

Using Genomics to Dissect Seed Development

Bob Goldberg & John Harada
UCLA & UC Davis



Why Seeds? Our Food is Derived From 14 Major Food Crops & Over Half Produce Seeds For Human and Animal Consumption

Seed Crops



- Wheat
- Rice
- Corn
- Barley
- Sorghum
- Soybean
- Common Bean
- Coconut

22,300 Seed-Bearing Plant Species
(90% of all known plants)

Non-Seed Crops

- Potato
- Sweet Potato
- Cassava
- Sugar Beet
- Sugar Cane
- Banana

\$36.5 Billion Dollars is the Value
of the World Seed Market (2010)

OVER THE NEXT 50 YEARS WE WILL NEED TO DOUBLE THE WORLD'S FOOD SUPPLY IN ORDER TO PRODUCE MORE FOOD THAN IN ALL OF HUMAN HISTORY - Yield, Yield Yield!!!

How Is a Seed Formed?

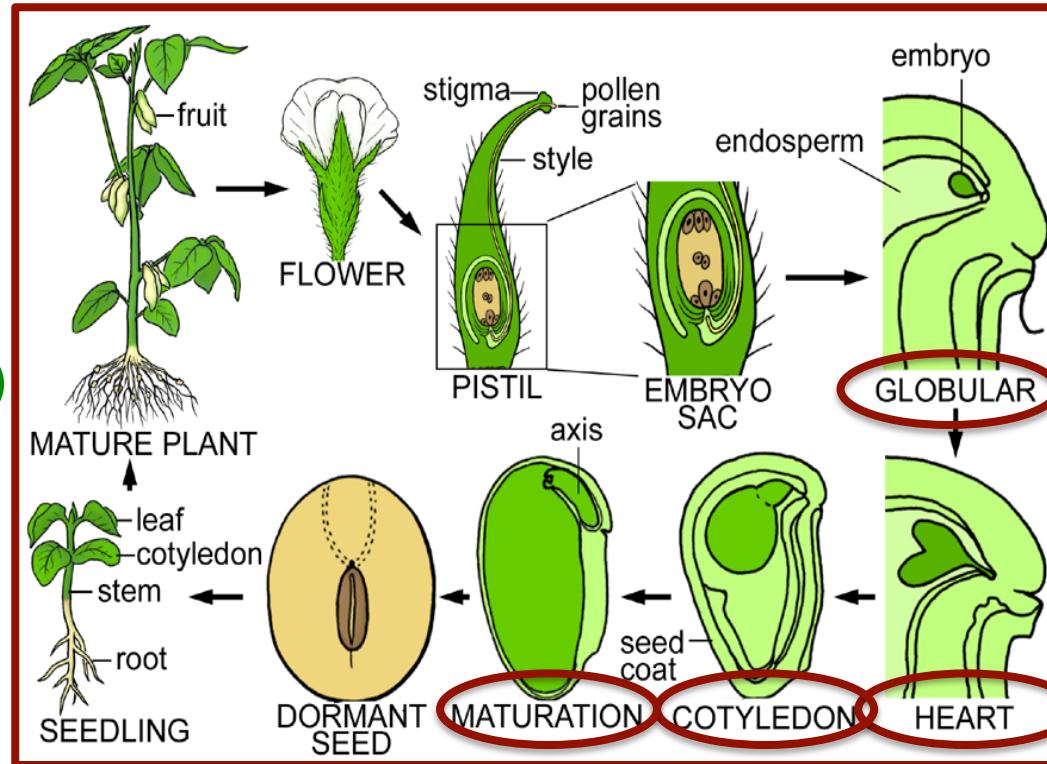


In the Beginning....



What Are the Questions Focused On In This Talk?

Seed Development



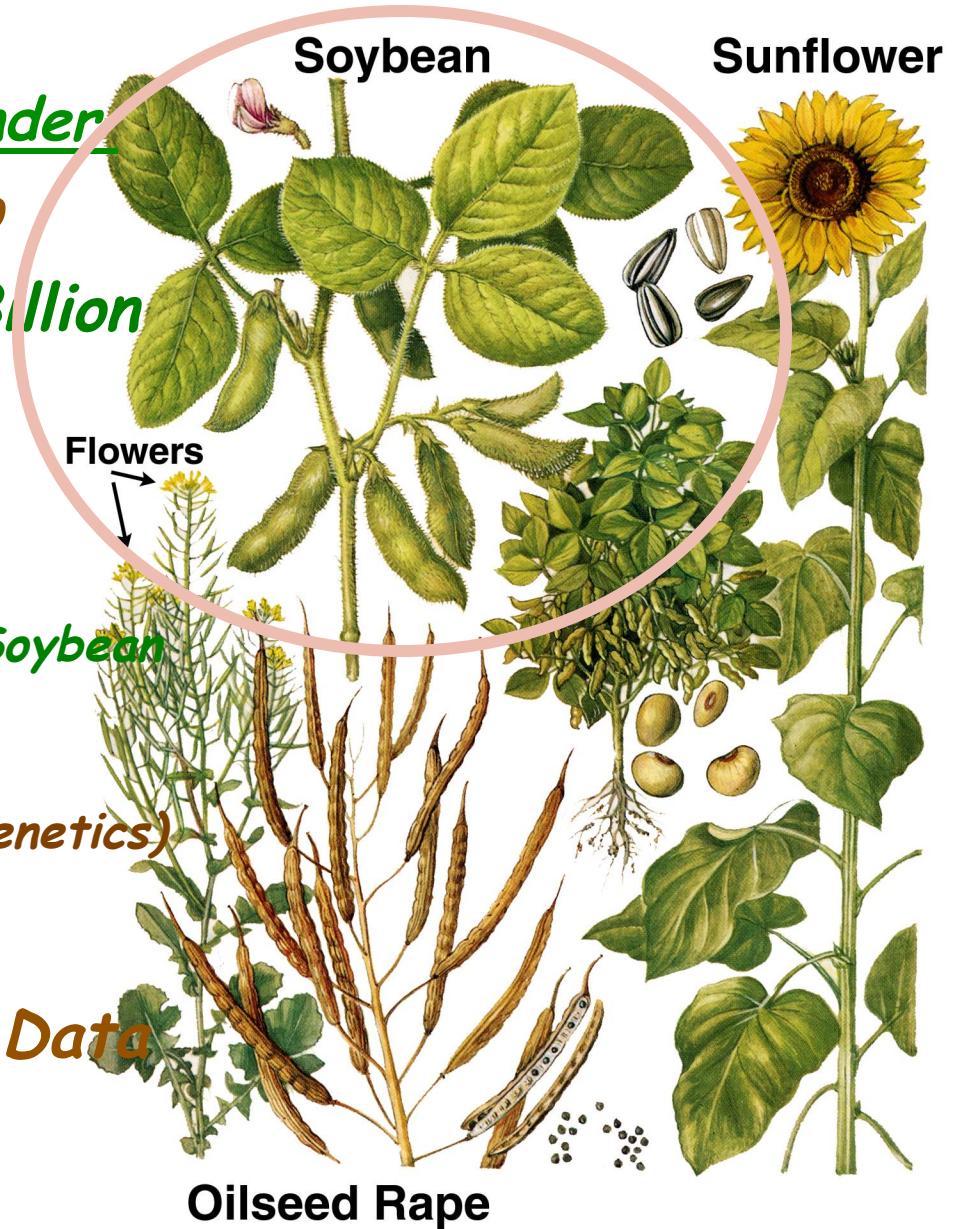
Seed Stages Studied

- What Are the Genes That Are Active In Specific Seed Compartments Throughout Soybean & Arabidopsis Seed Development?
- How Does Gene Activity Change During Seed Development?
- What Biological Processes Are Specific For Different Seed Compartments Throughout Development?
- What Are The Genes And Epigenetic Processes Required To Make A Seed?

Diversity of Oil Seed Plants

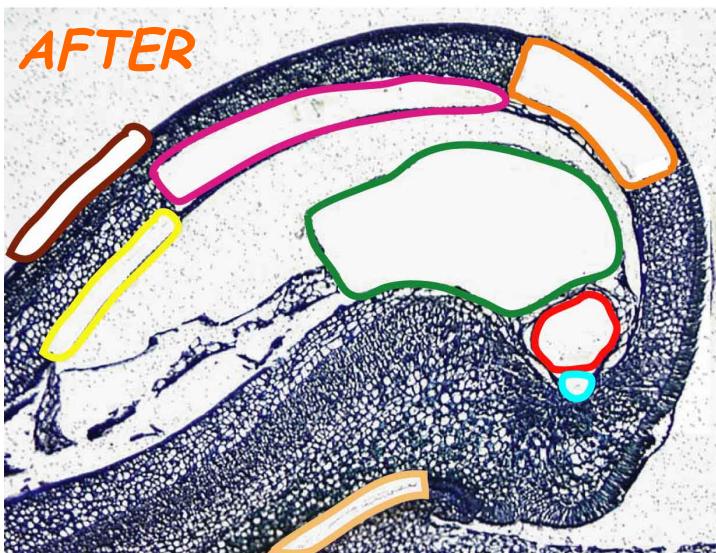
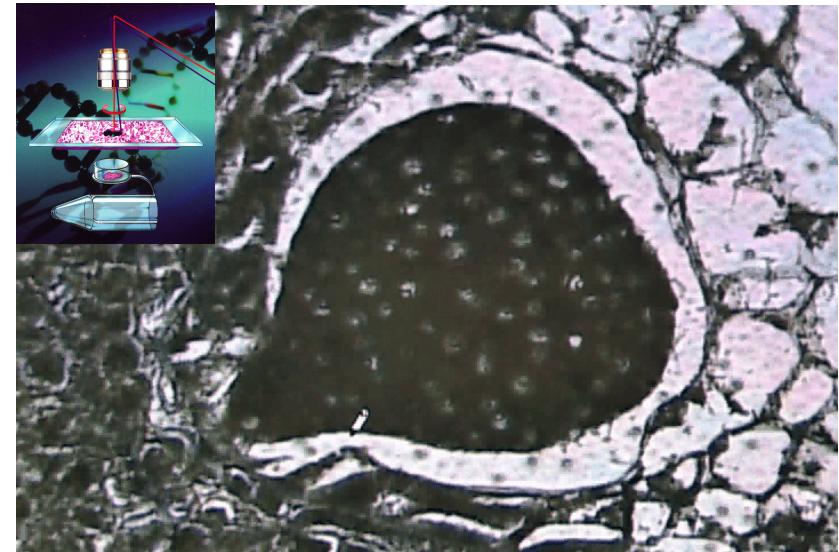
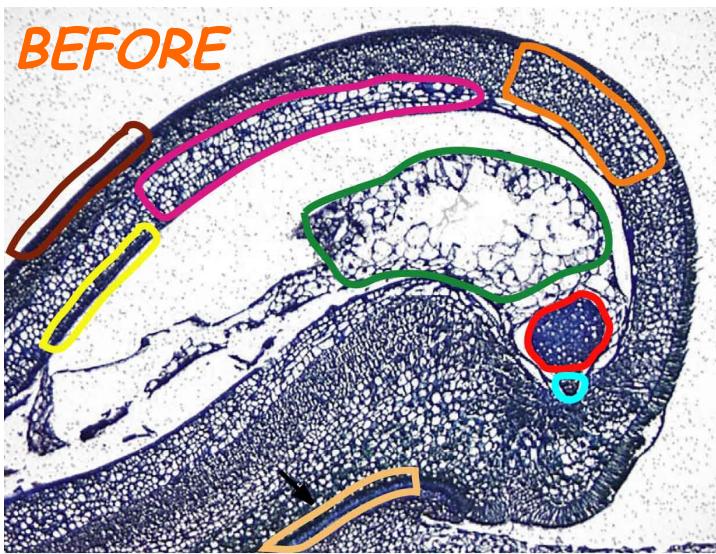
Why Soybean? - A Reminder

- Second Major US Crop
- Total Crop Value \$32Billion
(50% Value Exported)
- Major Food Source
- Important Biofuel Source(Biodiesel~20% of US Soybean Oil Production)
- Excellent Model Plant
(Transformation, Knockdowns, Genetics)
- Genome Sequenced
- Seed Gene Expression Data



How Study Gene Activity in Different Seed Compartments During Development?

Laser Capture Microdissection & Affymetrix GeneChip/RNASeq



All Compartments Of The Seed!!

- Embryo (Embryo-Proper & Suspensor)
- Endosperm
- Seed Coat (All Layers)

GLOBULAR-STAGE SEED AS AN EXAMPLE

How Did We Study Gene Activity in the Seed?

GENECHIP

1st Generation GeneChip



cDNA-Based Array
37,000 Probe Sets
>25,000 Gene Models
~40% of Genome

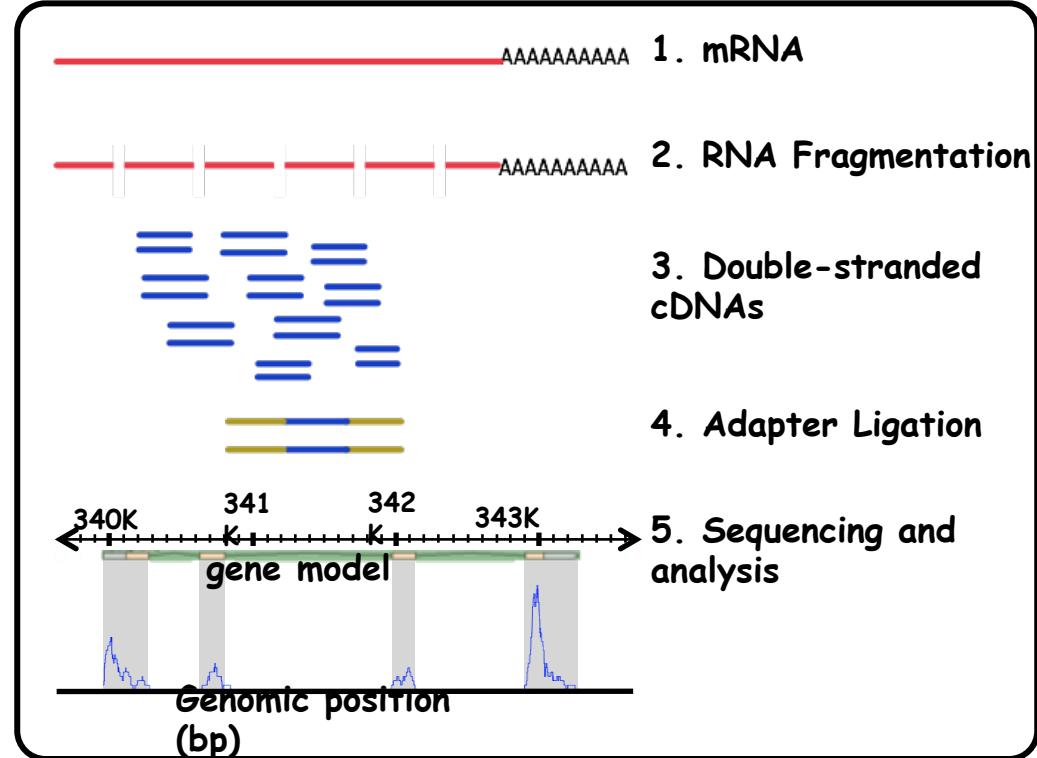
2nd Generation GeneChip*



Whole Gene Array
1.2 Million Probes
>66,000 Gene Models

SEQUENCING

RNASeq



*Note: 2nd generation GeneChips are available and library files can be found at:
<http://seedgenenetwork.net/annotate#soybeanWT>

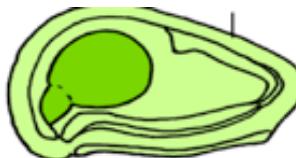
What Developmental Stages and Seed Compartments Studied?



