

Applications and issues of large scale transcriptome profiling experiments



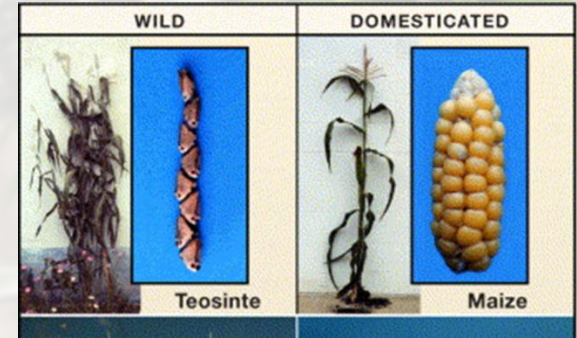
Outline

- **Co-expression and expression conservation**

- Reshaping of the maize transcriptome by domestication (Swanson-Wagner et al PNAS 2012)

- Variation among networks

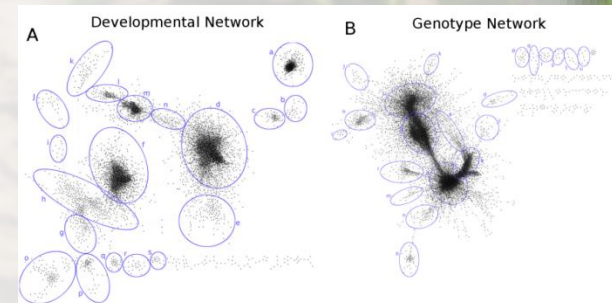
- RNAseq vs microarray



- Enabling usage of co-expression networks to study natural variation

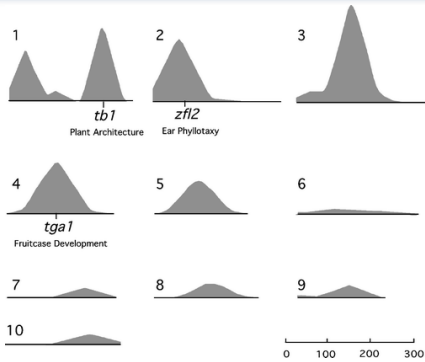
- eQTL hotspots

- Phenotypic QTL



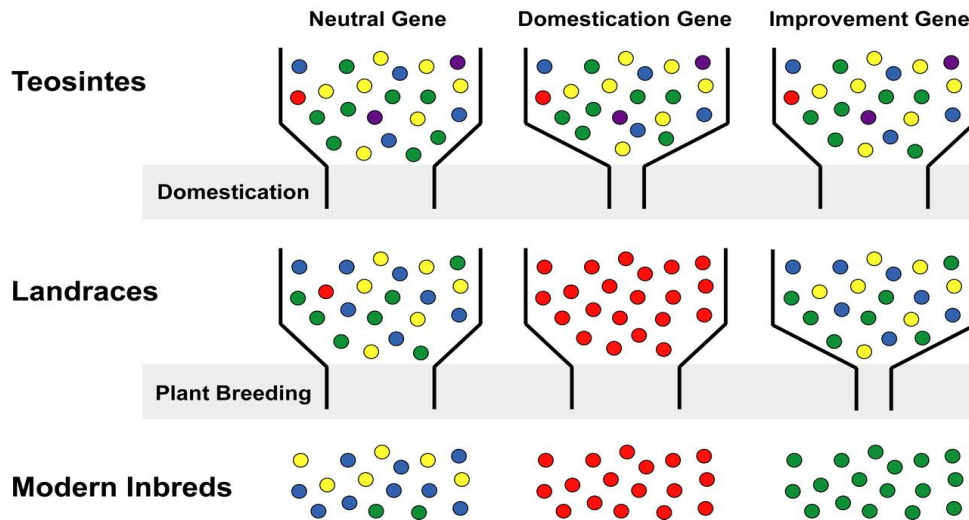
- Integration of transcriptome and epigenome

Identification of loci involved in domestication



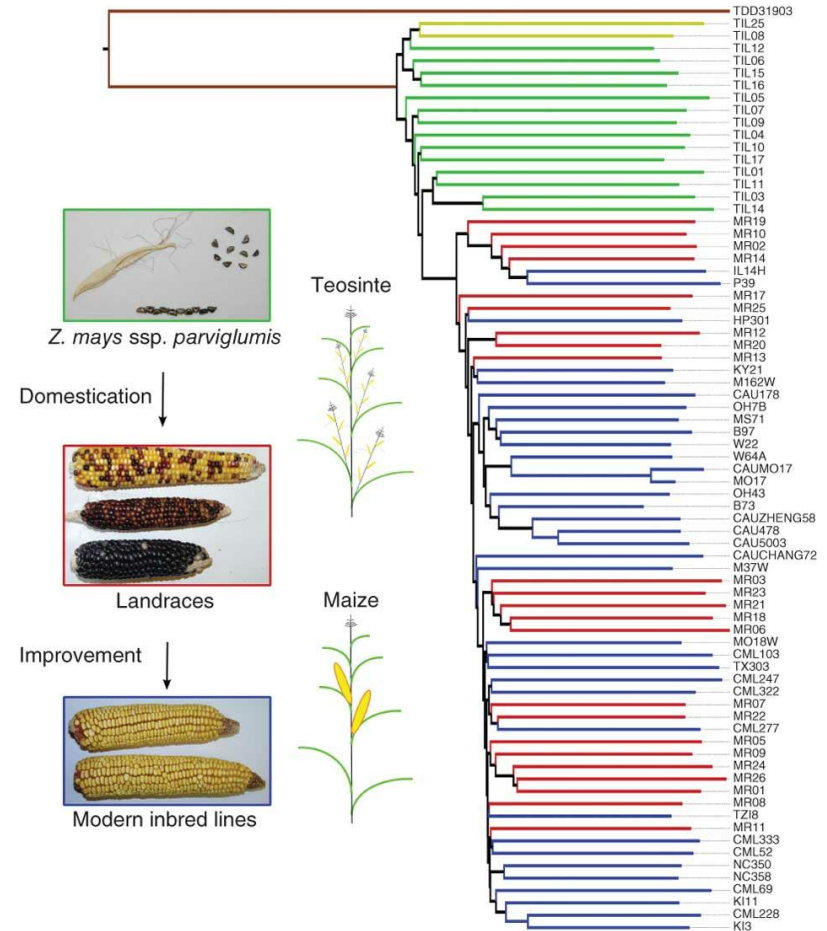
A QTL analysis focused on nine traits that measure plant and inflorescence architecture in a cross of maize vs teosinte find six major effect loci: Doebley 2004 Ann Rev Genetics

Genomic scans for selection (diversity scans)



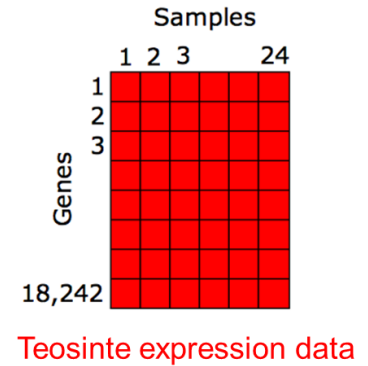
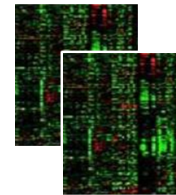
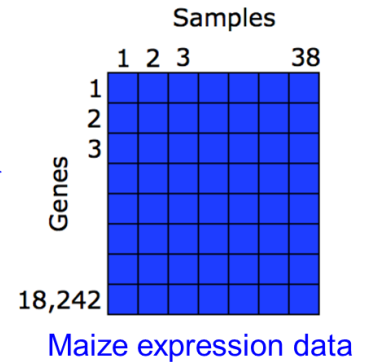
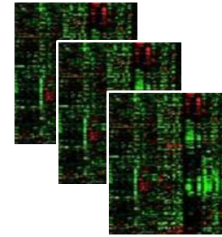
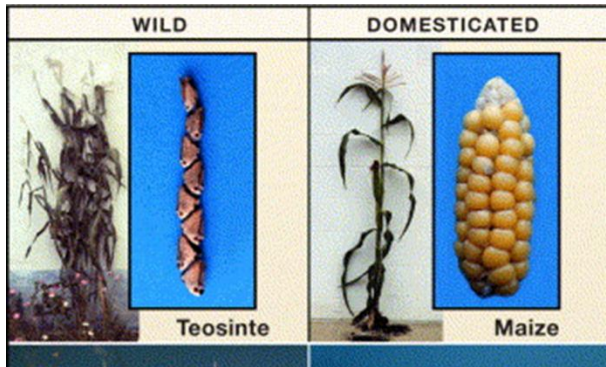
Wright et al., 2005: Identified ~30 targets of domestication

Yamasaki et al., 2007



Hufford et al., 2012
 Re-sequence 75 genomes
 Identified ~500 selected regions
 (1754 genes)

