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

Non-Seed Crops

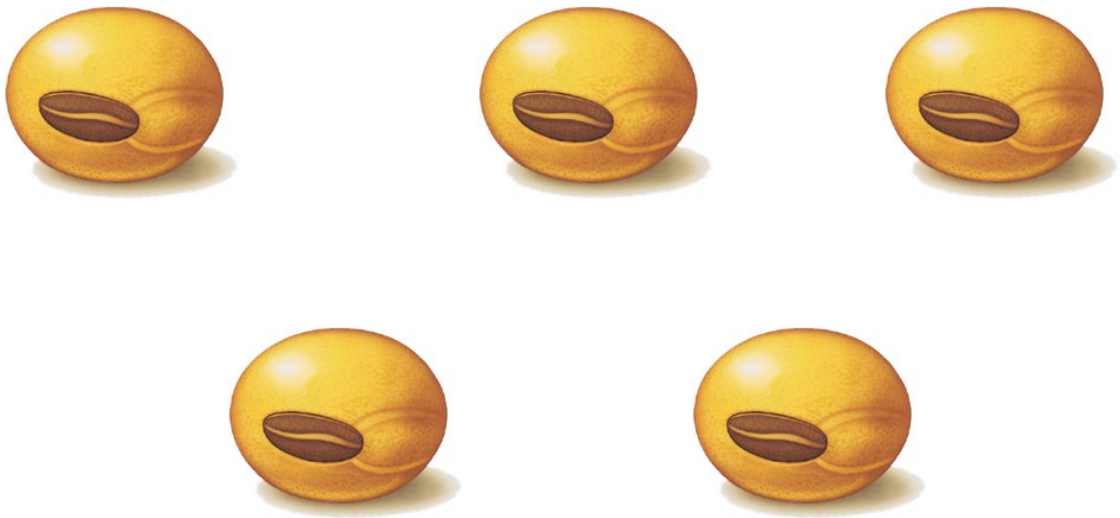
- Potato
- Sweet Potato
- Cassava
- Sugar Beet
- Sugar Cane
- Banana

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

$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!!!

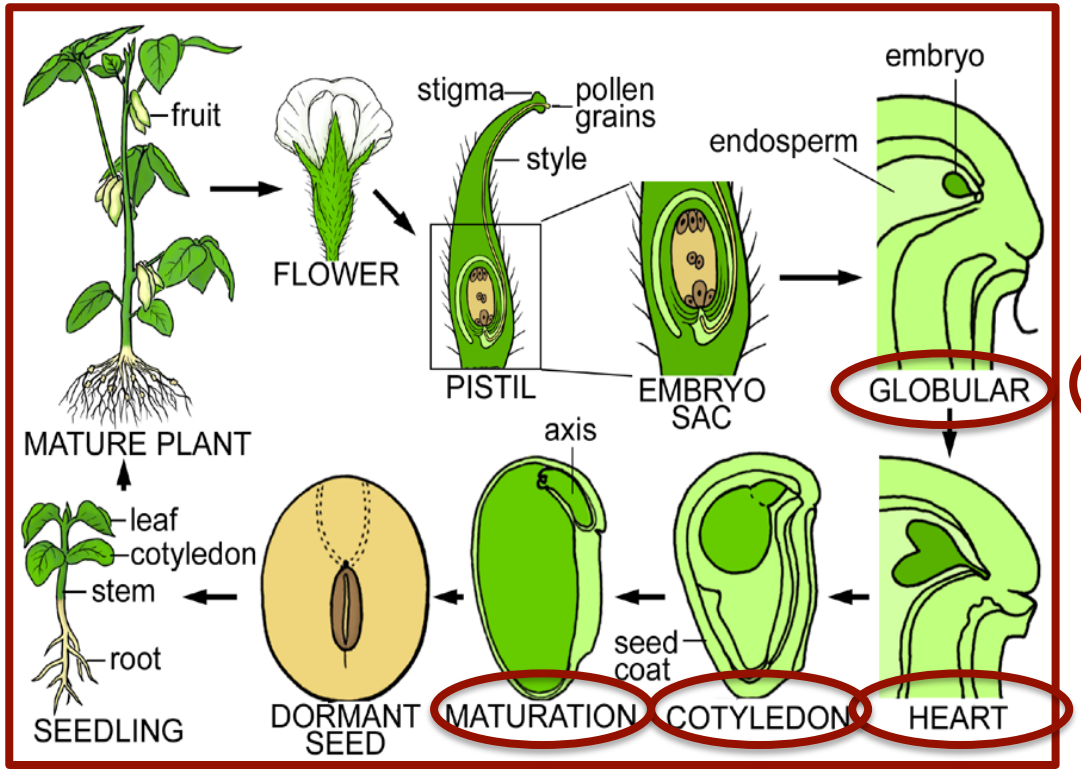
How Is a Seed Formed?



In the Beginning....

What Are the Questions Focused On In This Talk?

Seed Development



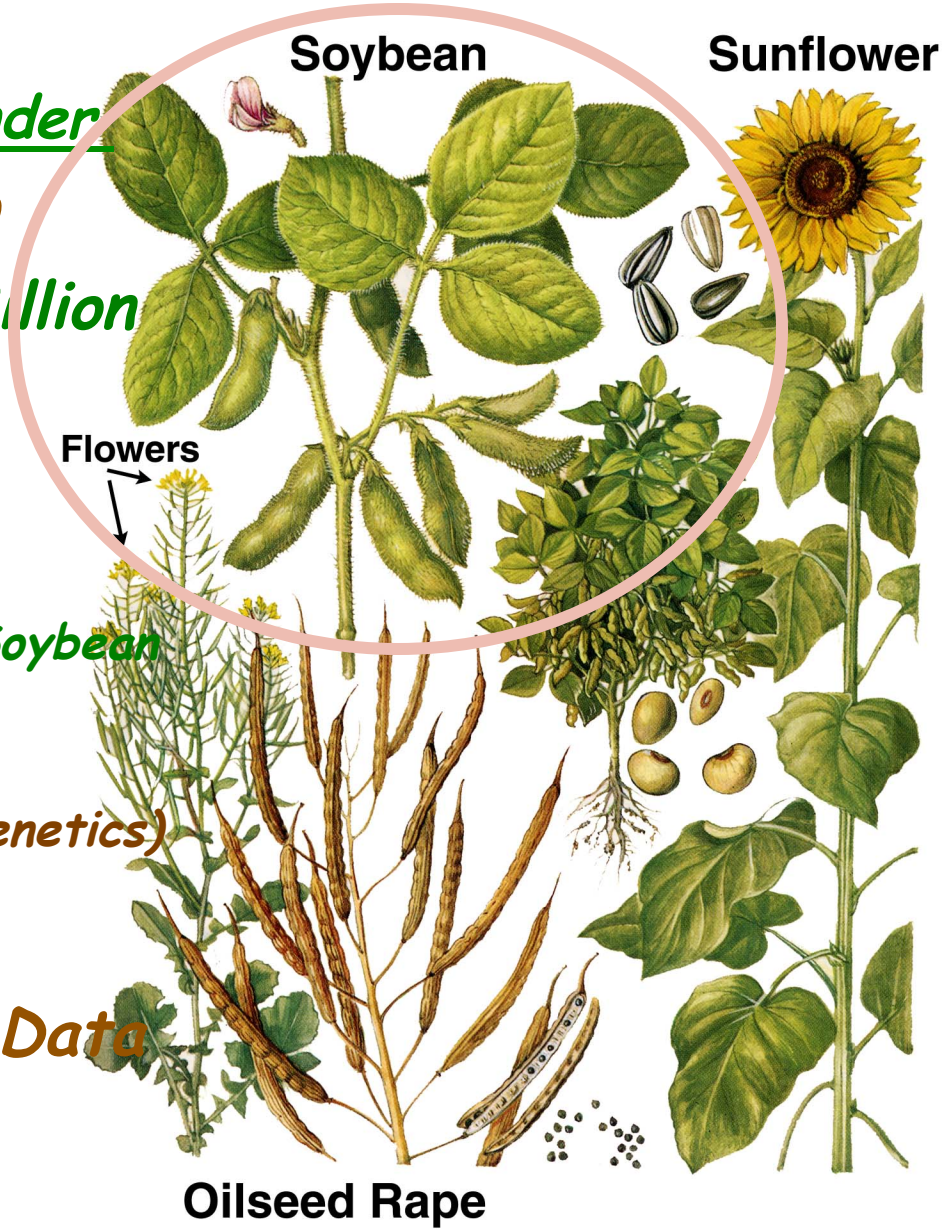
Seed Stages Studied

- 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

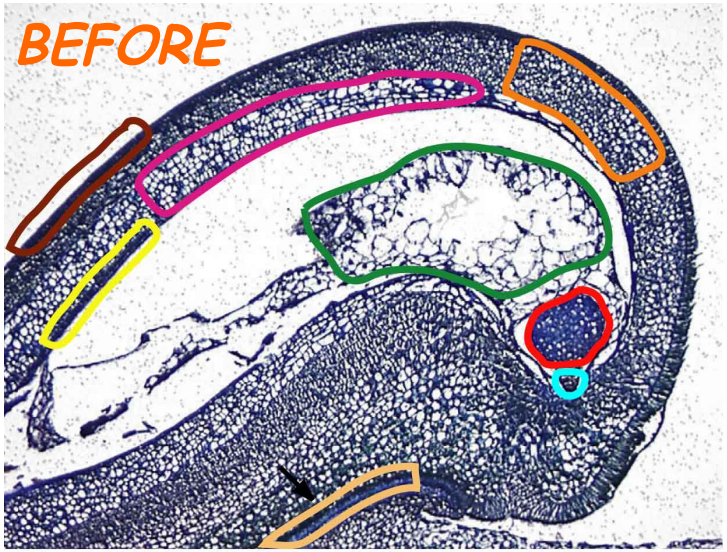
Why Soybean?-A Reminder

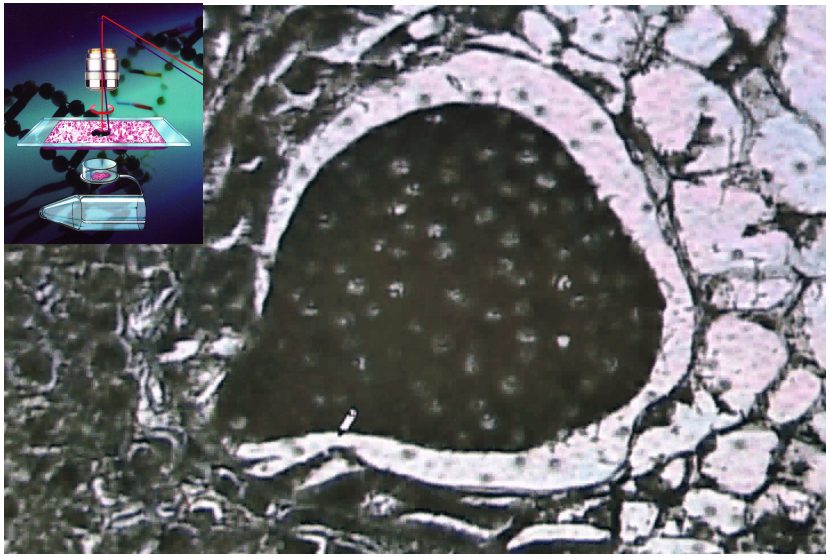
- Second Major US Crop
- Total Crop Value $32Billion
(50% Value Exported)
- Major Food Source
- Important Biofuel Source(Biodiesel~20% of US Soybean Oil Production)
- Excellent Model Plant
(Transformation, Knockdowns, Genetics)
- Genome Sequenced
- Seed Gene Expression Data

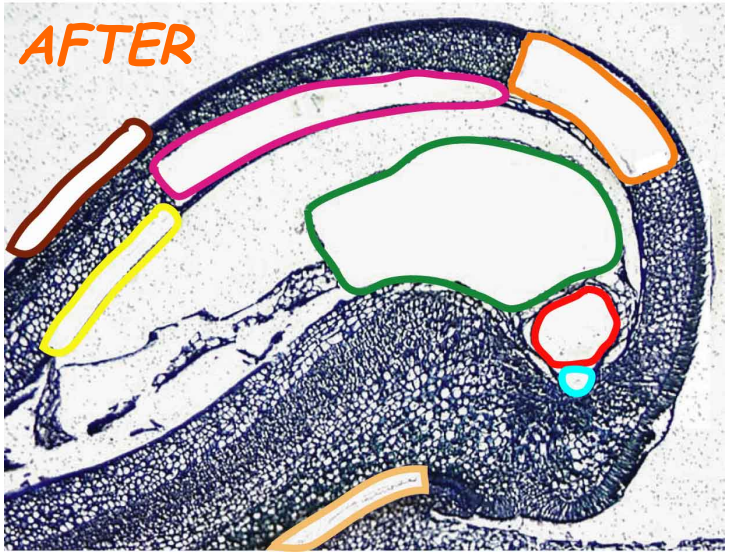


How Study Gene Activity in Different Seed Compartments During Development?

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

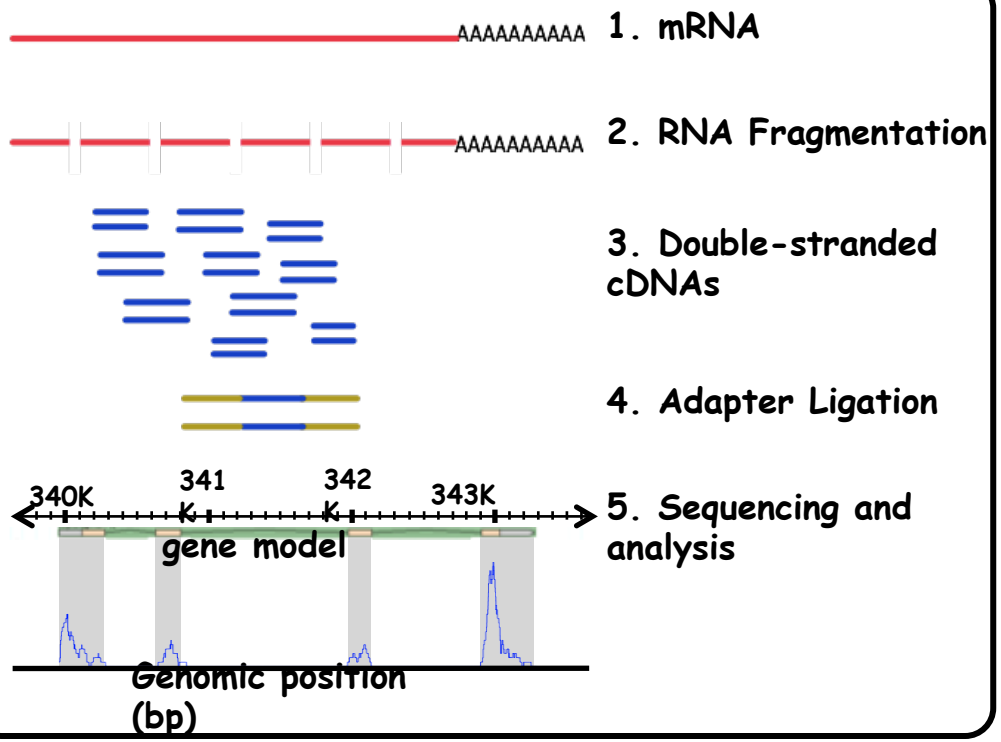
cDNA-Based Array
37,000 Probe Sets
>25,000 Gene Models
~40% of Genome

2nd Generation GeneChip*

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

SEQUENCING

RNASeq

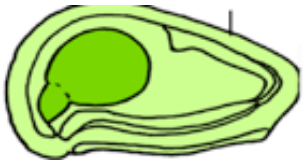


*Note: 2nd generation GeneChips are available and library files can be found at: http://seedgenenetwork.net/annotate#soybeanWT

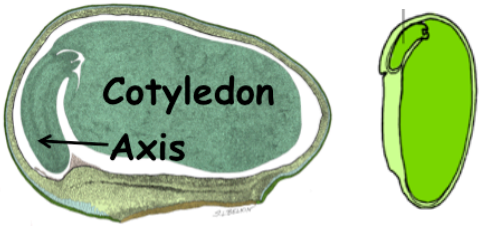
What Developmental Stages and Seed Compartments Studied?