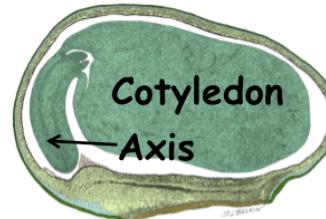
Globular



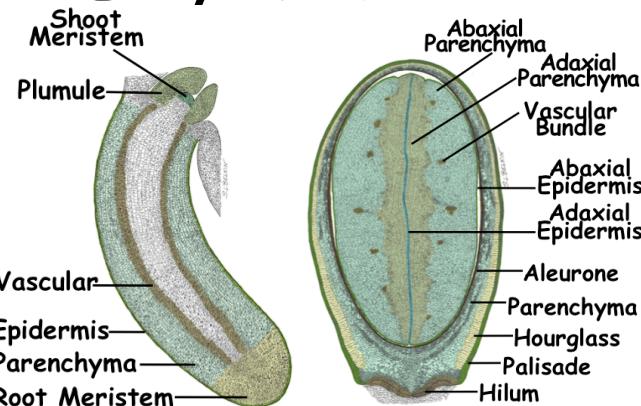
Heart



Cotyledon



Early Maturation



SEED STAGE	GLOB	HRT	COT	EM	TOTAL
# Compartments Studied	8	8	8	16	40
# GeneChip Experiments	26	21	18	34	99
# Genes Validated by qRT-PCR (TF)	42 (28TFs)	36 (20TFs)	55 (55TFs)	21 (21TFs)	154

Generated >100,000 Sections (> 3.4×10^6 Data Points)

Ran >10,000 qRT-PCR Reactions

The Interactive Seed Gene Network Website

<http://seedgenenetwork.net>

Supported by:  National Science Foundation

GENE NETWORKS IN SEED DEVELOPMENT

Identifying all the genes and gene networks required to "make a seed"

Home **Soybean** **Arabidopsis** **Annotations** **454ESTs** **People** **Presentations** **Genome Browser**

ANALYSIS TOOLS

Microarray Validation Analysis Tools Procedures Seed Images

Soybean Analysis Tools

- Browse**: Click here to browse the mRNA profiles of all compartments during soybean seed developments by probe identification, gene ontology, or function category.
- Analyze**: Click here to compare gene activity in different soybean seed compartments at different developmental stages.
- Blast**: Click here to BLAST DNA sequence against target sequences on the Affymetrix soybean array and view the seed expression pattern related to your gene-of-interest.

ARRAY ANNOTATIONS

Soybean Arabidopsis

Soybean GeneChip Array Annotation

Sequences used for BLAST came from the Affymetrix Soybean target sequences. Sequence information can be obtained directly from Affymetrix. The Affymetrix Soybean target sequence was based on the NCBI Unigene Build 13 (November, 2003). Probe design was based on the NCBI Unigene Build as well as the Affymetrix in-house clustering algorithm. Affymetrix in-house clustering probes are designated with the prefix GmaAffx.

BLASTX analysis was carried out using soybean target sequences searched against all *Arabidopsis* proteins (TAIR ATH1_pep_em_20040228). In our BLAST analysis, we filtered and removed any results with e-value greater than e-02. We selected the top *Arabidopsis* hit from each BLAST result (sometimes one Soybean sequence can hit many different *Arabidopsis* sequences) when identifying the corresponding *Arabidopsis* sequence. The e-value for that hit is displayed in the annotation file. Therefore, for each

EST SEQUENCING

Using 454 Sequencing to Survey The Transcriptome of Soybean Seeds Containing Globular-Stage Embryos

Background

The Affymetrix soybean genome array is being used to study the activity of genes in different compartments of the soybean seed at various stages of development (see the **Browse** link). The soybean array was designed using publicly available ESTs (Click here for more details about the soybean array). Most of the ESTs originate from reproductive and vegetative organs, but very few ESTs are from libraries constructed from soybean seeds throughout development. As such, genes active during many stages of seed development are most likely under-represented on the array. To uncover additional genes active during Soybean seed development, we carried out a pilot study using the high-throughput 454 sequencing (**454 Life Sciences**) to survey the transcriptome of a globular-stage soybean seed. We generated ~900,000 reads with the average length of 200 bases in one run representing a deep sampling of the globular-stage seed transcriptome.

GENOME BROWSER

Soybean (Glycine Max)
Showing 1.285 kbp from Gm14, positions 49,569,631 to 49,570,915

Instructions [Bookmark this] [Hide banner] [Share these tracks] [Link to image] [Help] [Search]

Landmark or Region: Gm14:49590.1 Search

Data source: Soybean

Overview Details

Reports & Analysis: Download Alignments Configure... Go | Flip

Scroll/Zoom: << >> Show 1.285 kbp << >> | Flip

Globular-Stage Soybean Seed ESTs

Probe Sequence Target Sequence Consensus Sequence

Search Gene of Interest

Gene Model

Illumina ESTs

Affymetrix Soybean GeneChip

PROBE SET ANNOTATION

Probe Set Identifier Gma.11213.2.S1_at

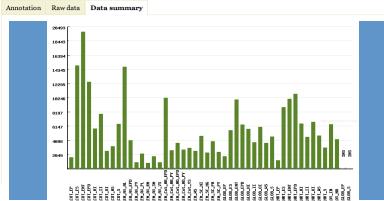
Annotation Raw data Data summary

GeneChip Array: Soybean Array
Closely Related: ATG10980
Arabidopsis Gene: 10-71
BLAST e-value to: 10-71
Gene Description: histone H3
Functional category: Cell Structure
Sub-category: Histone
Gene Family: G0: Biological Processes: GO:0007001 / chromosome_organization_and_biotogenesis_(semen_Eukaryota)
GO:0006534 / nucleosome_assembly

DATA SUMMARY PLOT

Probe Set Identifier Gma.11213.2.S1_at

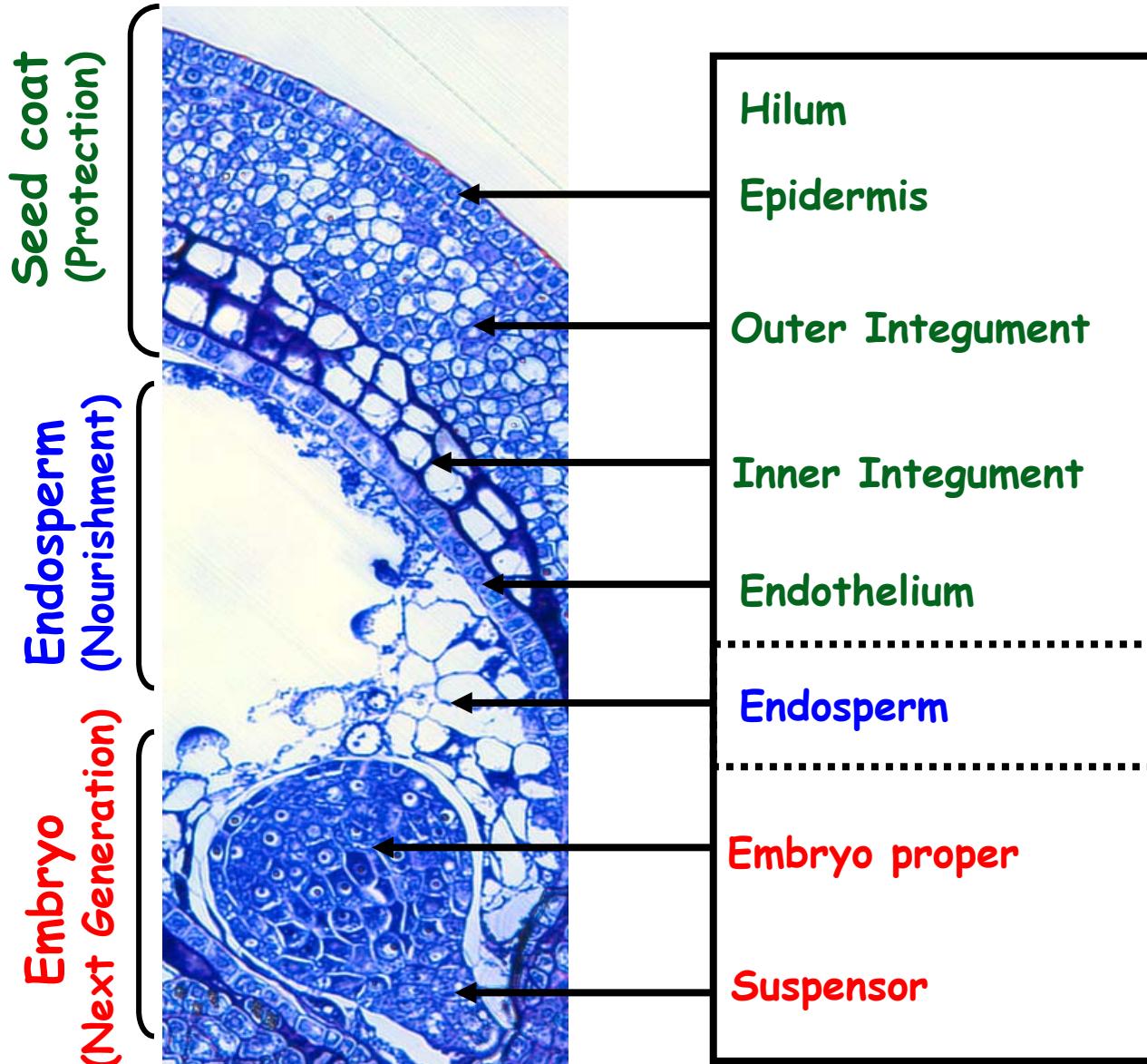
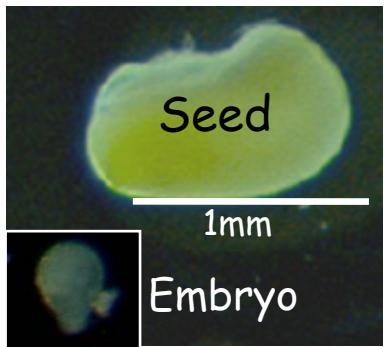
Annotation Raw data Data summary



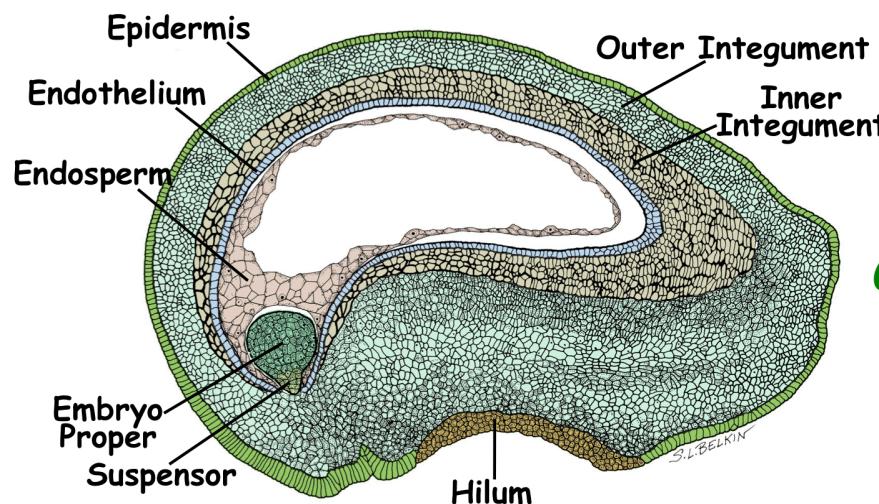
What Are The Genes Required To Make a Seed?