Collection of expression data

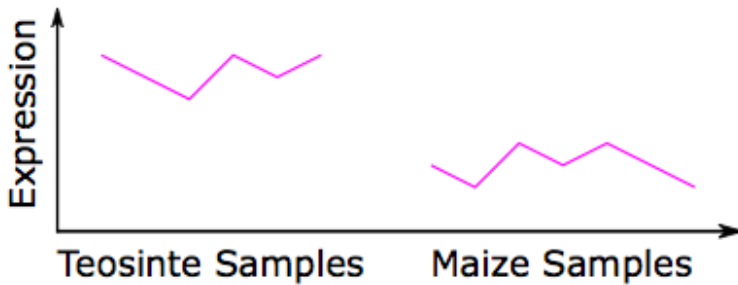
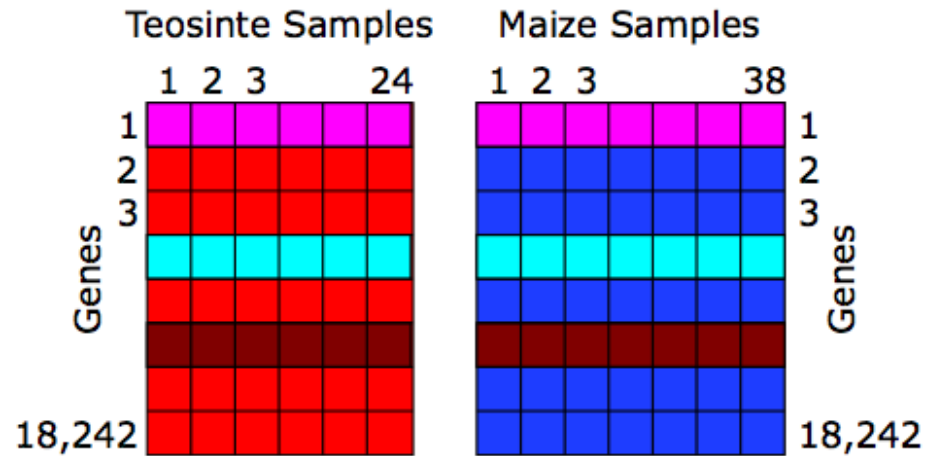


Ruth Swanson-Wagner

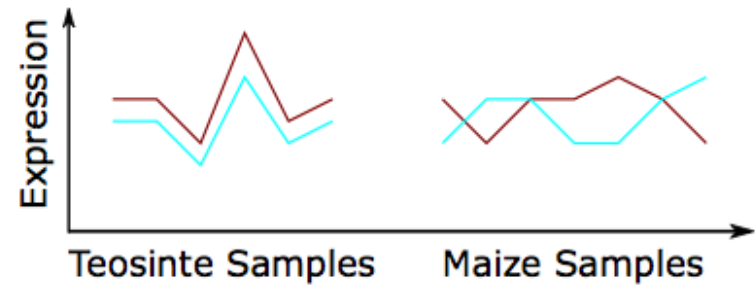
- Maize: 38 genotypes assayed (23 NAMs and other diverse maize inbreds)
- Teosinte: 24 genotypes profiled (7 TILs and 17 “wild” individuals)
- Seedling expression assayed by using custom NimbleGen array with 3-4 probes each for ~32,500 4a.53 filtered gene set

Finding differences in expression data

Expression Data



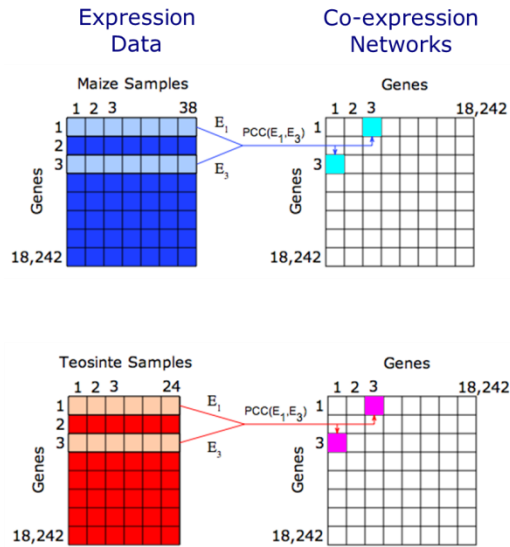
Differentially Expressed Genes



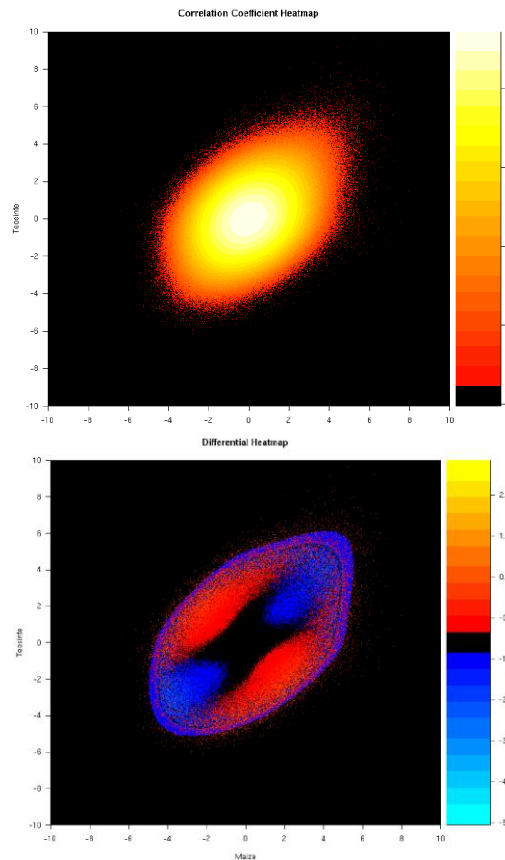
Differential Covariance

Re-wiring of transcriptome in maize

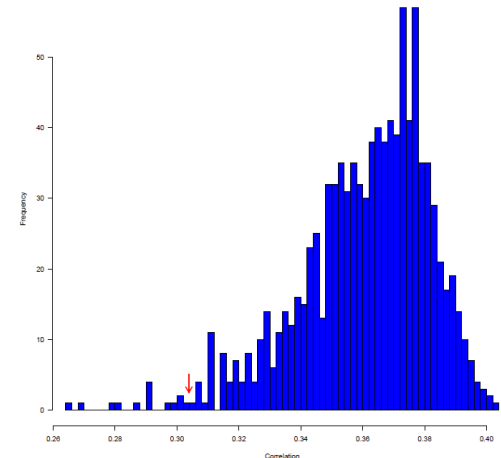
- Generate co-expression networks in maize and teosinte
- Assess network similarity and per-gene expression conservation (EC)



- Co-expression network records similarity between each pair of gene expression profiles.
- Fisher transformation and normalization (Huttenhower et al., 2006)