Globular

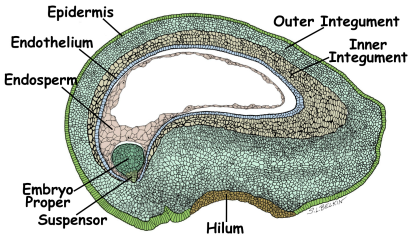
Heart

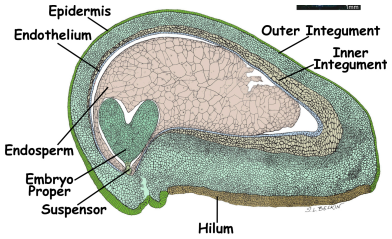


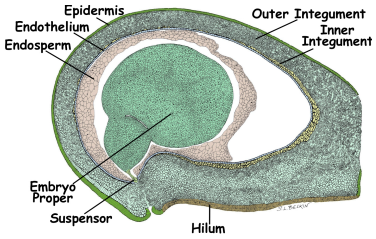
Cotyledon

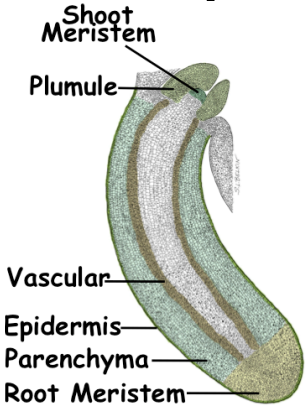


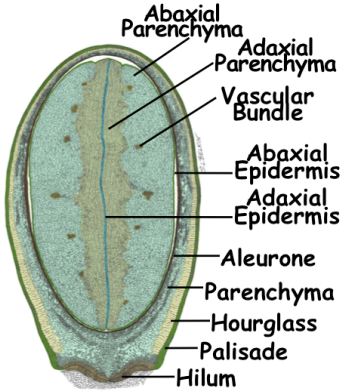
Early Maturation











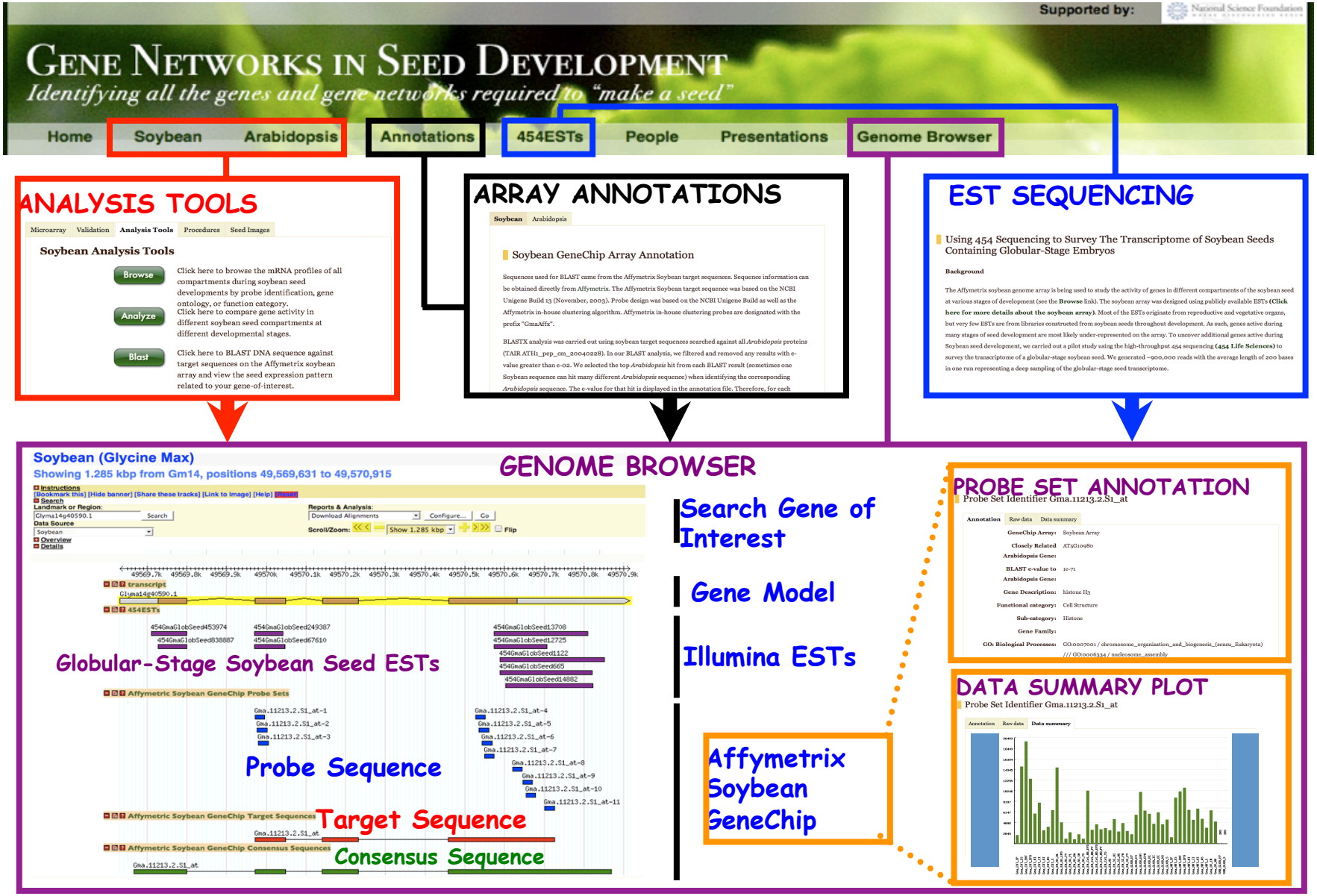
SEED STAGE	GLOB	HRT	COT	EM	TOTAL
# Compartments Studied	8	8	8	16	40
# GeneChip Experiments	26	21	18	34	99
# Genes Validated by qRT-PCR (TF)	42 (28TFs)	36 (20TFs)	55 (55TFs)	21 (21TFs)	154

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

Ran >10,000 qRT-PCR Reactions

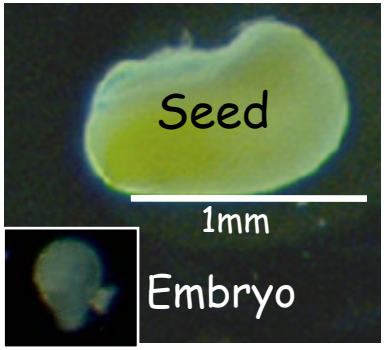
The Interactive Seed Gene Network Website

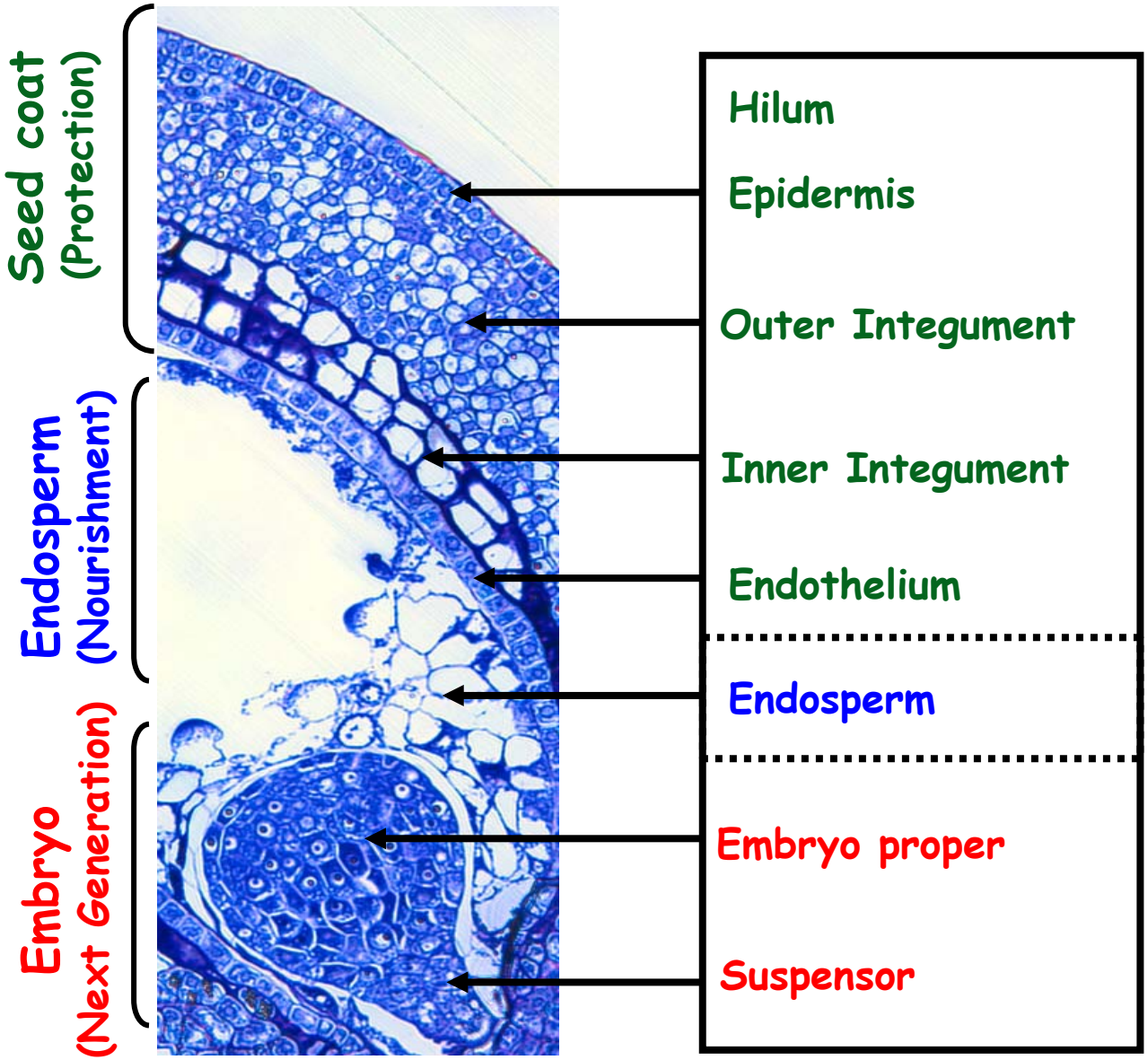
http://seedgenenetwork.net



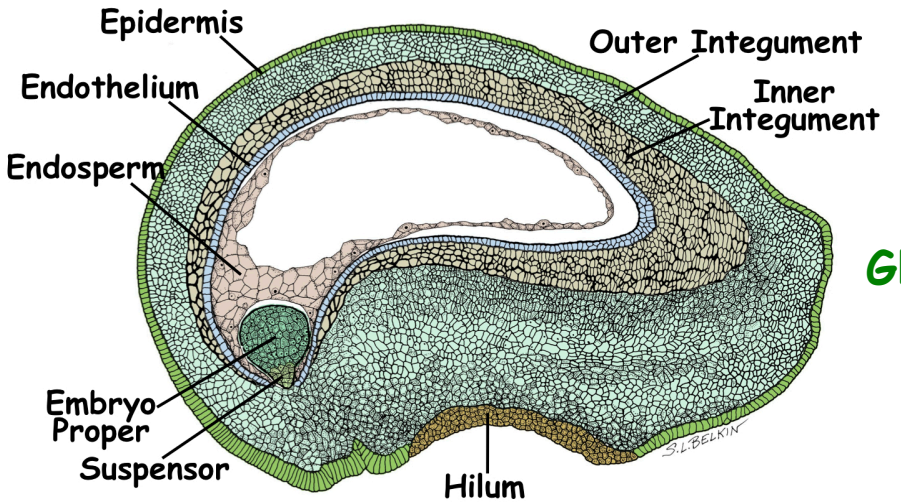
What Are The Genes Required To Make a Seed?
An Example-The Globular Stage