An Example-The Globular Stage

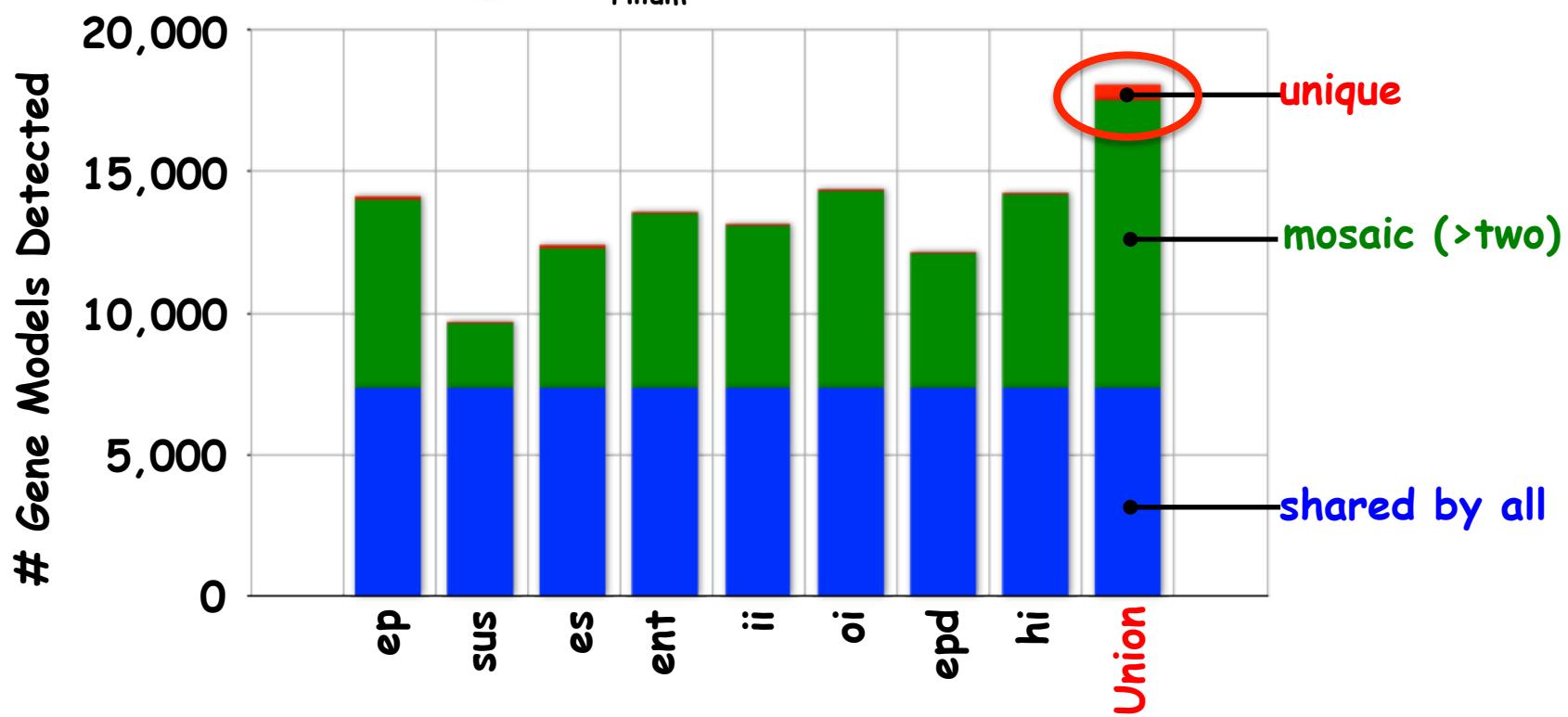
Globular Stage Seed
After Fertilization



Most Seed mRNAs Are Shared By Different Compartments & Regions

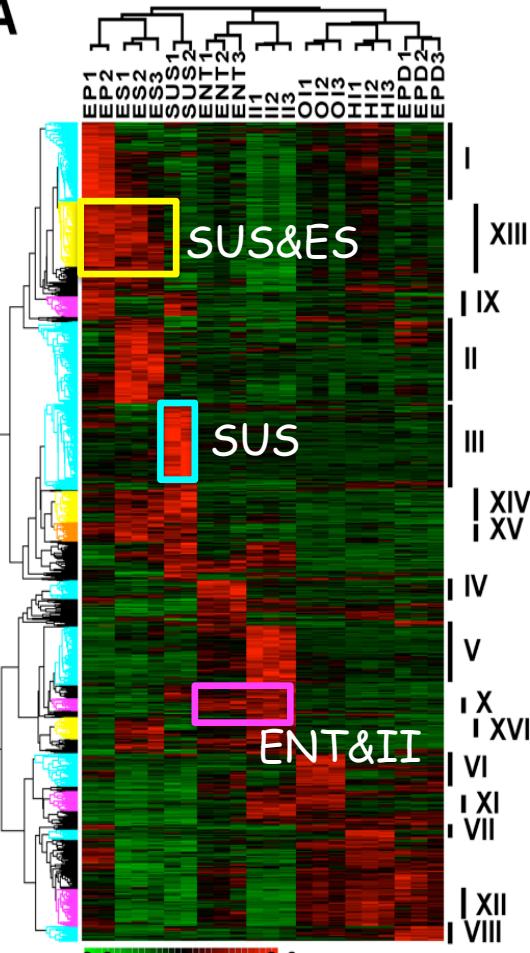


Globular Stage Seed

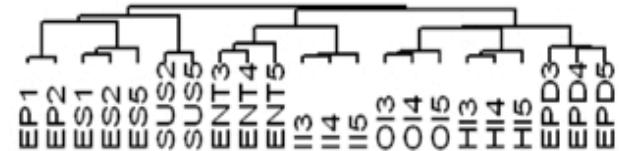
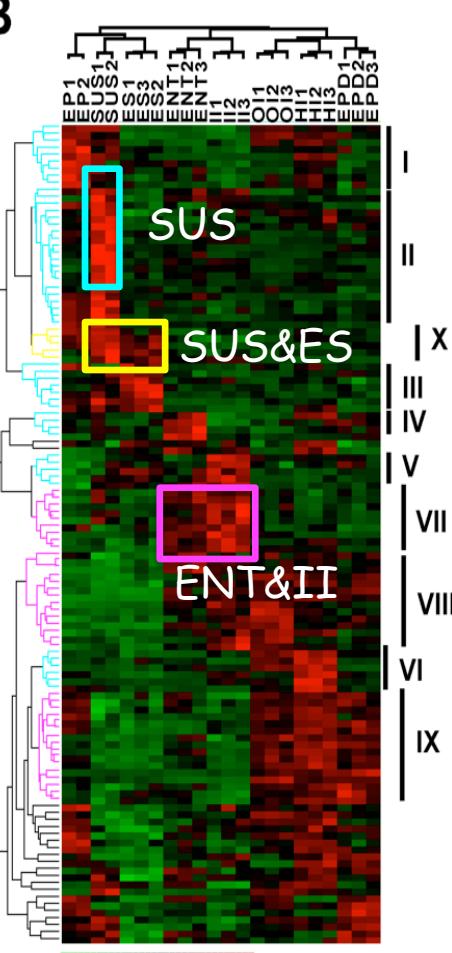


How Are *Shared* Transcripts Regulated in Globular-Stage Seed Compartments, Regions, and Tissues?

A



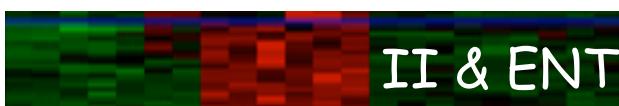
B



Tissue-Specific Enriched
Sets of mRNAs up-regulated in specific seed tissues



Compartment Enriched
Sets of mRNAs Co-upregulated in multiple seed tissues of the same origin (e.g., Embryo, Seed Coat)



Mosaic
Sets of mRNAs Co-upregulated in multiple seed tissues of the different origins (e.g., EP-ES)

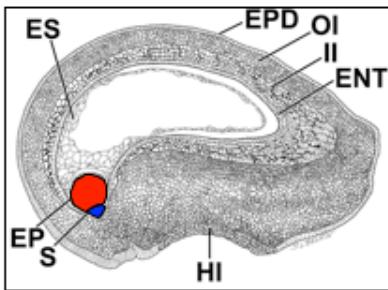


How Many Genes Are Active in the Globular Stage Seed?

Whole Seed

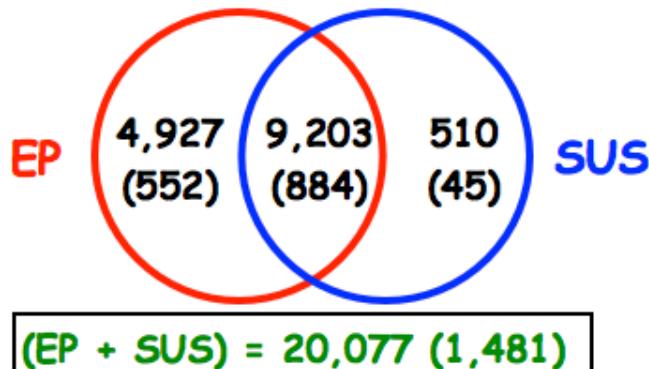


LCM

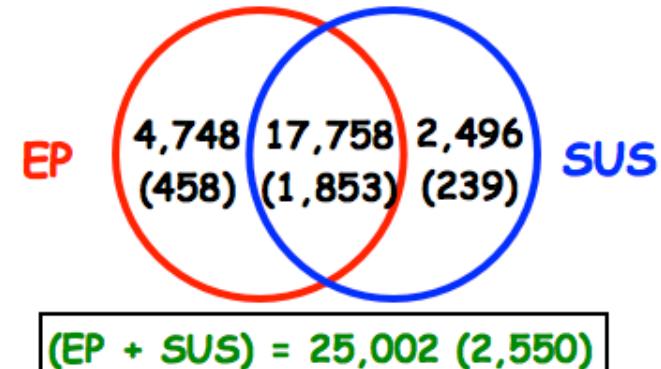


TECHNOLOGY	TISSUE/REGION	GENECHIP	RNA-SEQ
Whole Mount	Whole Seed	17,057 (1,748)	37,875 (3,926)
LCM	EP	14,130 (1,436)	22,506 (2,311)
	SUS	9,713 (929)	20,254 (2,092)
	EP + SUS	20,077 (1,481)	25,002 (2,550)

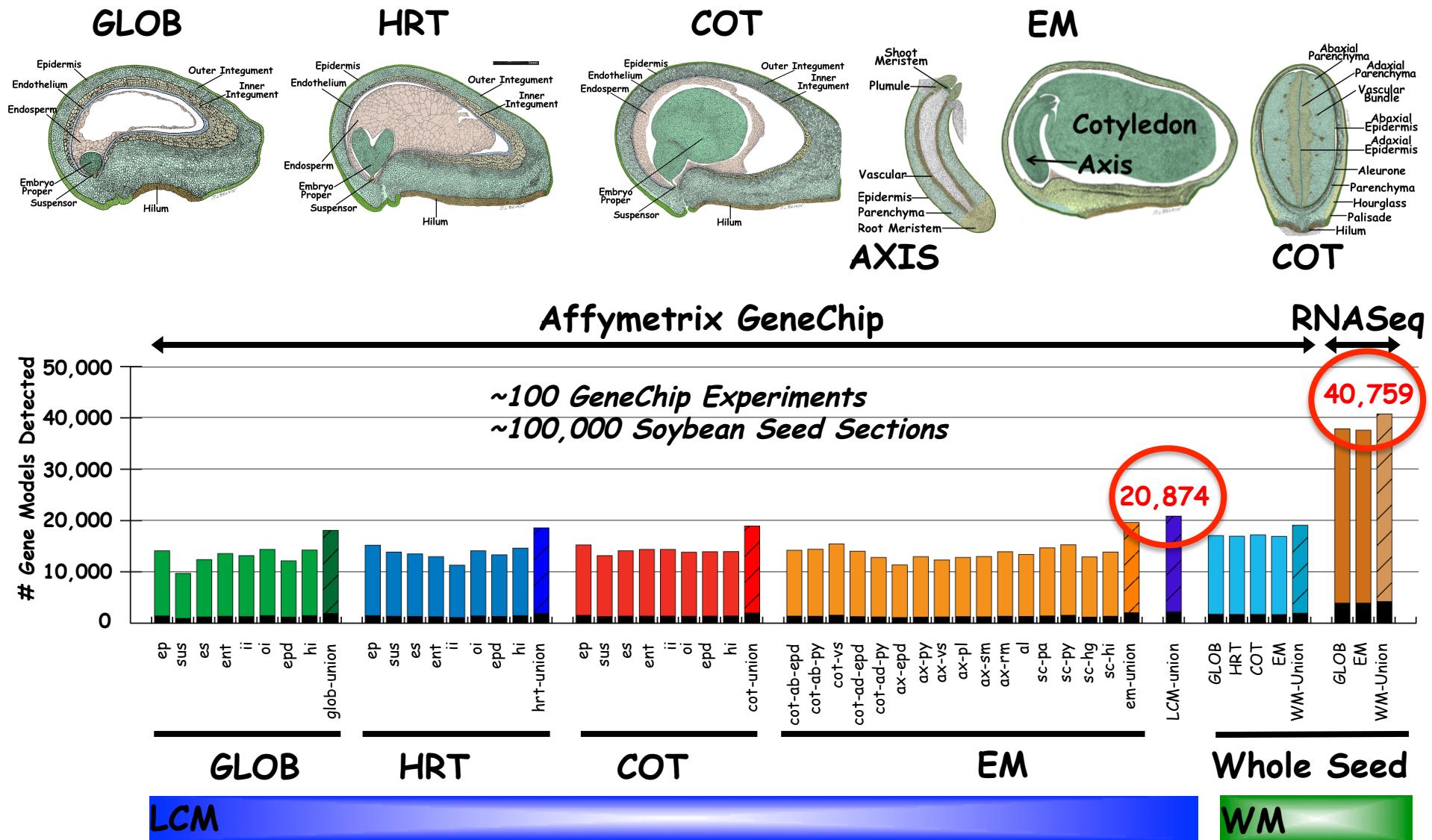
GeneChip



RNA-Seq

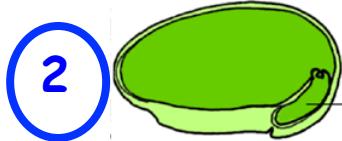


What Are the Genes Active in Every Soybean Compartment, Region, & Tissue Throughout Development?



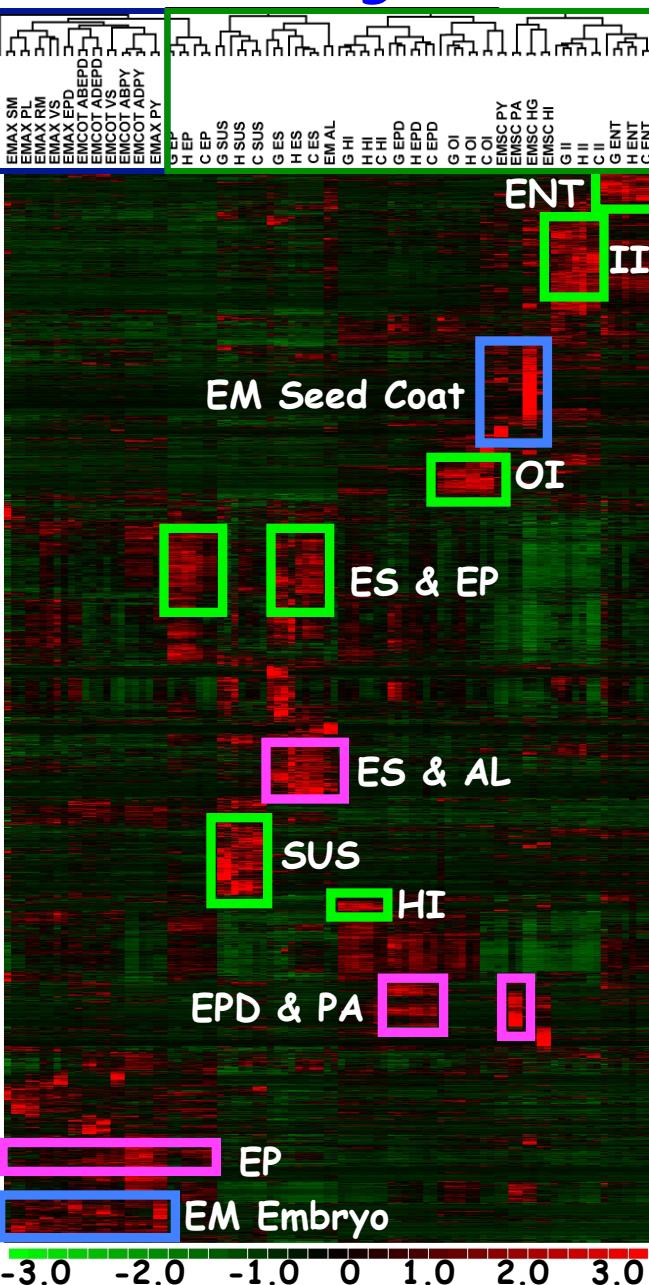
What Are the Temporal and Spatial mRNA Accumulation Patterns Throughout Seed Development?

Maturation phase



A TEMPORAL pattern of mRNA accumulation is apparent (Early Maturation)

Top 4K Varying mRNAs
ANOVA $p < 0.0001$



Morphogenesis phase



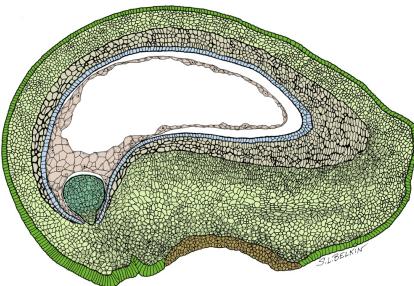
A SPATIAL pattern of mRNA accumulation is apparent (GLOB-HRT-COT)



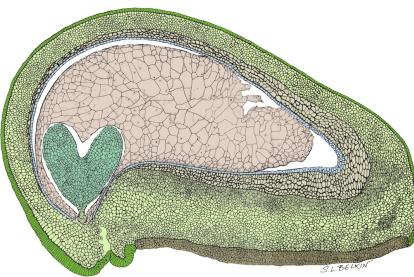
mRNA sets up-regulated throughout development (GLOB to EM)

Are Seed Compartments Specialized For Specific Metabolic Processes?