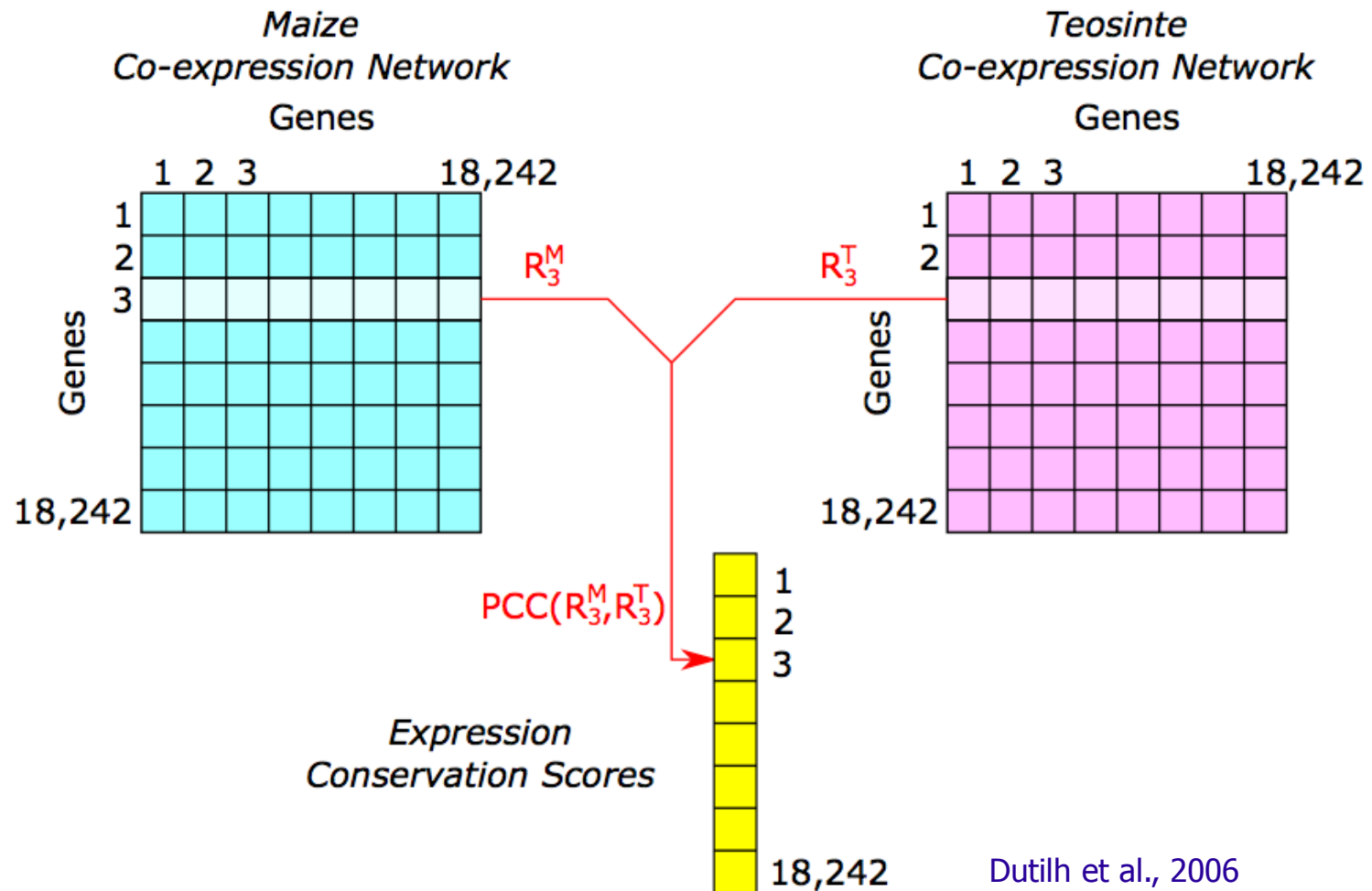


Roman Briskine



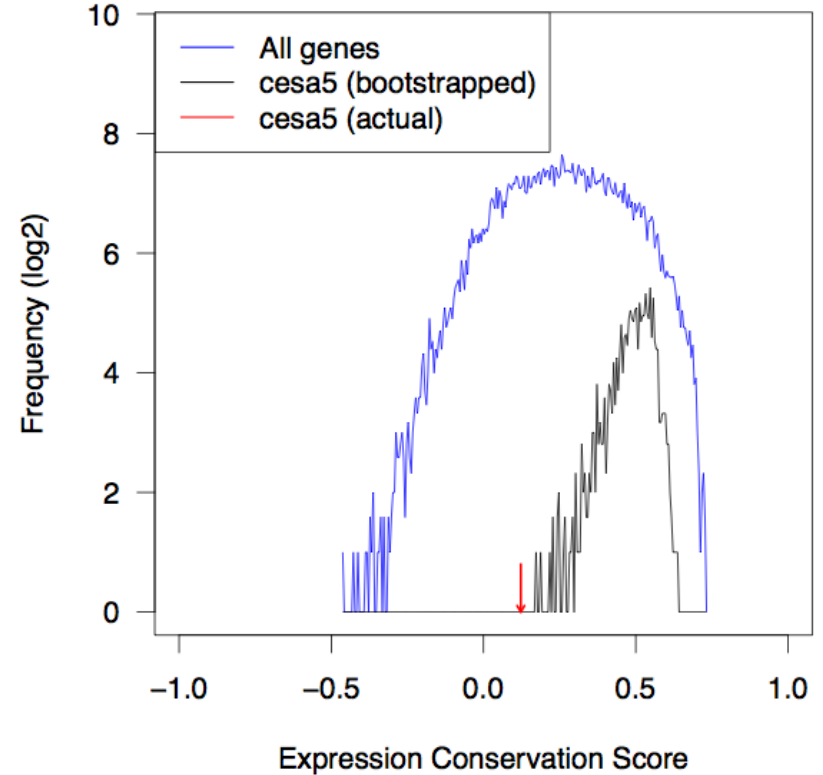
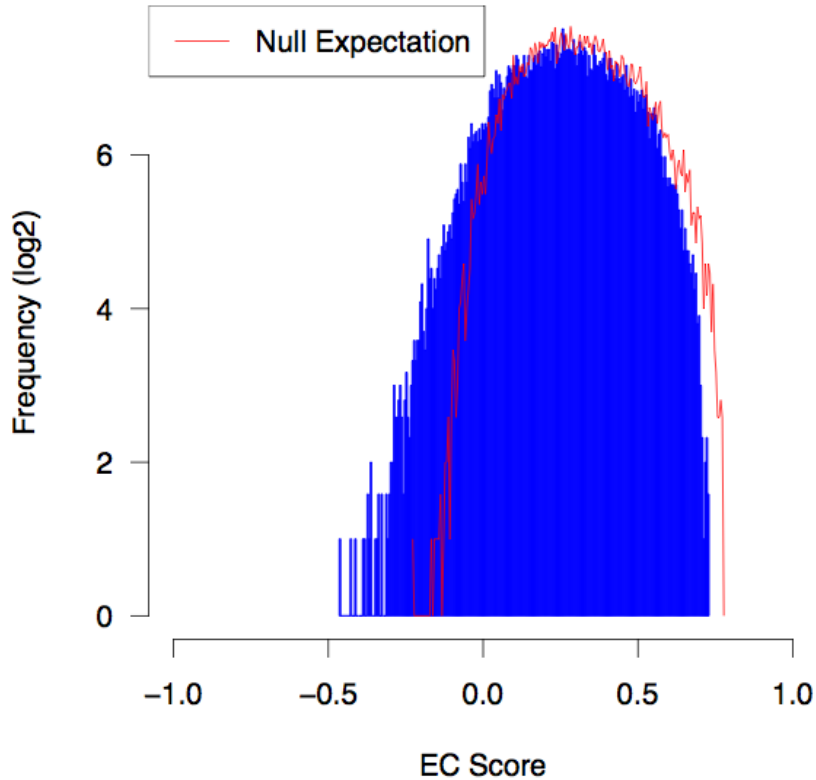
Transcriptome is significantly re-wired

Expression conservation score measures similarity between gene's co-expression profiles in two networks



Dutilh et al., 2006

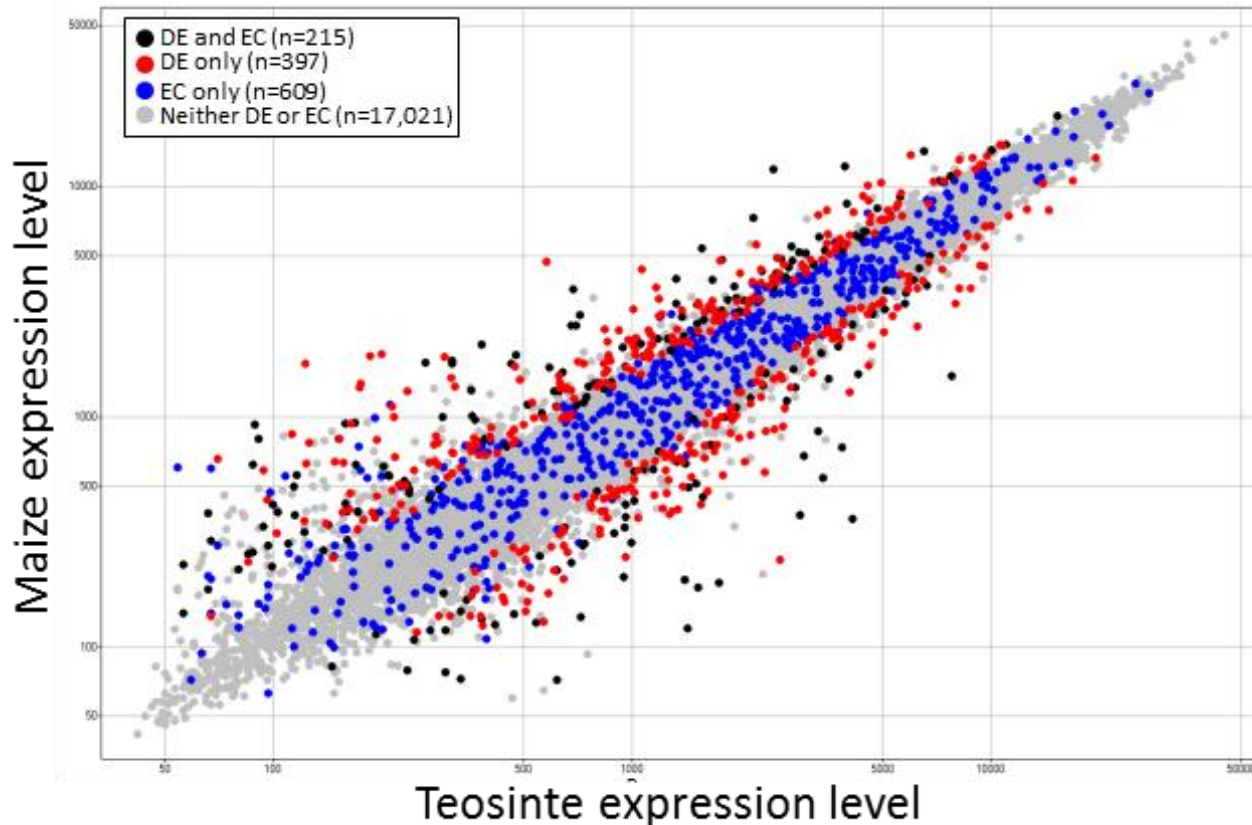
Identification of genes with significant differences in EC