Globular Stage Seed
After Fertilization

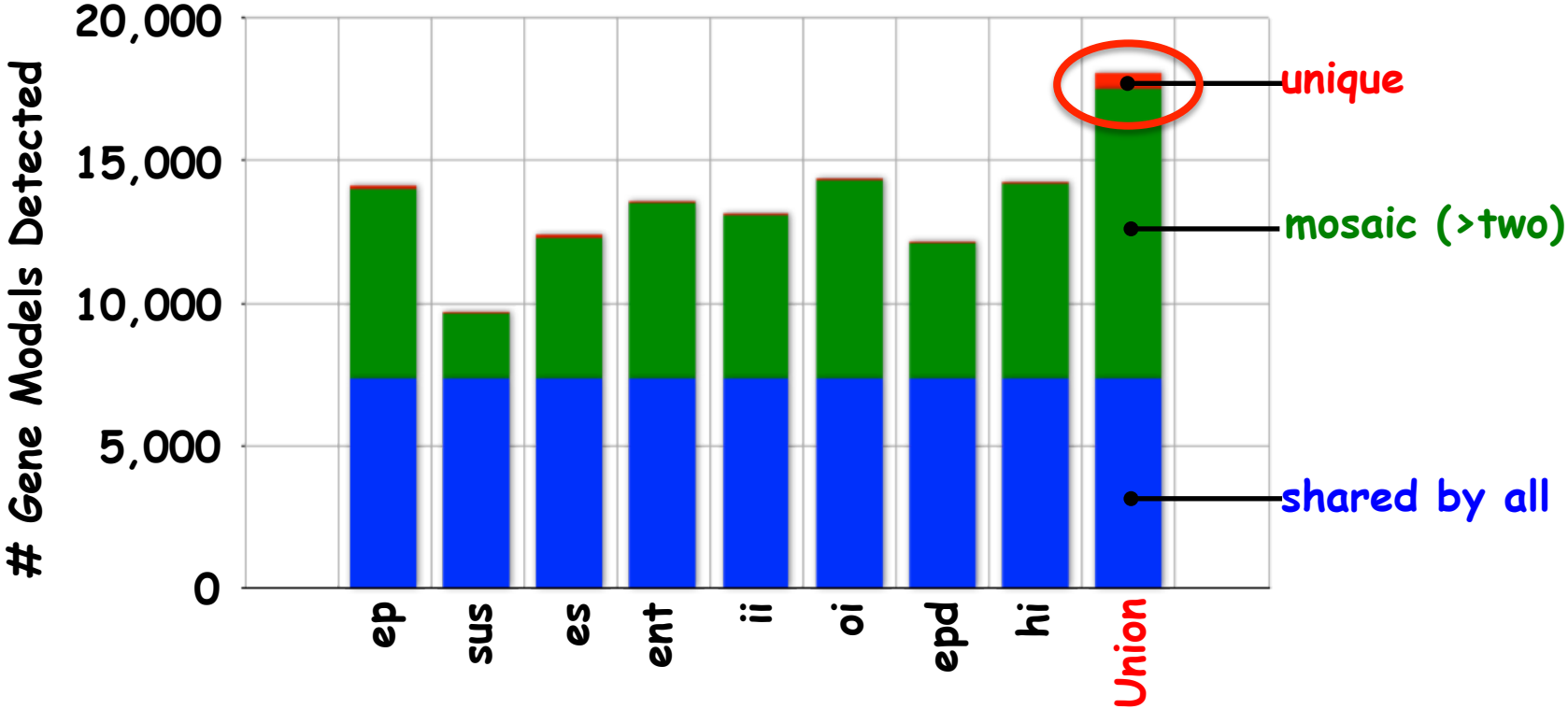




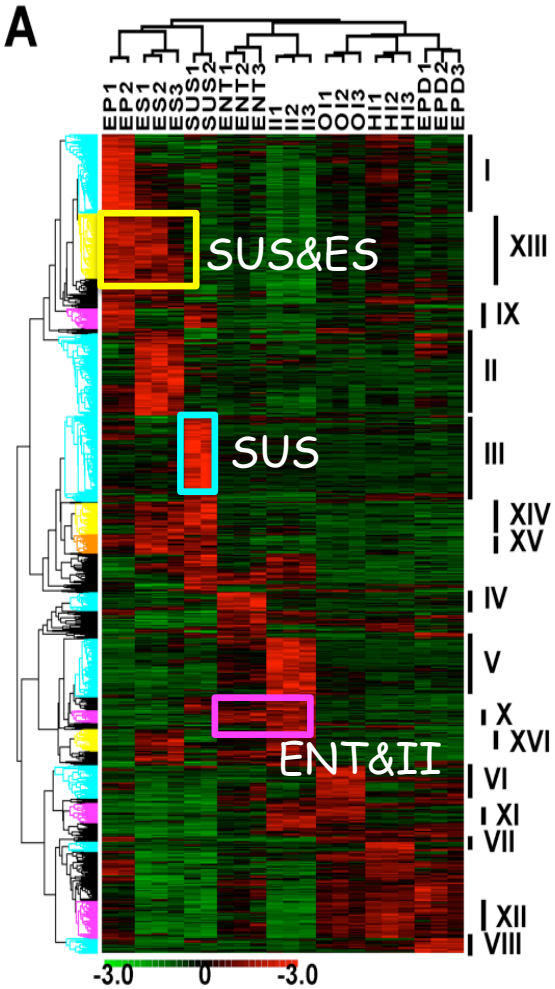
Most Seed mRNAs Are Shared By Different Compartments & Regions



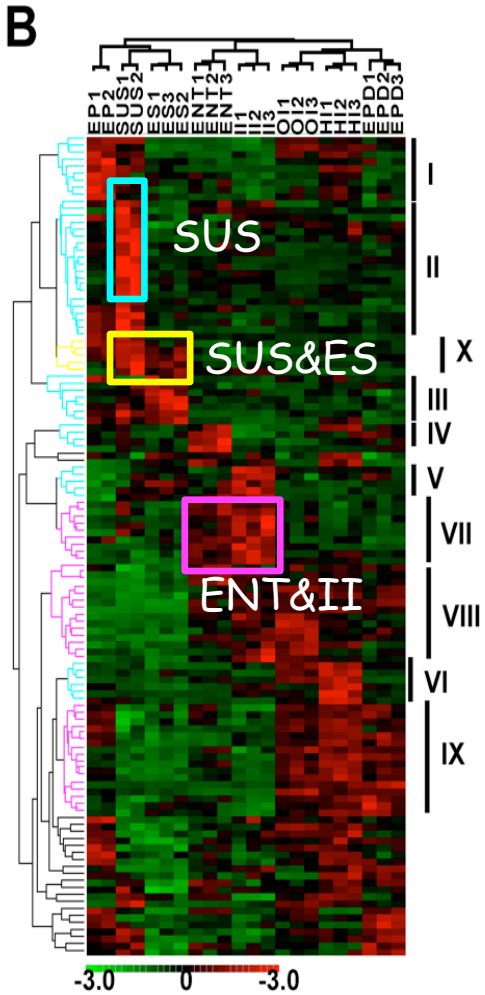
Globular Stage Seed



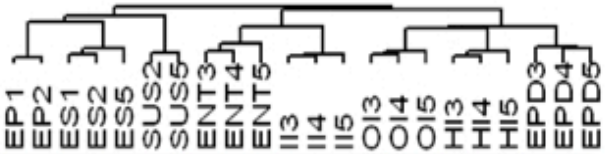
How Are Shared Transcripts Regulated in Globular-Stage Seed Compartments, Regions, and Tissues?



Hierarchical Clustering of the Top 2000 Varying mRNAs, p<0.001



Hierarchical Clustering of the Varying TF mRNAs mRNAs, p<0.001

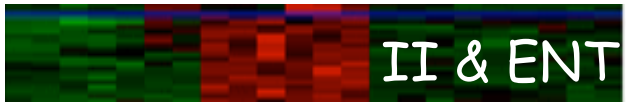


Tissue-Specific Enriched

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)

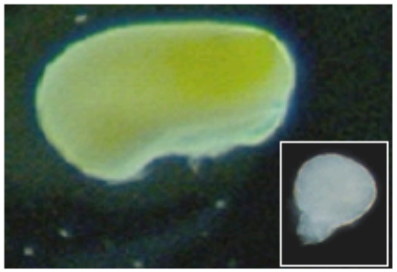


Mosaic

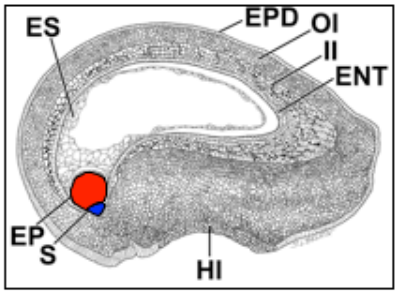
Sets of mRNAs Co-upregulated in multiple seed tissues of the different origins (e.g., EP-ES)

How Many Genes Are Active in the Globular Stage Seed?

Whole Seed



LCM

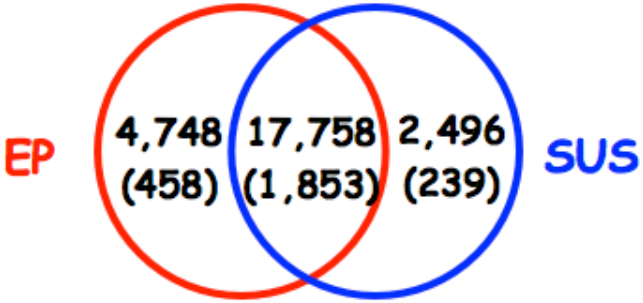


TECHNOLOGY	TISSUE/REGION	GENECHIP	RNA-SEQ
Whole Mount	Whole Seed	17,057 (1,748)	37,875 (3,926)
LCM	EP	14,130 (1,436)	22,506 (2,311)
	SUS	9,713 (929)	20,254 (2,092)
	EP + SUS	20,077 (1,481)	25,002 (2,550)

GeneChip

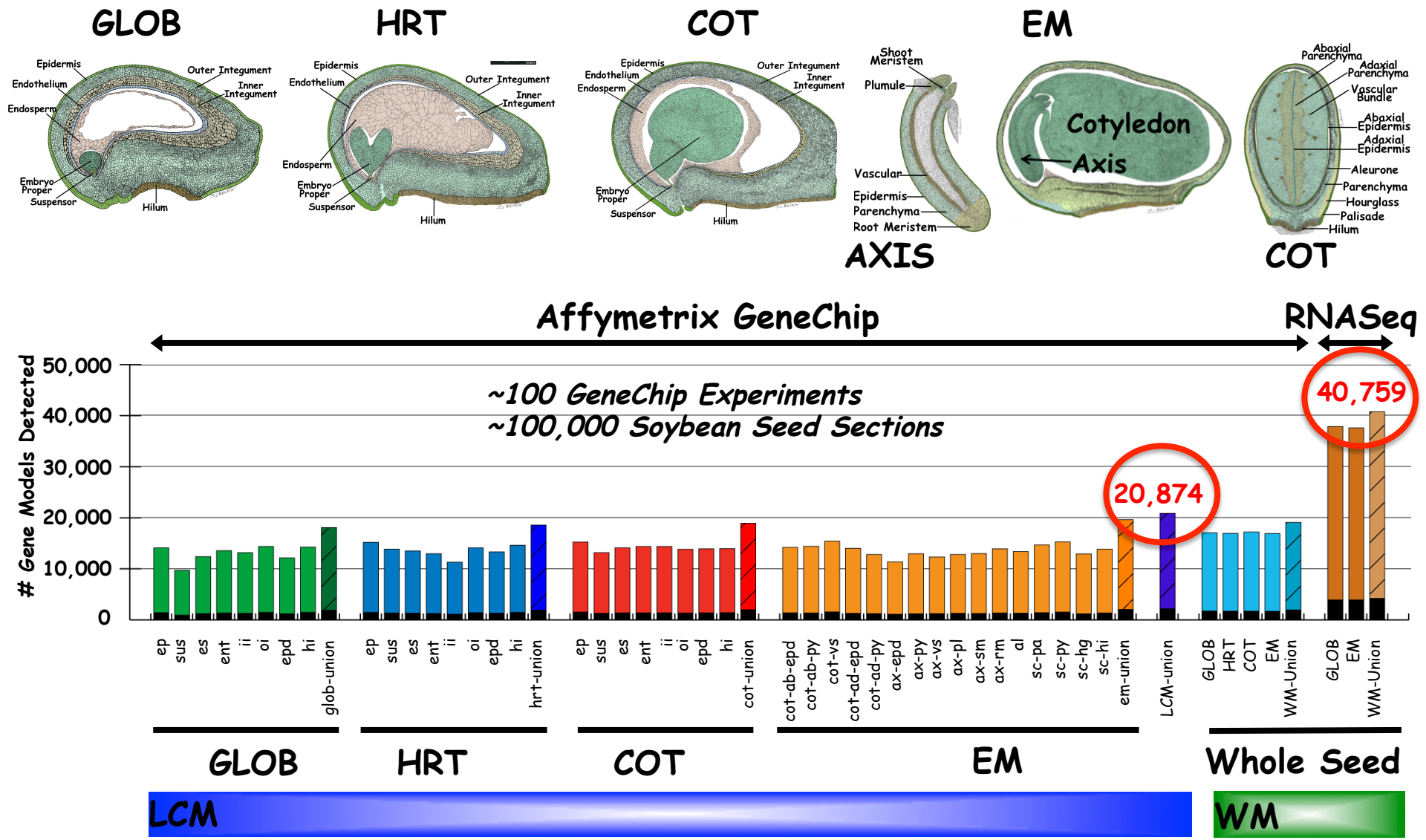
(EP + SUS) = 20,077 (1,481)

RNA-Seq



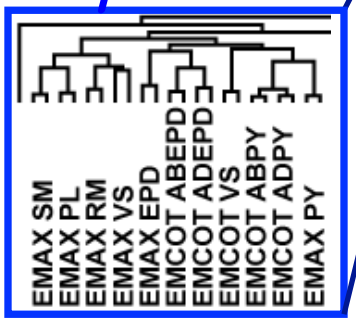
(EP + SUS) = 25,002 (2,550)

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



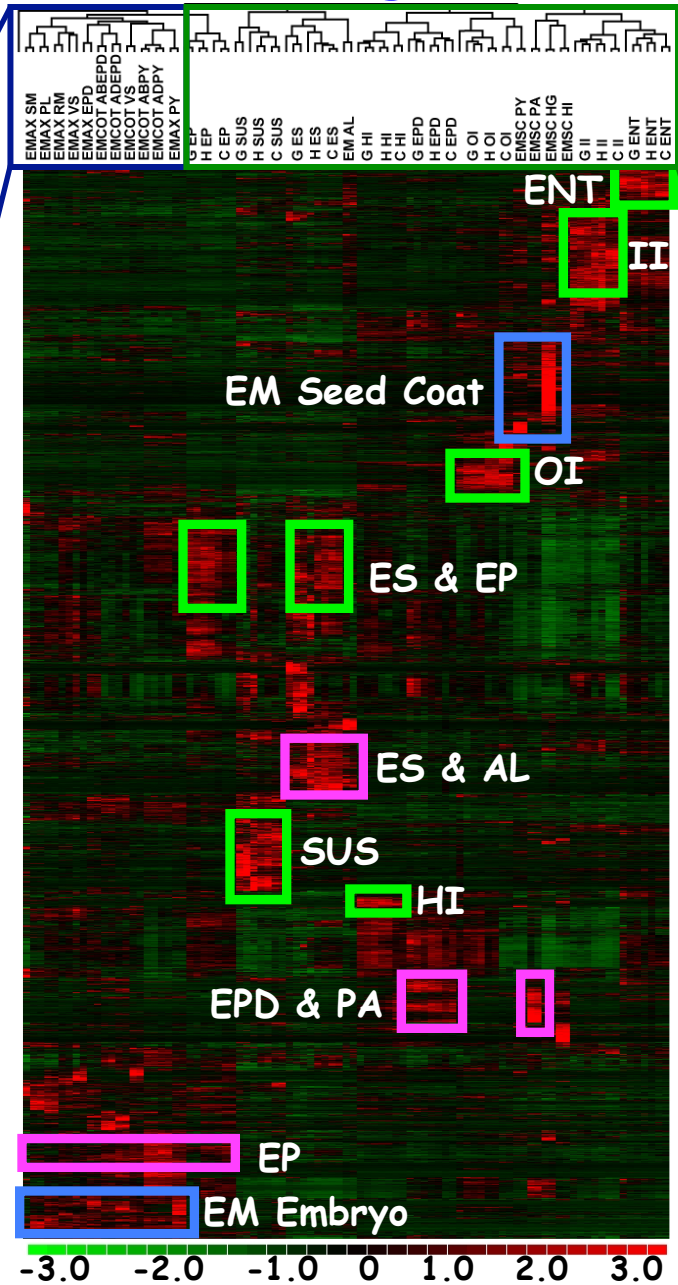
What Are the Temporal and Spatial mRNA Accumulation Patterns Throughout Seed Development?

