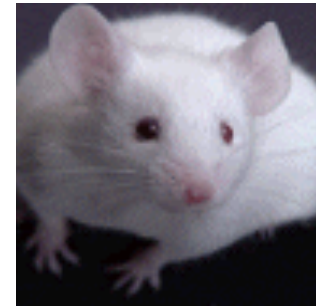


Variation Regulatory Information Within and Between Species

Michael Snyder

March 18, 2014



Conflicts: Personalis, Genapsys, AxioMx

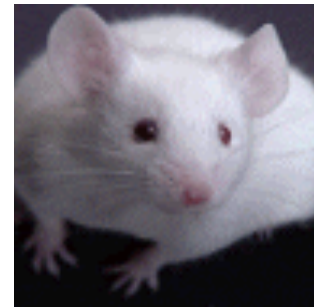
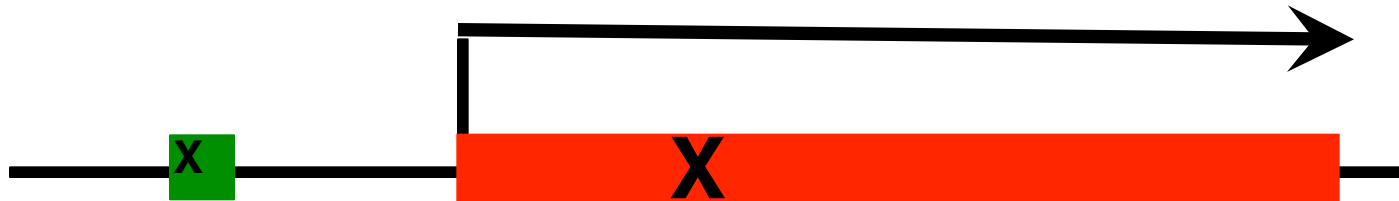
Topics

1. Variation in regulatory information among humans
 - Chromatin
 - Protein levels
2. Master variators: linking variation to phenotypes
3. Differences between mouse and human
4. ENCODE: Relationship of biochemical activity and conservation



What is the Difference Between People and Between Species?

Coding Sequences vs Regulation



Many Studies Have Demonstrated that Gene Expression Differs Among Individuals and Species

Among humans

- 1) **Microarrays:** Stranger, Dermitzakis et al. Science (2007), PloS Genetics (2005), Nature Genetics (2009)
Cheung et al Nature (2004)
- 2) **RNA-Seq:** Pritchard 2010; Dermitzakis 2010

Among Species

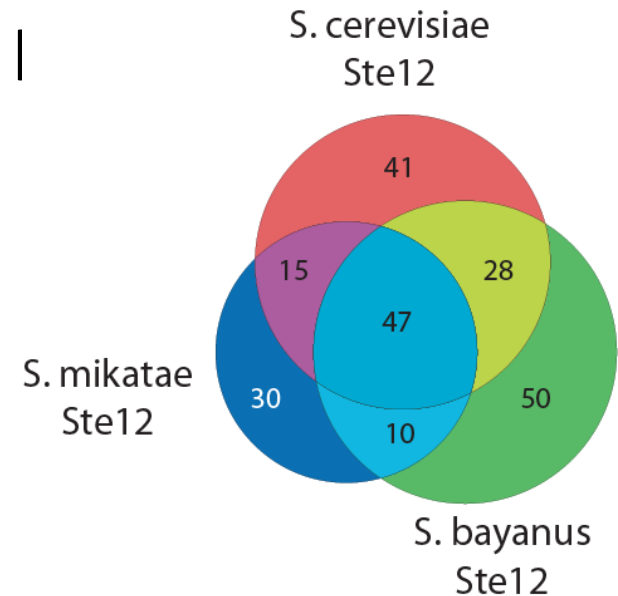
- 1) White et al Nature 2006; Gilad, Pritchard et al 2013

Extensive Differences in TF Binding Between Species

Extensive differences in binding |
yeast species

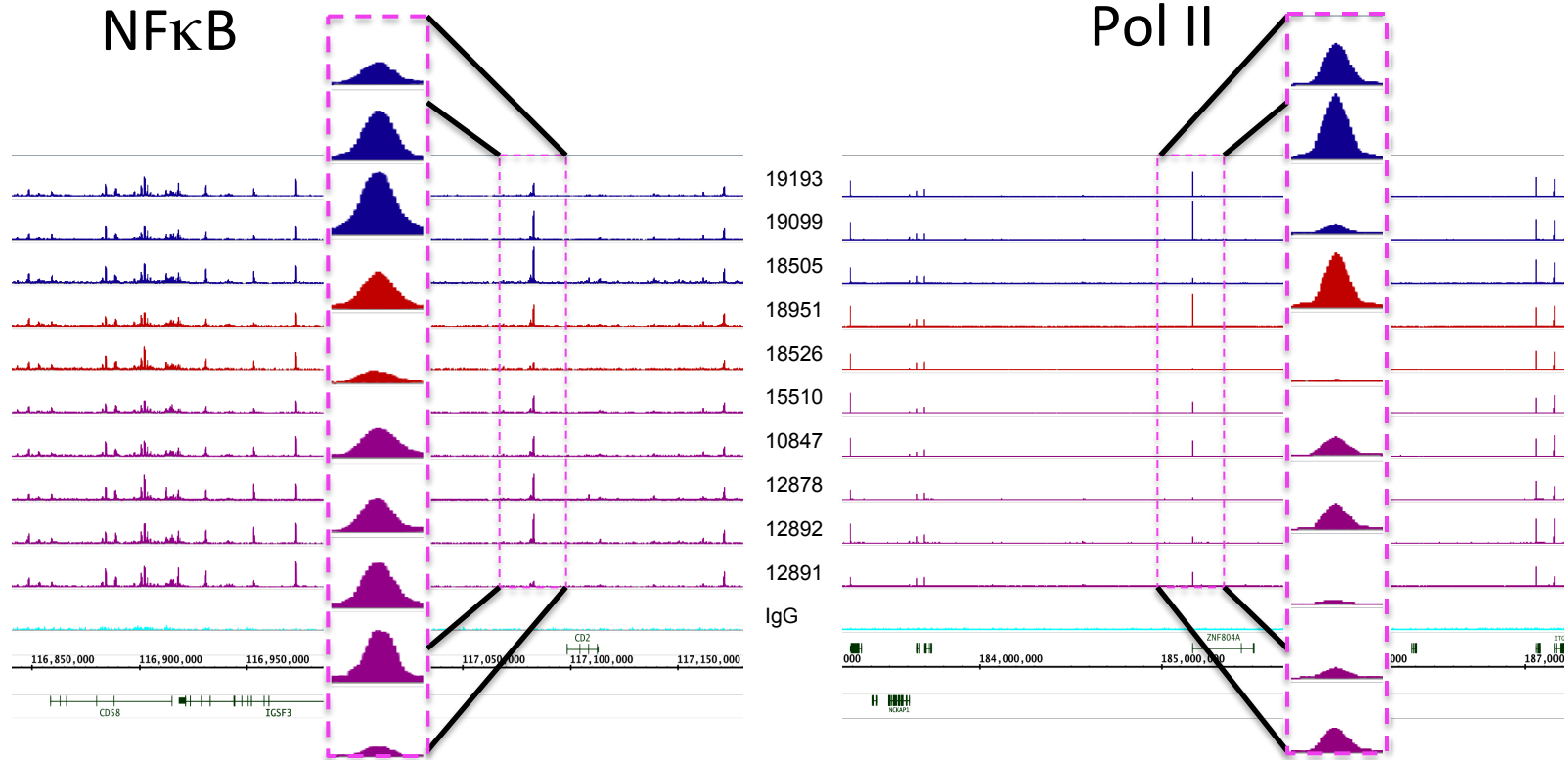
Borneman et al 2007
Science.

Parallel study Mice-Humans
Odom et al 2007



20-25% of Binding Sites Shared in All Three Species

Variability in NFκB and Pol II loci in Humans



7.5% and 25% of NFκB and Pol II binding regions vary

Additional Fundamental Questions?

- 1) What types of elements differ and which ones differ most (e.g. promoters, enhancers, etc)?
- 2) How do these elements vary among people?
- 3) Do differences in gene expression match differences in protein expression