$$z = \frac{EC - \mu_{null}}{\sigma_{null}}$$

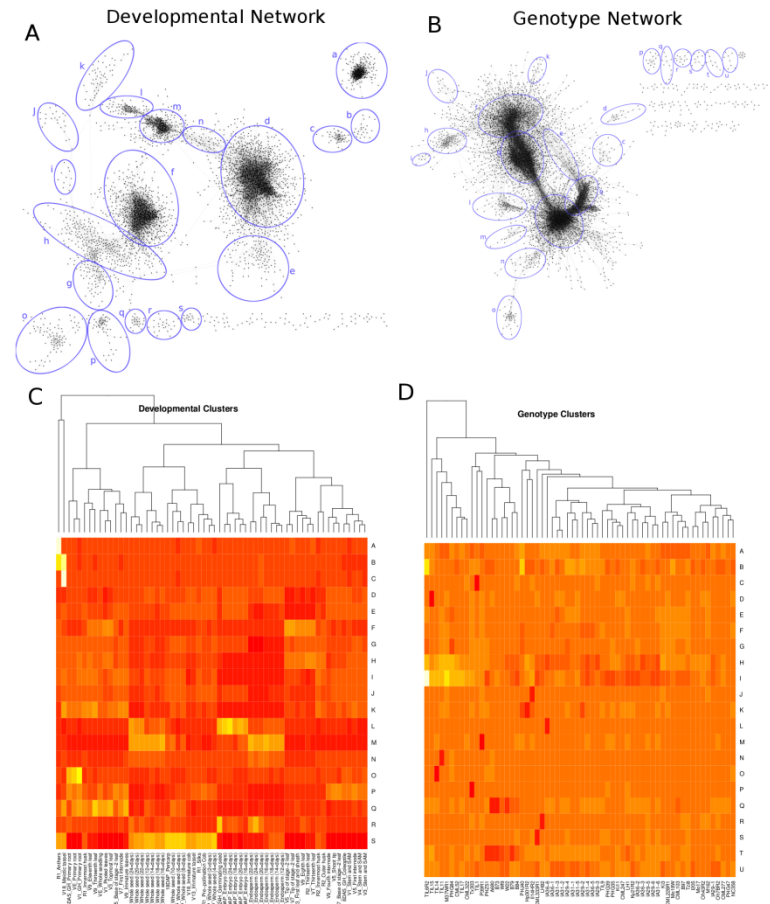
EC and DE approaches identify different expression changes

- 18,224 expressed genes assessed
- 612 DE genes (enriched for targets of selection)
- 824 AEC genes
- 215 in common (enriched for targets of selection)



Co-expression networks

- Co-expression analysis identifies genes with similar patterns of expression: **Relies upon variation in gene expression**
- Should we be using the “kitchen sink” approach or developing multiple networks?

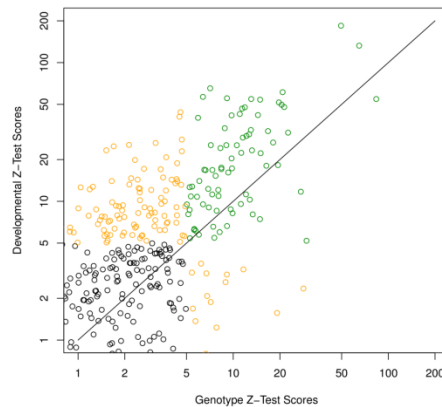
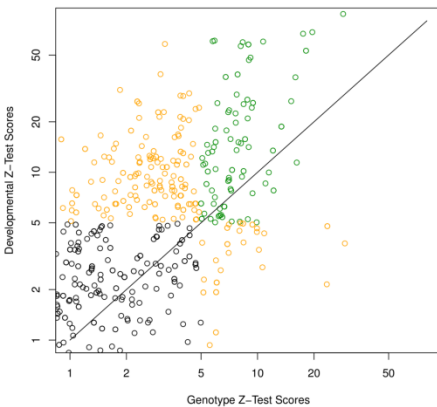


60 tissues of B73
 Sekhon et al., 2011
 protein catabolic processes
 organic substance transport
 cell wall modification

62 genotypes
 Swanson-Wagner et al 2012
 electron transport chain
 glucose metabolism
 response to biotic stimulus

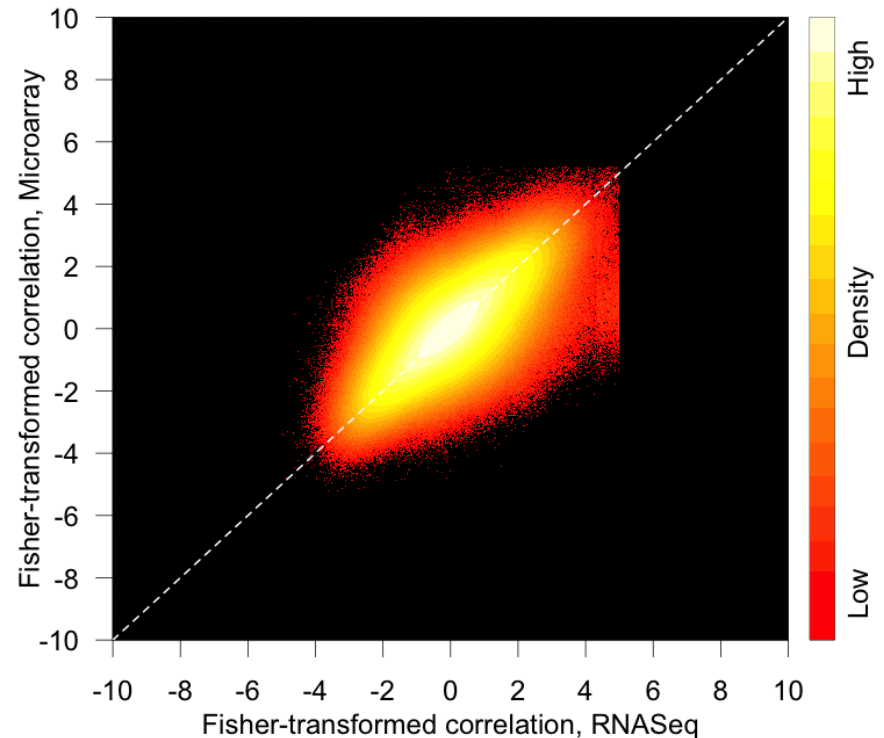
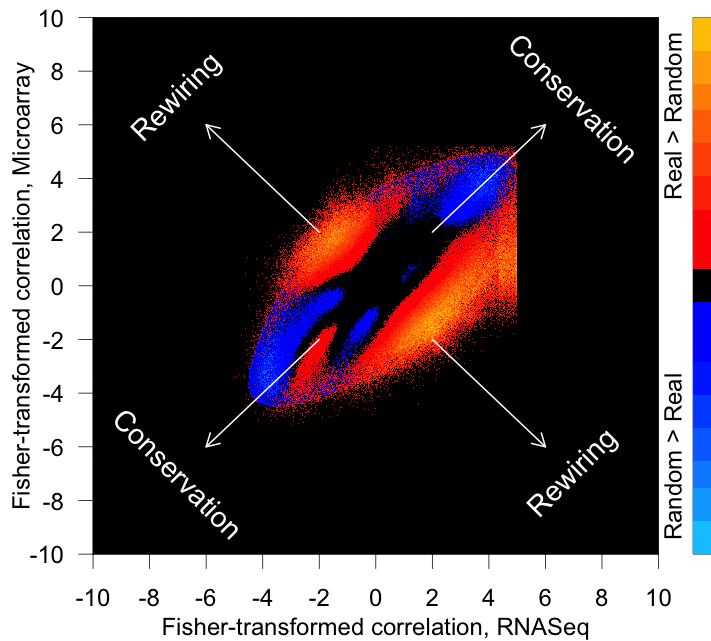
Enrichment For GO Terms