Maturation phase



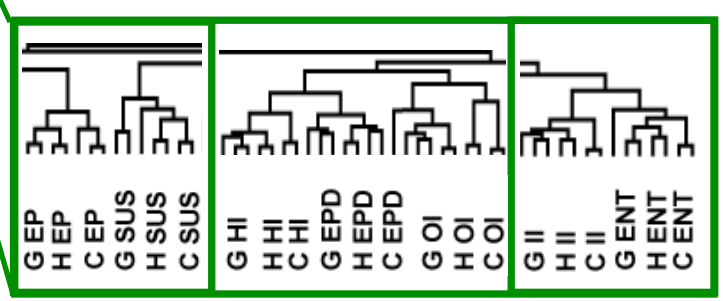
2

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



Top 4K Varying mRNAs
ANOVA p<0.0001

Morphogenesis phase



1

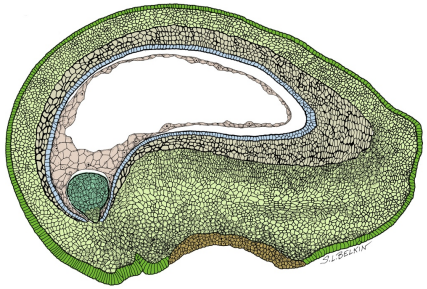
A SPATIAL pattern of mRNA accumulation is apparent (GLOB-HRT-COT)

3

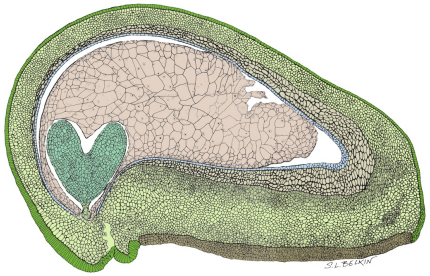
mRNA sets up-regulated throughout development (GLOB to EM)

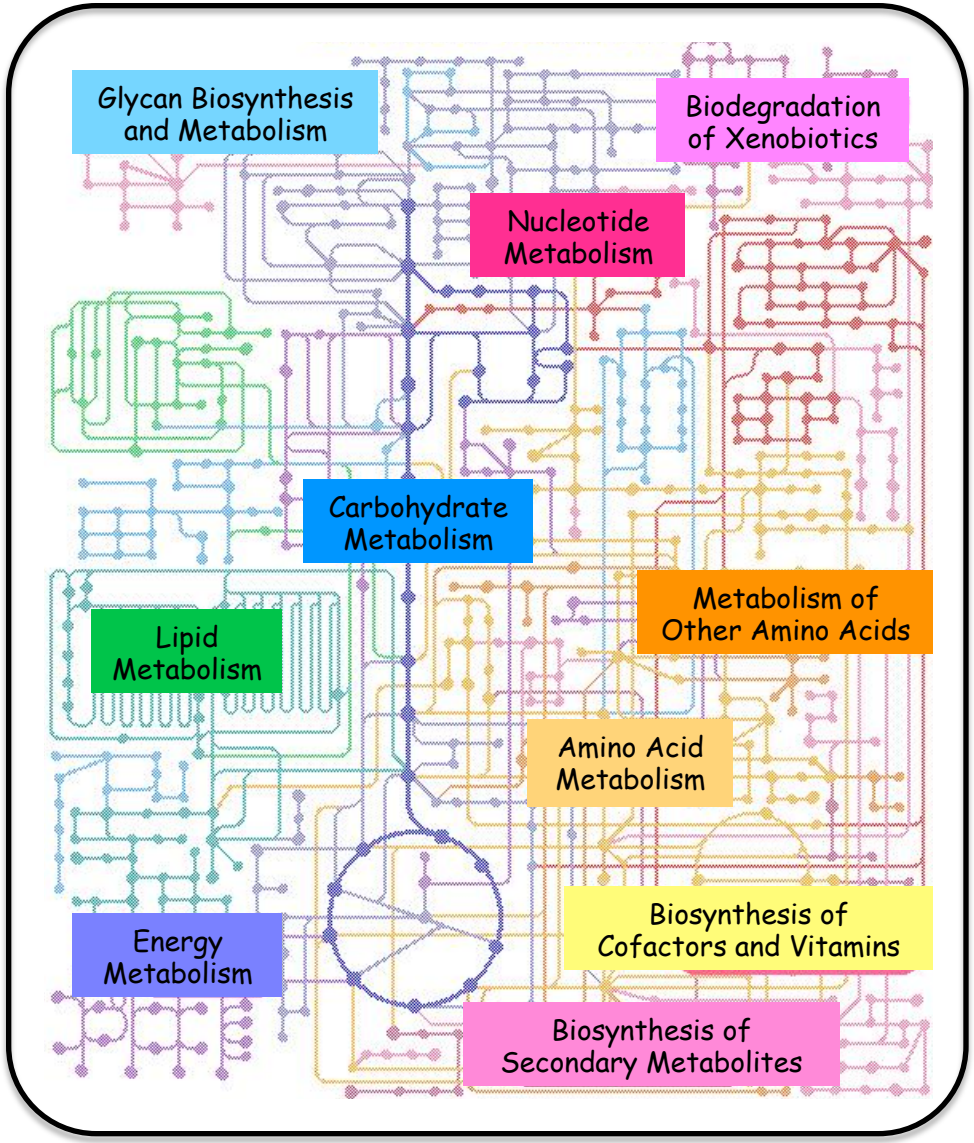
Are Seed Compartments Specialized For Specific Metabolic Processes?

Globular

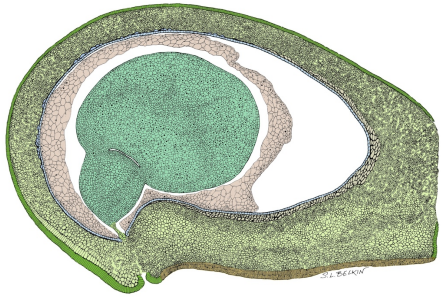


Heart

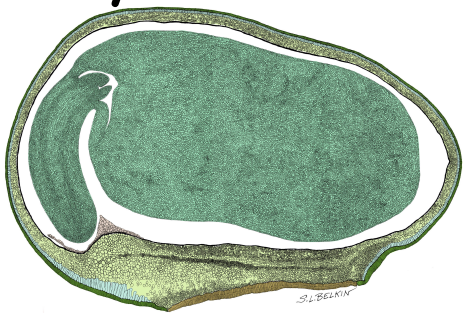




Cotyledon

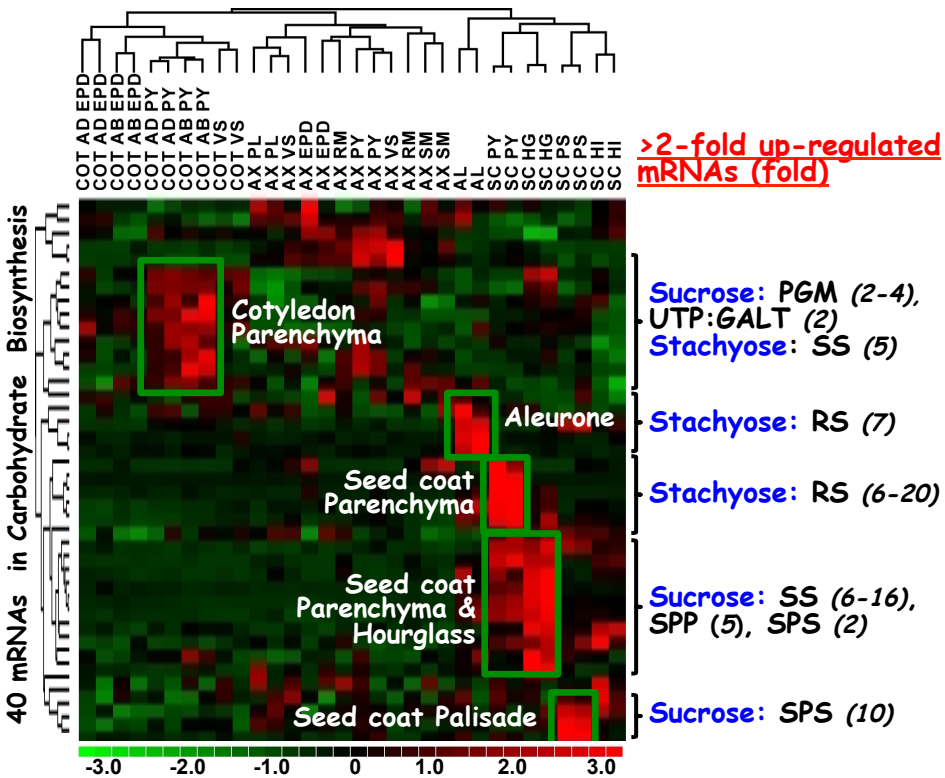


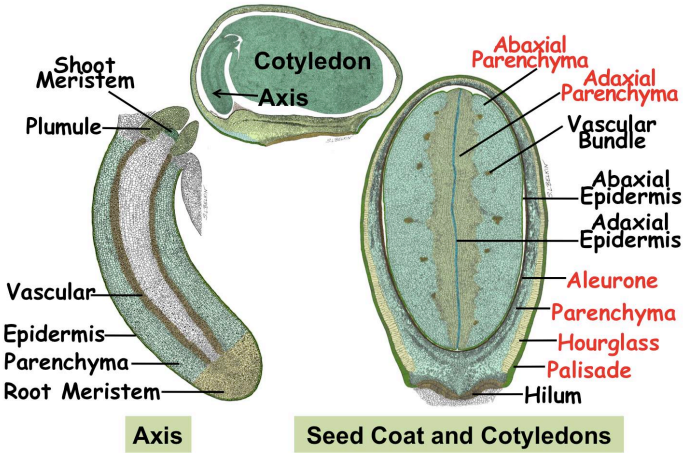
Early Maturation

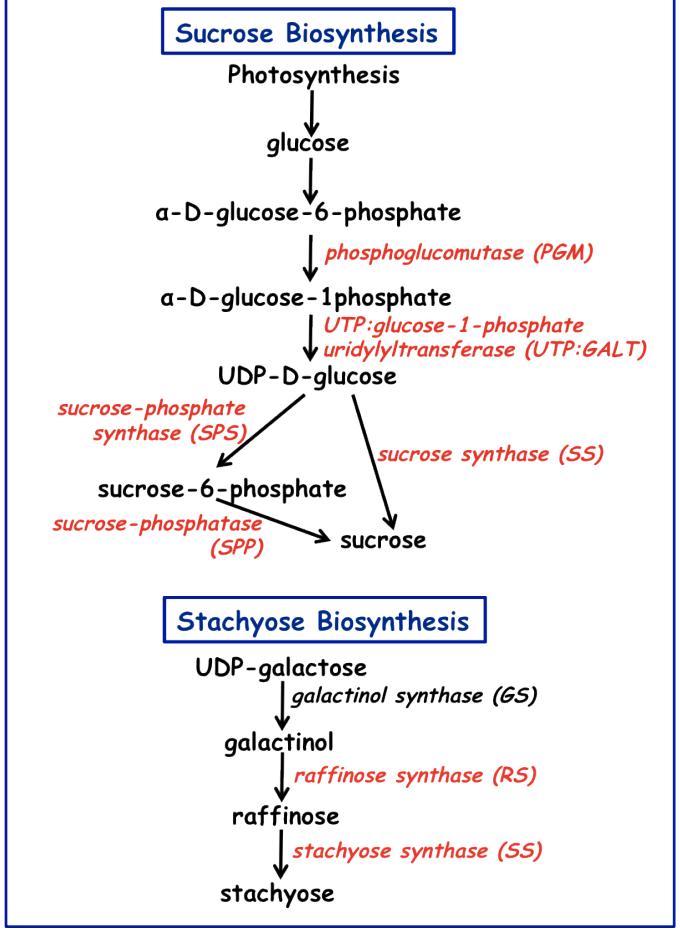


Examined 325 Metabolic Pathways in 40 Different Seed Compartments

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?

AXIS

PLUMULE
Storage Protein (β-conglycinin)

SHOOT MERISTEM
Storage Protein (β-conglycinin)
Amino acids (Ser, Trp)

PARENCHYMA
Isoflavones (Genistein, Daidzein)
Storage Protein (β-conglycinin)

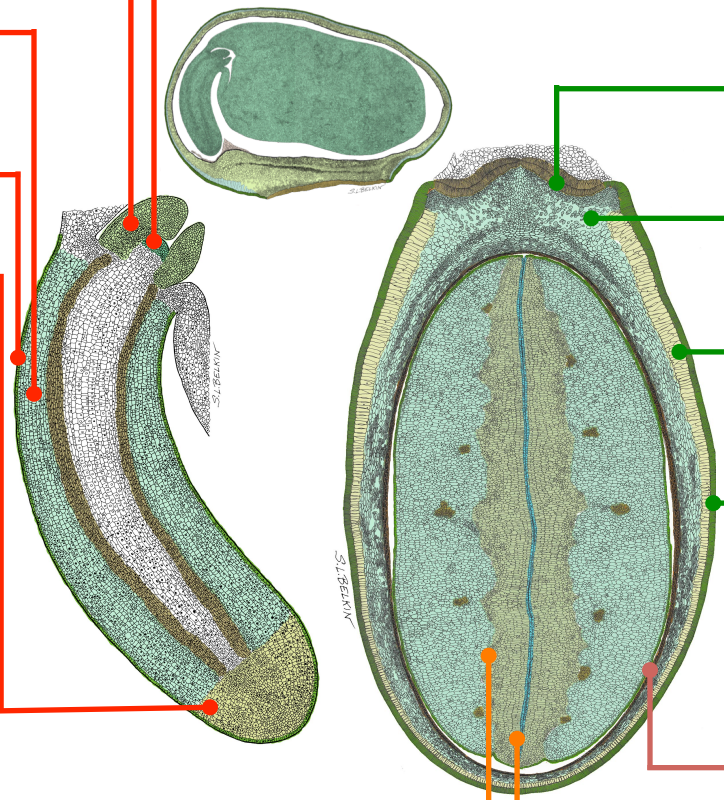
EPIDERMIS
Vitamin B1

ROOT MERISTEM
Isoflavones (Genistein, Daidzein)

COTYLEDON

PARENCHYMA
Isoflavones (Genistein,Daidzein)
Vitamin B1, B2, B5, B9, C, E
Fatty acids (Linoleate, Oleate)
Amino acids (Asn, Asp, Cys, His, Ile, Leu, Lys, Met, Thr, Val)
Storage Proteins (β-conglycinin, Glycinin)
Sucrose, Stachyose

ADAXIAL EPIDERMIS
Storage Protein (Glycinin)



SEED COAT

HILUM
Amino acids (Asn, Phe)

PARENCHYMA
Sucrose, Stachyose
Amino acid (Gln)

HOUR GLASS
Vitamin B9,C, E
Sucrose
Amino acids (Cys, Gln, Gly, Lys)