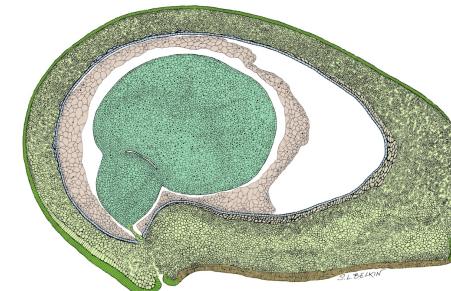
Globular



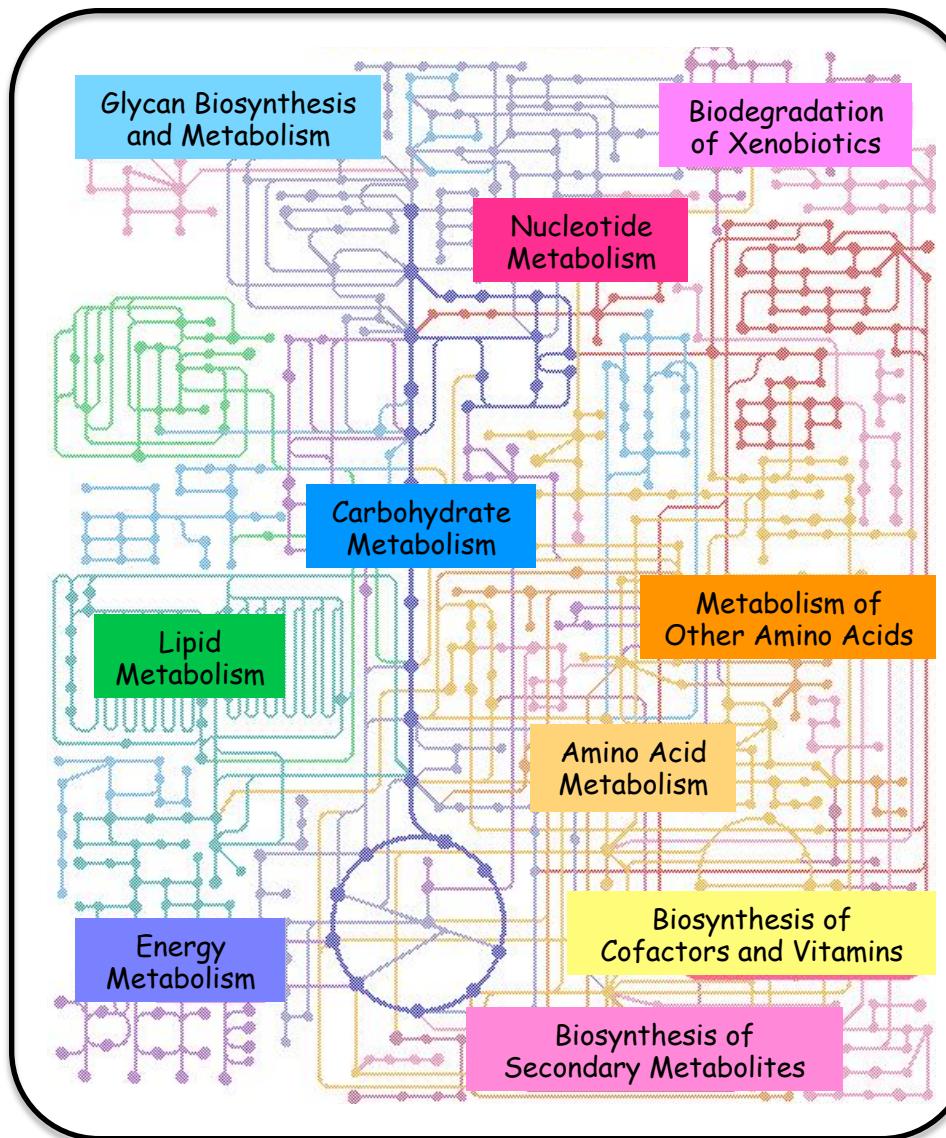
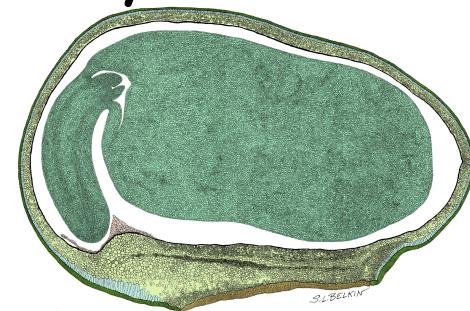
Heart



Cotyledon

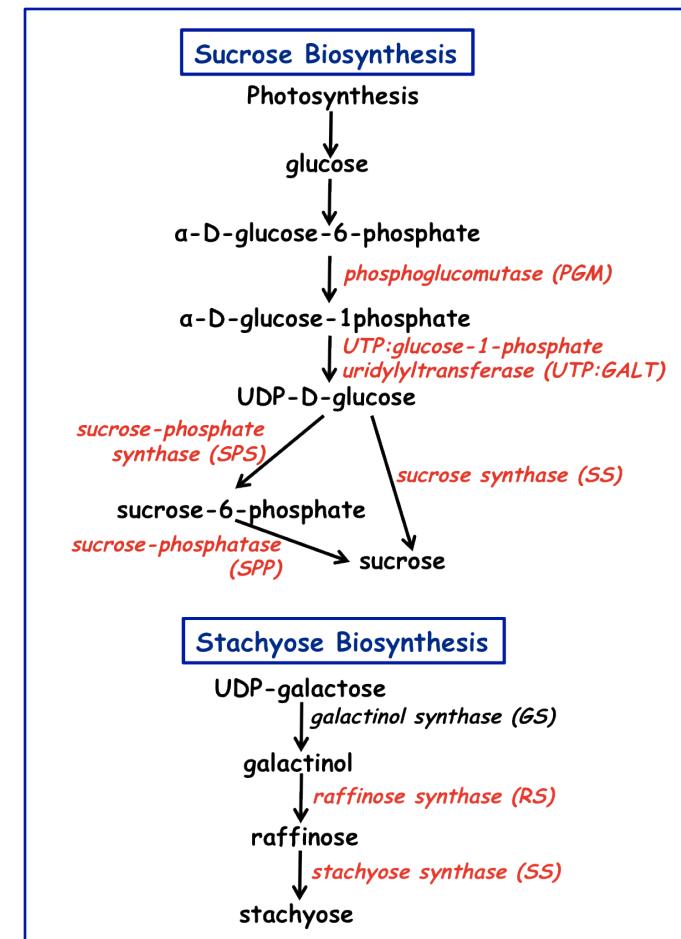
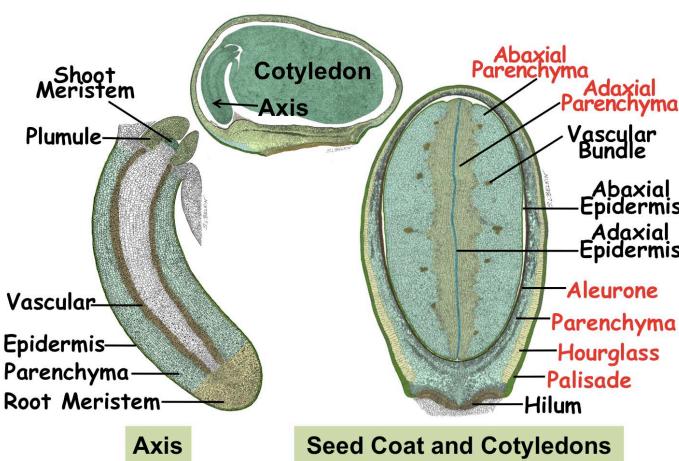
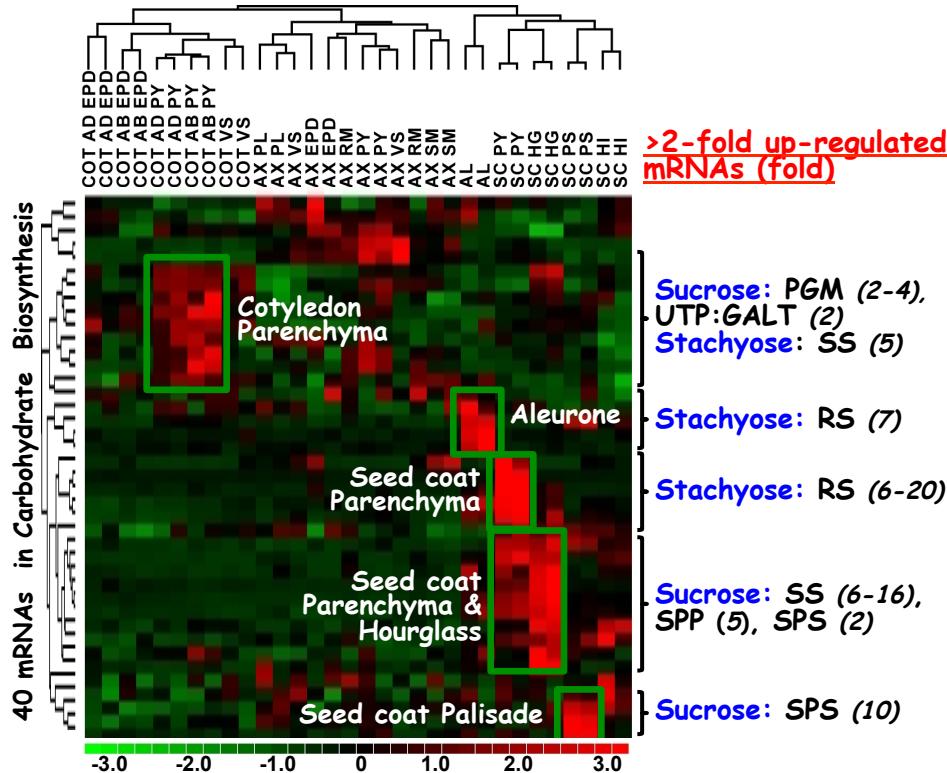


Early Maturation

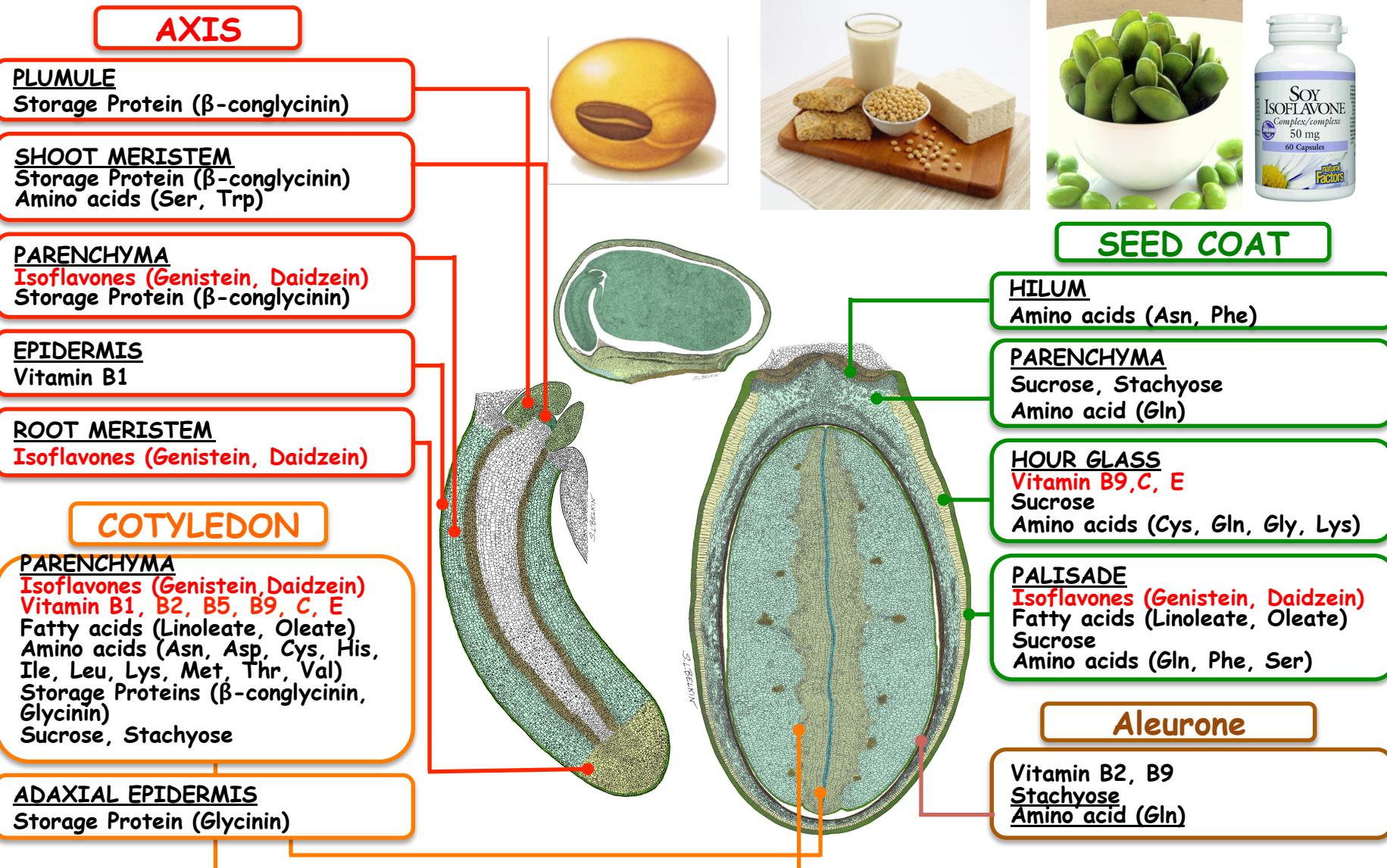


Examined 325 Metabolic Pathways in 40 Different Seed Compartments

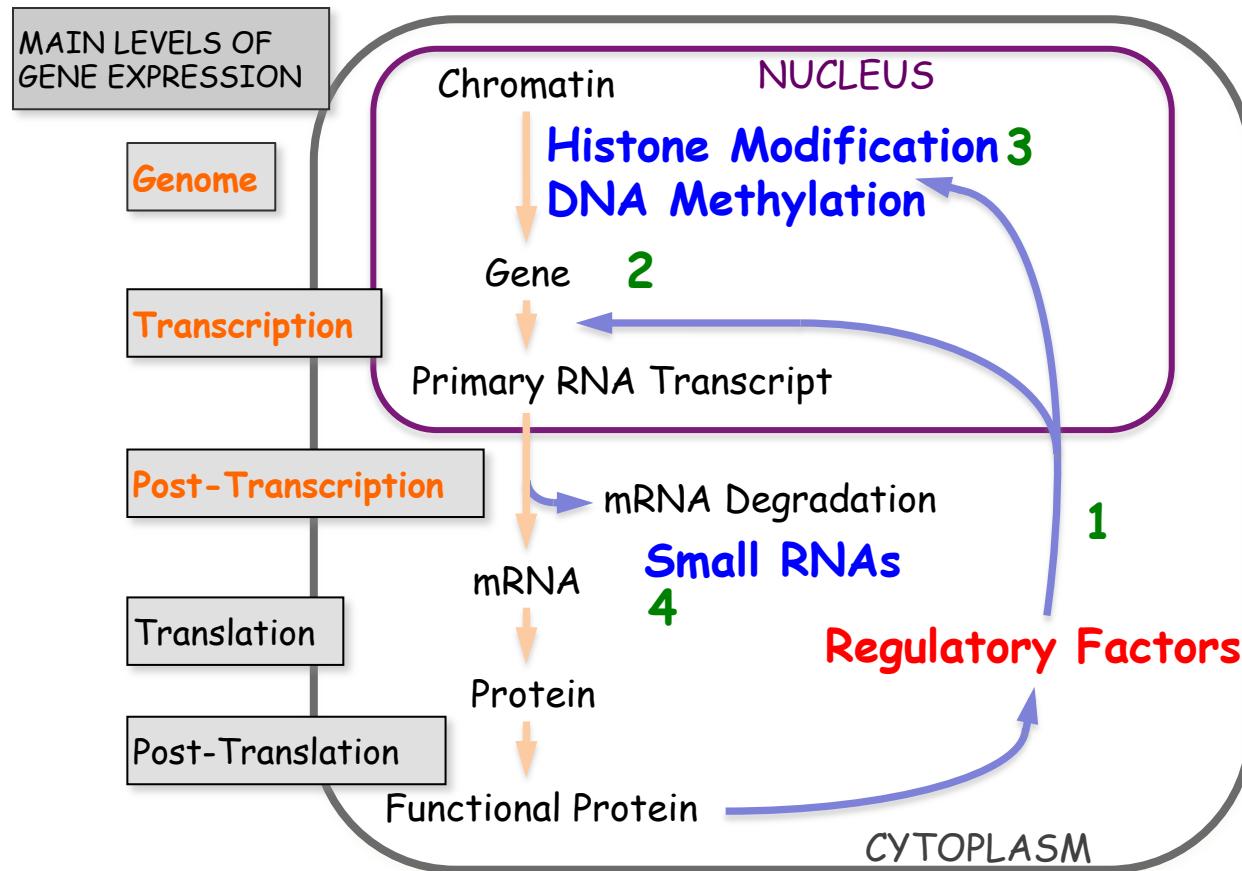
YES.....mRNAs Encoding Enzymes in Metabolic Pathways Are Up-Regulated in Specific Seed Compartments!!!



Where Are the Pathways For Soybean "Health" Products Localized Within the Seed?



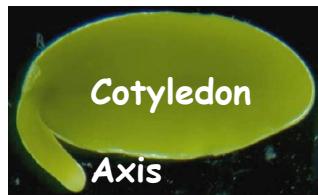
Unraveling the Gene Regulatory Processes that Are Required to "Make a Seed" (The Next Step!)