Enrichment For MapMan Annotations

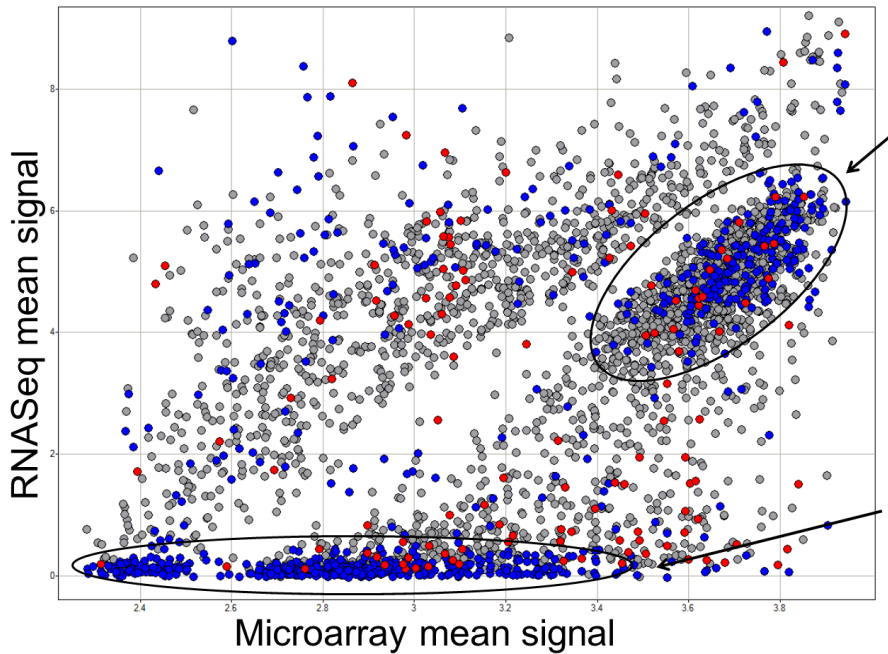


Co-expression networks: RNAseq vs microarray

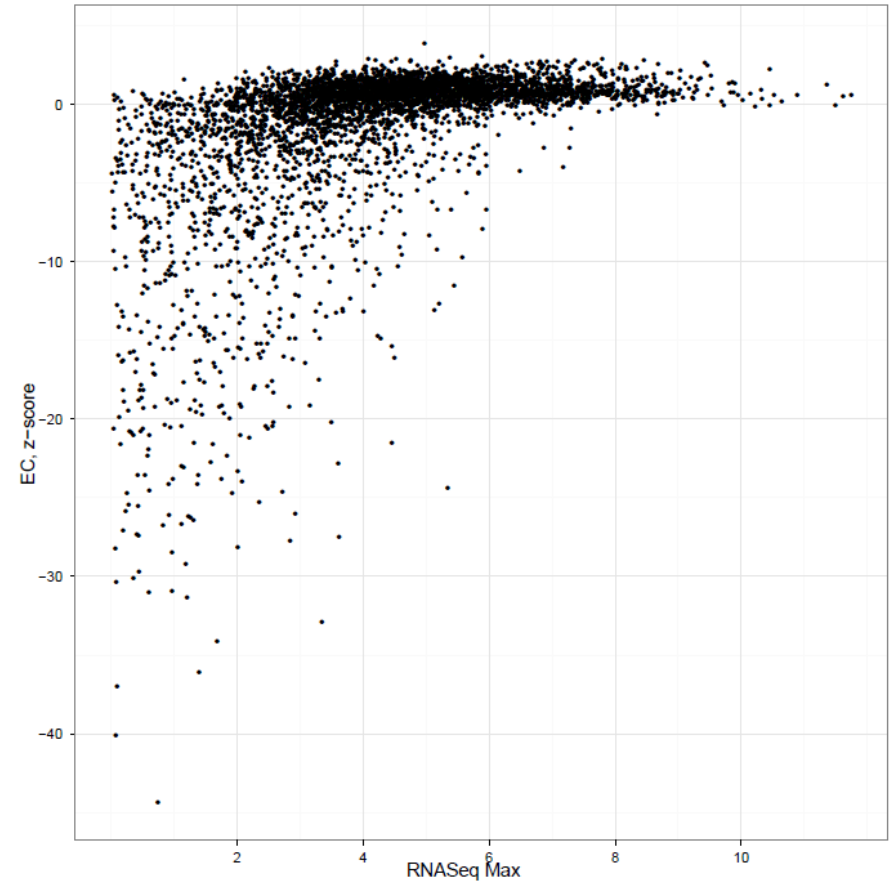
- Comparison of microarray and RNAseq data: Co-expression
 - 18 samples from different tissues of B73
 - Selected 19,328 “expressed” genes from microarrays



Expression conservation: RNAseq vs microarray



Maximum expression level vs. Expression Conservation z-score



➤ 3,354 genes with different expression co the RNAseq and microarray networks ($p < C$)

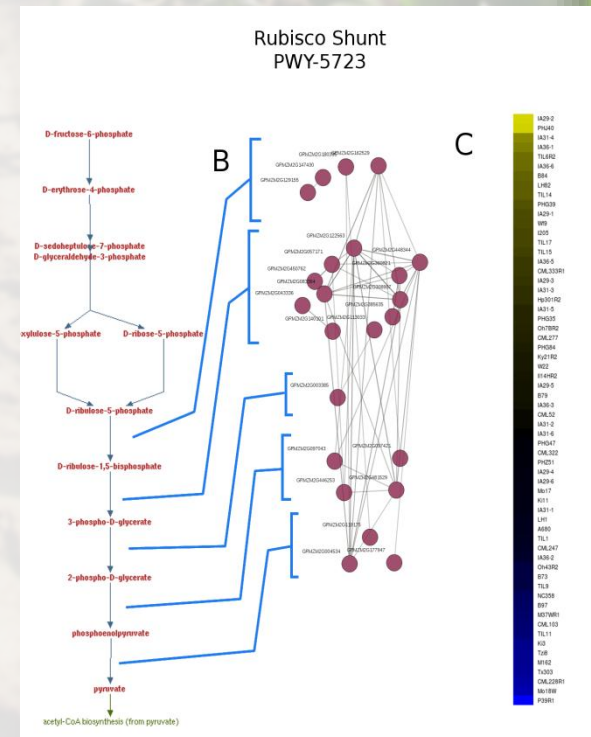
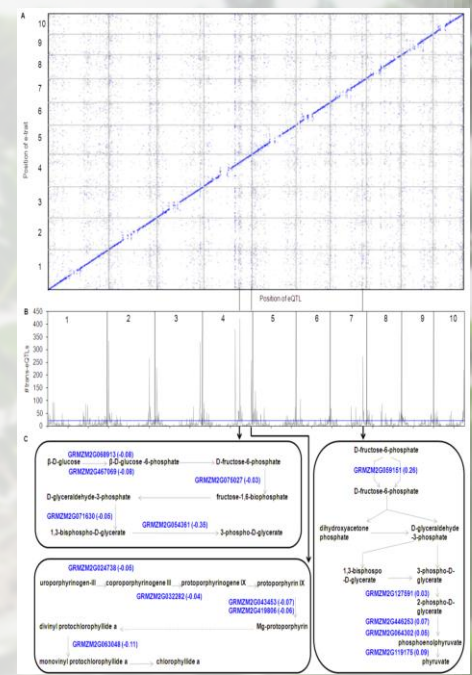
➤ **Red**: 122 more connected in the microarray network

➤ **Blue**: 796 genes connected in the RNAseq network (circled)

➤ **Grey**: Equally connected in both networks

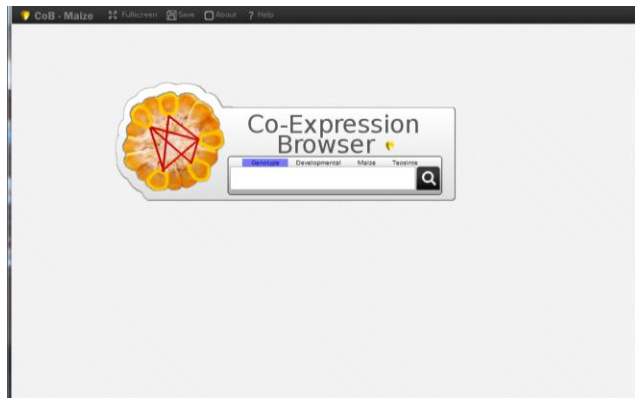
Outline

- Co-expression and expression conservation
- Enabling usage of co-expression networks to study natural variation
 - Simple user queries of networks
 - eQTL hotspots
 - Phenotypic QTL
- Integration of transcriptome and epigenome