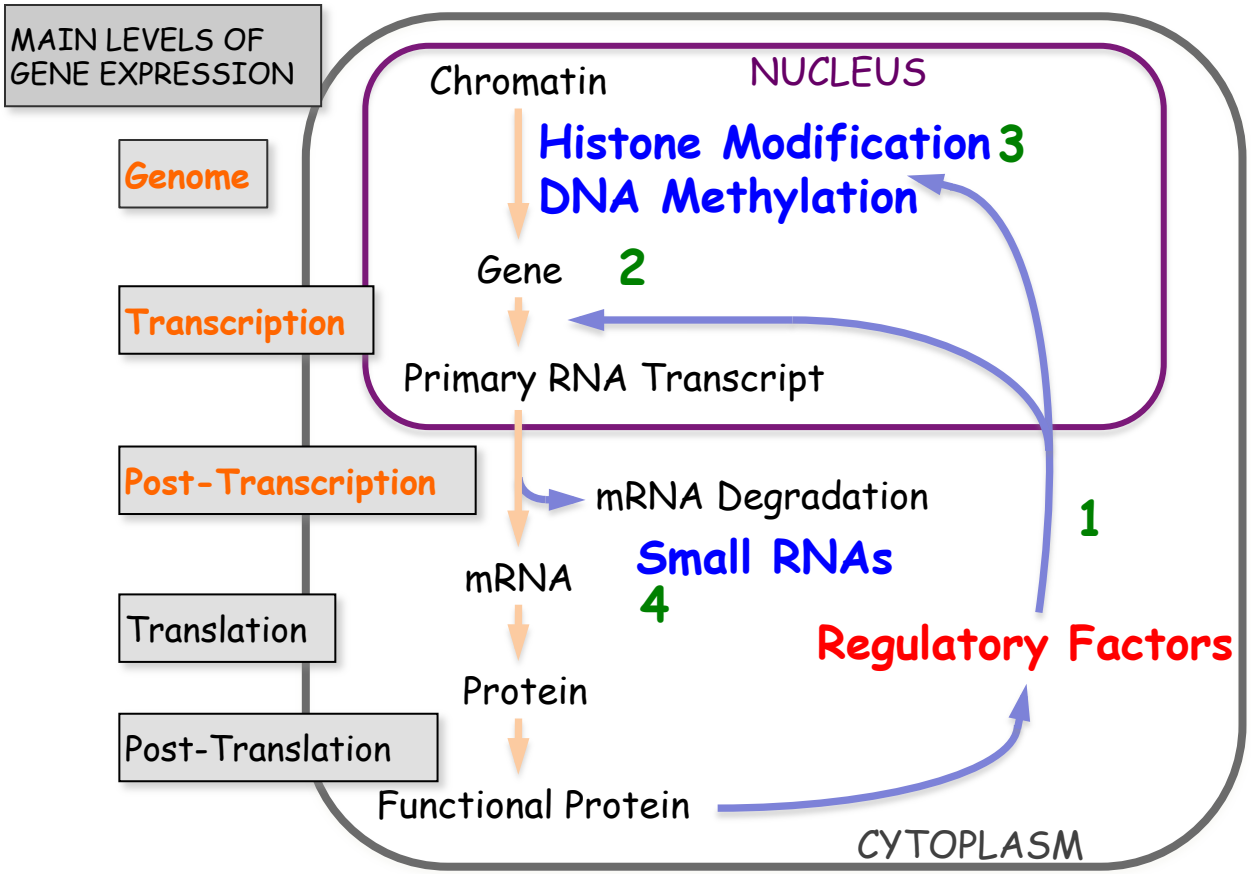
PALISADE
Isoflavones (Genistein, Daidzein)
Fatty acids (Linoleate, Oleate)
Sucrose
Amino acids (Gln, Phe, Ser)

Aleurone

Vitamin B2, B9
Stachyose
Amino acid (Gln)

*Biosynthetic pathways having mRNAs more than 2-fold up-regulated in the indicated tissue compared with all other tissues in early maturation-stage are listed

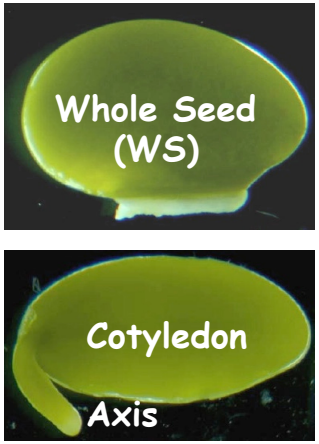
Unraveling the Gene Regulatory Processes that Are Required to "Make a Seed" (The Next Step!)



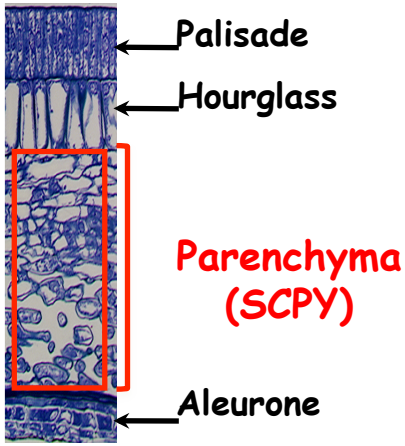
1. Using **RNAi** to knock down compartment-specific TF mRNAs during seed development
2. Identifying **Downstream Targets** of Compartment-Specific Transcription Factors
3. Characterization of compartment-specific **microRNA** populations during seed development
4. Characterization of **DNA methylation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

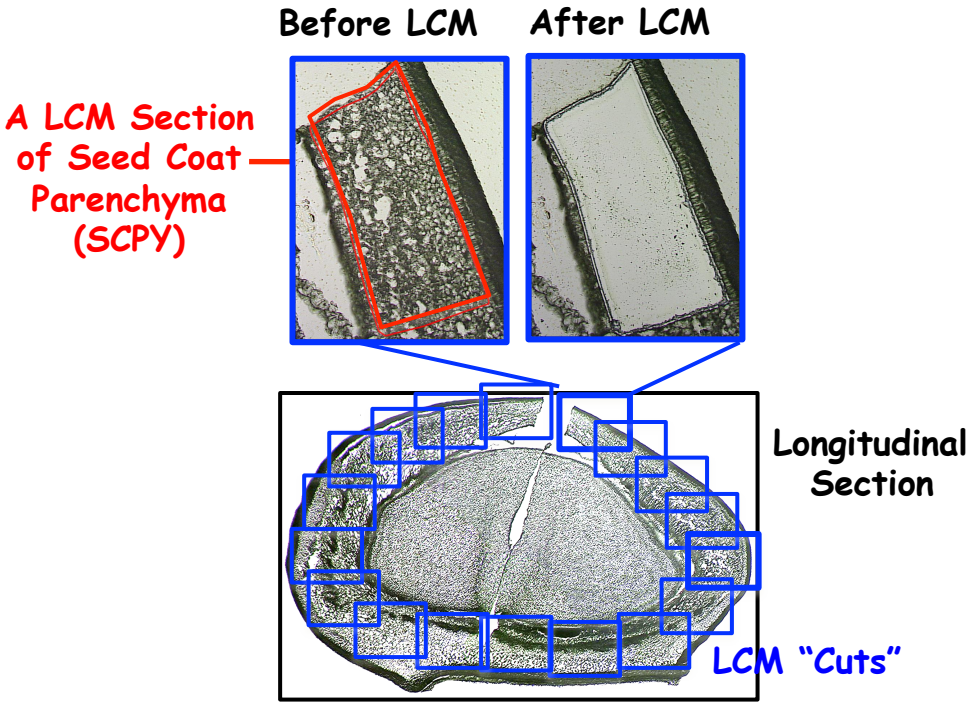
Using LCM to Identify Seed Compartment-Specific MicroRNAs & Methylation Patterns

Early Maturation Stage Seed



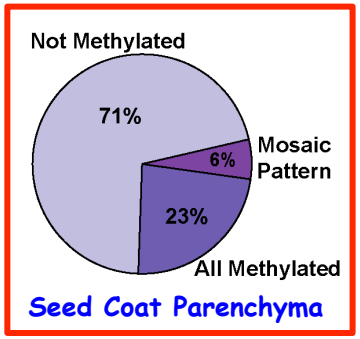
Early Maturation Seed Coat Layers



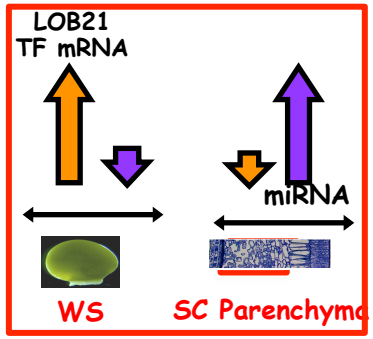


	Small RNA	DNA
# Longitudinal sections (10μm)	23	190
Avg. # LCM sections per longitudinal section	20	20
Total # LCM sections captured	460	3,800
Amount of Enriched RNA or DNA (ng)	~125	~450

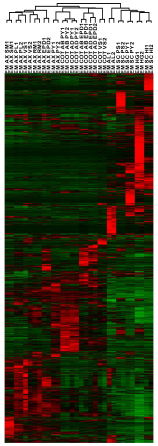
Methylome (1X)

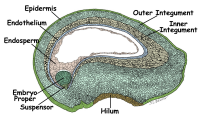


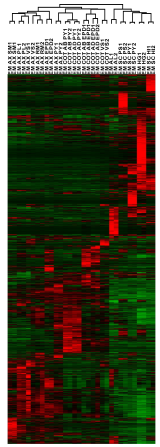
MicroRNAs (21M Reads)

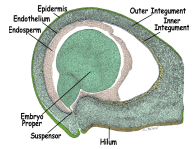


What Are the Functions of Seed Compartment-Specific Transcription Factor mRNAs?