Map Chromatin Marks Across Different Individuals

Histone Marks Representing a Variety of Functional Elements

promoter

H3K4me2
H3K4me3
Acetylation
H2A.Z

RNAPII

H3K36me3
H3K79me2

gene body

Exon Intron Exon

enhancer

p300

Enhancer

H3K4me1
H3K4me2
H3K27ac

CTCF

Boundary

Cohesin

Exon Intron Exon

structural proteins

H3K9me2
H3K9me3

H3K27me3

repressed regions

Data Sets

19 Diverse LCLs

4 San

7 Yoruban*

6 Caucasian*

2 Asian

*includes a trio

9-13 Histone/Regulatory marks

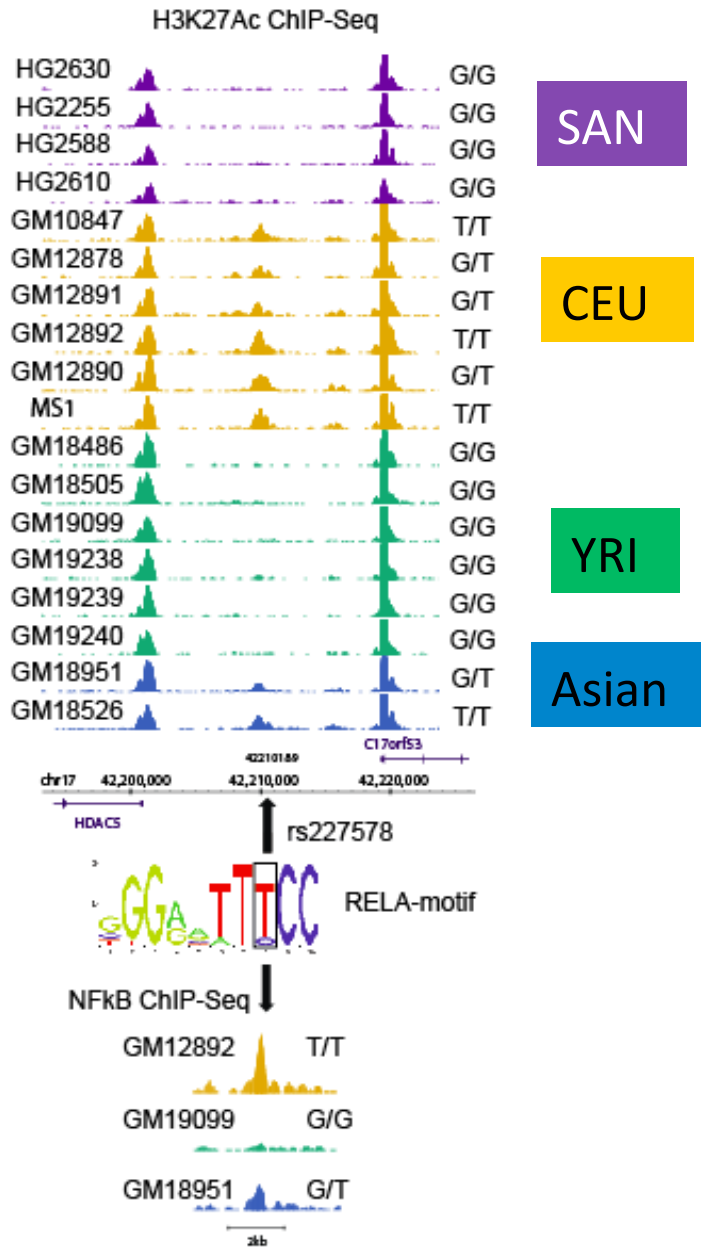
2 or more Replicates

~56M reads/mark

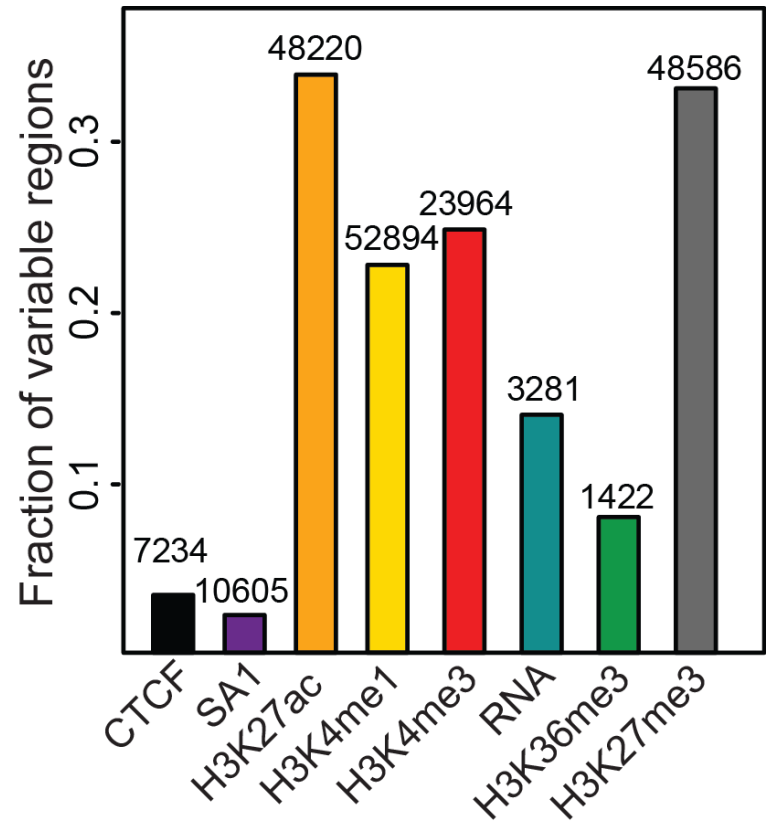
2X100b Paired-end Reads



Chromatin variation across individuals

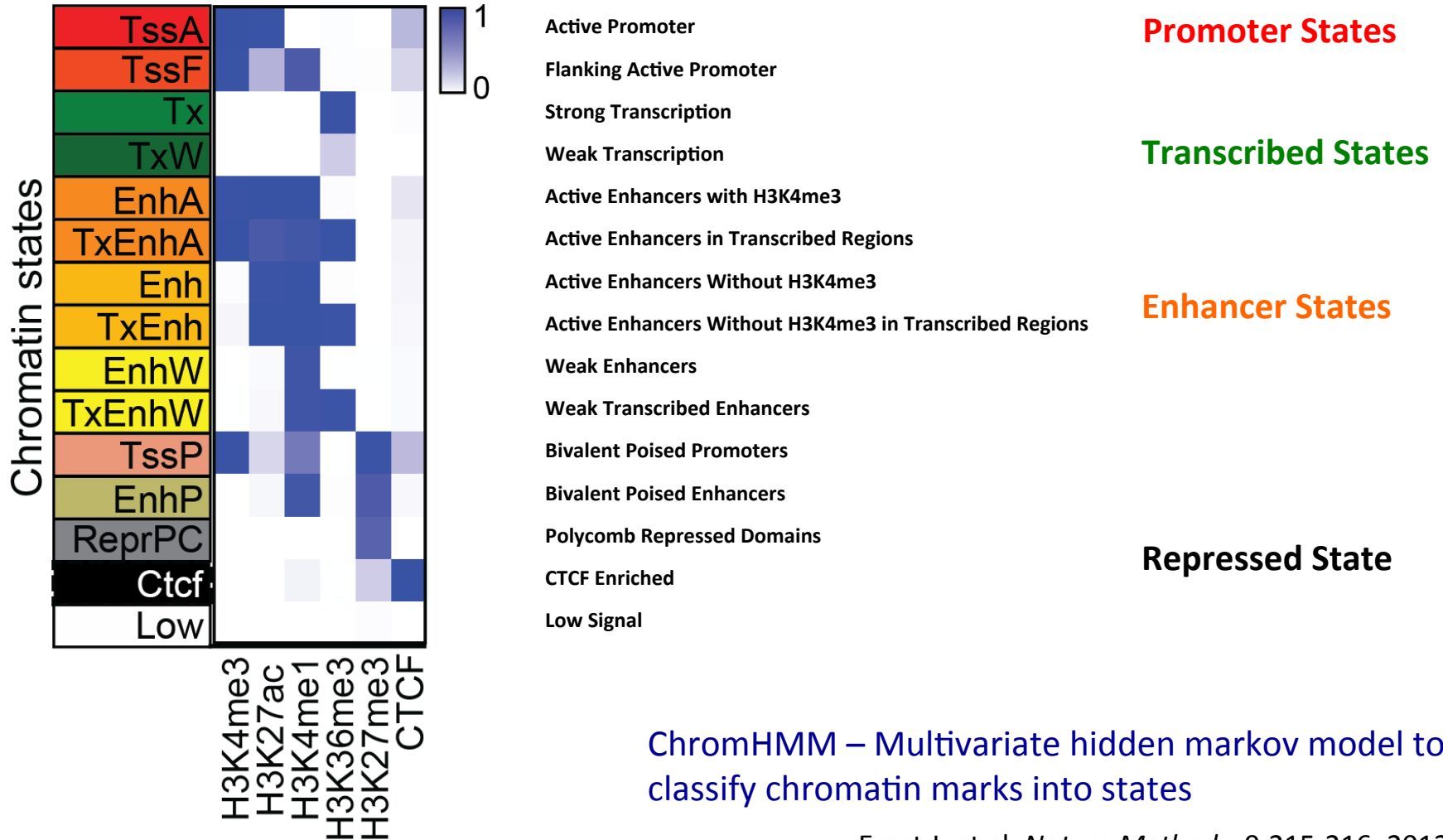


Enhancer related marks
are more variable than RNA



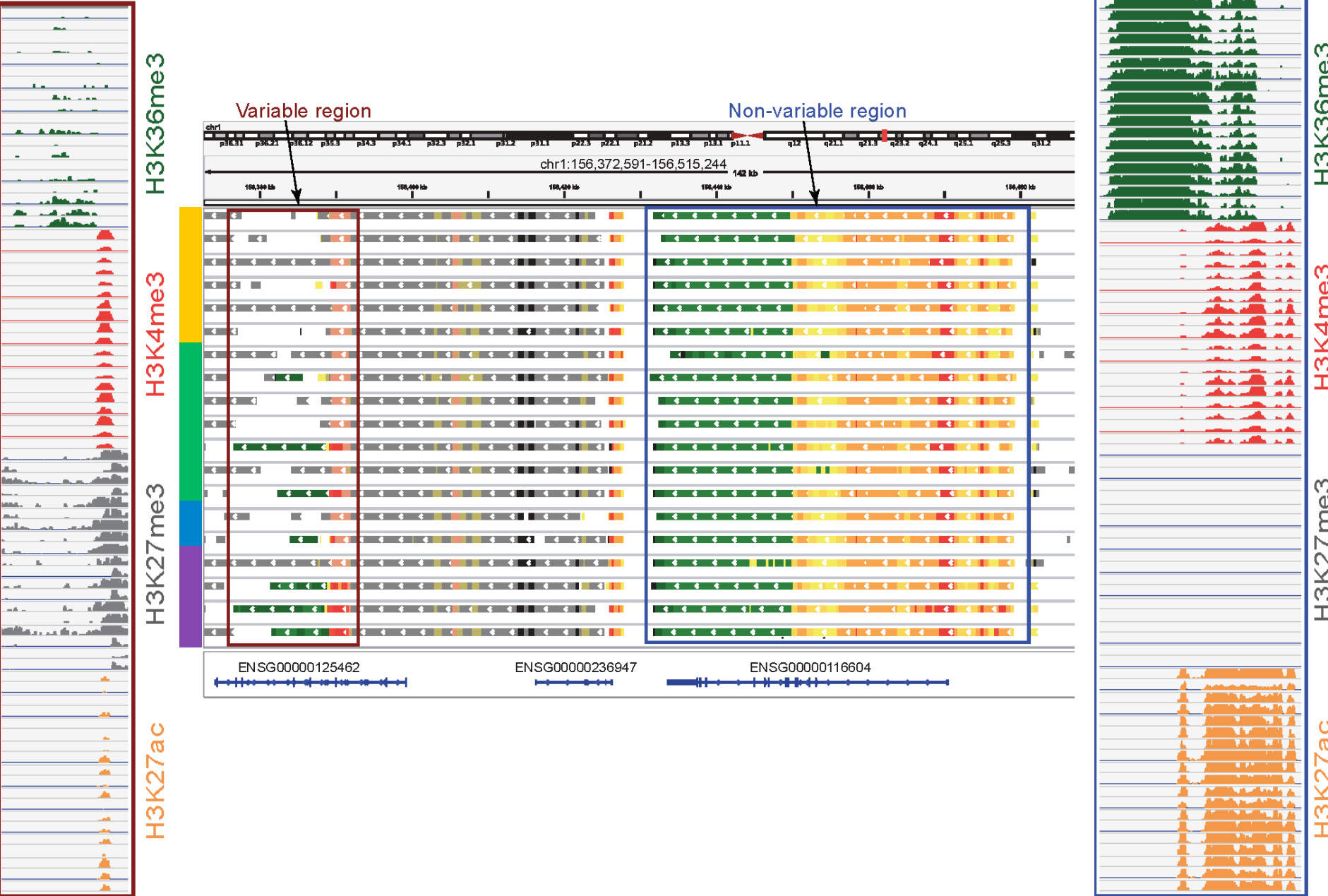
Kasowski et al 2013 Science

Chromatin States Learned on 19 Individuals

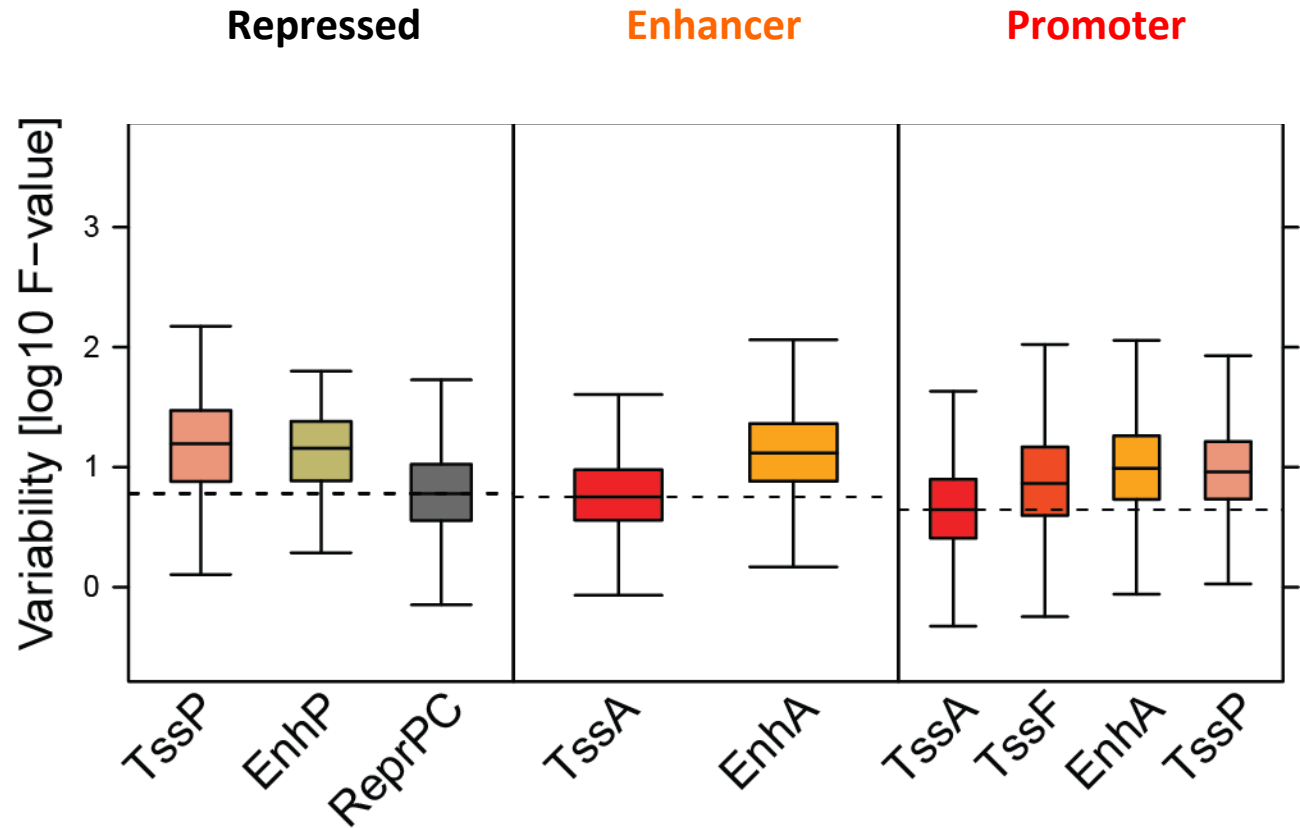
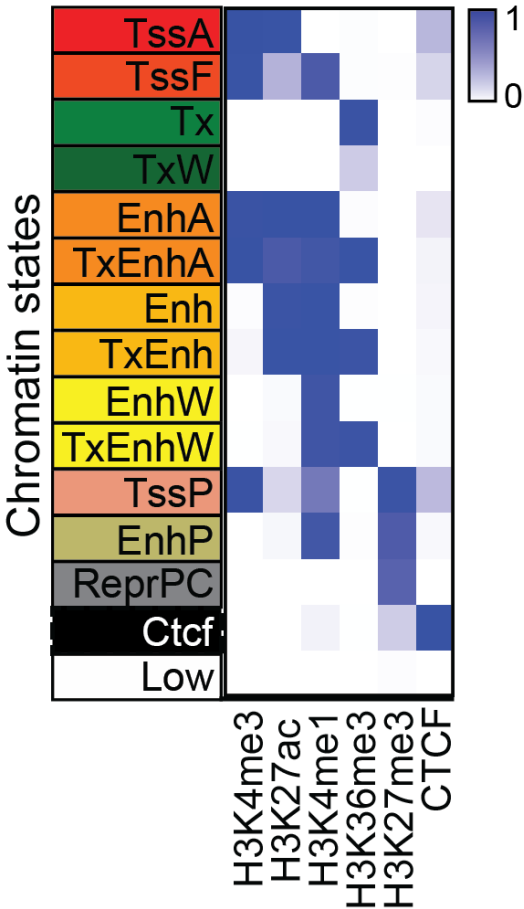


Ernst J. et al. *Nature Methods*, 9:215-216, 2012.