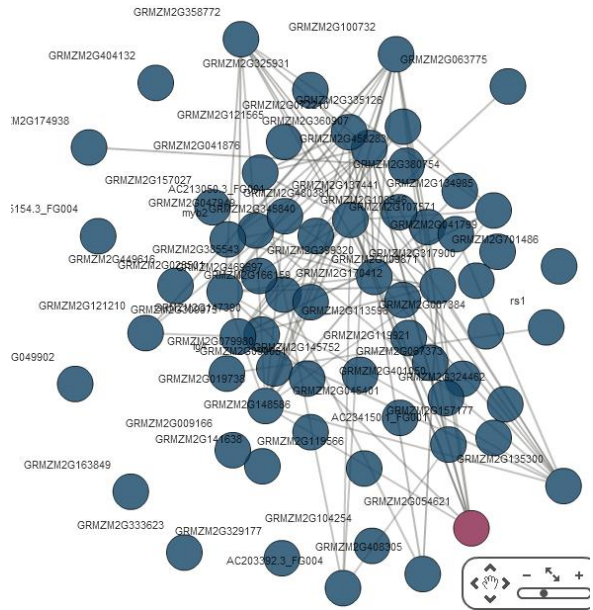
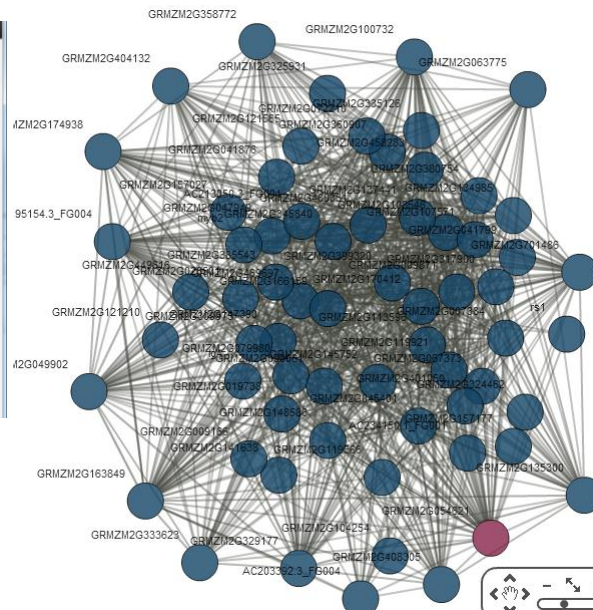


COB: A viewer to query co-expression networks with genes and coordinates

- <http://csbio.cs.umn.edu/cob/>
- Allows user to query various networks with gene(s) and then to visualize genomic coordinates or overlap between networks

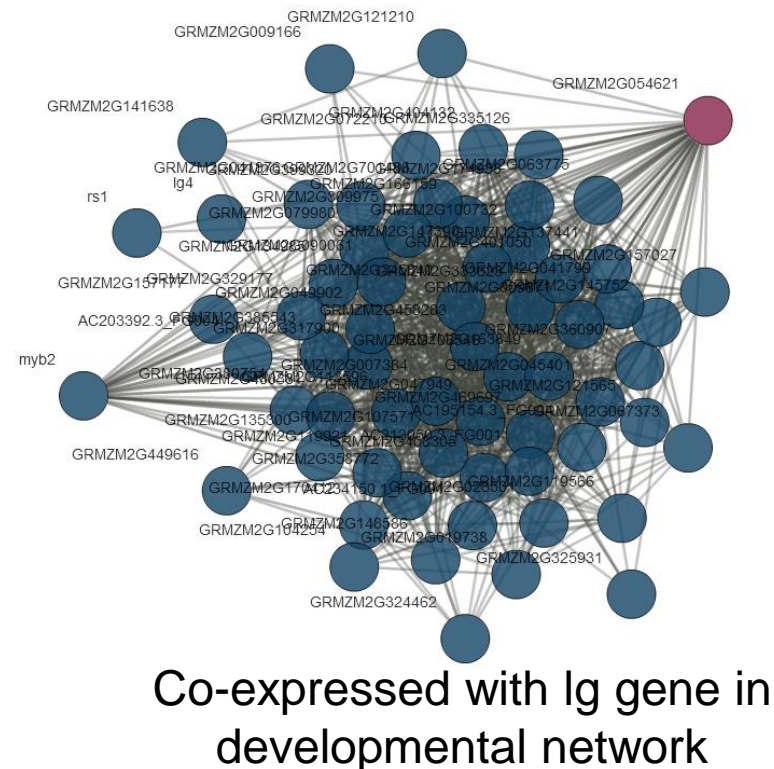
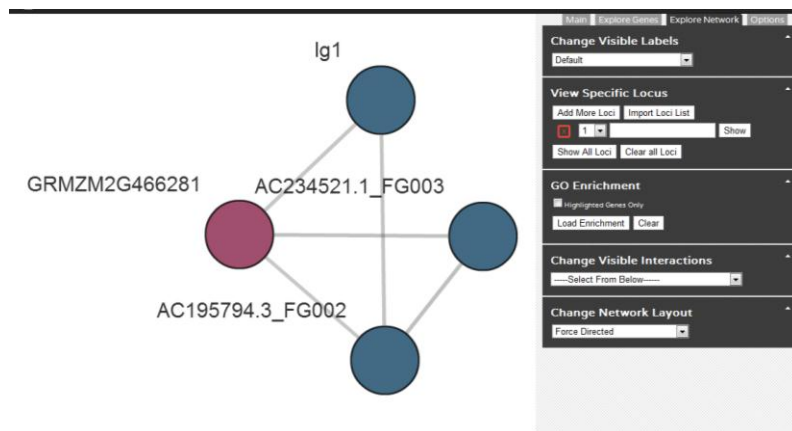


Rob Schaefer



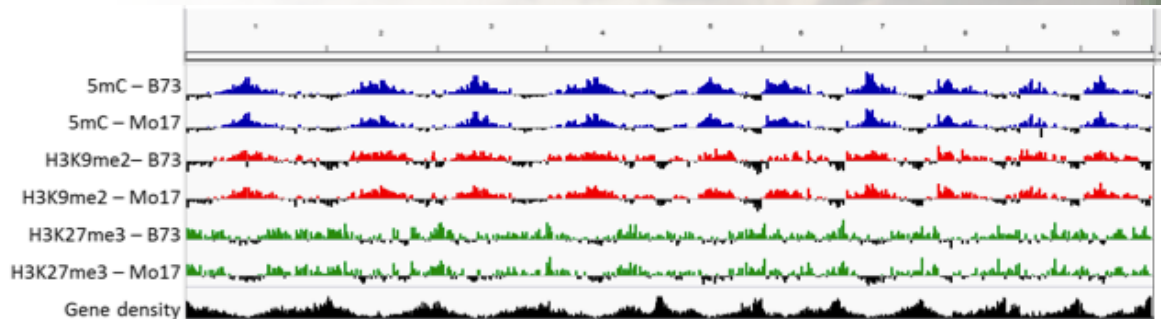
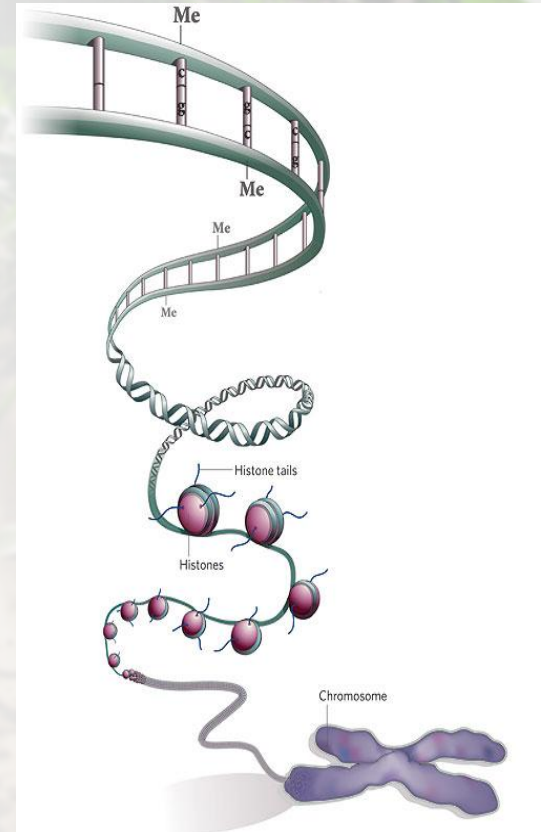
Co-expression networks: Phenotypic QTL

- Tian et al (2011) identified ~30 QTL for leaf angle by joint linkage analysis
 - Also performed GWAS
- Two classical maize mutants; lg1 and lg2 likely are molecular bases for two of the QTL (and have significant SNP associations)
- Rest are unknown
- Query co-expression networks to identify genes co-expressed with lg genes and located within QTL