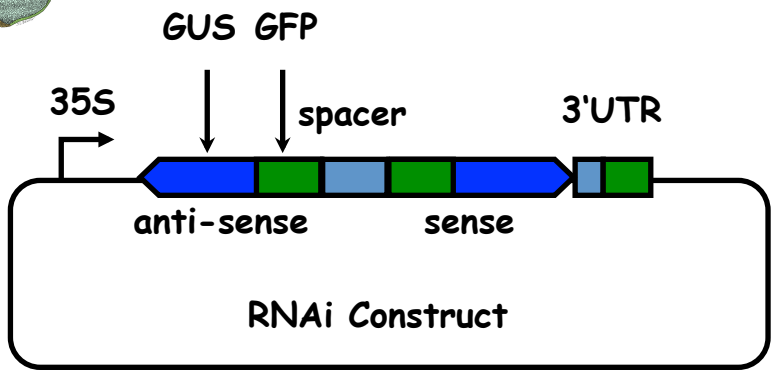






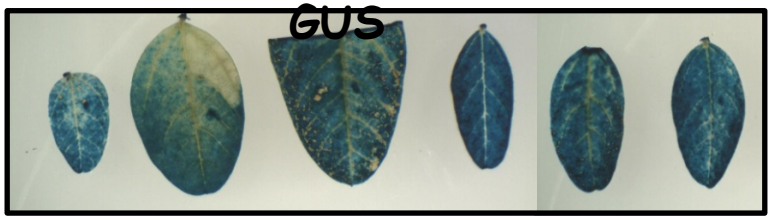


500 Transgenic Soybean RNAi Lines Generated and 61 TF mRNAs Knocked Down



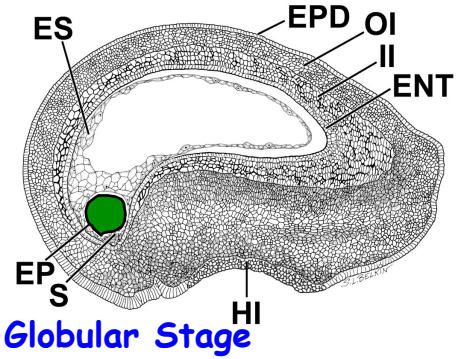
35S-GUS

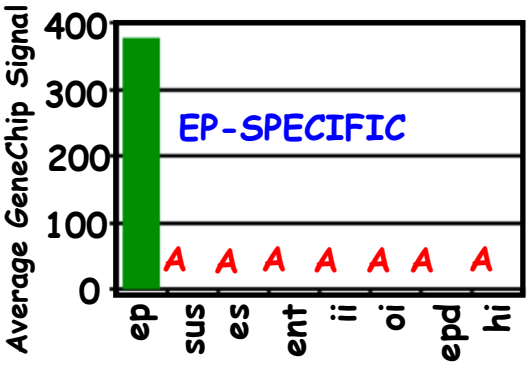
35S-GUS

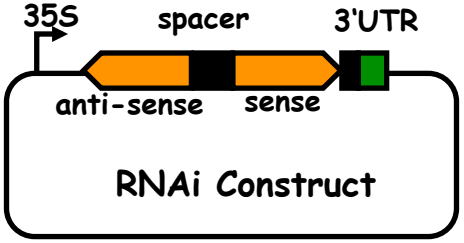


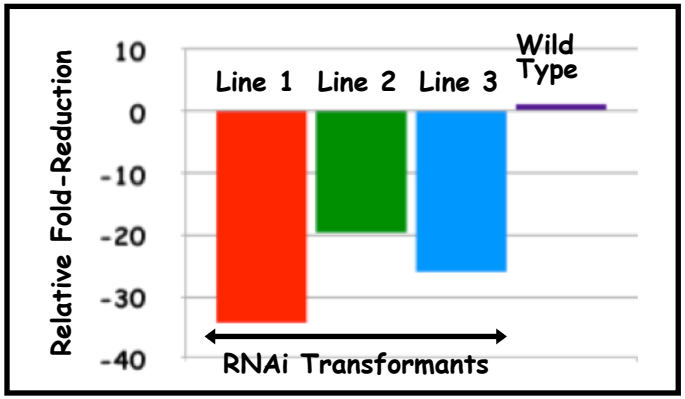
35S-GUS; 35S-RNAi

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

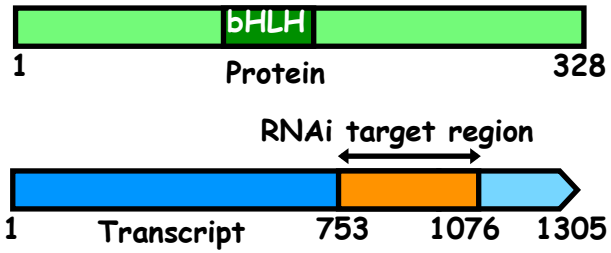


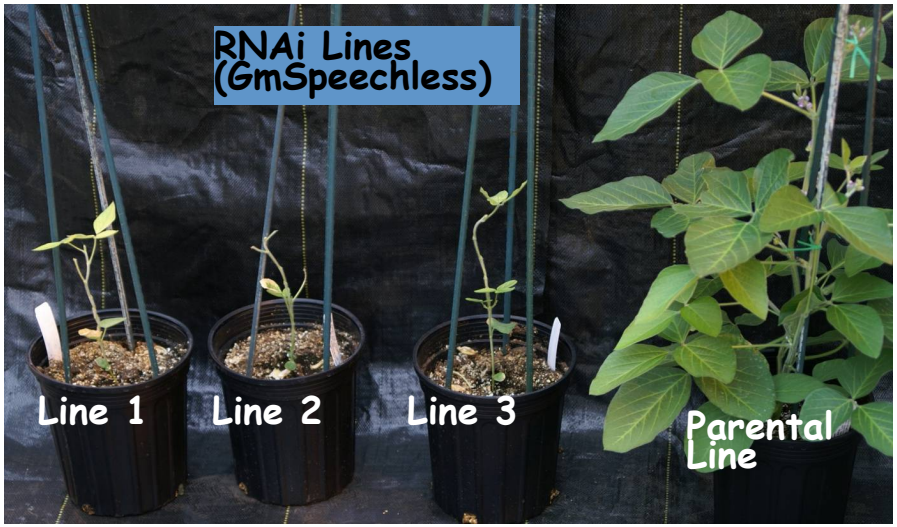


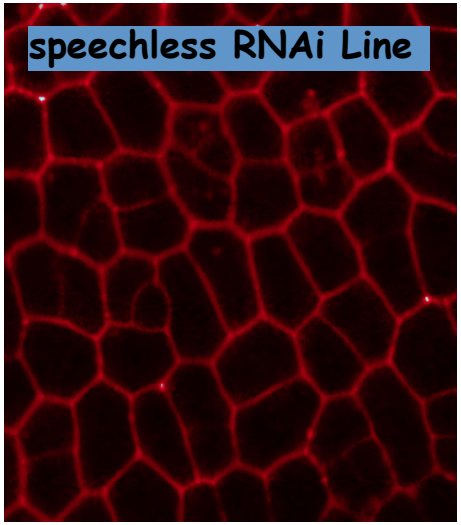


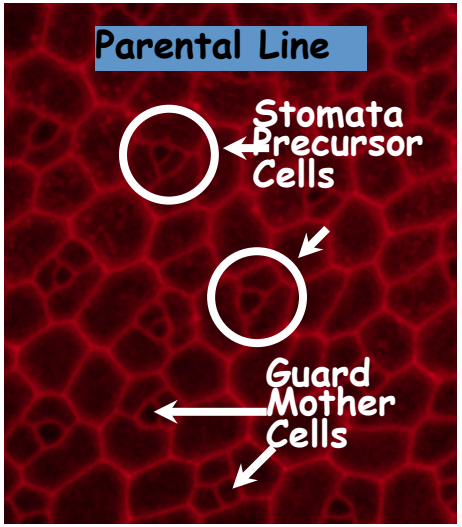


GmSpeechless
(Glyma04g41710)

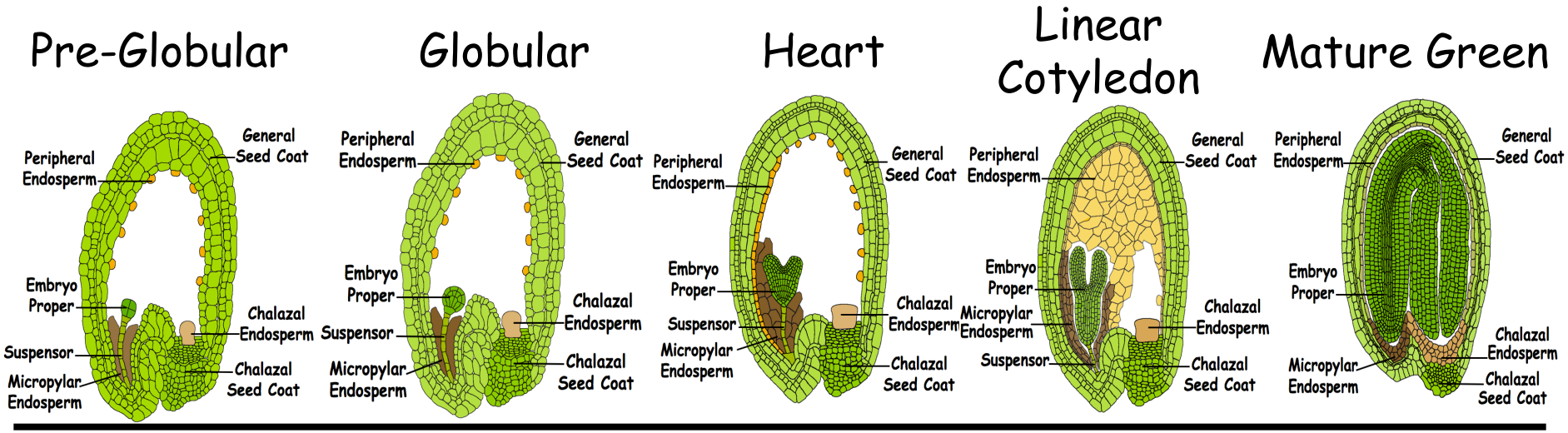








What About Another Seed……Arabidopsis?



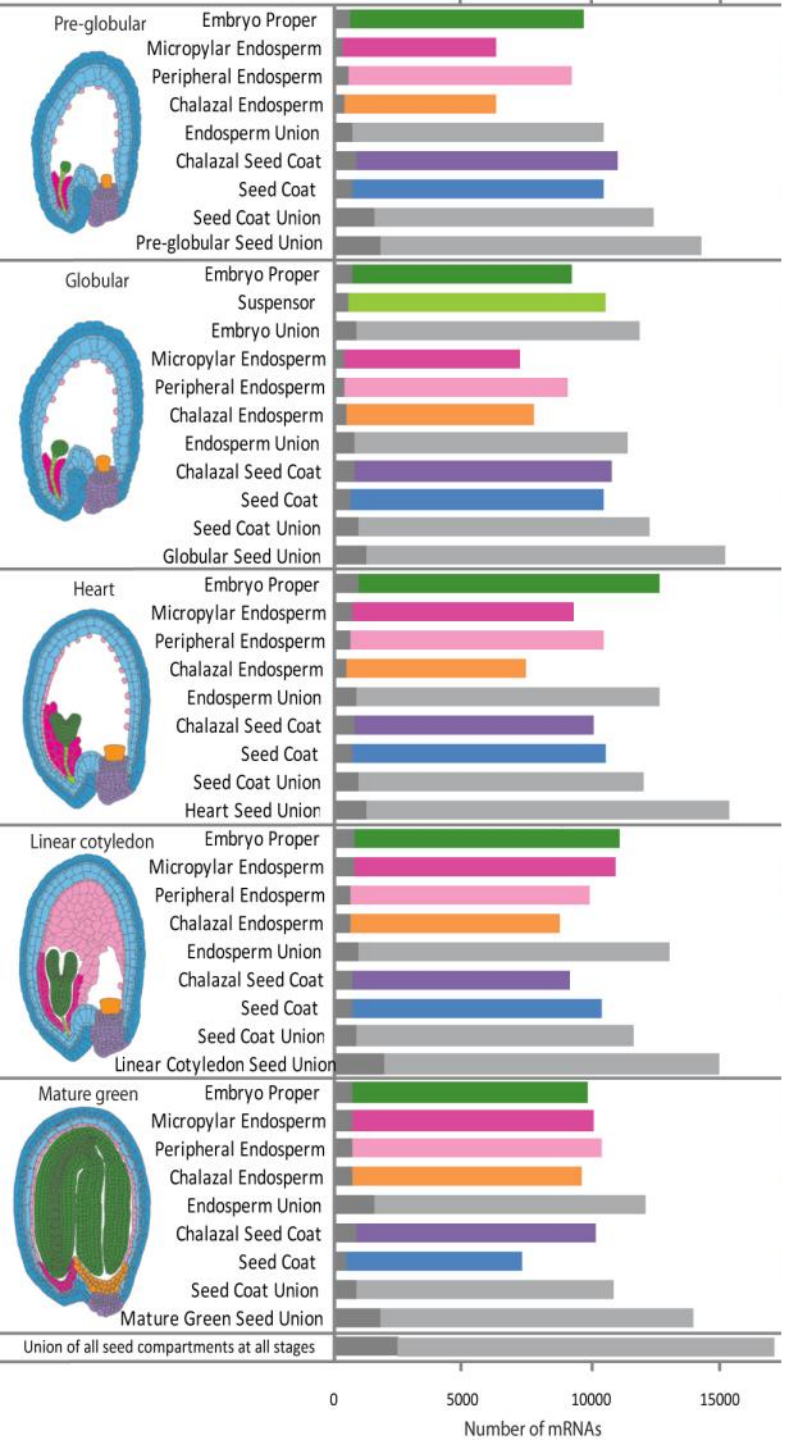
SEED STAGE	PREGLOB	GLOB	HRT	LCOT	BENT	MG	TOTAL
# Compartments Studied	6	7	6	6	5	6	76
# GeneChip Experiments	12	15	14	12	10	12	166
# Genes Validated by qRT-PCR (TF)	A total of 148 genes (>10,000 reactions)						

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

Ran >10,000 qRT-PCR Reactions

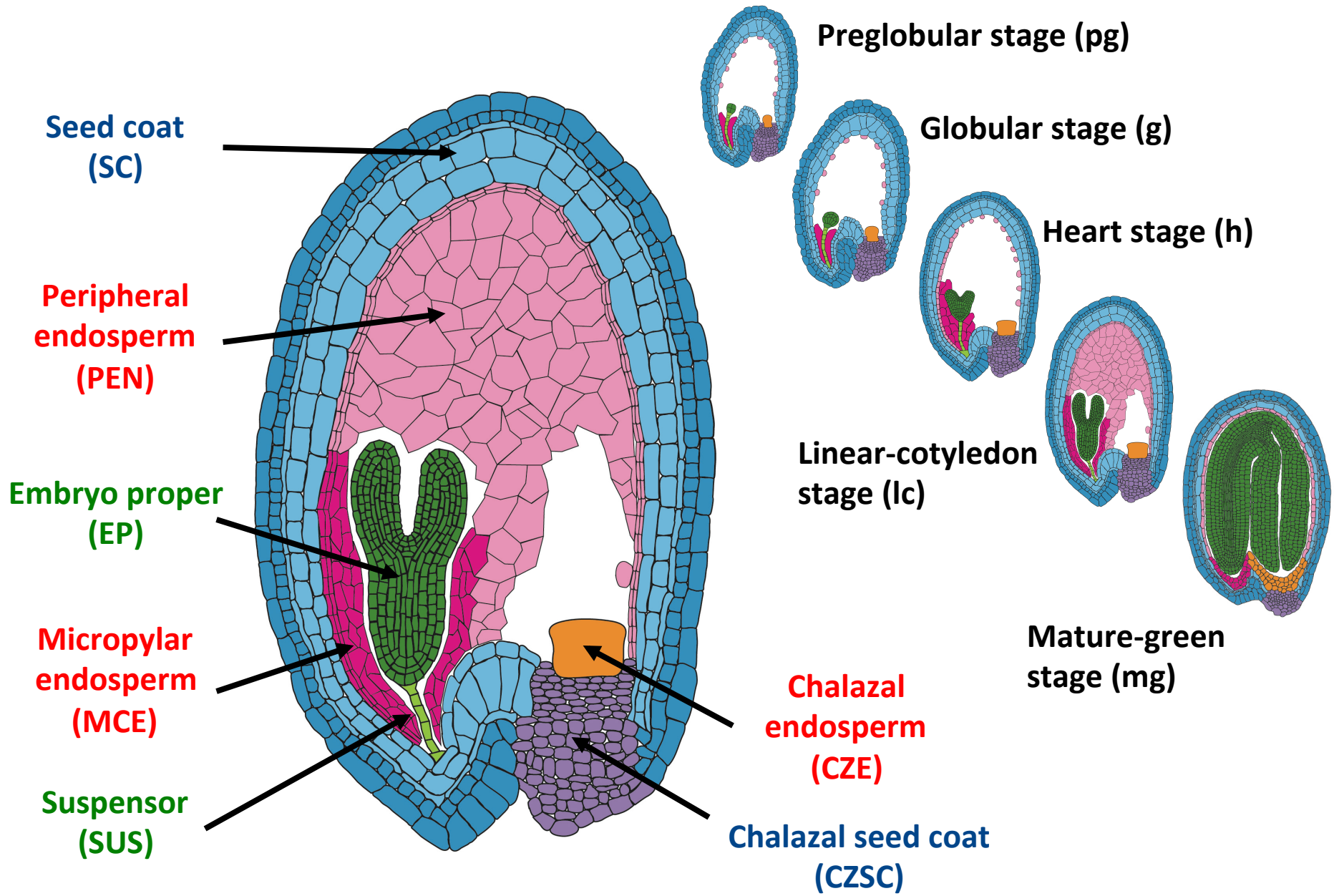
Soybean ***Arabidopsis***

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



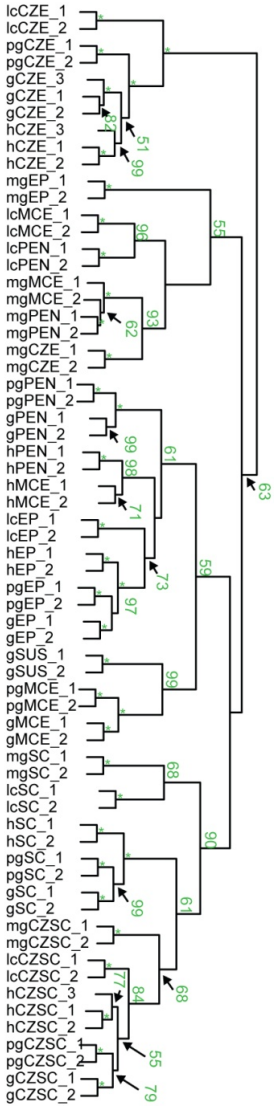
Data available @ GEO & http://www.seedgenenetwork.net/