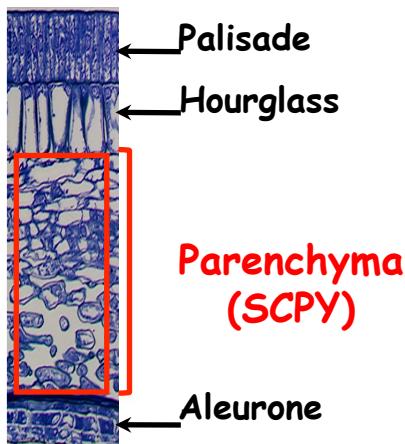
1. Using **RNAi** to knock down compartment-specific TF mRNAs during seed development
2. Identifying **Downstream Targets** of Compartment-Specific Transcription Factors
3. Characterization of compartment-specific **microRNA** populations during seed development
4. Characterization of **DNA methylation & histone modification** patterns in seed compartments

Using LCM to Identify Seed Compartment-Specific MicroRNAs & Methylation Patterns

Early Maturation Stage Seed

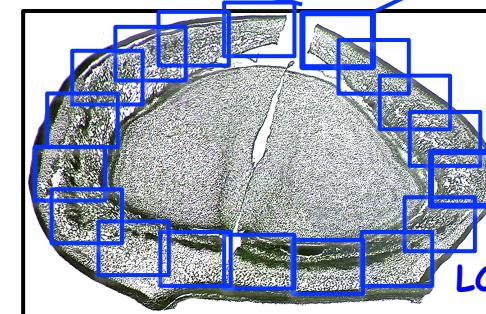
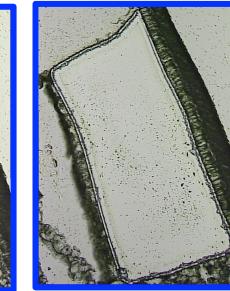
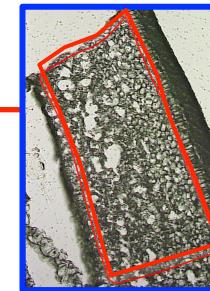


Early Maturation Seed Coat Layers



A LCM Section of Seed Coat Parenchyma (SCPY)

Before LCM After LCM

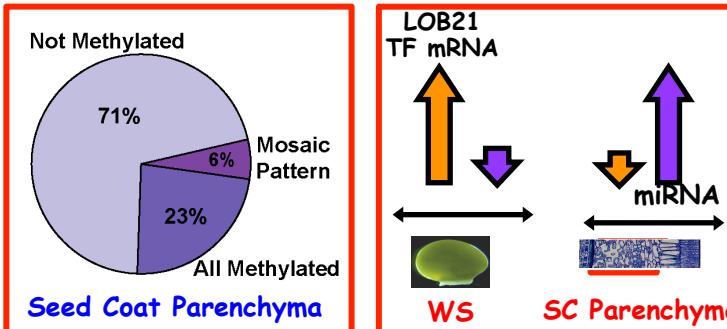


Longitudinal Section

Small RNA DNA

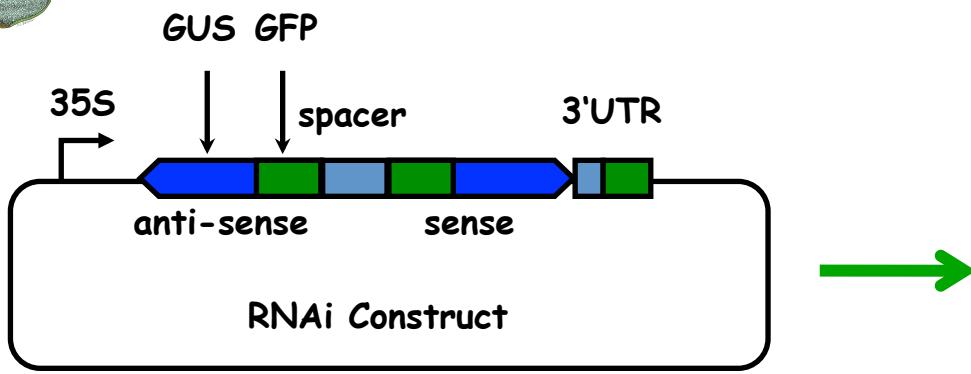
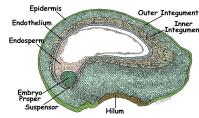
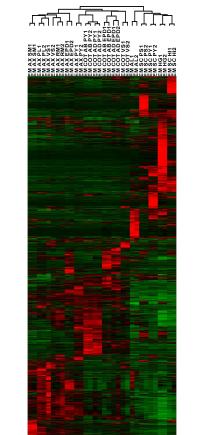
Methylome (1X)

MicroRNAs (21M Reads)

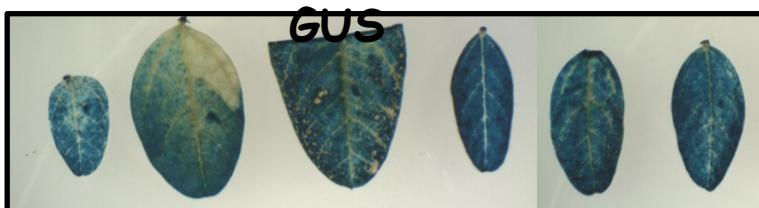


What Are the Functions of Seed Compartment-Specific Transcription Factor mRNAs?

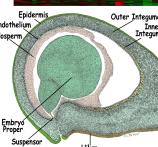
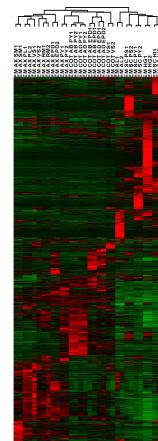
500 Transgenic Soybean RNAi Lines Generated and 61 TF mRNAs Knocked Down



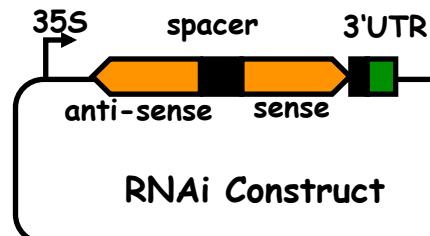
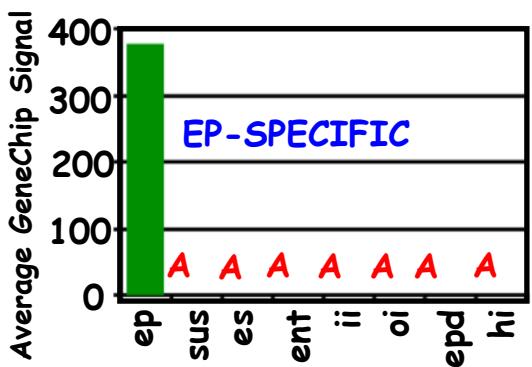
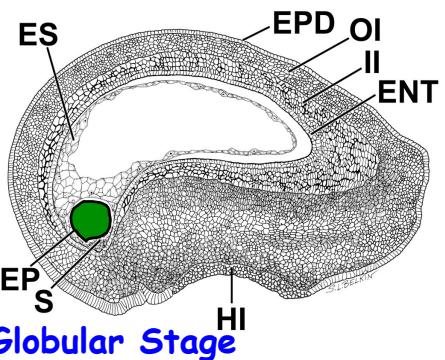
35S-GUS



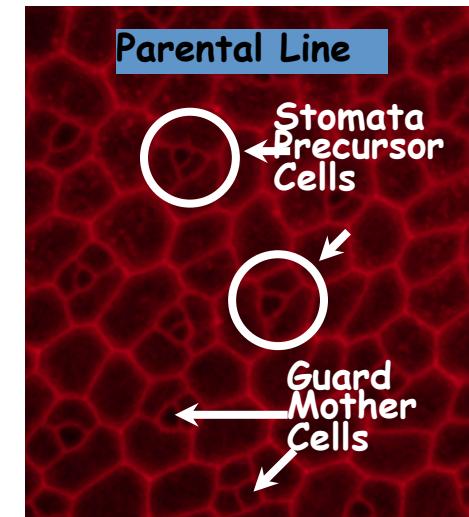
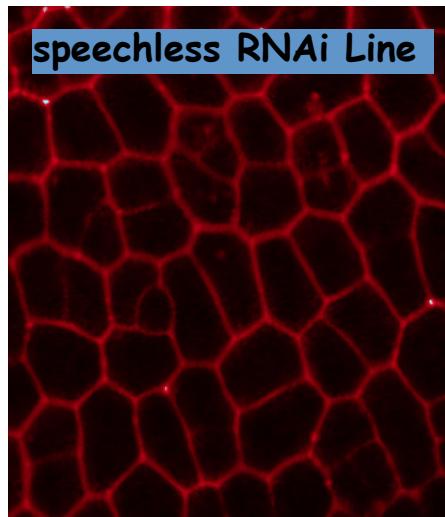
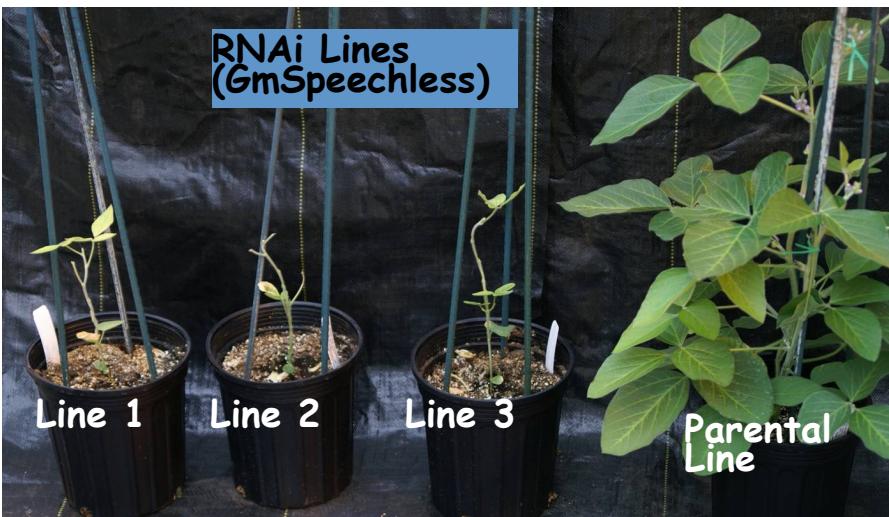
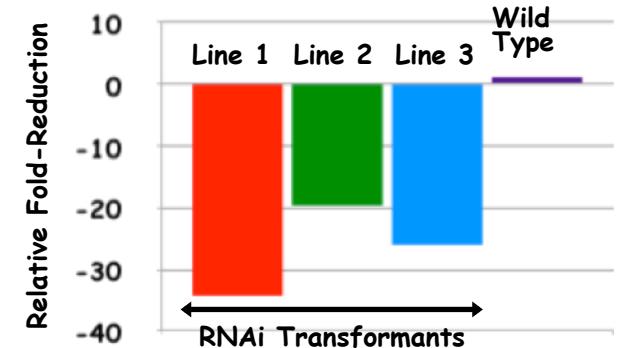
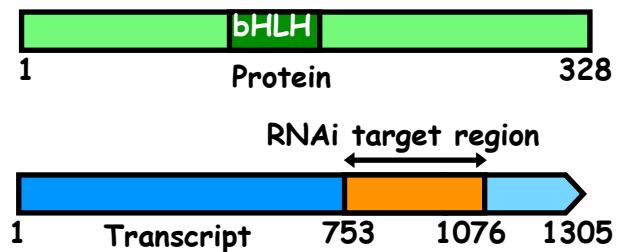
35S-GUS; 35S-RNAi



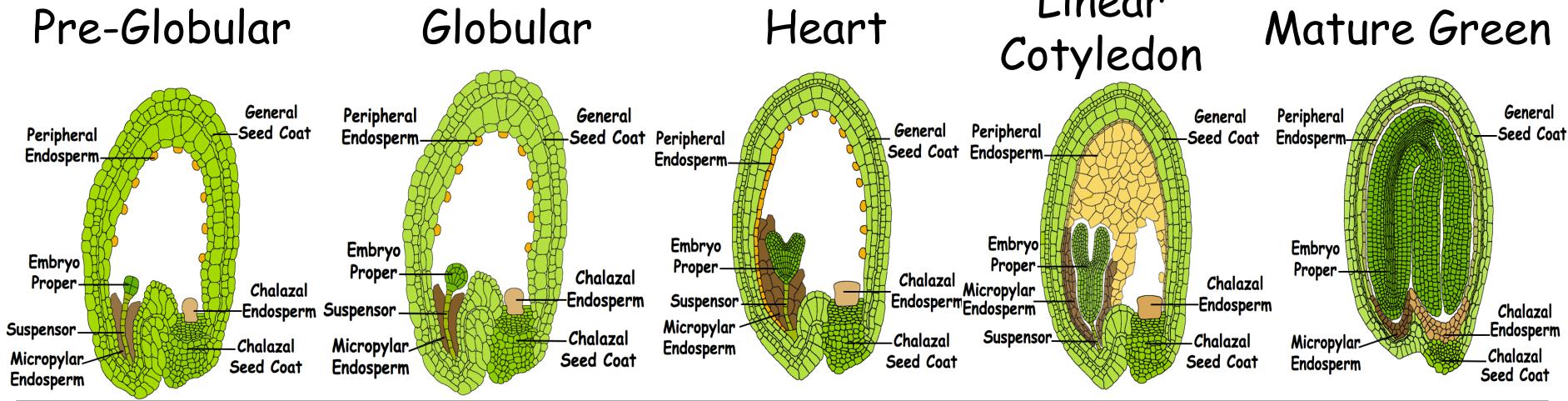
Soybean Speechless-Like TF mRNA Prevents Stomata Formation on Developing Cotyledons...A Knock-Down Example



GmSpeechless
(*Glyma04g41710*)



What About Another Seed.....Arabidopsis?



SEED STAGE	PREGLOB	GLOB	HRT	LCOT	BENT	MG	TOTAL
------------	---------	------	-----	------	------	----	-------

# Compartments Studied	6	7	6	6	5	6	76
------------------------	---	---	---	---	---	---	----

# GeneChip Experiments	12	15	14	12	10	12	166
------------------------	----	----	----	----	----	----	-----

# Genes Validated by qRT-PCR (TF)	A total of 148 genes (>10,000 reactions)					
-----------------------------------	--	--	--	--	--	--

Generated >10,000 Sections (> 3.7×10^6 Data Points)