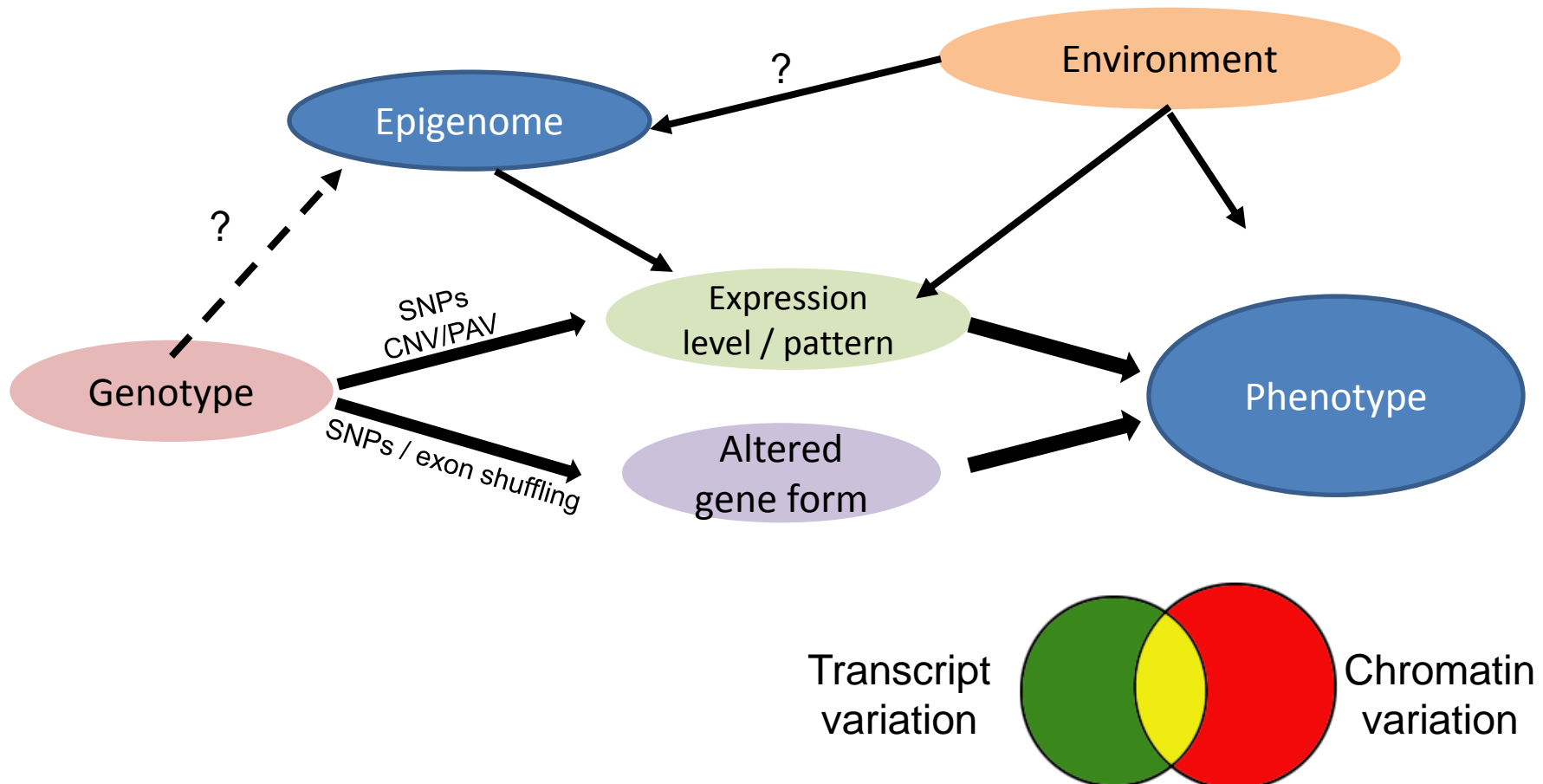
Outline

- Co-expression and expression conservation
- Enabling usage of co-expression networks to study natural variation
- **Integration of transcriptome and epigenome**
 - Different data types
 - How to isolate contribution of epigenome to transcriptome variation



Transcriptome profiling provides critical information for understanding phenotype

- What proportion of expression level variation is attributable to epigenome?



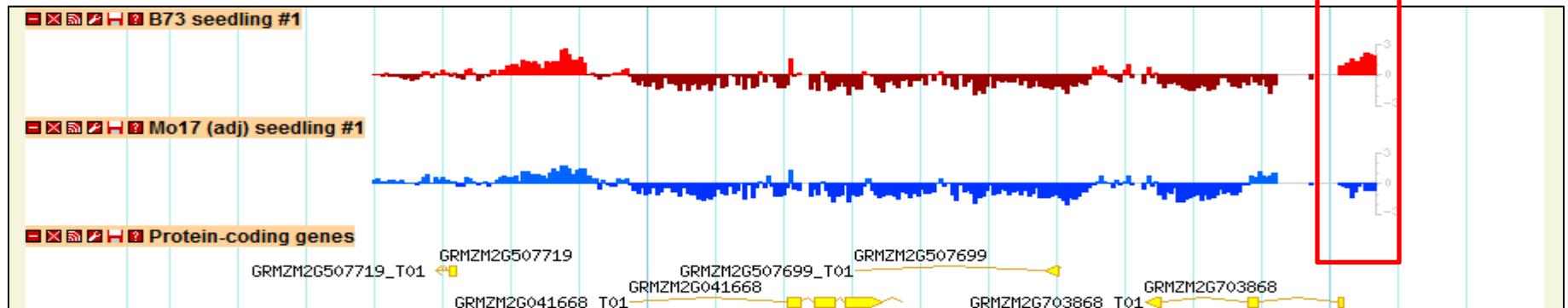
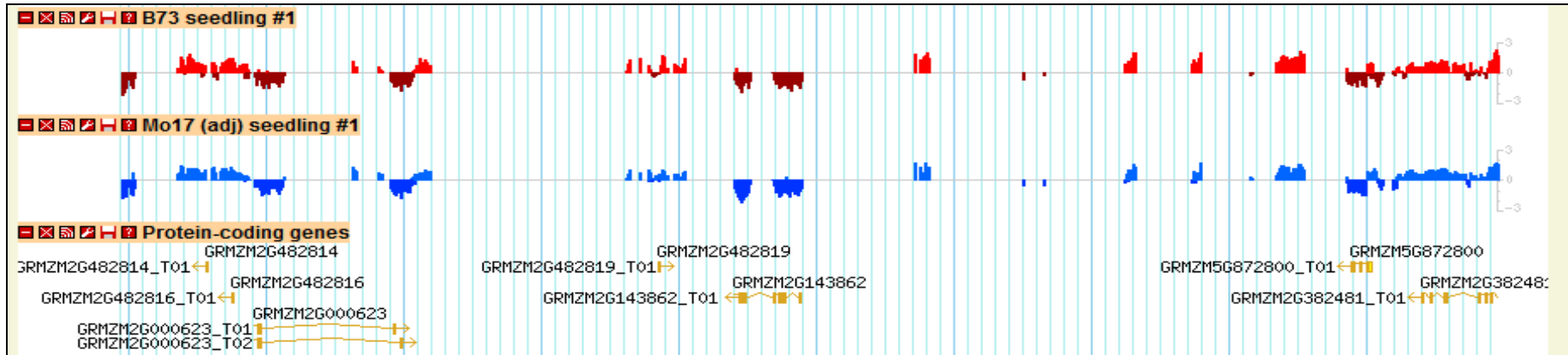
Data and questions

- Data types:
 - meDIP-chip (DNA methylation) [n=~140 profiles]
 - ChIP-chip (H3K9me2; H3K27me3) [n=~75 profiles]
 - RNAseq 120 samples (20-25 million reads each)
- Samples
 - Five tissues for two genotypes
 - 1 tissue for 25 genotypes
- Identification of initial variation (two samples with replicates) easy
- How to collapse and classify variation in large population more difficult
- Overlap? (lots of samples, not requiring complete correlation)
 - Chromatin marks and expression
 - Chromatin marks and SNPs

Limited variation for DNA methylation patterns

Generally very similar patterns of DNA methylation

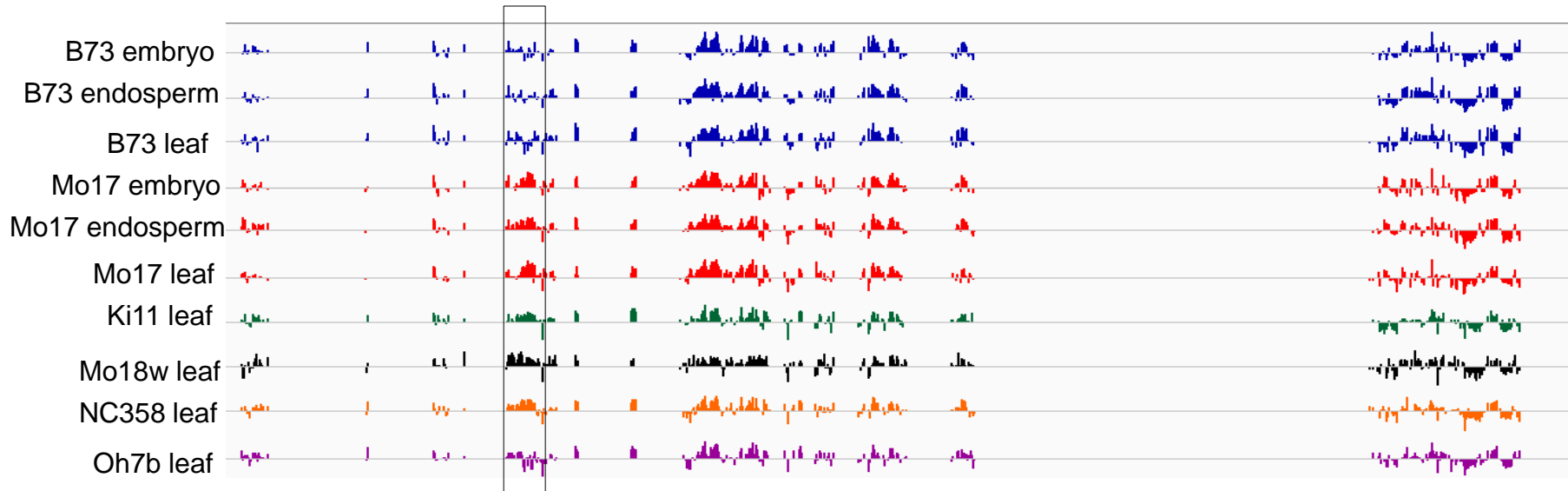
Increased methylation near repetitive sequences; decreased methylation near genes