Arabidopsis Seed Development

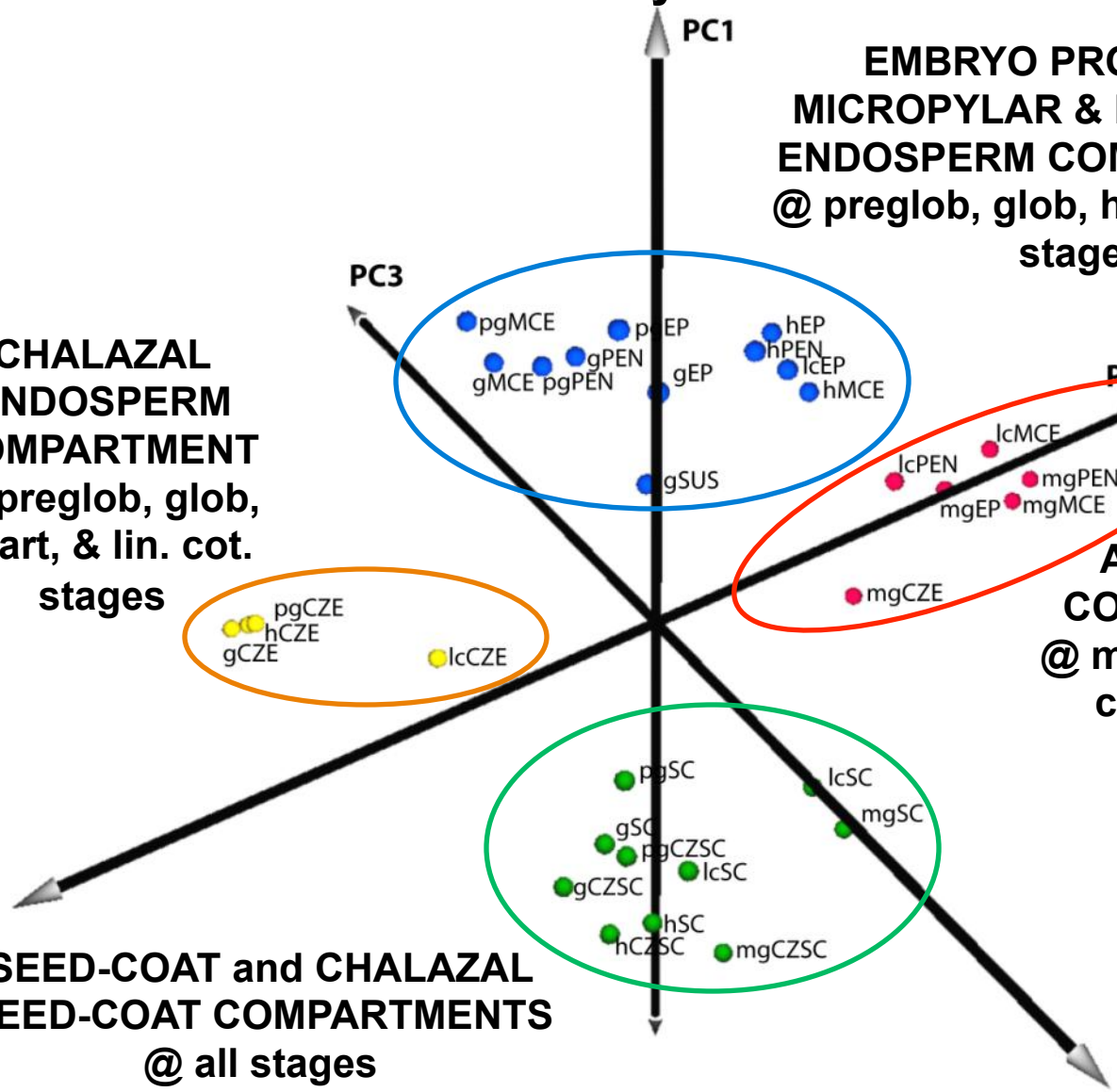


Global Analyses of Seed Compartment RNA Populations

Hierarchical Clustering



Principal Component Analysis



EMBRYO PROPER and MICROPYLAR & PERIPHERAL ENDOSPERM COMPARTMENTS @ preglob, glob, heart, & lin cot-stages

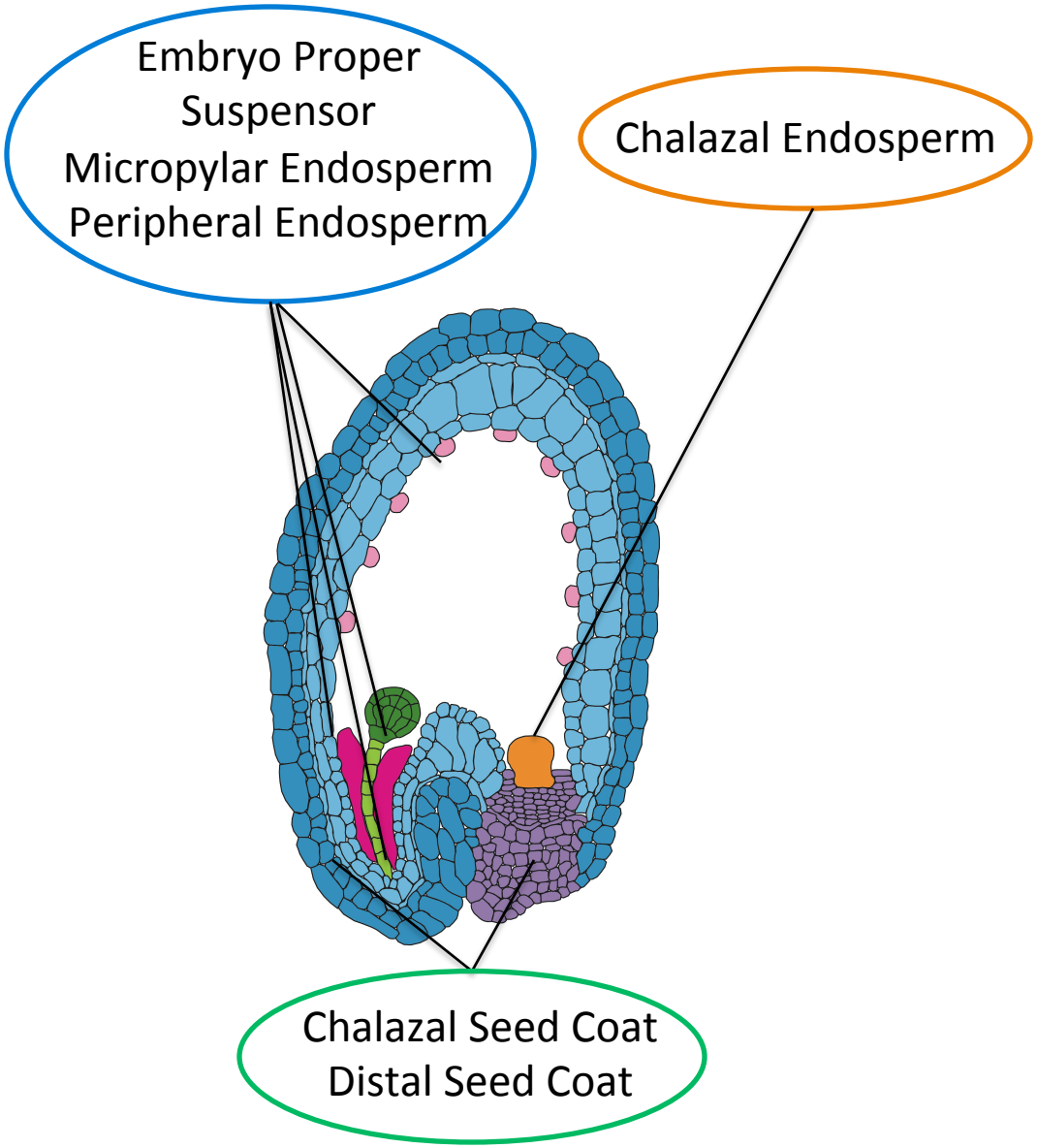
CHALAZAL ENDOSPERM COMPARTMENT @ preglob, glob, heart, & lin. cot. stages

ALL ZYGOTIC COMPARTMENTS @ maturation phase compartments

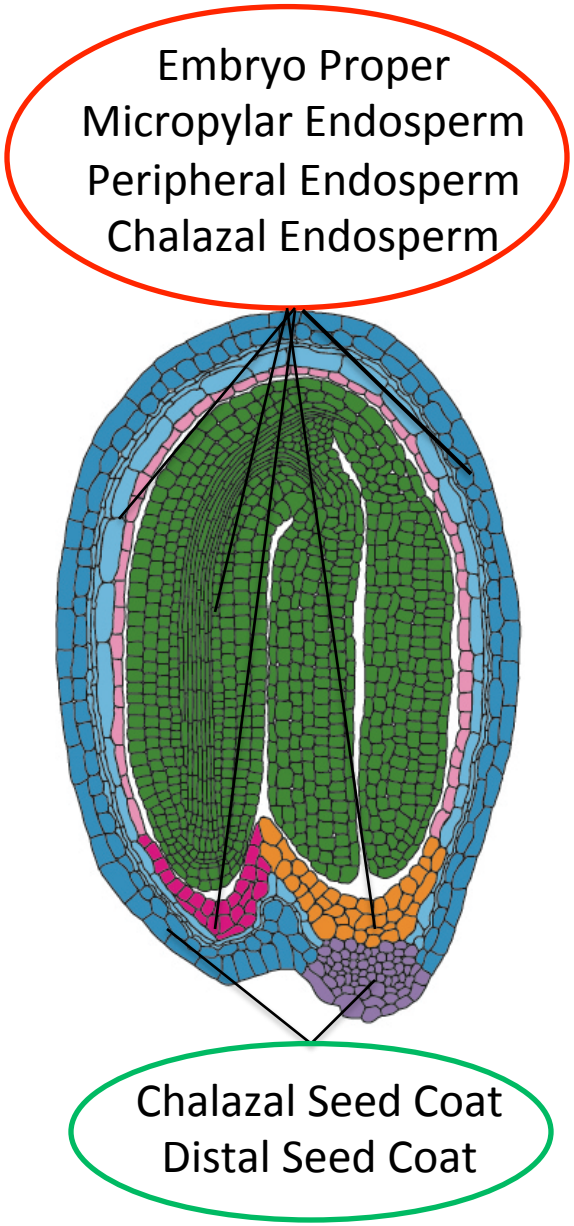
SEED-COAT and CHALAZAL SEED-COAT COMPARTMENTS @ all stages