Chromatin state variation across 19 individuals



State Specific Variability of Chromatin Marks

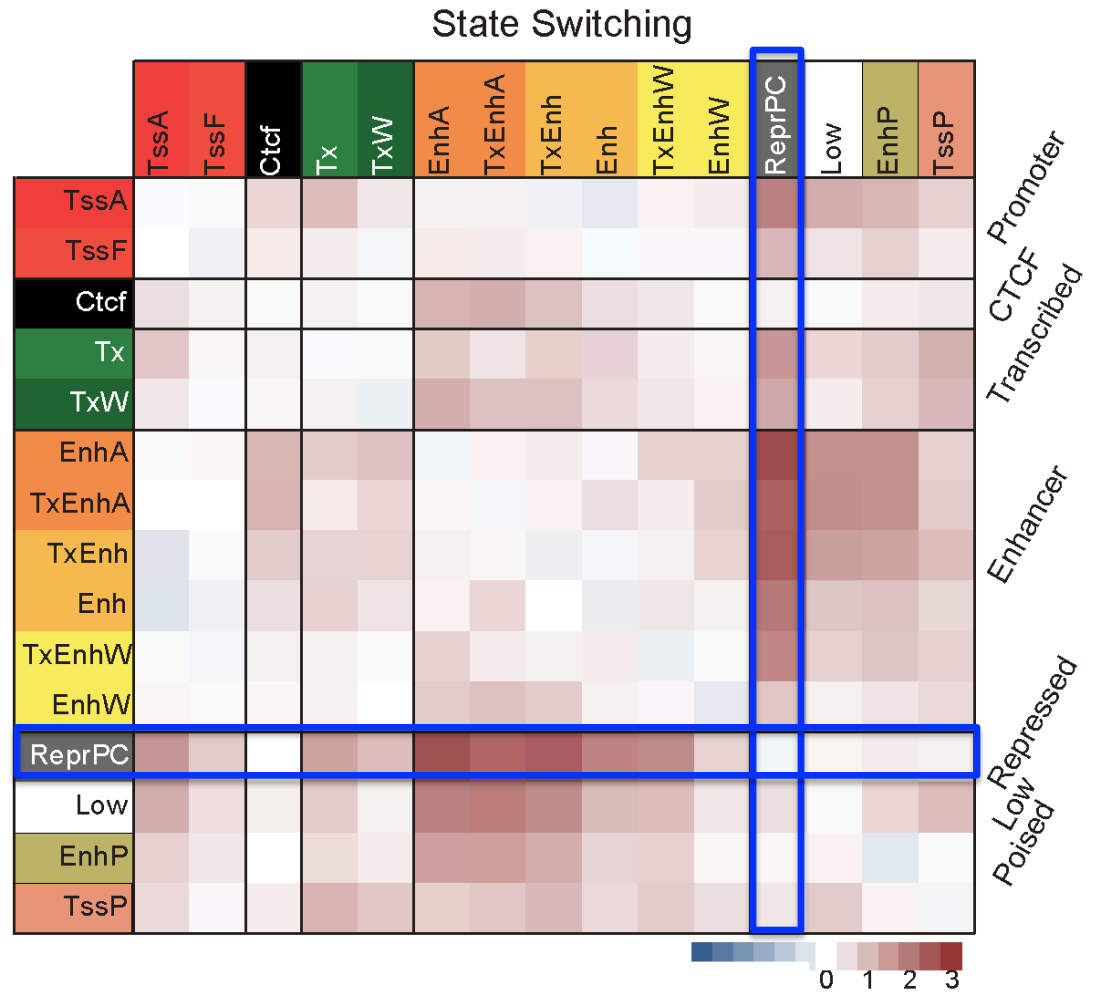


- Active Enhancer state is associated with higher variability than active promoter state

Chromatin state switching across 19 individuals

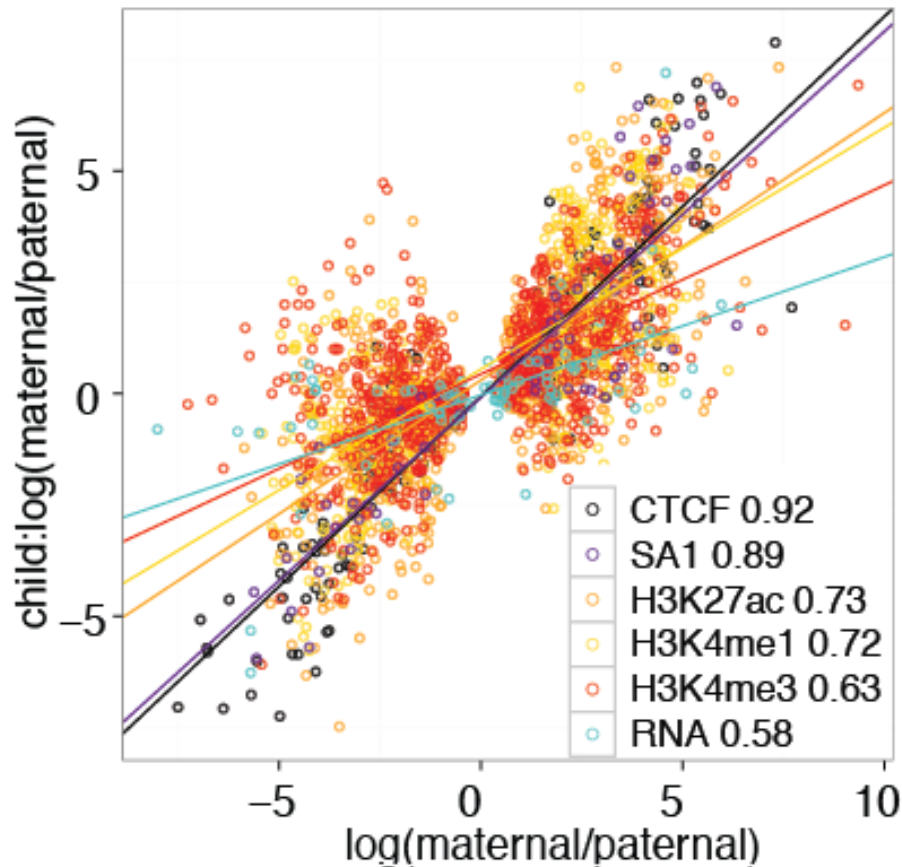
log10 (Observed/Expected)

Observed = Between Individuals
Expected = Between Replicates

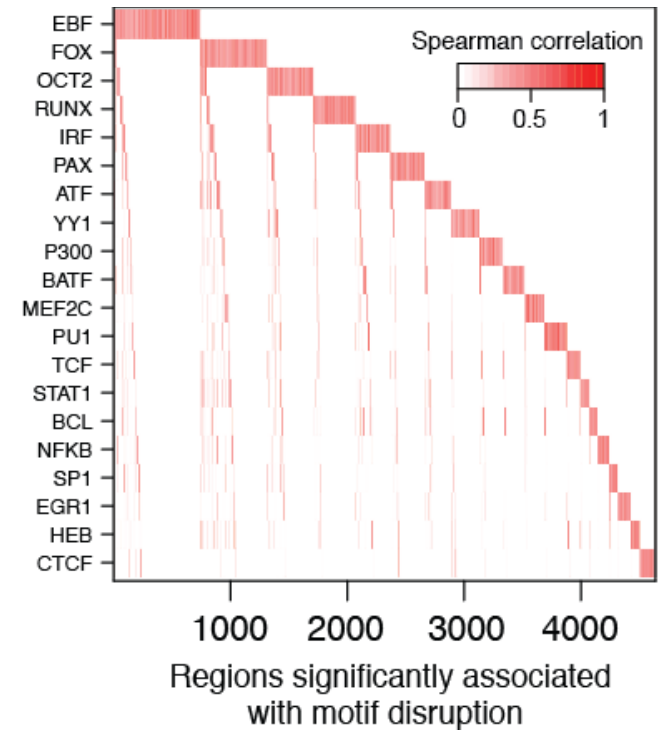


Active (Enhancer/Promoter) <-> Repressed

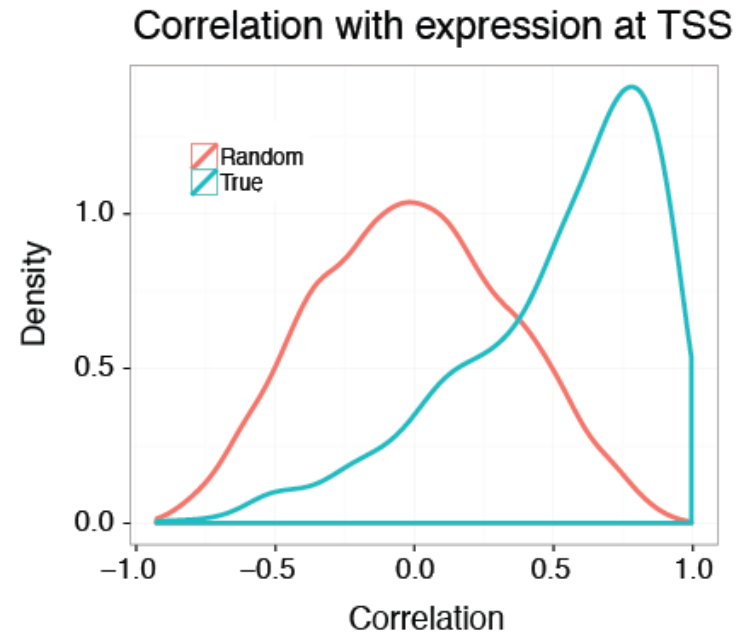
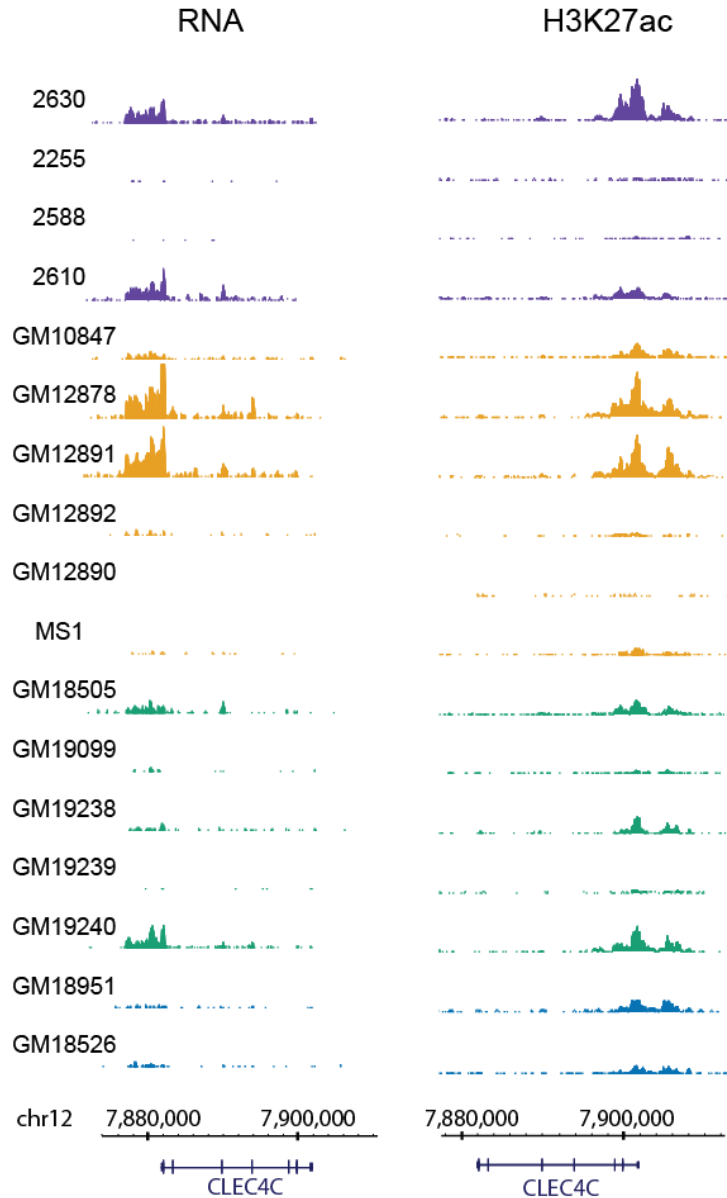
Chromatin Marks are Highly Heritable



Extent of differential signal in parents correlates with extent of allelic bias in daughters