Regions with extremely different methylation profiles can also be found
~1000 DMRs in B73 vs Mo17

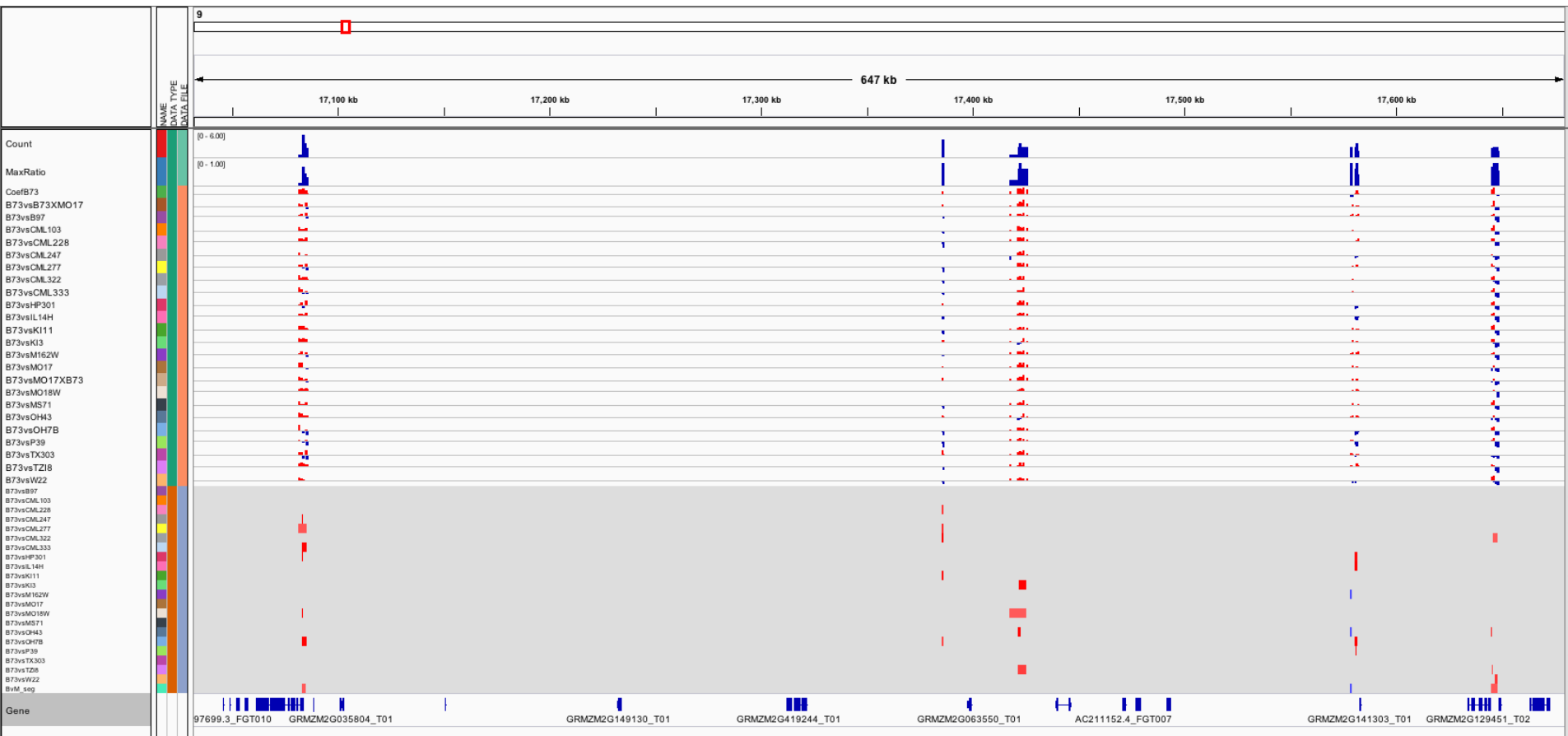
What happens in other maize genotypes?

- DNA methylation patterns are generally quite similar among genotypes and tissues.
- However, there are ~1000 DMRs between any two genotypes.
- Variation frequently acts equally upon all tissues.

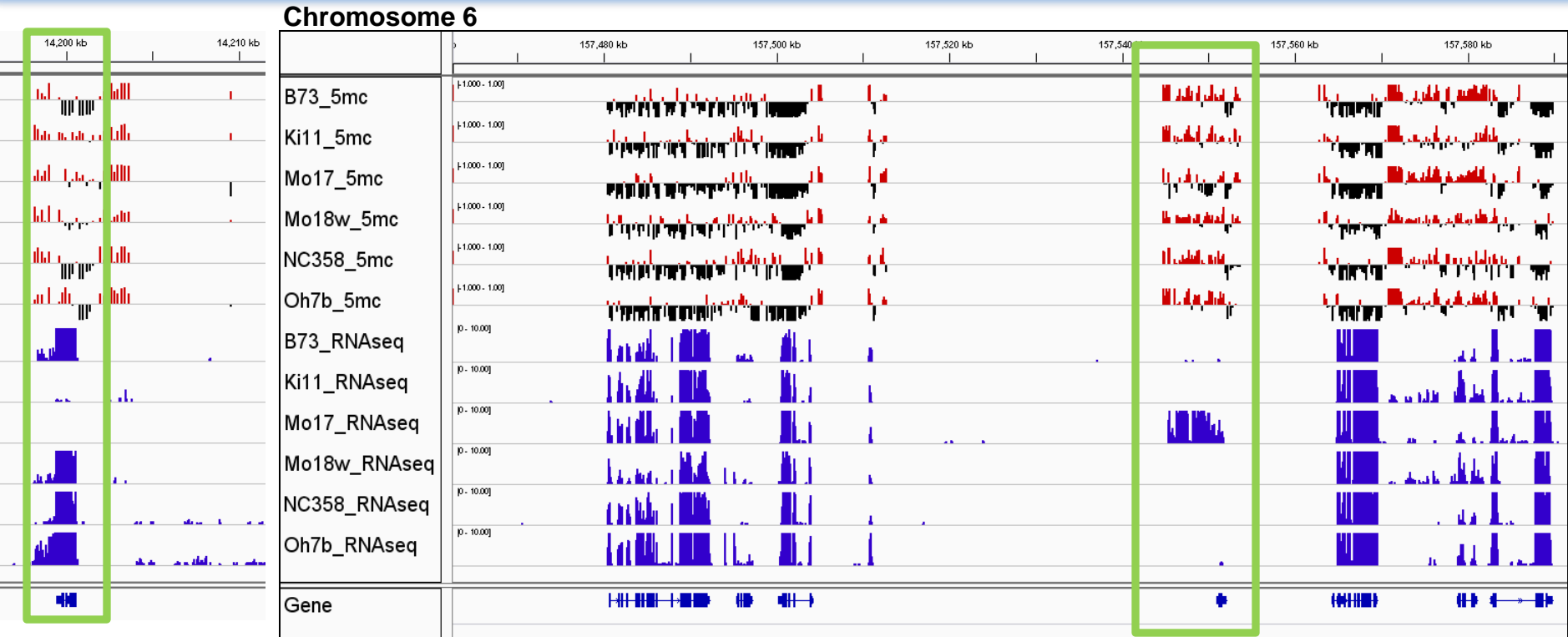


What happens in other maize genotypes?

- Call DMRs between two genotypes
- Need tools for simultaneously defining regions and classifying among all genotypes



Once DMRs are found: Causes and Effects



- What is causing chromatin change? Is it associated with SNPs? Rare phenotype problem
- Does the chromatin change cause an expression change? What about partial correlations?

Summary

- **Making sense of differences among populations**
 - Co-expression and expression conservation
- **Enabling usage of –omics datasets (transcriptome, epigenome, etc)**
 - Interrogation tools
 - Visualization tools
- **When is enough enough?**
 - Allele-specific expression analysis

Thanks!

- Steve Eichten
 - Irina Makarevitch
 - Amanda Waters
 - Ruth Swanson-Wagner
 - Peter Hermanson
 - Matthew Vaughn
 - NSF DBI# 0922095
- Chad Myers
 - Roman Briskine
 - Rob Schaefer
 - Shawn Kaeppler
 - Robin Buell
 - Rajandeep Sekhon
 - Candy Hansey
 - Lin Li
 - Gary Muehlbauer
 - Patrick Schnable
 - Mary Gehring
 - Jeffrey Ross-Ibarra
 - Matthew Hufford
 - Peter Tiffin