Ran >10,000 qRT-PCR Reactions

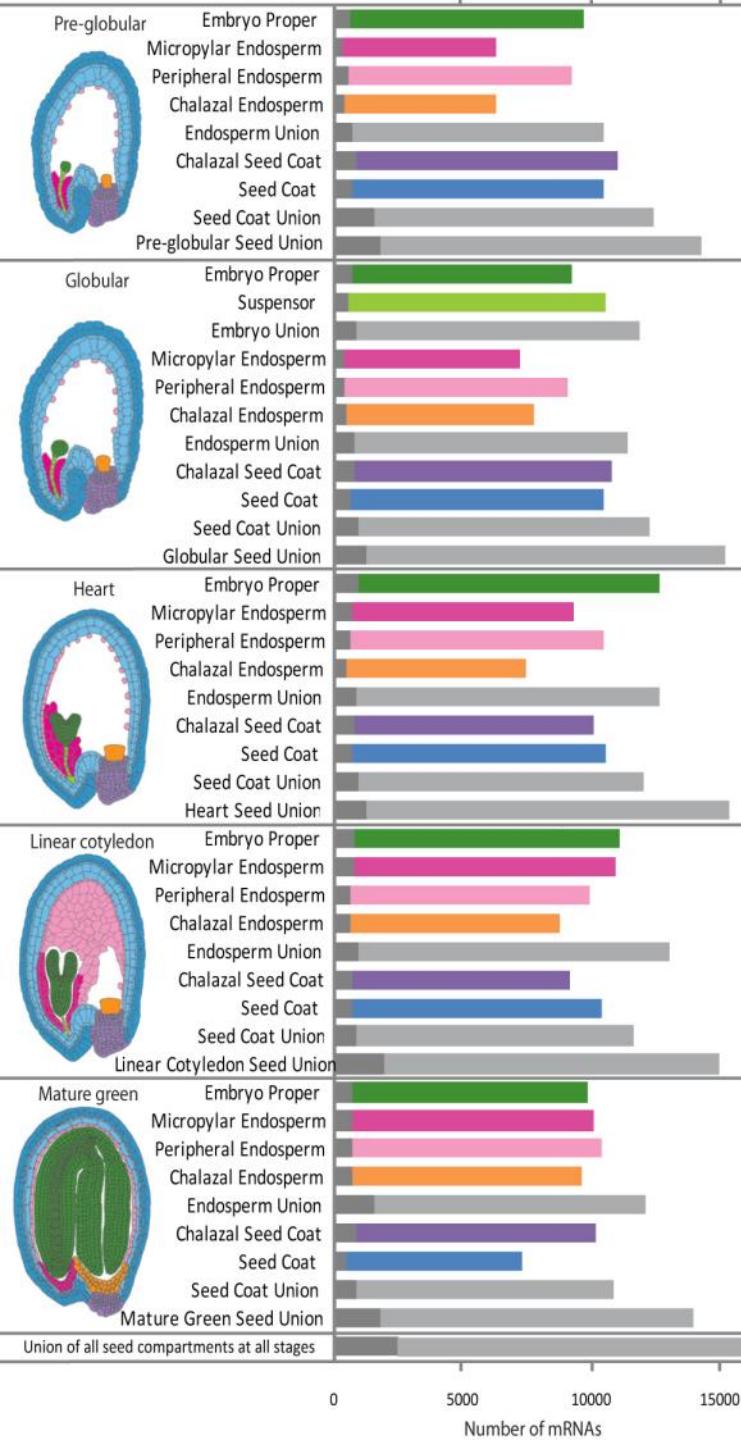


Soybean



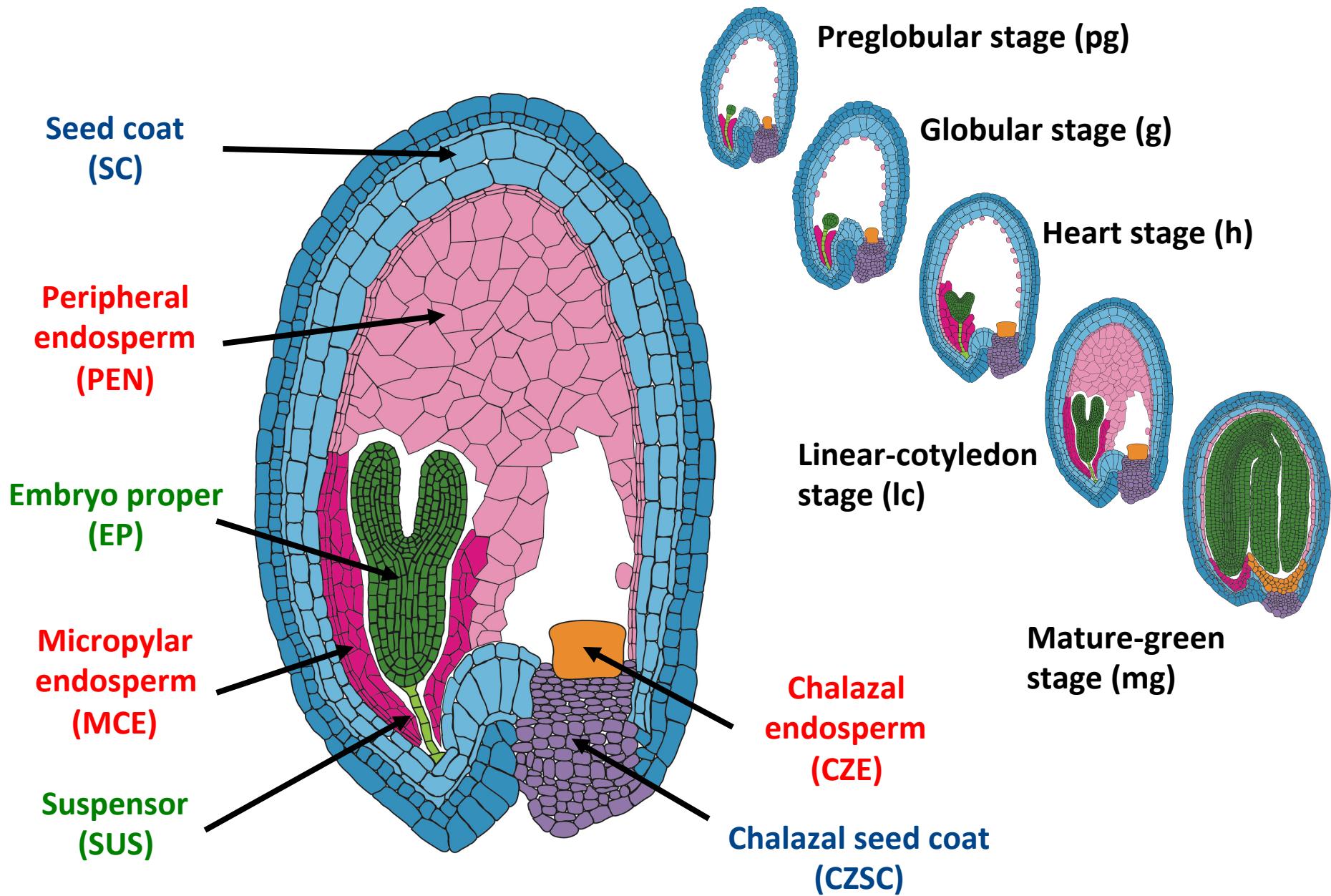
Arabidopsis

Numbers of Distinct mRNAs Detected in Each Seed Region and Compartment at All Stages of Development



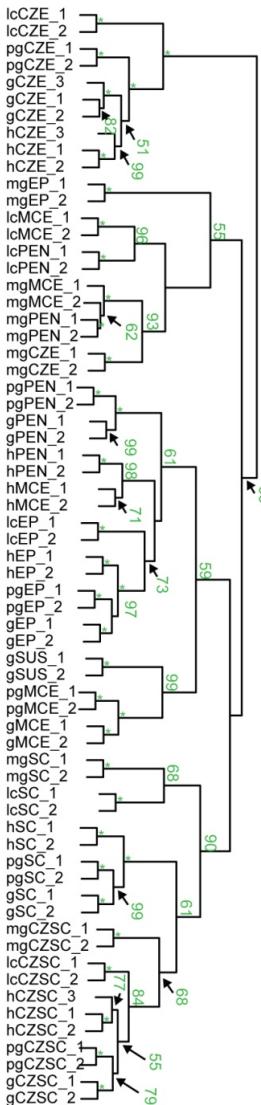
Data available @ GEO &
<http://www.seedgenenetwork.net/>