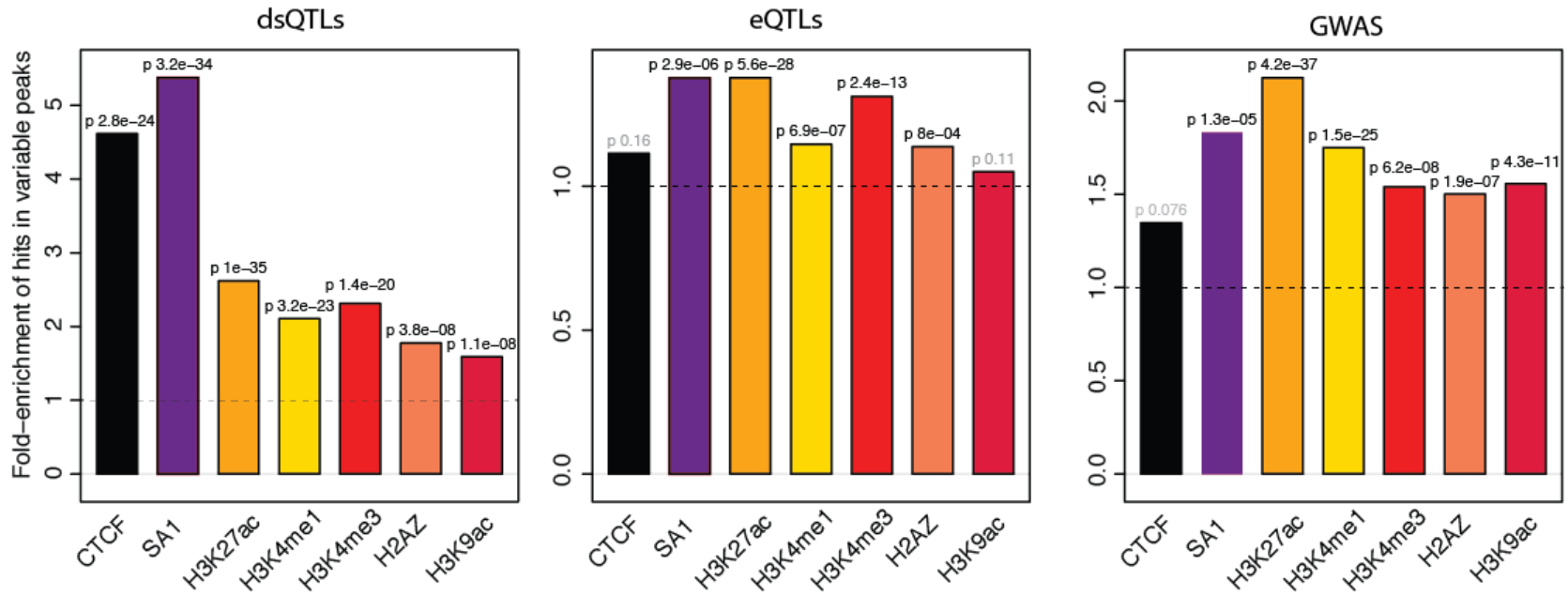


Enhancer Variation Correlates with Expression



Enhancer-gene pairs show strong correlation

Functional Consequences of Chromatin Variation



Variable regions are enriched in QTLs and GWAS SNPs

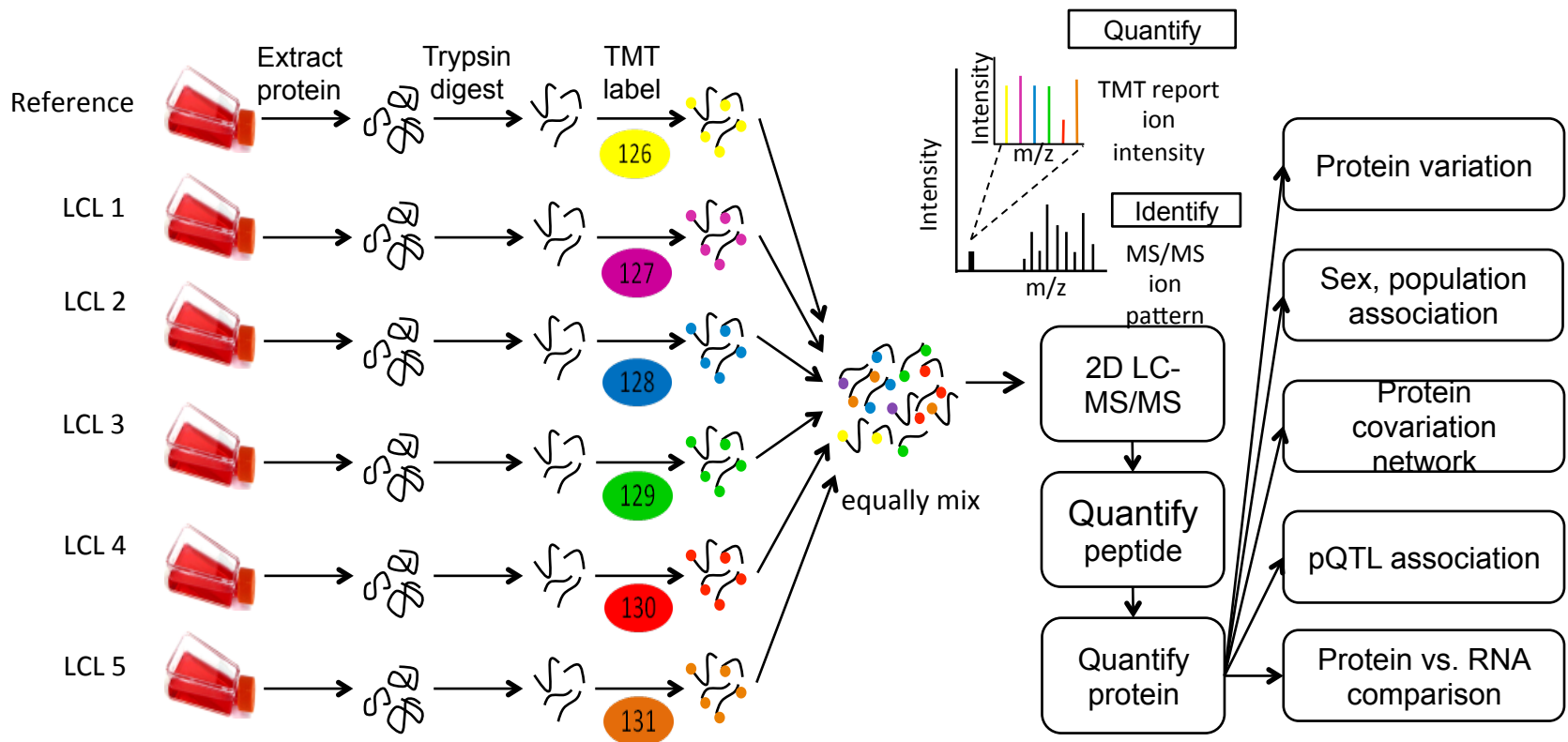
TF and Chromatin States Vary

Gene Expression Varies

What About Proteins?

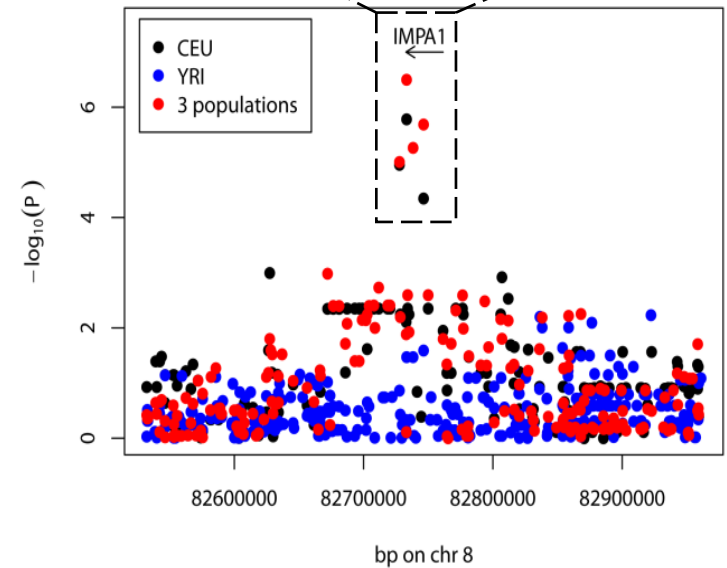
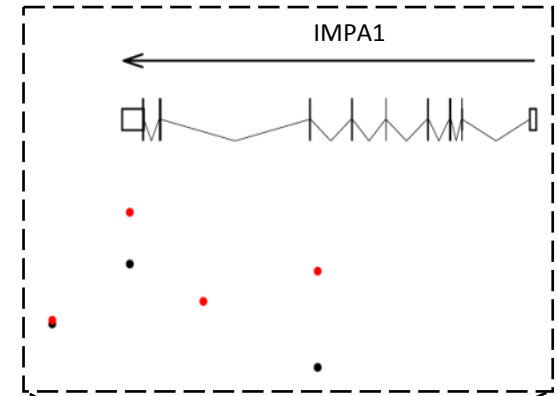
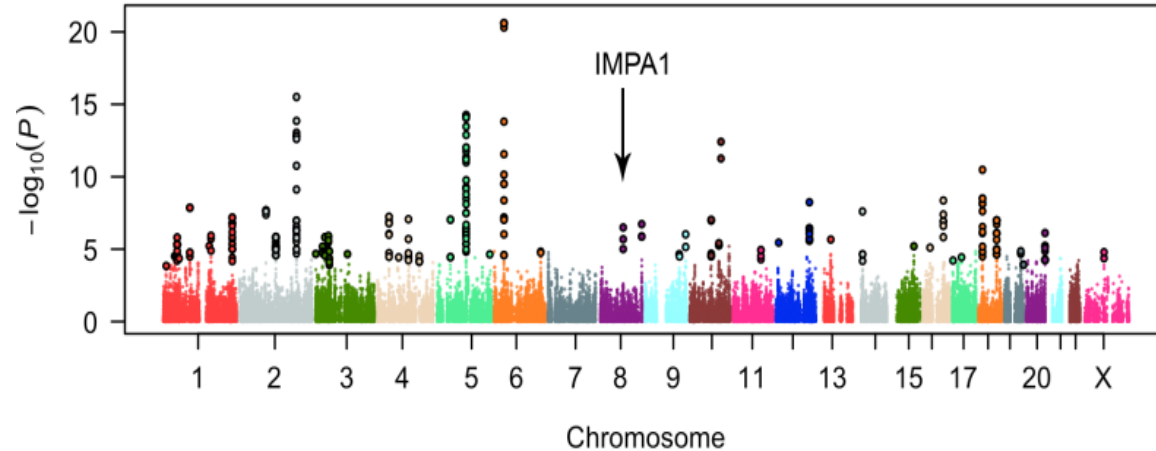
Map Protein Variation Across Individuals

Map Relative Protein Levels Across 90 LCL Lines Using TMT Labeling



Follow 5,963 proteins; 5.7% vary 1.5 fold or more

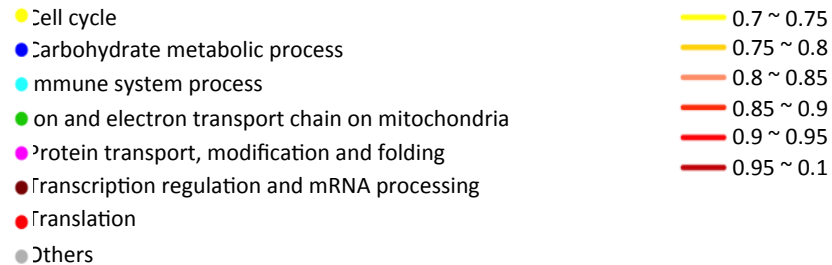
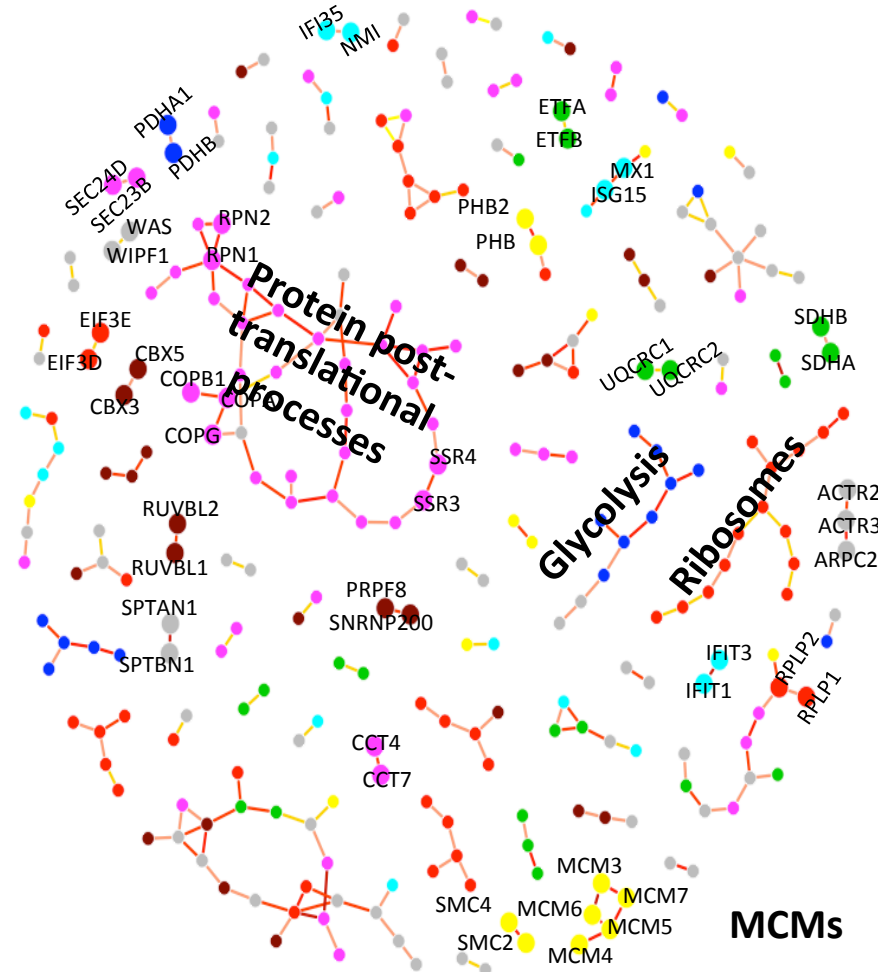
Identify 77 pQTLs with 10%FDR; 273 with 30% FDR



