4me3 levels converge at Day 14



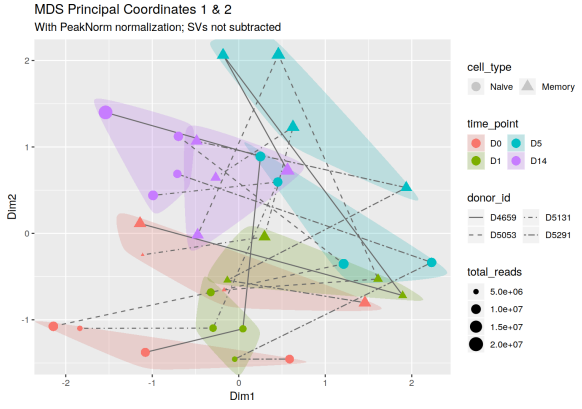
Promoter H3K27me3 levels converge at Day 14?



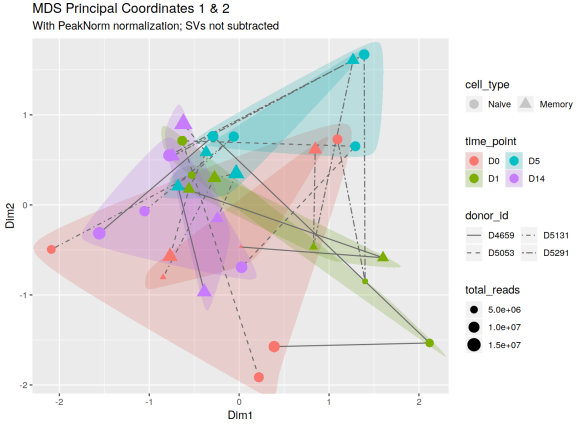
Expression converges at Day 14 (in PC 2 & 3)



But the data isn't really that clean...

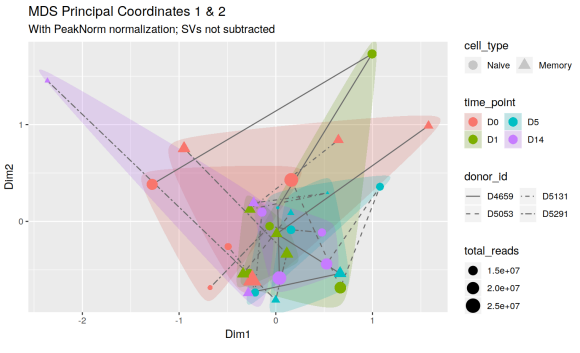


H3K4me2

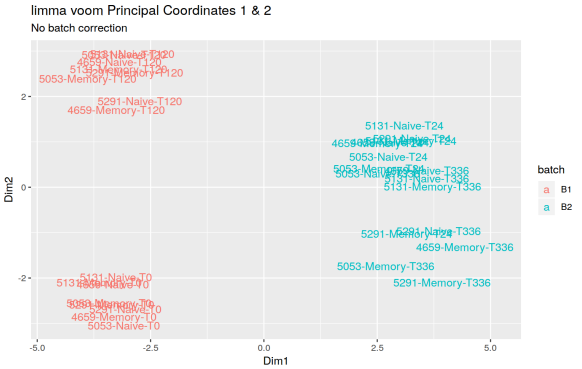


H3K4me3

But the data isn't really that clean...

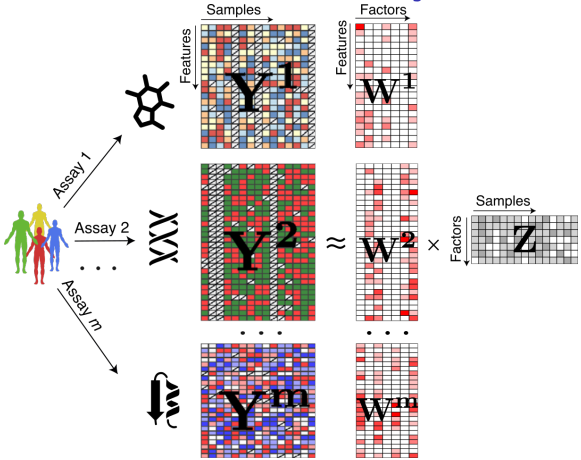


H3K27me3



RNA-seq

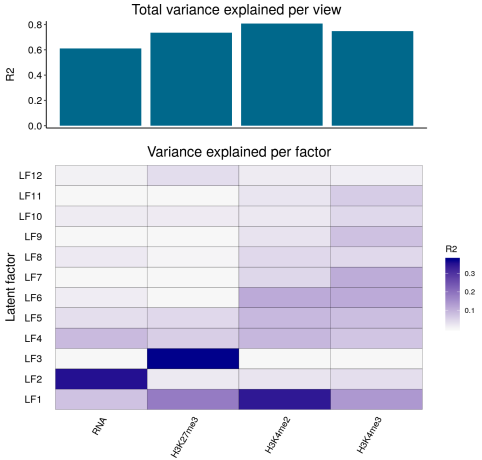
MOFA: cross-dataset factor analysis



MOFA factor analysis schematic[9]