Arabidopsis Seed Development

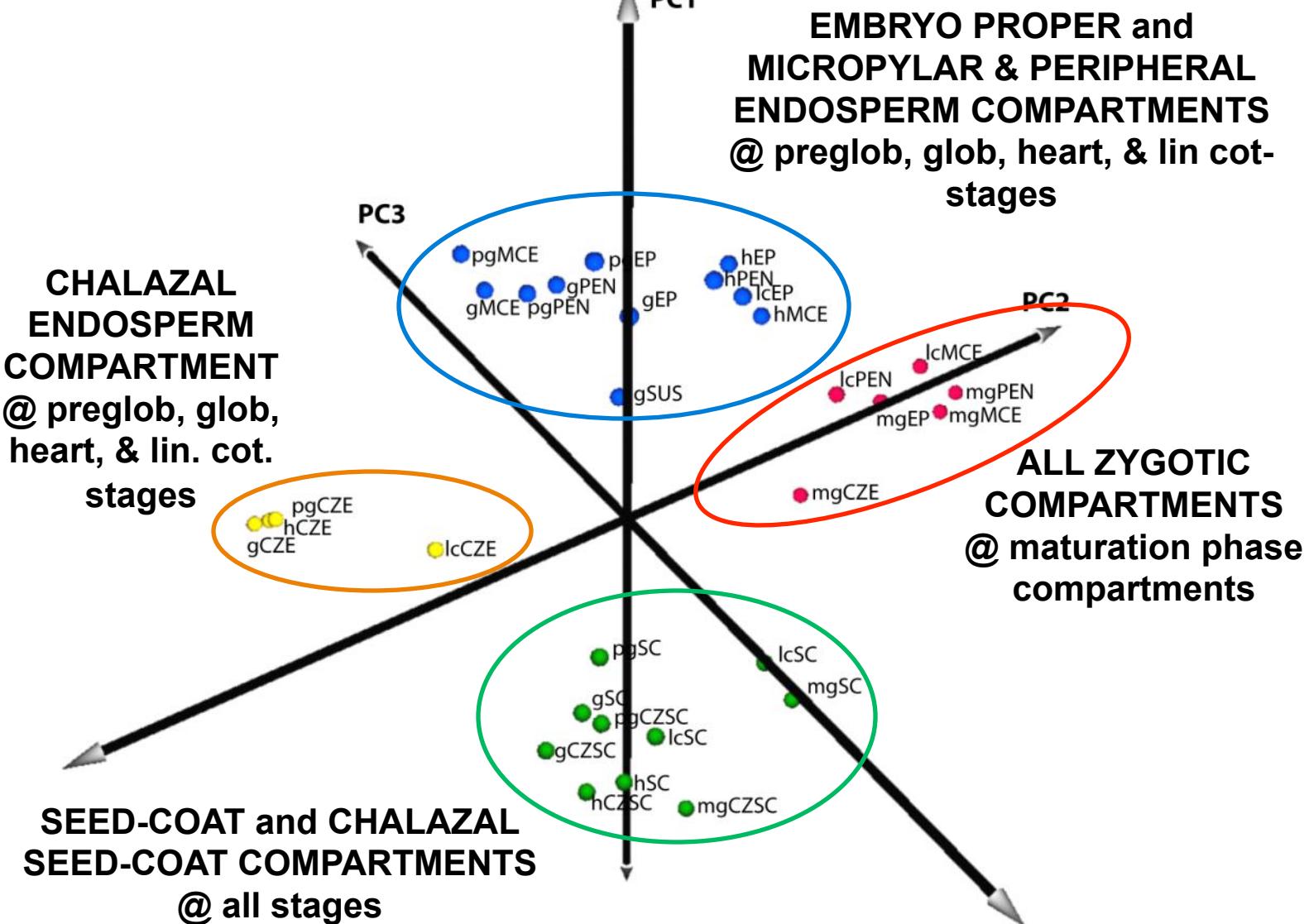


Global Analyses of Seed Compartment RNA Populations

Hierarchical Clustering



Principal Component Analysis

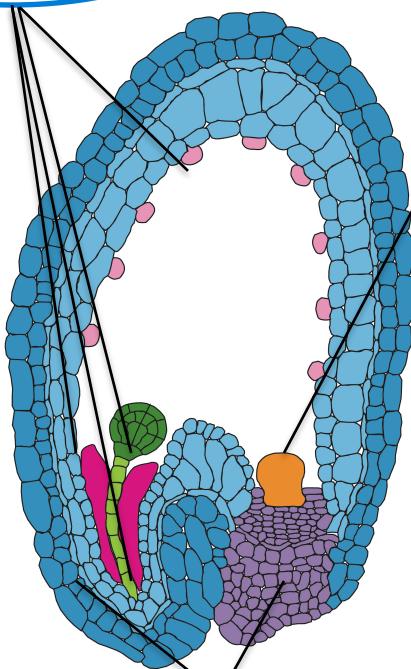


Global Analyses of Seed Compartment RNA Populations

Early Seed Development

Embryo Proper
Suspensor
Micropylar Endosperm
Peripheral Endosperm

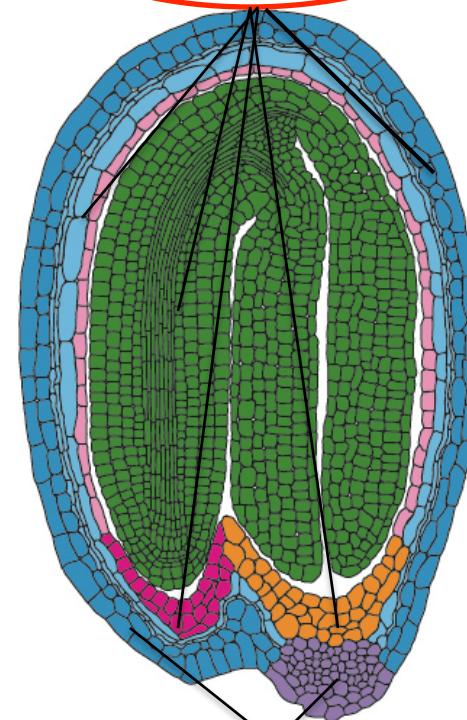
Chalazal Endosperm



Chalazal Seed Coat
Distal Seed Coat

Late Seed Development

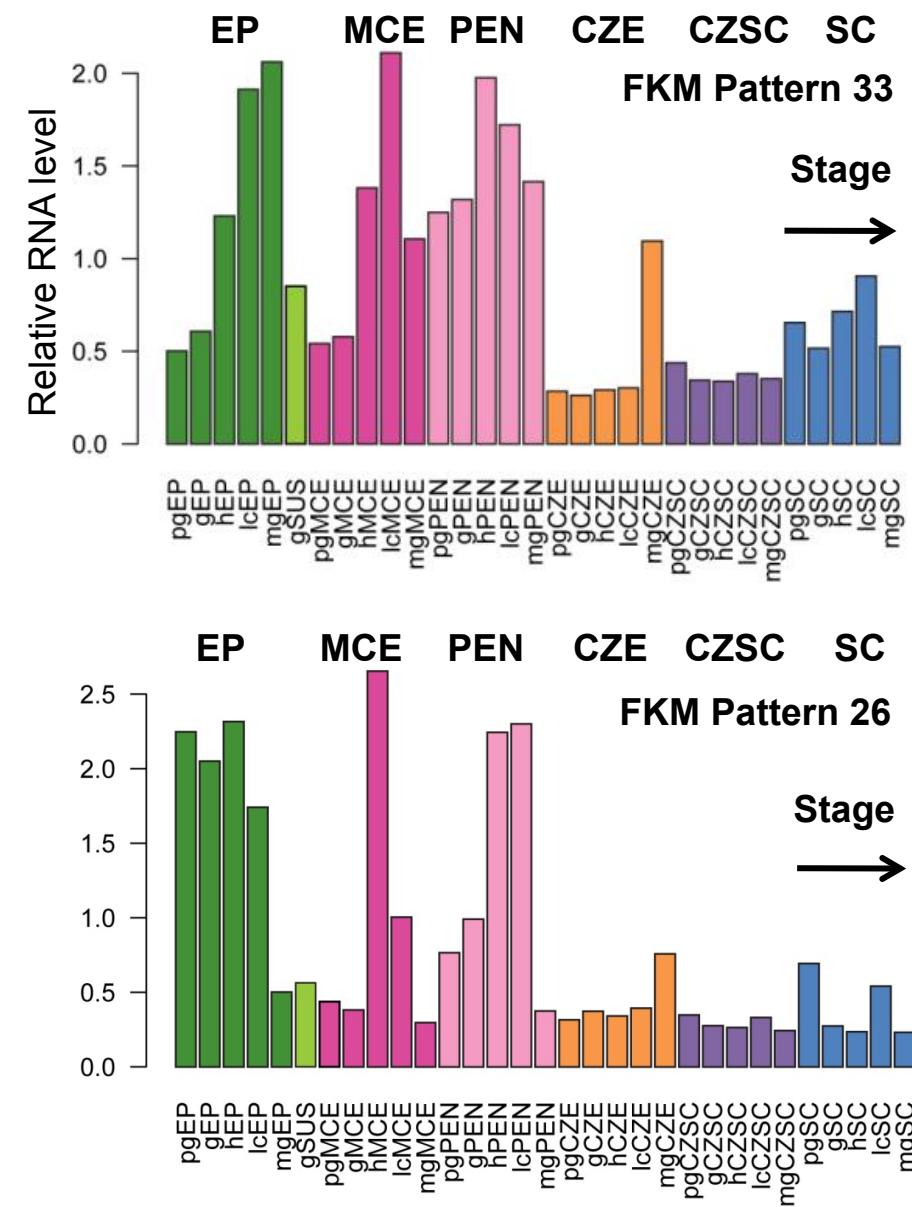
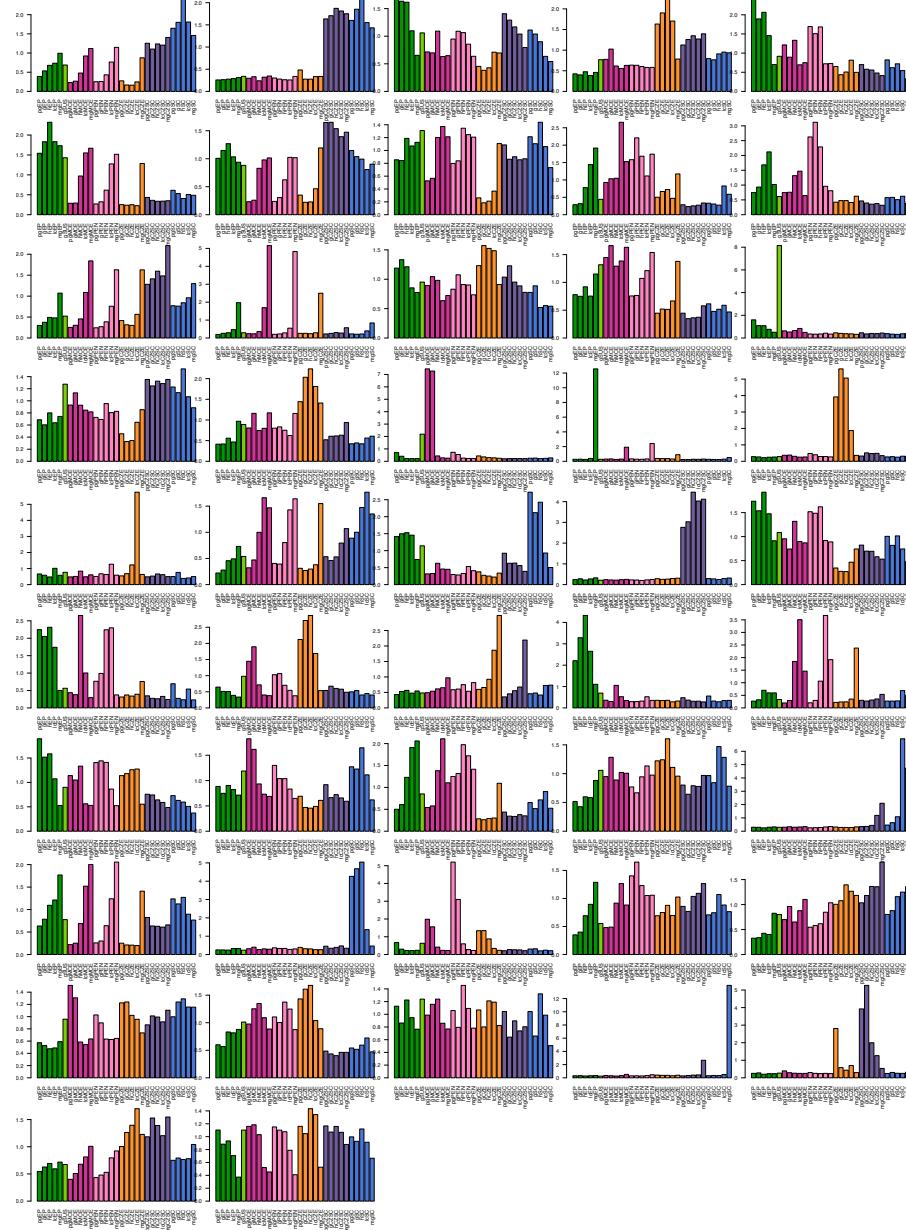
Embryo Proper
Micropylar Endosperm
Peripheral Endosperm
Chalazal Endosperm



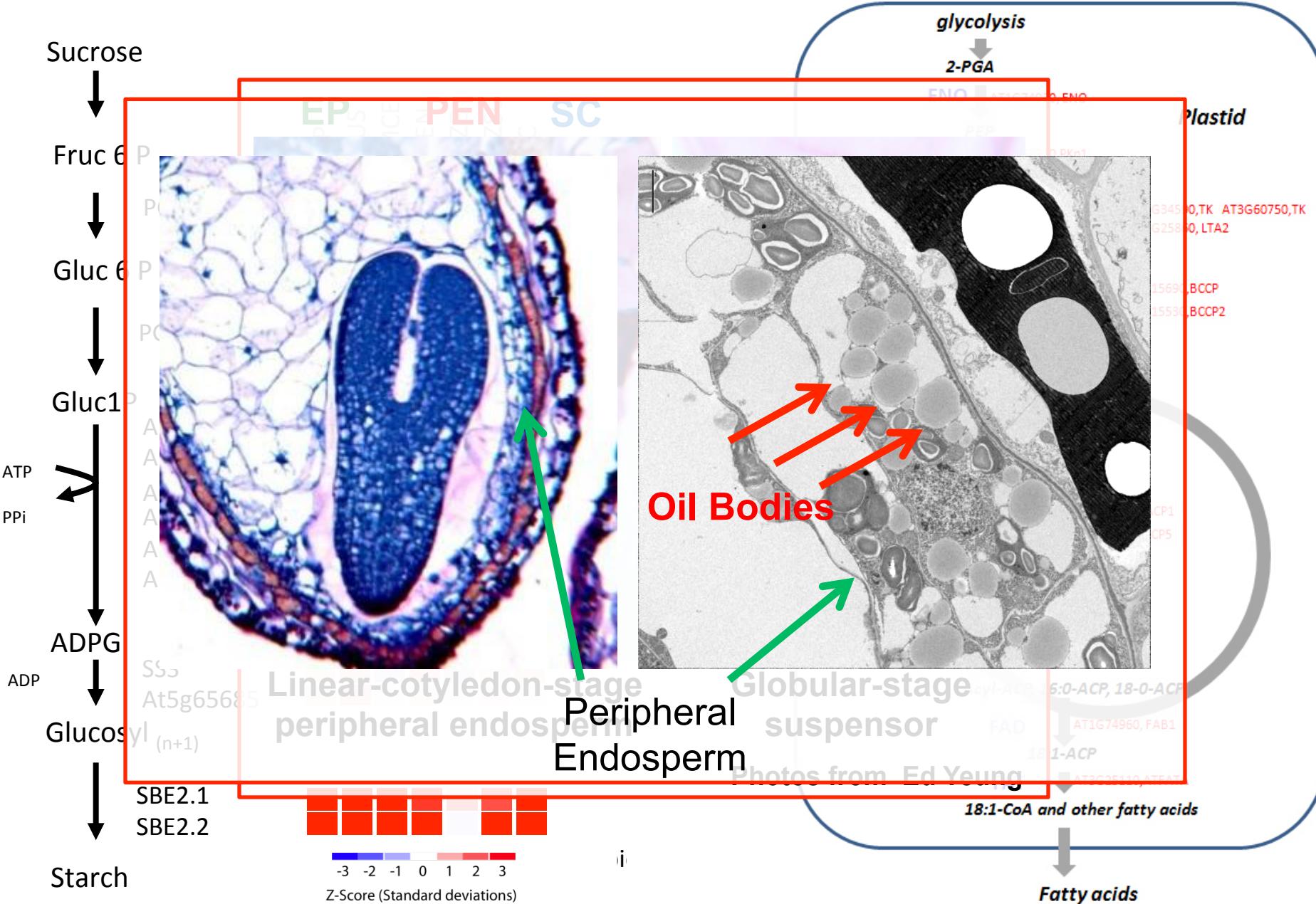
Chalazal Seed Coat
Distal Seed Coat

Identification of Coexpressed Gene Sets

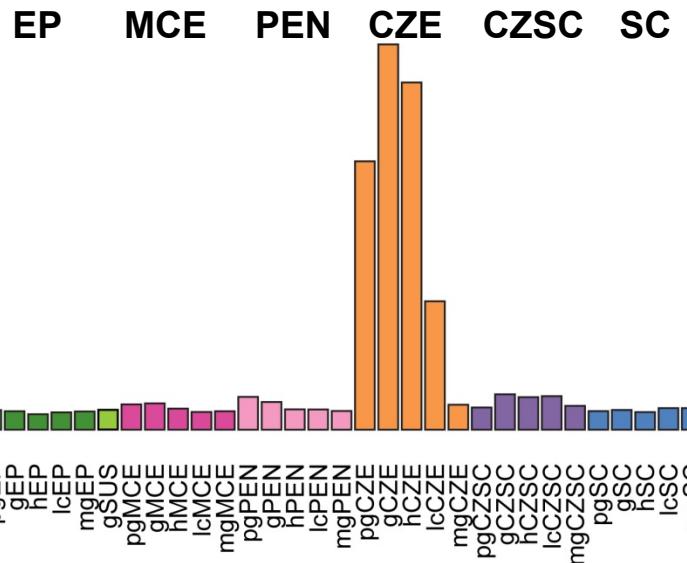
Fuzzy K-means Cluster Analysis (temporal & spatial)



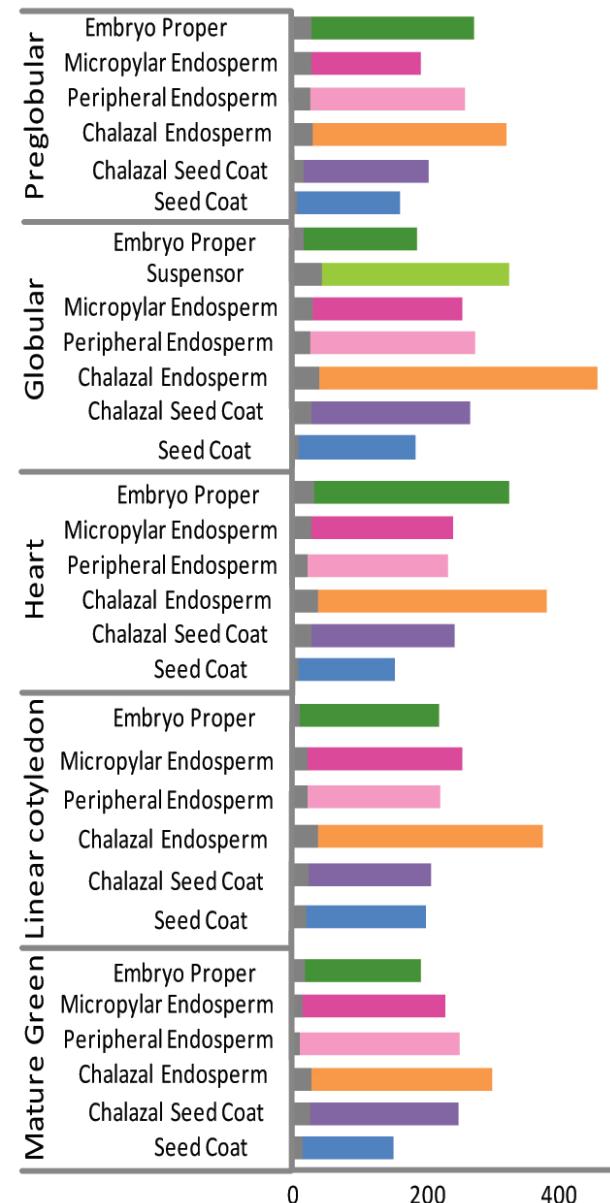
Carbon Metabolism in the Embryo and Endosperm



Chalazal Endosperm



Seed-specific RNAs

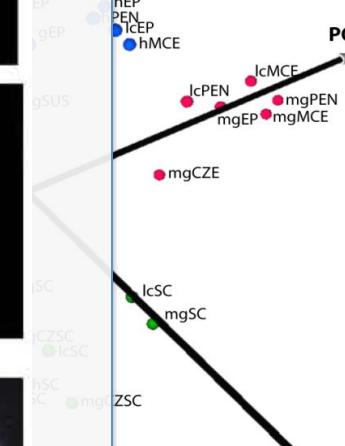


Union of all
seed RNAs

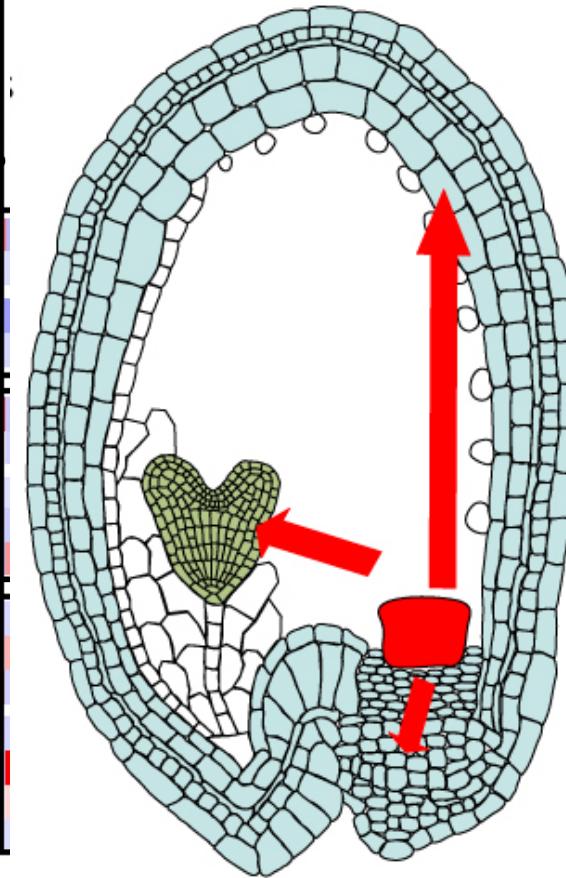
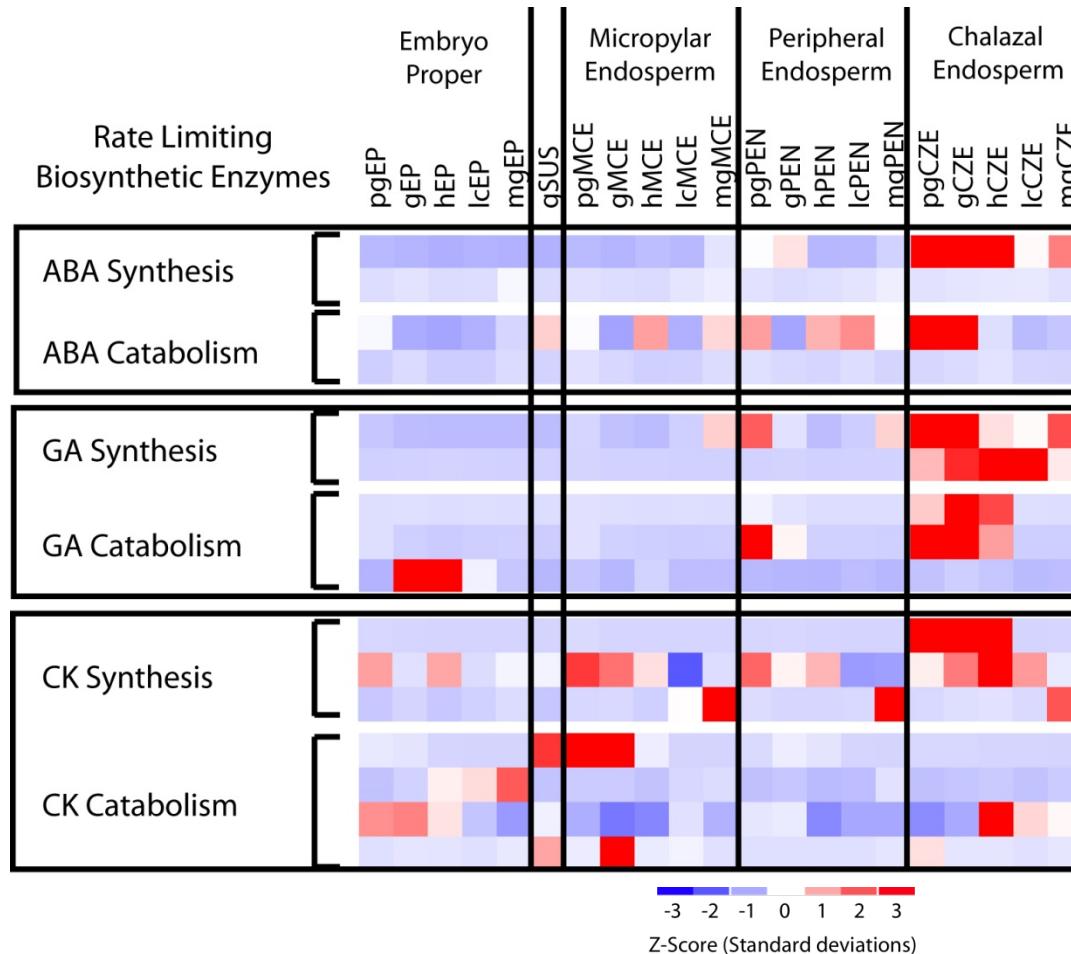
1,297