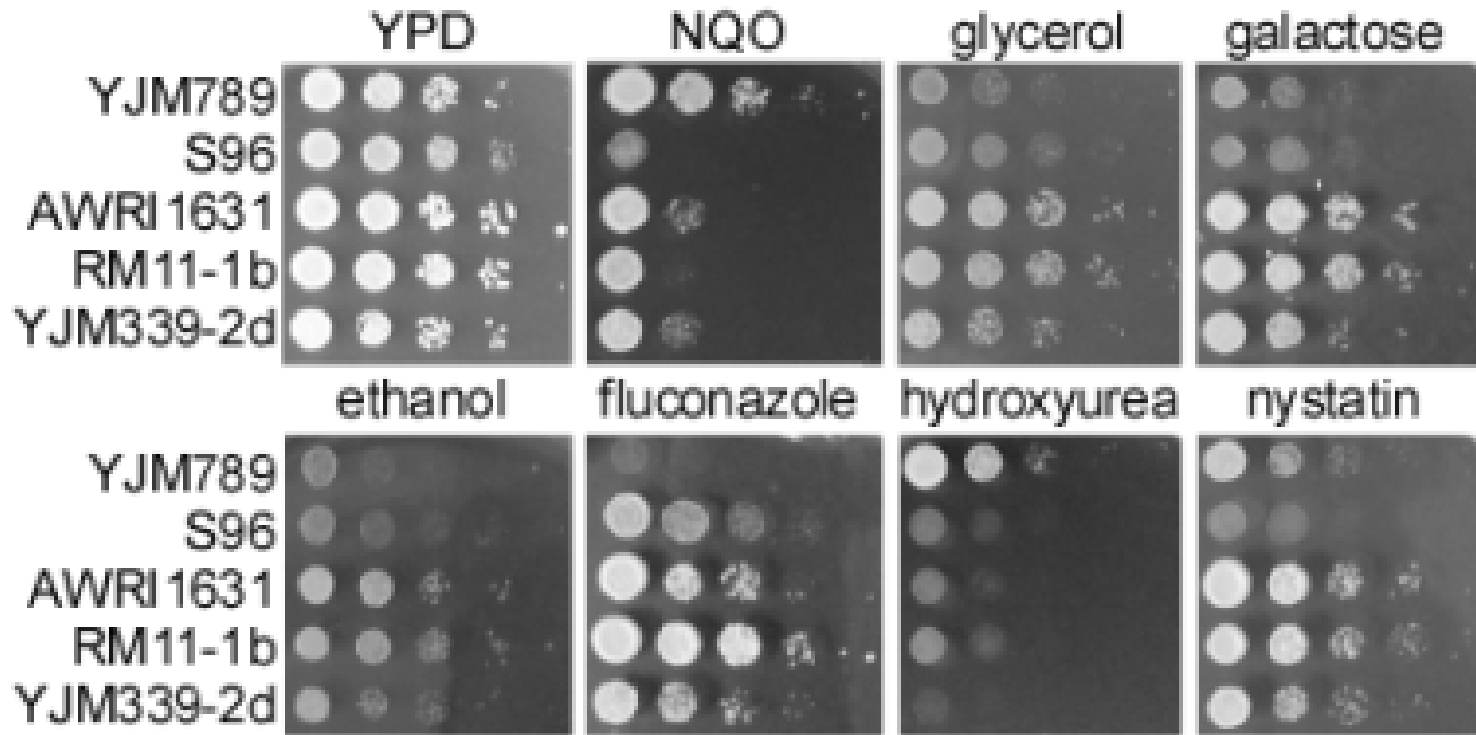
~50% of pQTLs overlap with eQTLs

50% unique

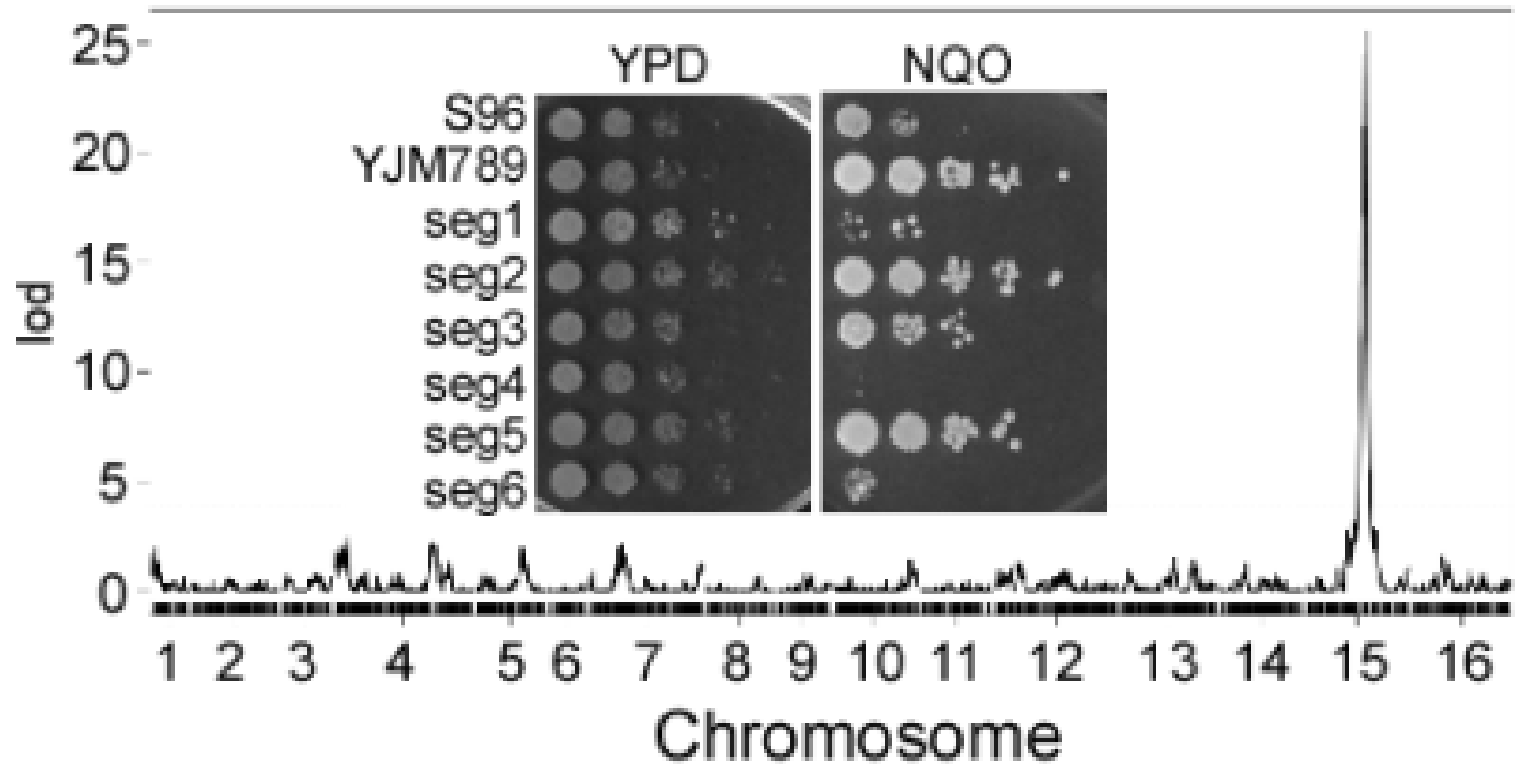
Covarying Proteins Group in Functional Clusters



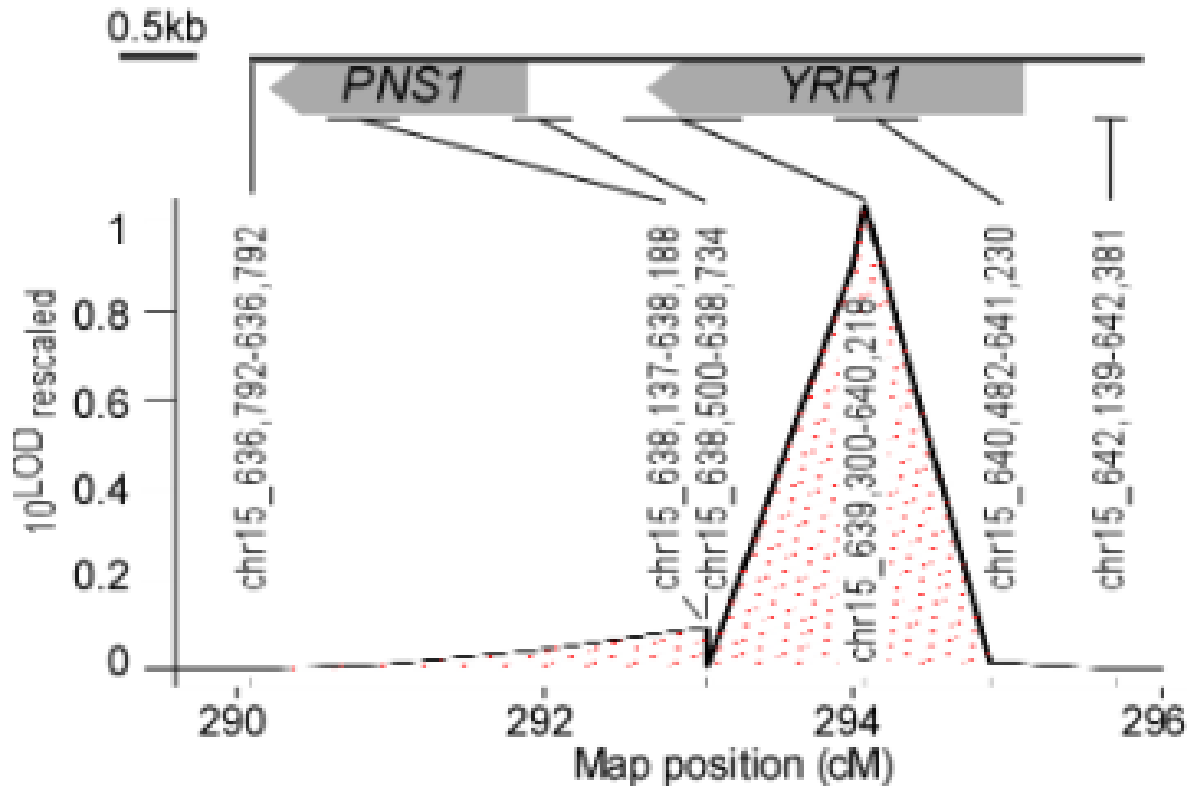
Personal Phenotypes: Variation in Chemical Response



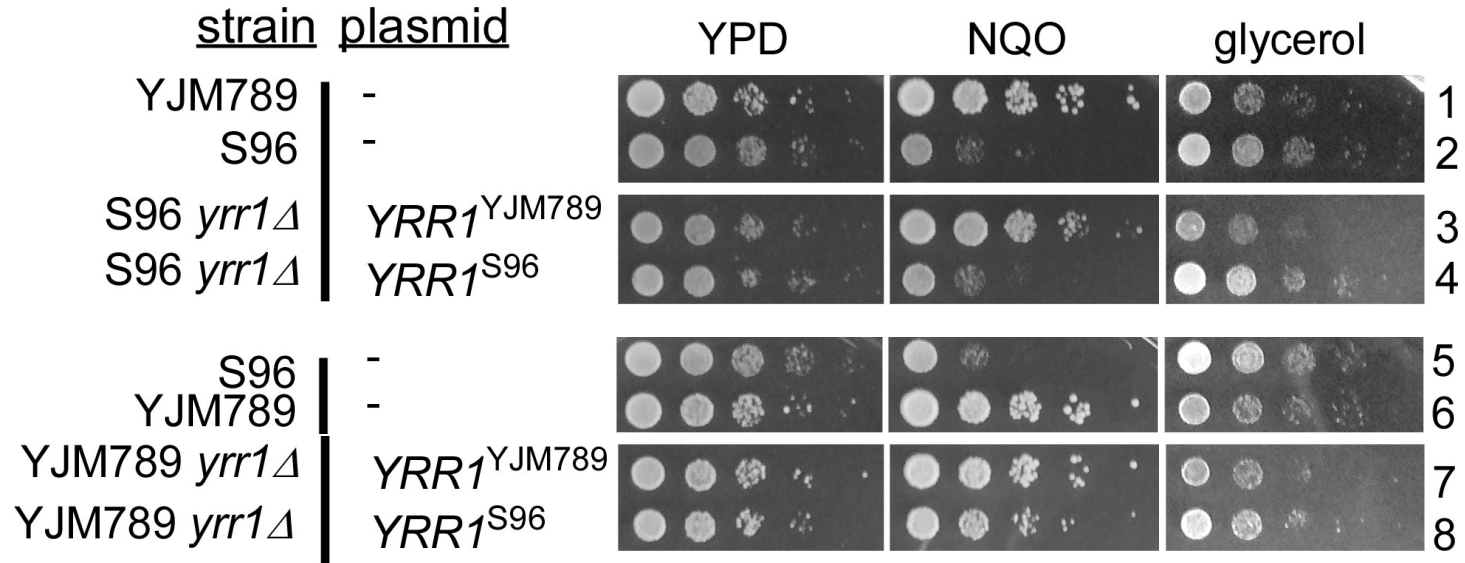
Genome-wide analysis study finds one region linked to NQO response