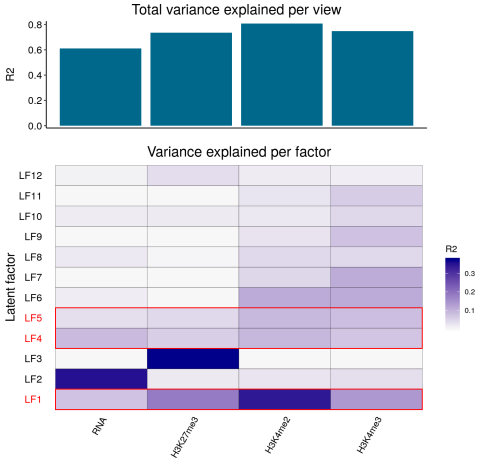
[9] Argelaguet, Velten, et. al. (2018)

Some factors are shared while others are not



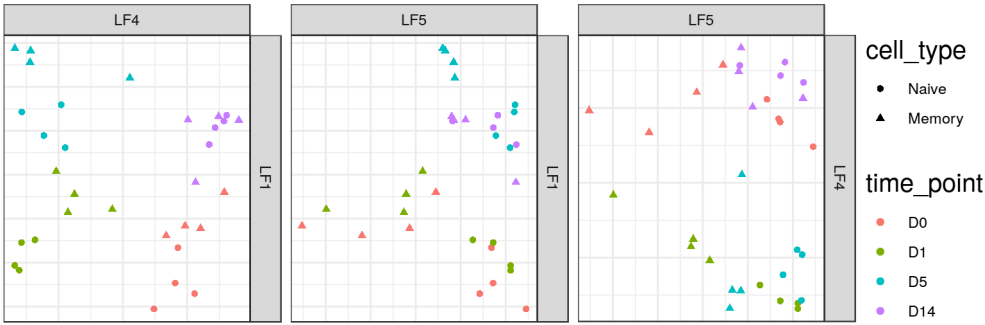
Variance explained in each data set by each LF

3 factors are shared across all 4 data sets



LFs 1, 4, and 5 explain variation in all 4 data sets

MOFA LF5 captures convergence pattern



LF1 & LF4: time point effect; LF5: convergence

Last question

What can these histone marks tell us about T-cell activation and differentiation?

Answer: Epigenetic convergence between naïve and memory!

- Almost no differential histone modification observed between naïve and memory at Day 14, despite plenty of differential modification at earlier time points.
- Expression and 3 histone marks all show "convergence" between naïve and memory by Day 14 in the first 2 or 3 principal coordinates.
- MOFA captures this convergence pattern in a single latent factor, indicating that this is a shared pattern across all 4 data sets.

Answers to key questions

How do we define the "promoter region" for each gene?

Define empirically using peak-to-promoter distances; validate by correlation with expression.

Answers to key questions

How do we define the "promoter region" for each gene?

Define empirically using peak-to-promoter distances; validate by correlation with expression.

How do these histone marks behave in promoter regions?

Location matters! Specific coverage patterns correlated with elevated expression.

Answers to key questions

How do we define the "promoter region" for each gene?

Define empirically using peak-to-promoter distances; validate by correlation with expression.

How do these histone marks behave in promoter regions?

Location matters! Specific coverage patterns correlated with elevated expression.

What can we learn about T-cell activation and differentiation?

Epigenetic & expression state of naïve and memory converges late after activation, consistent with naïve differentiation into memory.

Further conclusions & future directions

- "Effective promoter region" is a useful concept but "radius" oversimplifies: seek a better definition
- Coverage profiles were only examined in naïve day 0 samples: further analysis could incorporate time and cell type
- Coverage profile normalization induces degeneracy: adapt a better normalization from peak callers like SICER
- Unimodal distribution of promoter coverage profiles is unexpected

Further conclusions & future directions

- Experiment was not designed to directly test the epigenetic convergence hypothesis: future experiments could include cultured but un-activated controls
- High correlation between H3K4me3 and H3K4me2 is curious given they are mutually exclusive: design experiments to determine the degree of actual co-occurrence

Implications for transplant biology

- Epigenetic regulation through histone methylation is surely involved in immune memory

Implications for transplant biology

- Epigenetic regulation through histone methylation is surely involved in immune memory
- Can we stop memory cells from forming by perturbing histone methylation?

Implications for transplant biology

- Epigenetic regulation through histone methylation is surely involved in immune memory
- Can we stop memory cells from forming by perturbing histone methylation?
- Can we disrupt memory cell function during rejection by perturbing histone methylation?

Implications for transplant biology

- Epigenetic regulation through histone methylation is surely involved in immune memory
- Can we stop memory cells from forming by perturbing histone methylation?
- Can we disrupt memory cell function during rejection by perturbing histone methylation?
- Can we suggest druggable targets for better immune suppression by looking at epigenetically upregulated genes in memory cells?

Acknowledgements

- My mentors, past and present: Drs. Terry Gaasterland, Daniel Salomon, and Andrew Su
- My committee: Drs. Nicholas Schork, Ali Torkamani, Michael Petrascheck, and Luc Teyton.
- My many collaborators in the Salomon Lab
- The Scripps Genomics Core
- My parents, John & Chris Thompson

Questions?