Union of all
reproductive
RNAs

Union of all
vegetative
RNAs

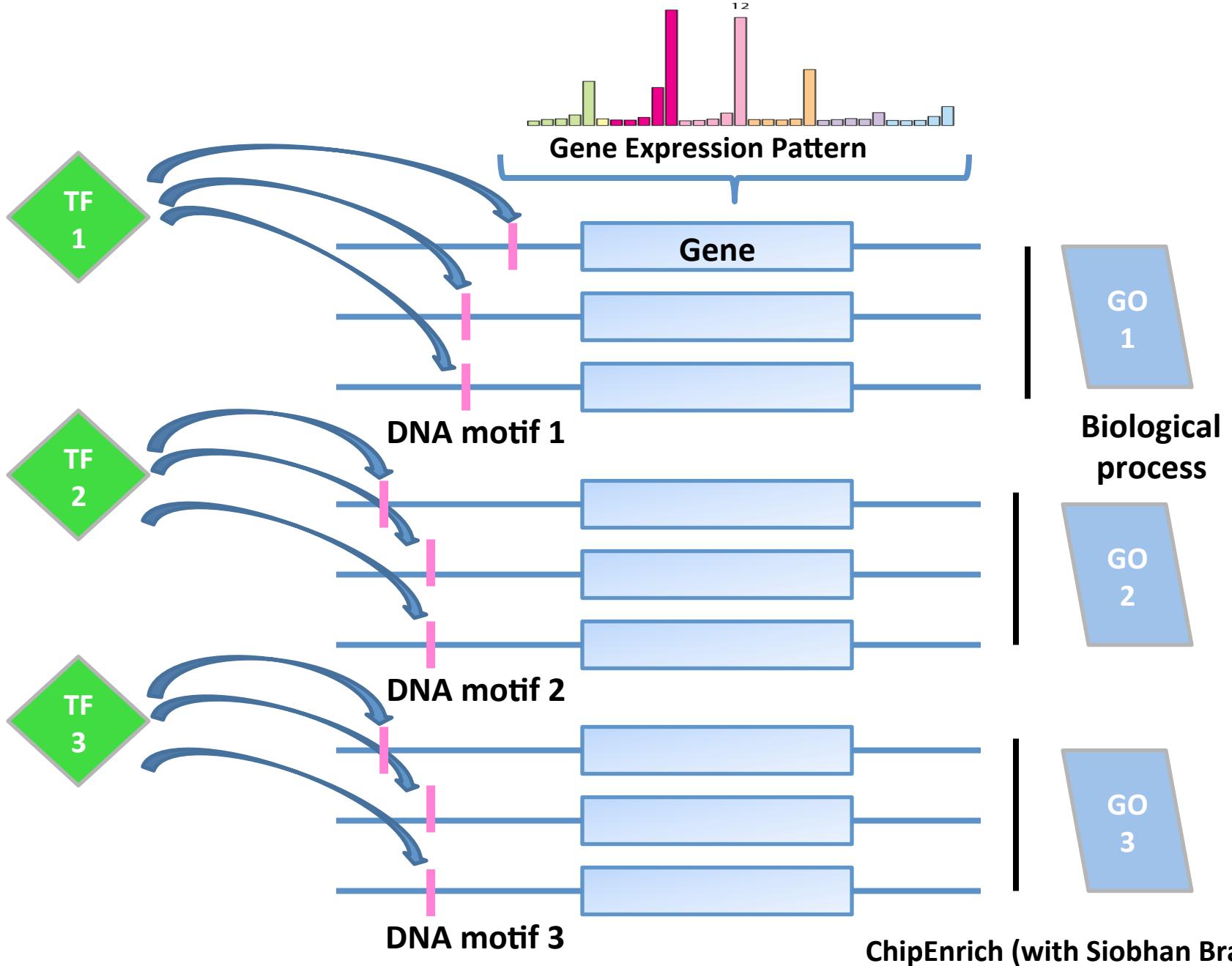


RNAs Encoding Hormone Biosynthetic Enzymes Accumulate Primarily in the Chalazal Endosperm



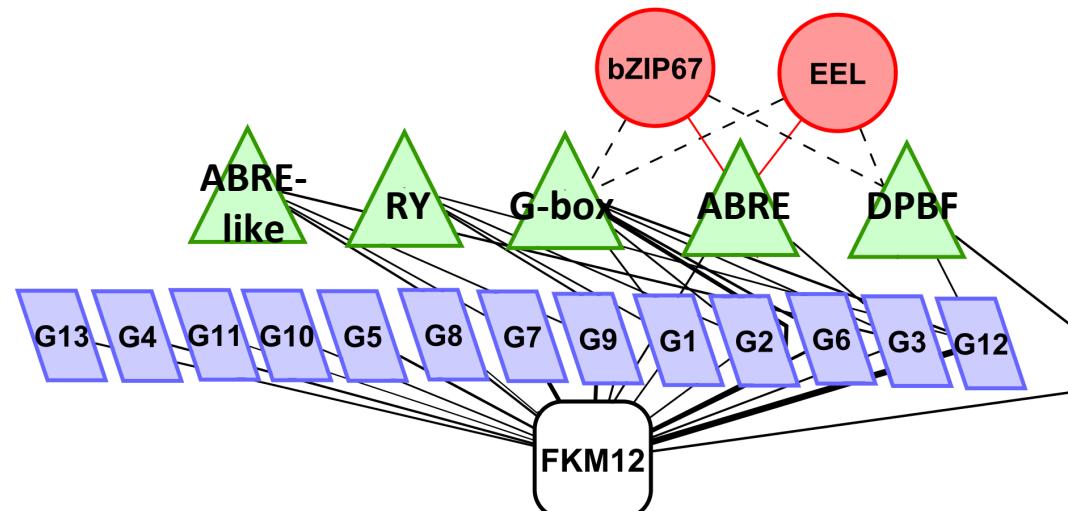
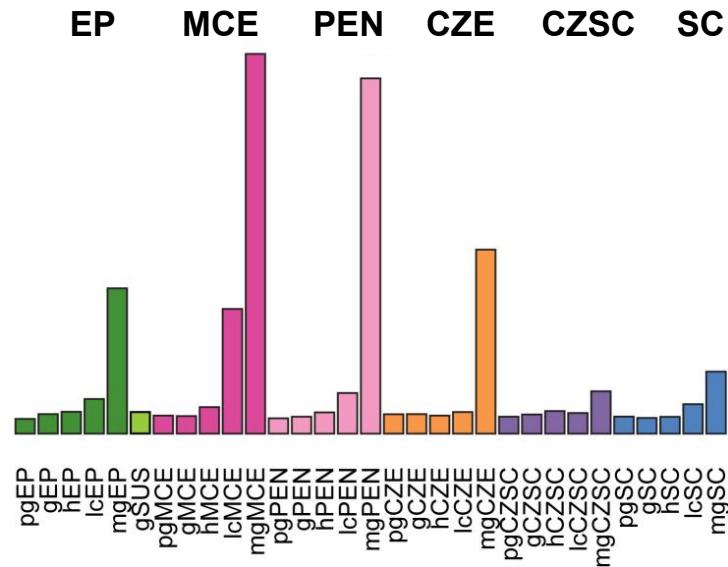
- Seed coat region-specific gene list is overrepresented for the GO terms Response to Gibberellin (@ pg g h lC) and Response to Abscisic Acid Stimulus (pg h mg)
- Embryo proper gene list is overrepresented in the GO terms for response to cytokinin

Transcriptional Networks Controlling Coexpressed Gene Sets

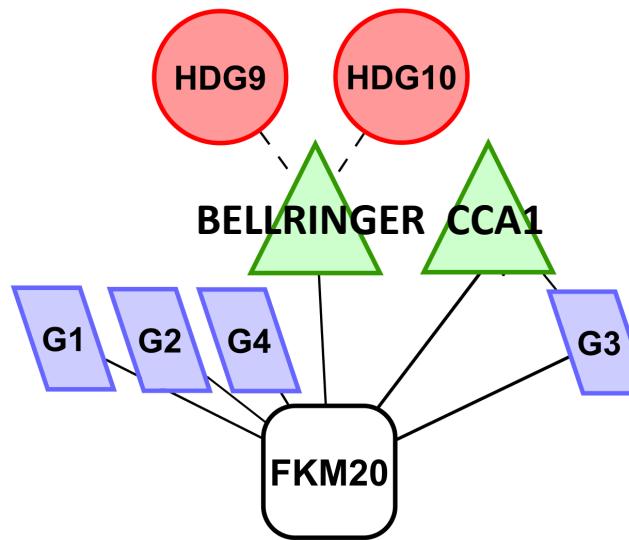
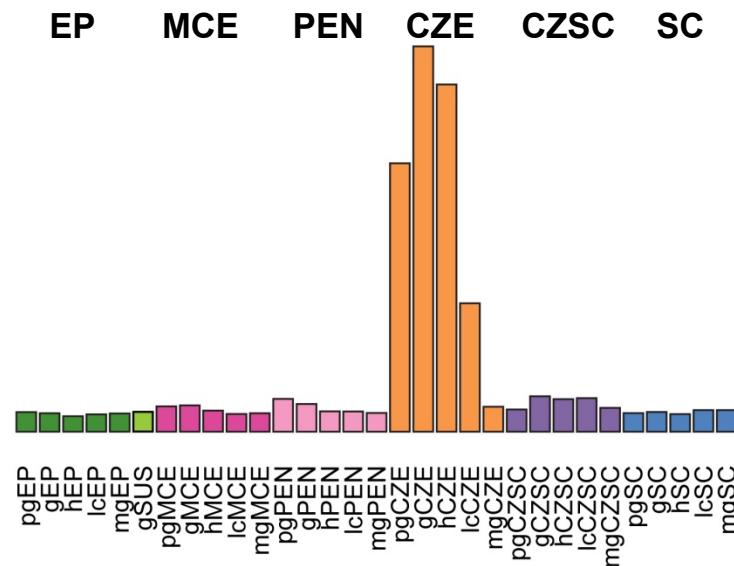


Coexpressed Gene Sets Define Transcriptional Modules

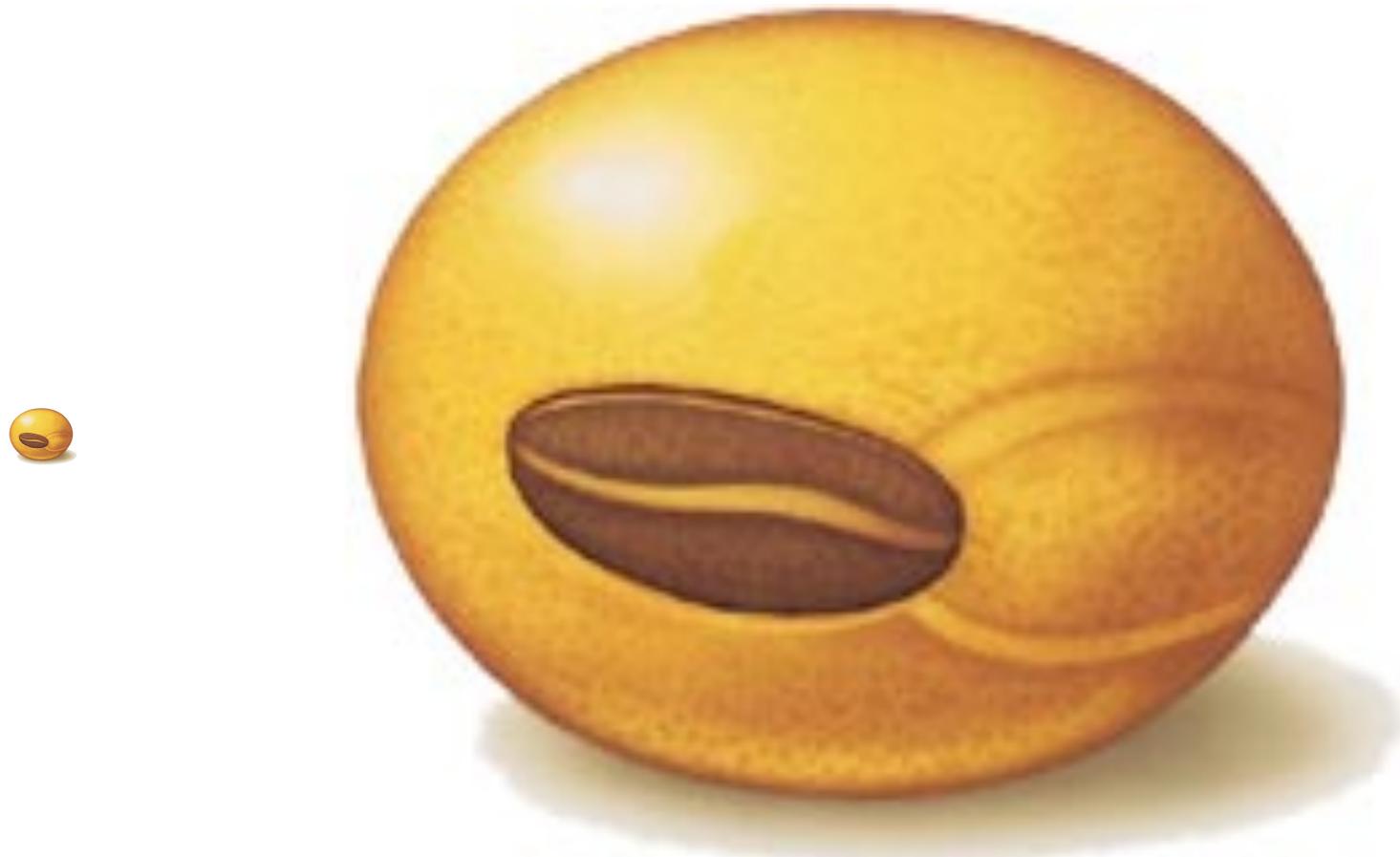
Maturation module



Chalazal endosperm module



The End.....



.....or is it the Beginning?



UCLA
Bob Goldberg
Brandon Le
Chen Cheng
Min Chen
Jungim Hur
Kelli Henry
Zidian Xie
Weihong Yang

Former UCLA
Anh thu Bui
Javier Wagmister
Shundai Li
Xingjun Wang
Harry Hahn
Daisy Robinton
Tomo Kawashima

Monsanto Collaborators
John Danzer
Dave Somers

UC Davis
John Harada
Julie Pelletier
Ryan Kirkbride
Mark Belmonte
Sandra Stone
Jiong Fei
Meryl Hashimoto

UC Berkeley Collaborators
Bob Fischer
Tzung-Fu Hsieh

UCLA Collaborators
Matteo Pellegrini



*Funded By an NSF Plant Genome Grant To
Bob Goldberg & John Harada*