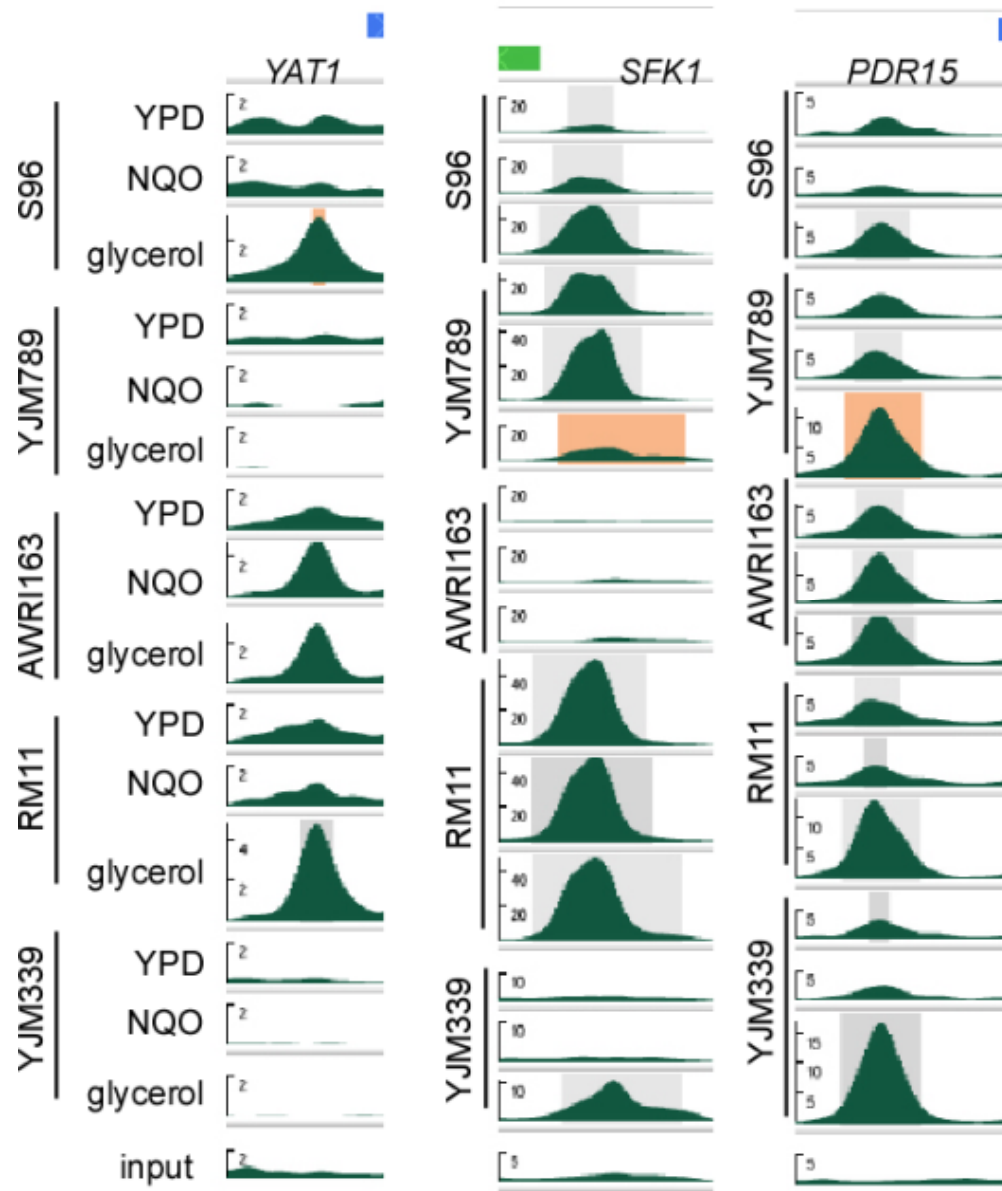
Within this region lies *YRR1*



Allele swap



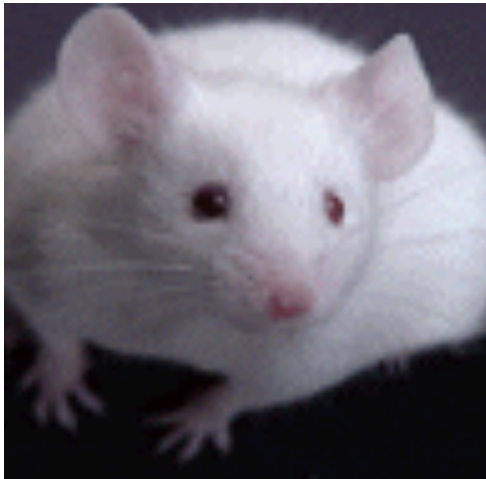
Variation of Yrr1 Binding Regions



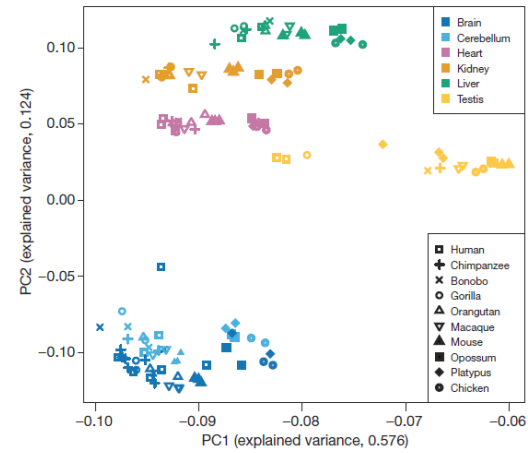
Yrr1: Master Variator

- 1) Regulatory proteins whose genetic alteration can switch on a whole new transcriptional program and phenotype
- 2) Explain how large significant phenotypic changes can be rapidly acquired.

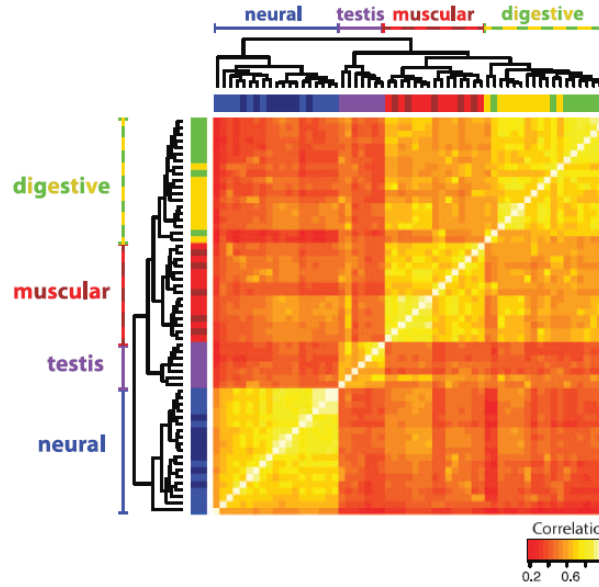
Comparing Mouse and Human with Mouse ENCODE Data



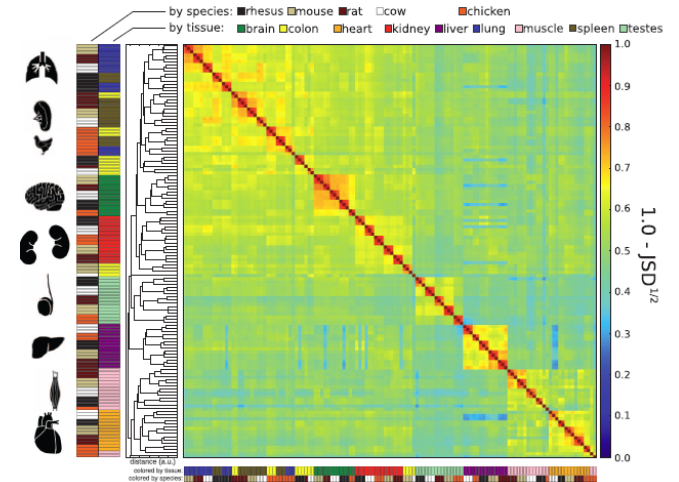
Literature: Clustering by tissue using **RNA-seq** expression data from mouse and human



20 OCTOBER 2011 | VOL 478 | NATURE |

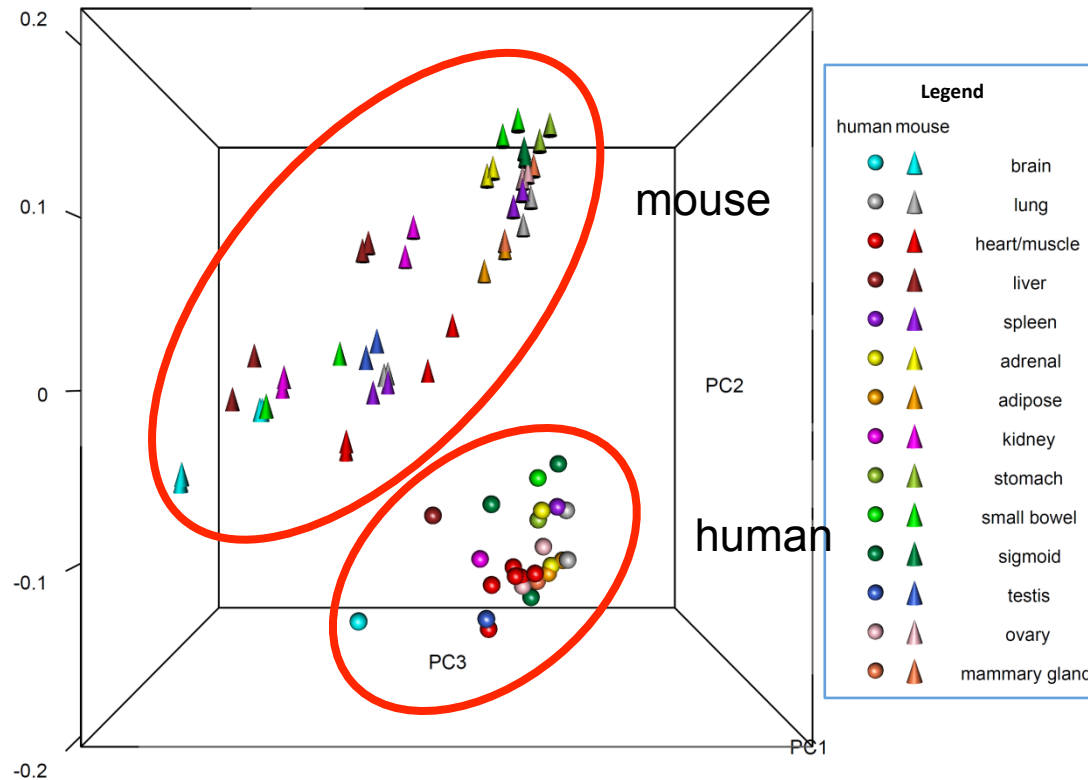


21 DECEMBER 2012 VOL 338 SCIENCE



21 DECEMBER 2012 VOL 338 SCIENCE

Gene Expression Clusters Better By Species, Not Tissues

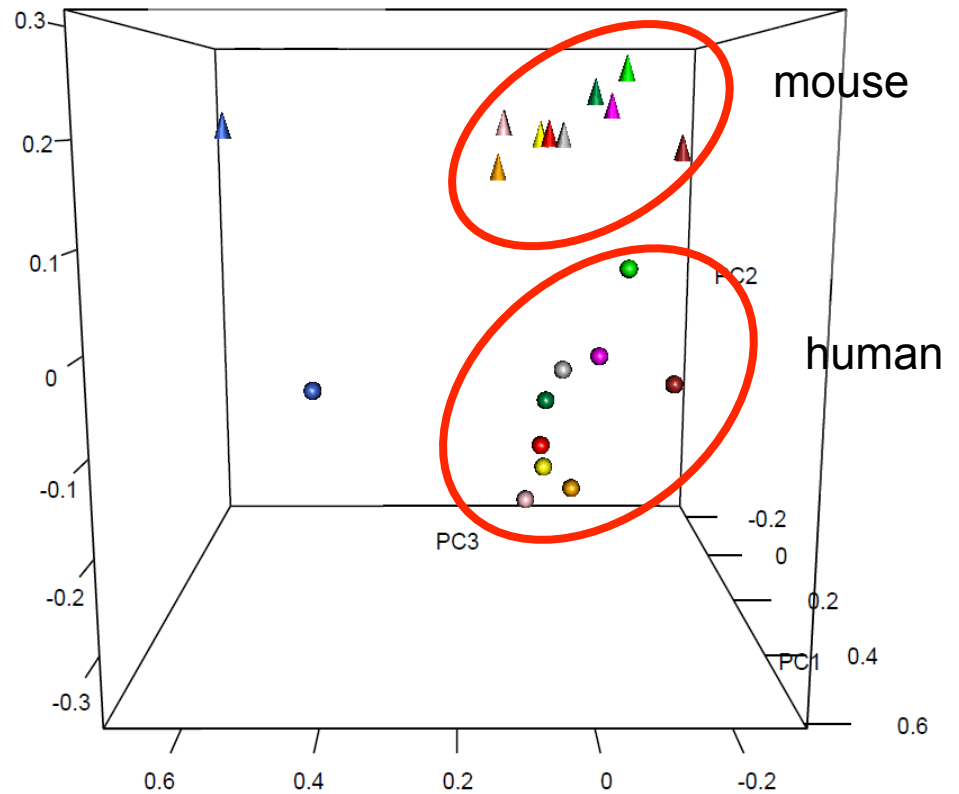


-14 tissues Data From Four Different Labs

cones: mouse

spheres: human

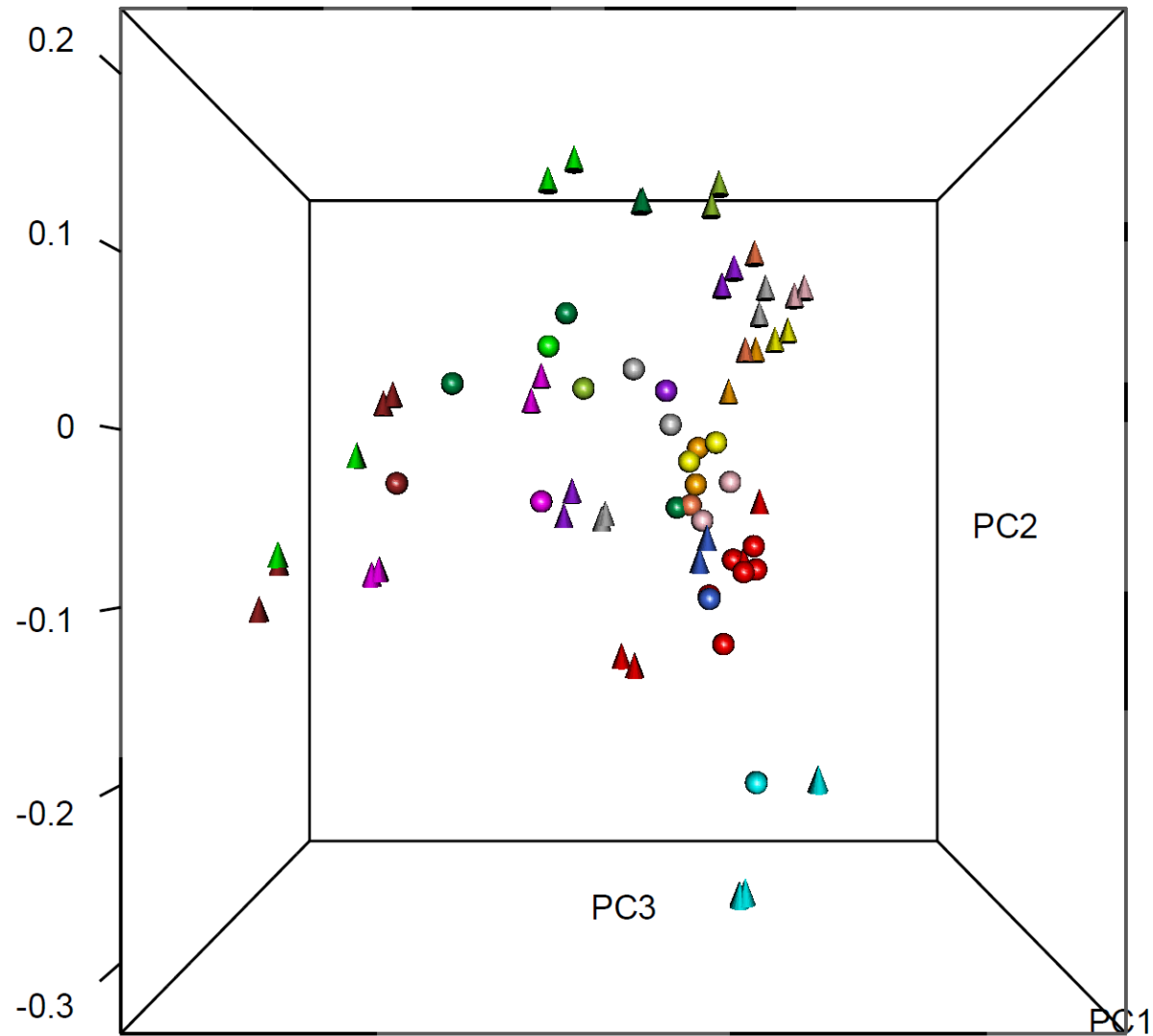
**Same Result When All
Samples Done by the Same
Lab**



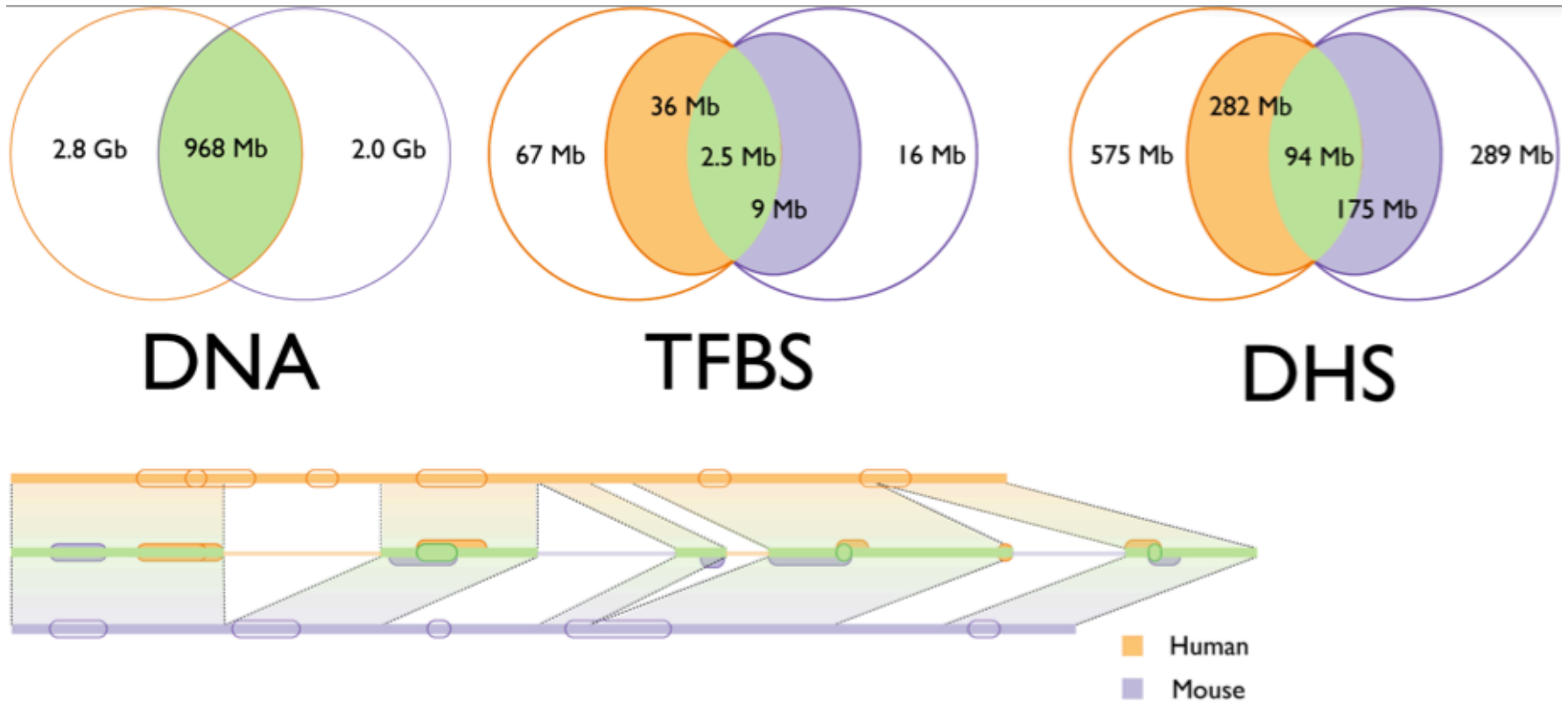
-10 tissues
cones: mouse
spheres: human

For Genes More Highly Expressed in Human over Mouse (2706 genes)	No. of Genes	For Genes More Highly Expressed in Mouse over Human (2391 genes)	No. of Genes
cellular nitrogen compound metabolic process	981	cellular nitrogen compound metabolic process	732
biosynthetic process	833	biosynthetic process	703
signal transduction	591	signal transduction	535
anatomical structure development	539	anatomical structure development	506
transport	502	transport	467
response to stress	491	small molecule metabolic process	392
cellular protein modification process	413	cell differentiation	371
catabolic process	355	response to stress	358
small molecule metabolic process	350	cellular protein modification process	345
cell differentiation	338	catabolic process	281
cell cycle	319	lipid metabolic process	209
cellular component assembly	308	immune system process	207
cell death	290	cell death	202
immune system process	280	cellular component assembly	194
macromolecular complex assembly	222	neurological system process	175
cell proliferation	220	homeostatic process	168
reproduction	195	cell cycle	165
DNA metabolic process	188	cell proliferation	161
nucleobase-containing compound catabolic process	164	reproduction	157
vesicle-mediated transport	157	cell-cell signaling	149
homeostatic process	155	macromolecular complex assembly	131
protein complex assembly	154	DNA metabolic process	130

PCA after removal of 2706 + 2391 differentially expressed genes



Conservation: Sequence-level and activity-level

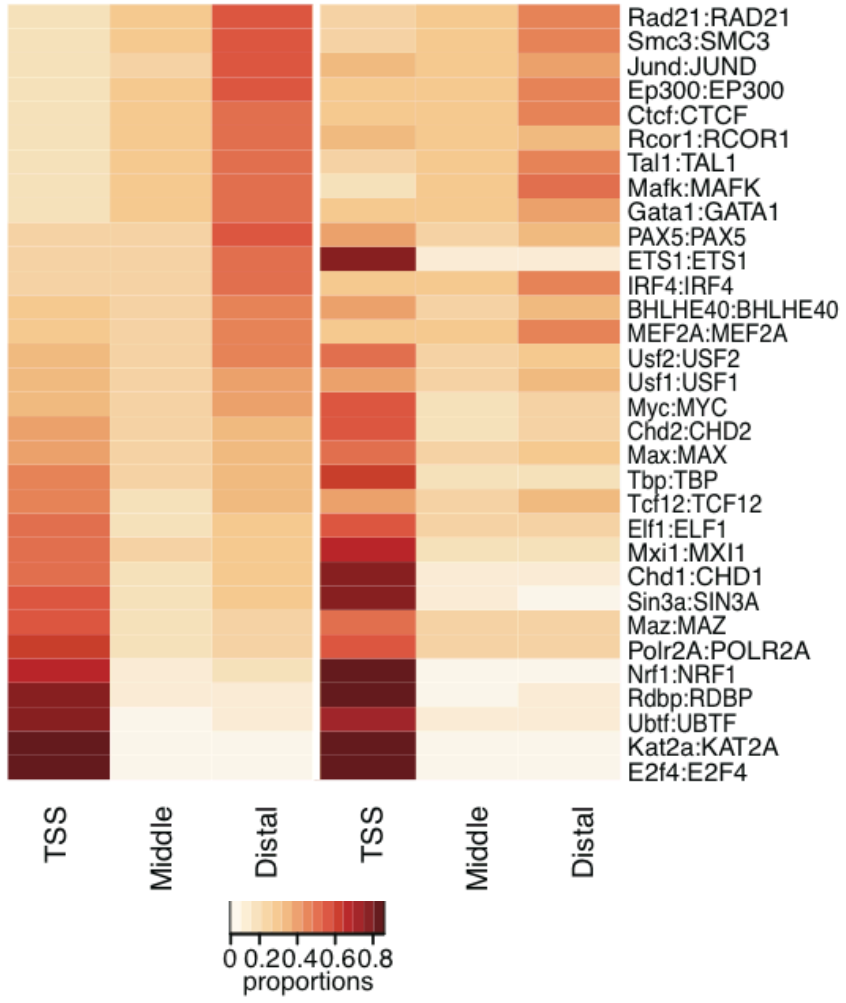


About 40% of regulatory DNA (TFBS, DHS) in mouse maps to aligning DNA in human.
About 10% of TF-bound DNA in mouse is also bound by the same TF in human.

Olgert Denas, Richard Sandstrom, Yong Cheng, Kathryn Beal, Javier Herrero, Ross Hardison, **James Taylor**, submitted.
Genome-wide comparative analysis reveals human-mouse regulatory landscape and evolution.

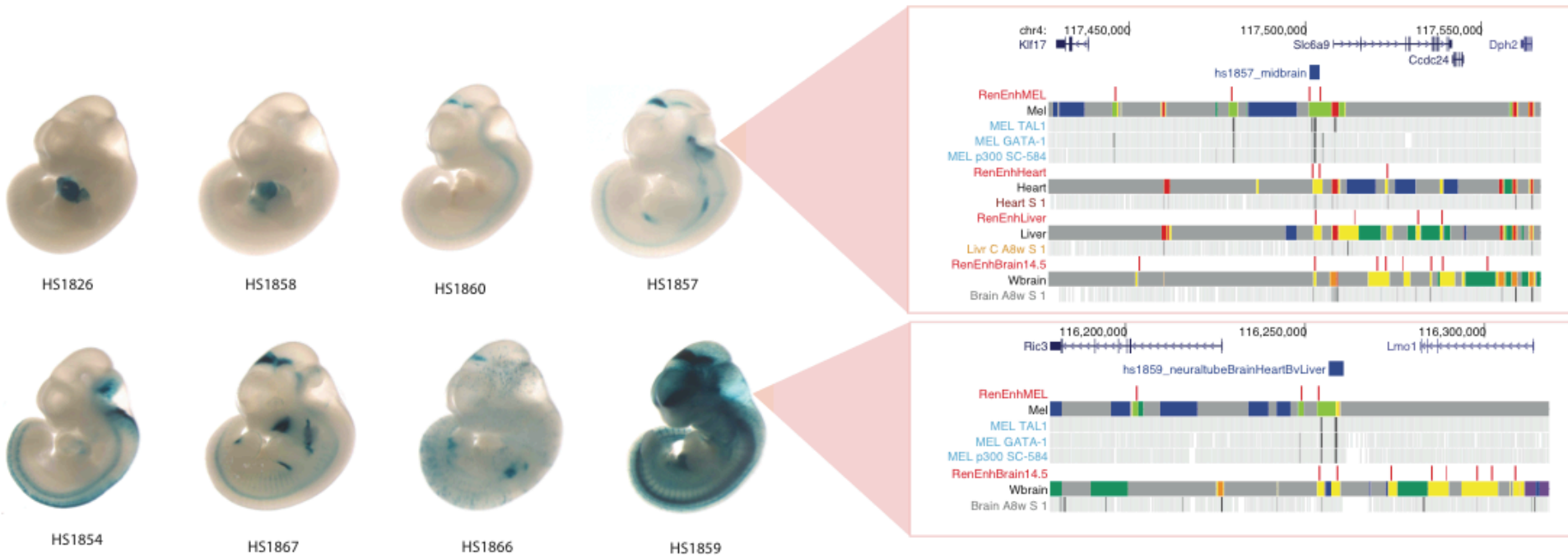
Genomic DNA segments occupied by orthologous pairs of TFs

Conserved locations relative to TSS



Yong Cheng et al., Mouse ENCODE Consortium, submitted. Principles of Regulatory Information Conservation Revealed by Comparing Mouse and Human Transcription Factor Binding Profiles. Snyder, Hardison, Pennacchio labs, 35

Conservation of occupancy by GATA factors predicts enhancers active in *multiple* tissues



Model: **Pleiotropic** functions (multiple tissues, multiple TFs binding) are subject to stronger constraint, leading to preservation of occupancy despite tendency of regulatory regions to “turn over”

Possible Measures of Function?

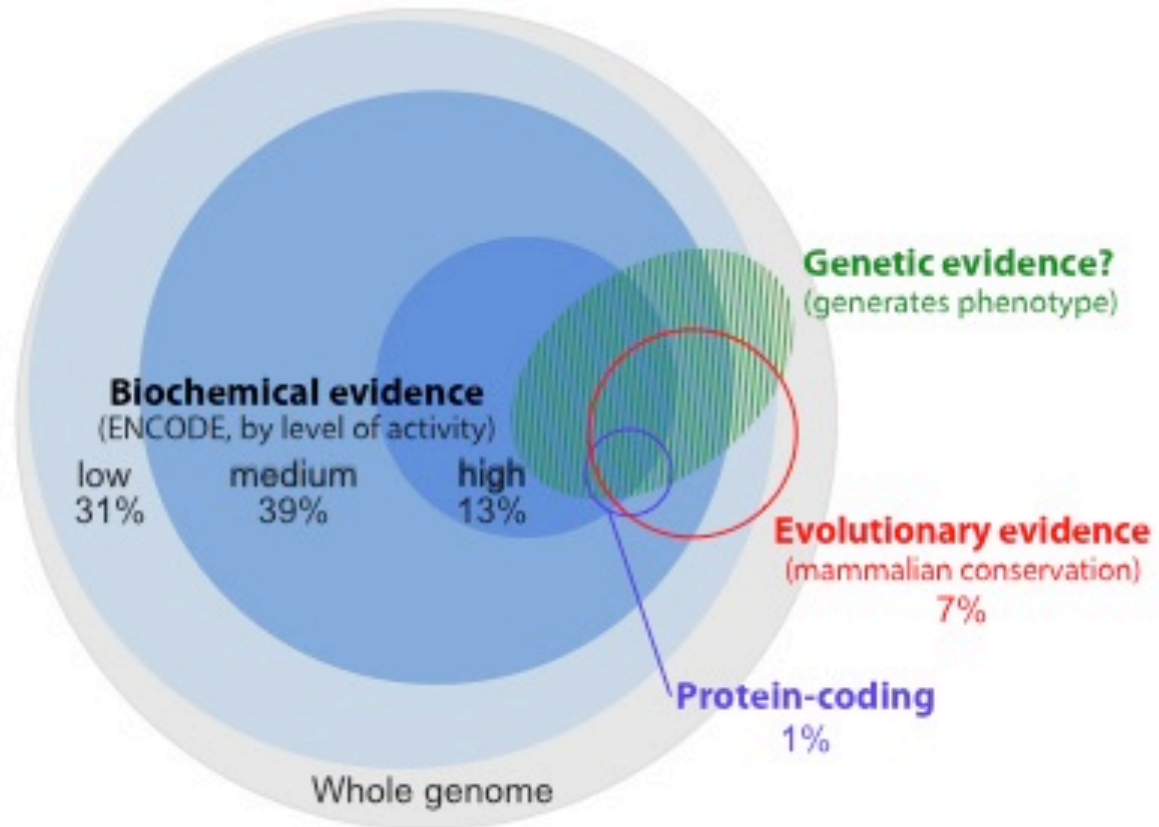
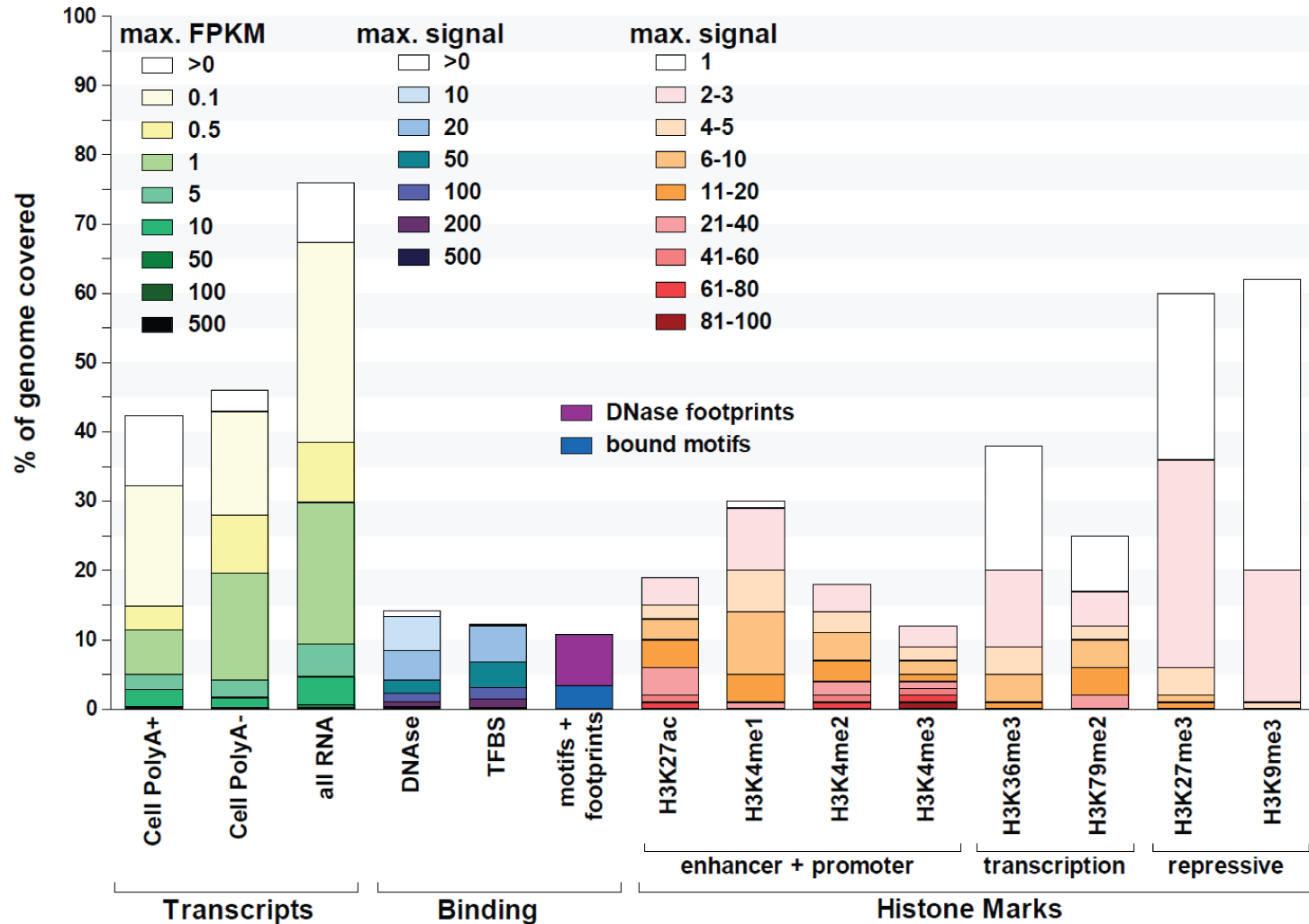
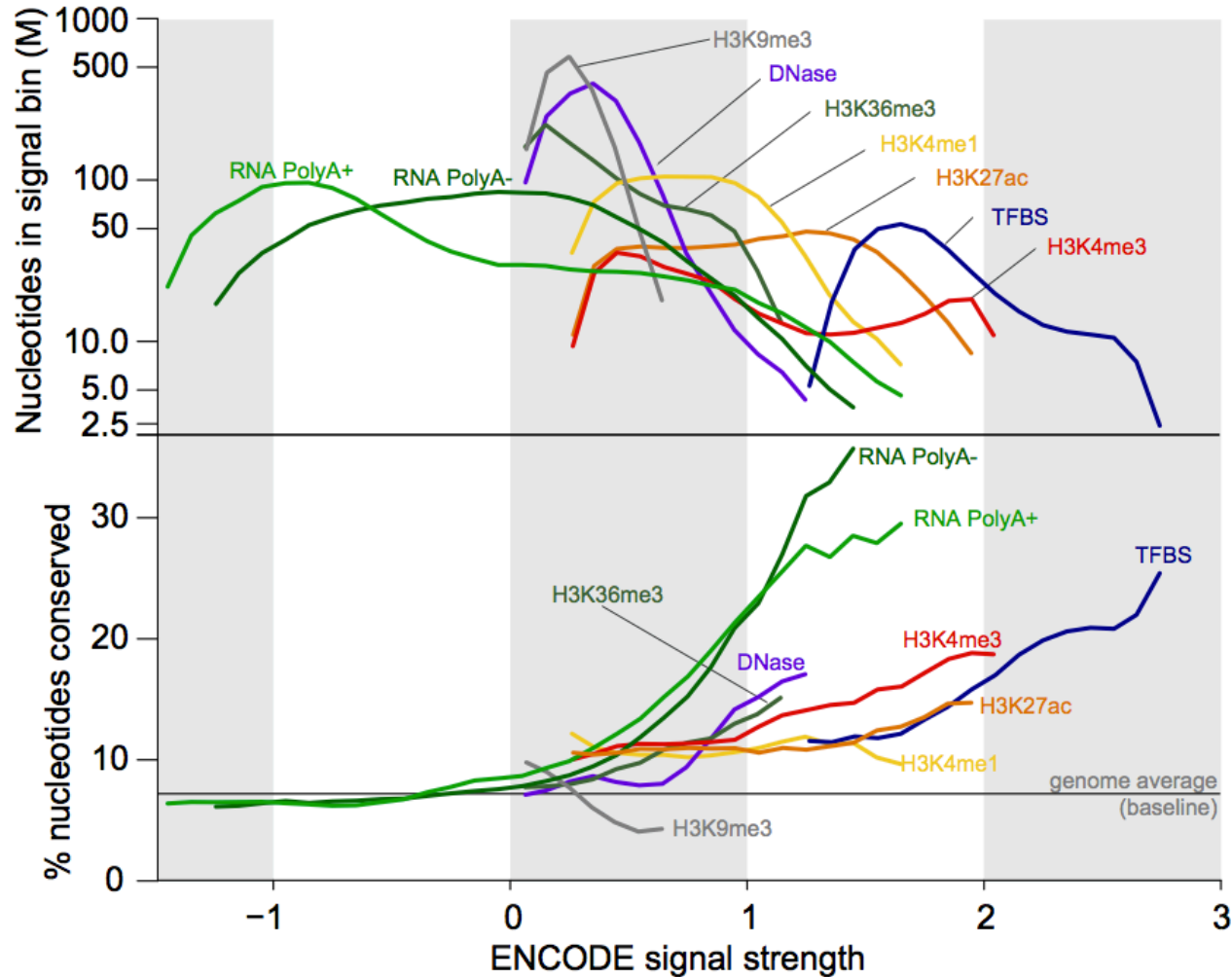


Figure 1 - The complementary nature of evolutionary, genetic, and biochemical evidence

Much of the biochemical activity covering a substantial portion of the genome is low level



Higher signals for biochemical activity in more conserved genomic regions



Conclusions

- 1) There is extensive variation in regulatory information among humans
- 2) Enhancers are the most variable
- 3) Protein variation is also extensive in humans
- 4) Master variators distinct control regulatory programs in different individuals/species
- 5) Mice and humans different more at the level of species than tissues (by gene expression)
- 6) Biochemical activity and conservation is not equal

Acknowledgements

TF and Chromatin Regulatory Variation:

Maya Kasowski, Sofia Kyriazopoulou-Panagiotopoulou, Fabian Grubert, Judith Zaugg, Anshul Kundaje, Yuling Liu, Alan Boyle, Cliff Zhang, Fouad Zakharia, Damek Spacek, Jingjing Li, Dan Xie, Lars M Steinmetz, John Hogenesch, Manolis Kellis, Serafim Batzoglou

Protein Variation:

Linfeng Wu, Sophie Candlie, Yoonha Ma, Lihua Jiang, Jennifer Li-Pook Than, Hua Tang

Master Variator

Jen Gallagher, Wei Cheng

Mouse ENCODE

Shin Lin, Yiing Lin, Yong Cheng, Ross Hardison, Mouse ENCODE Consortium