Global Analyses of Seed Compartment RNA Populations

Early Seed Development

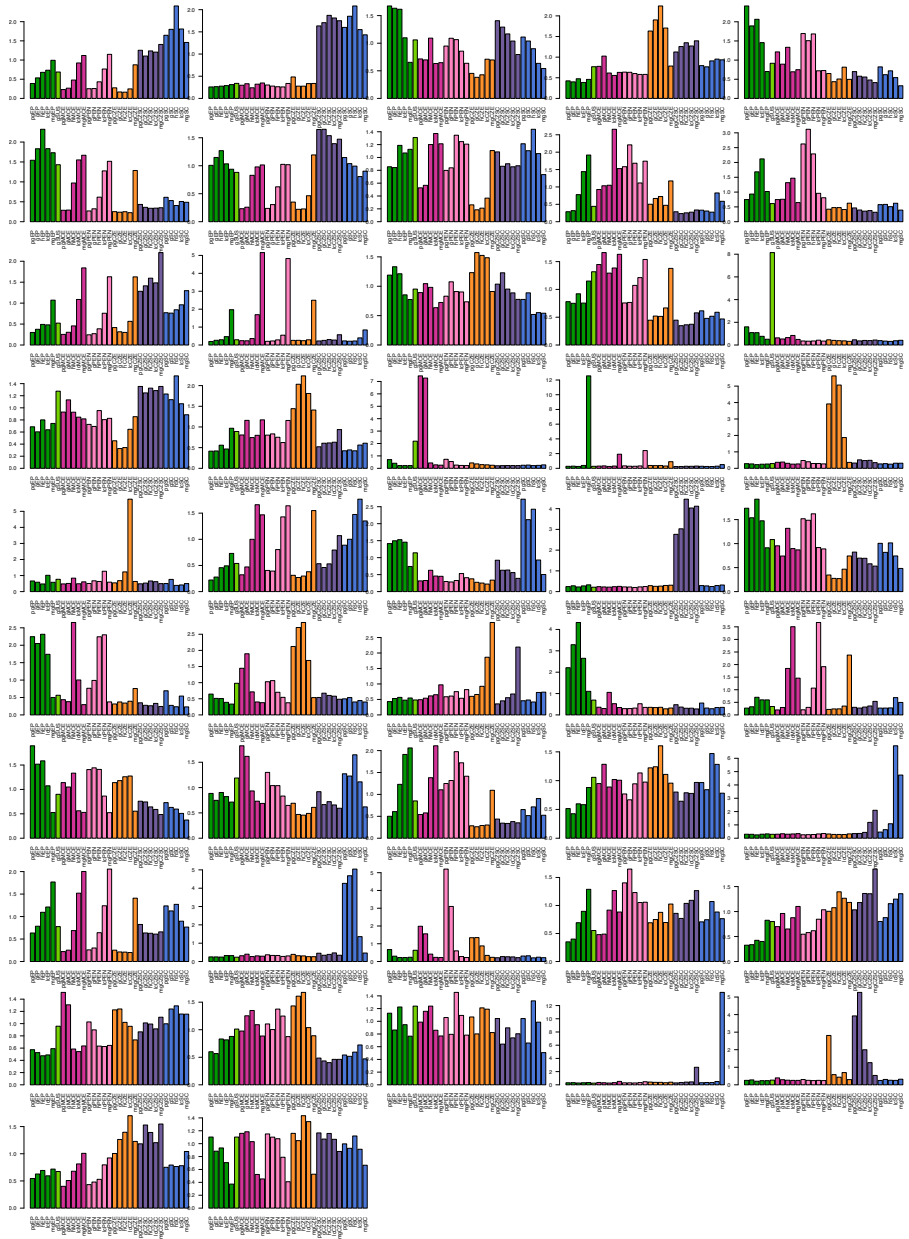


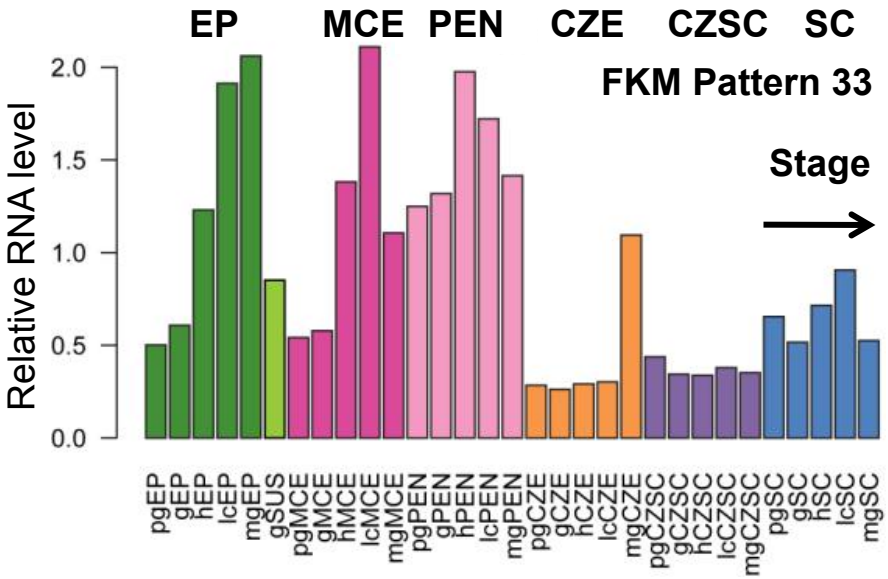
Late Seed Development

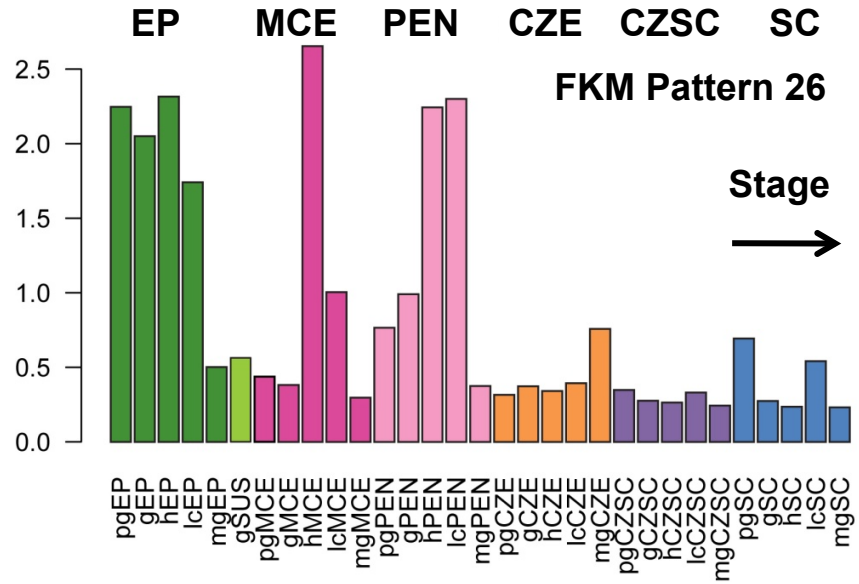


Identification of Coexpressed Gene Sets

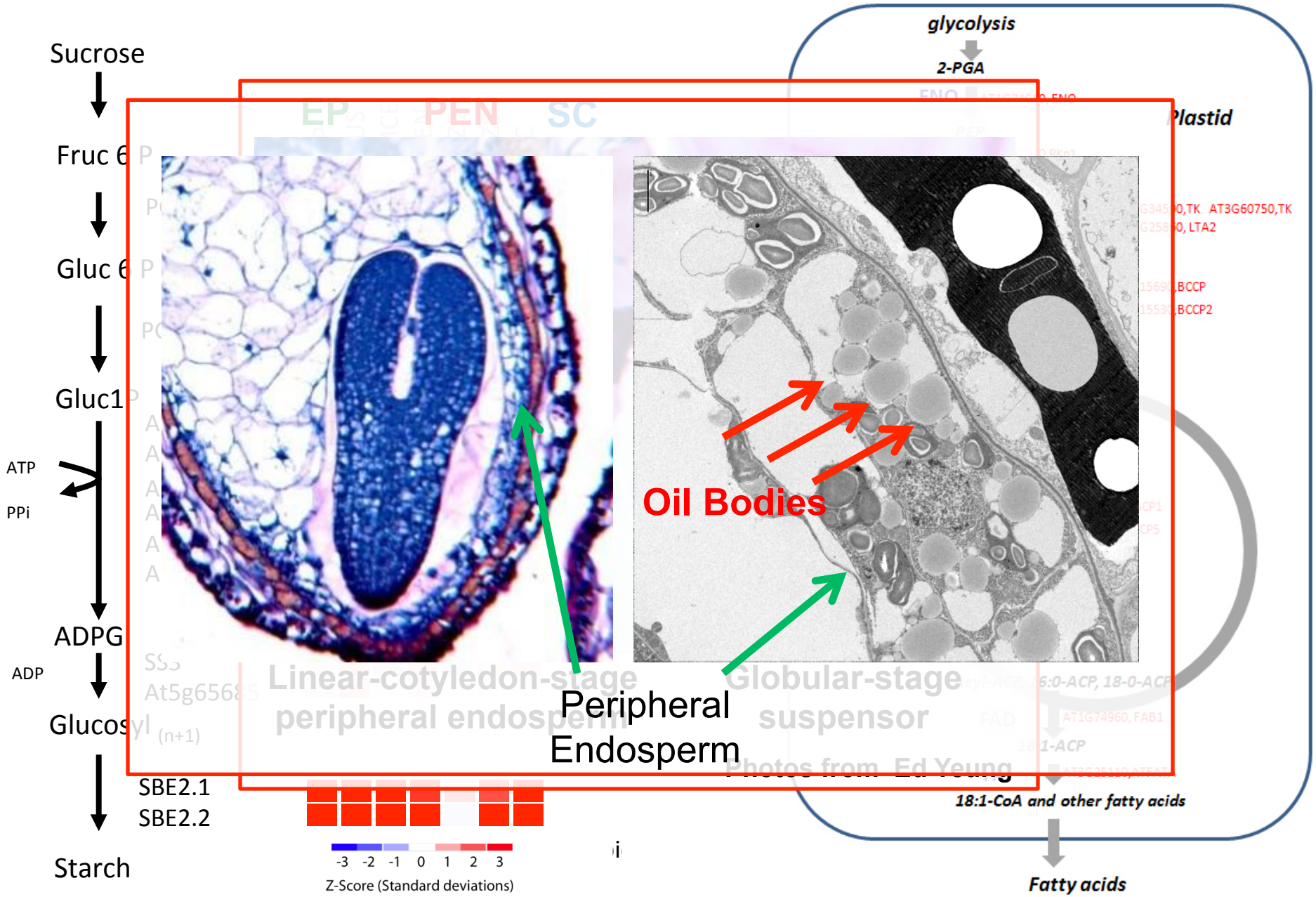
Fuzzy K-means Cluster Analysis (temporal & spatial)



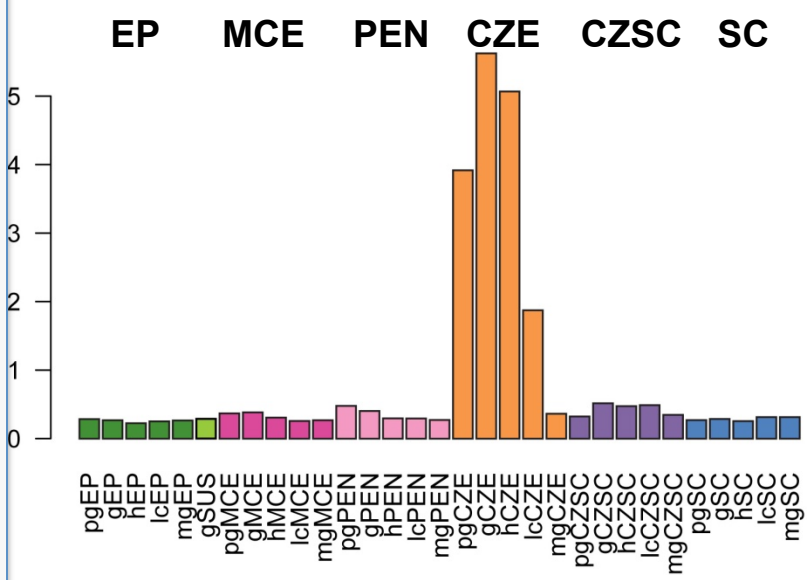


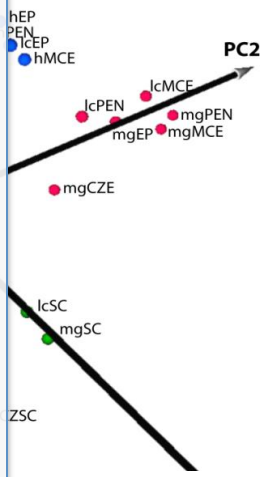


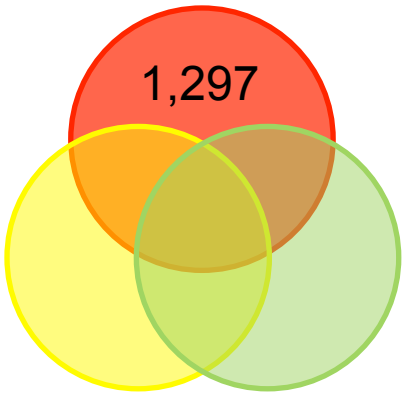
Carbon Metabolism in the Embryo and Endosperm



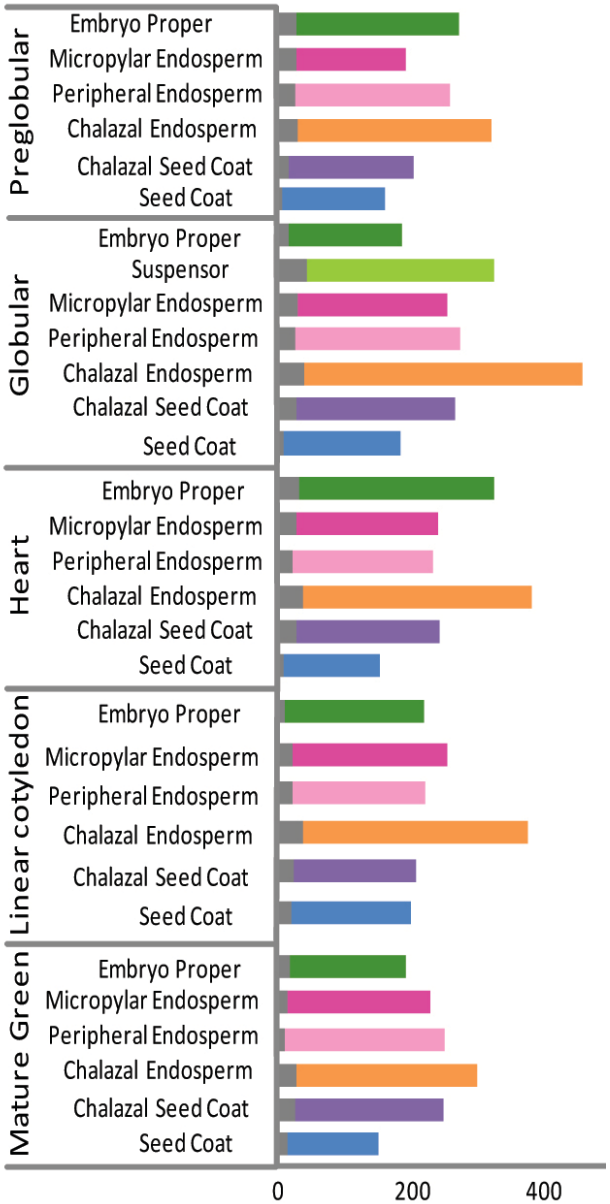
Chalazal Endosperm



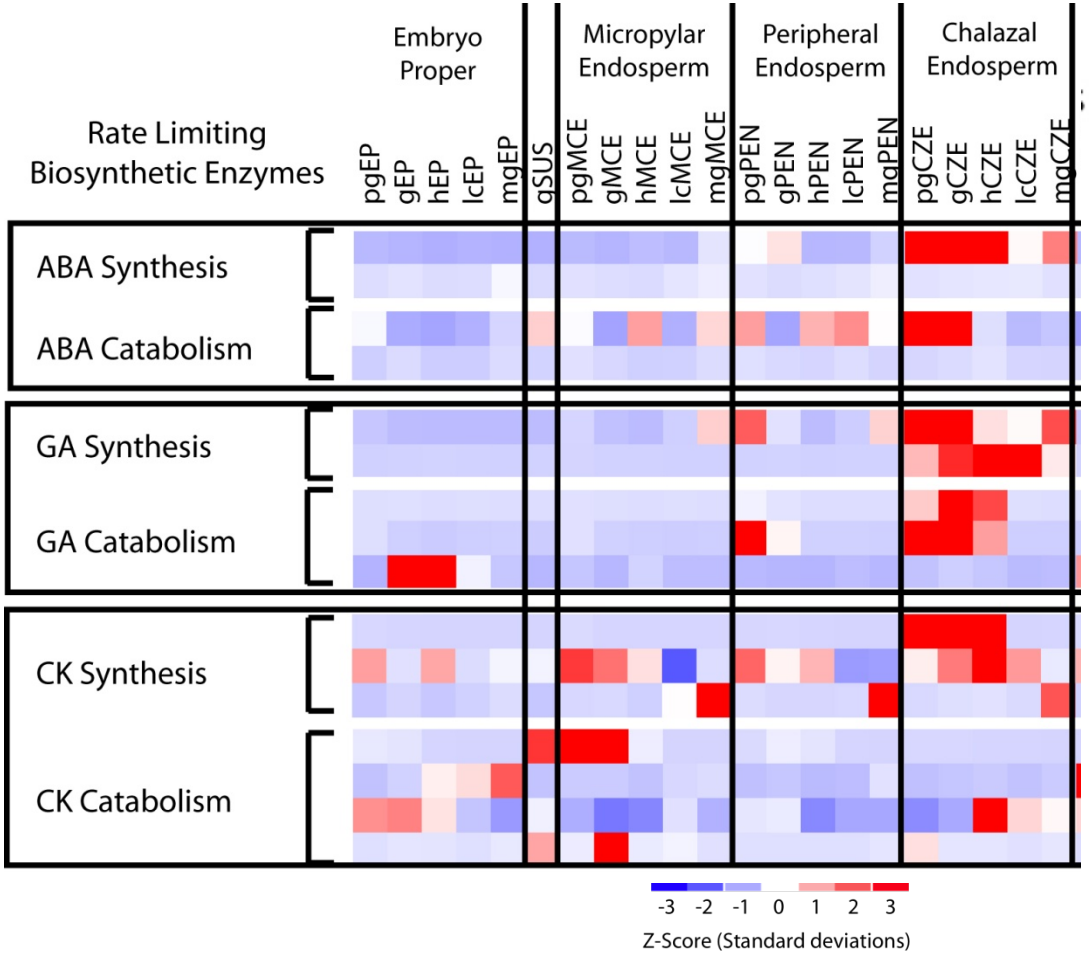


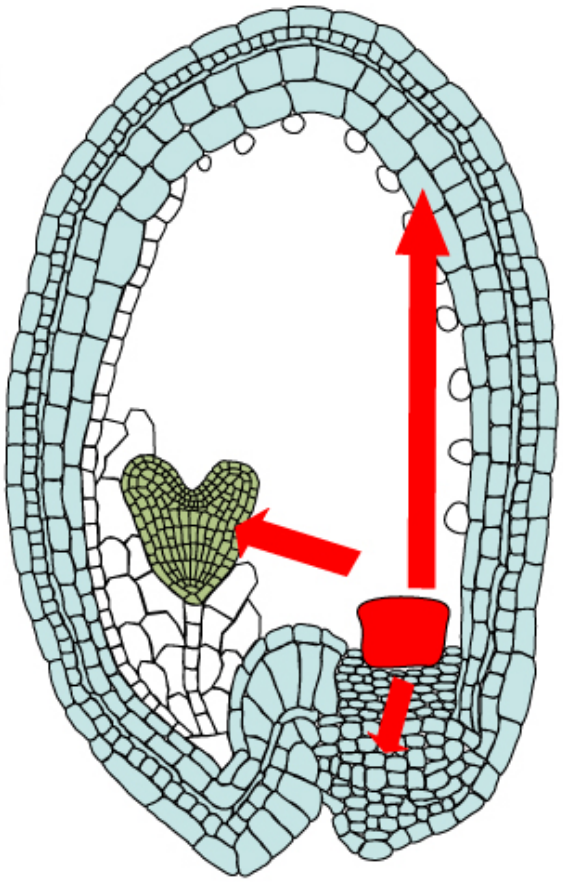


Seed-specific RNAs



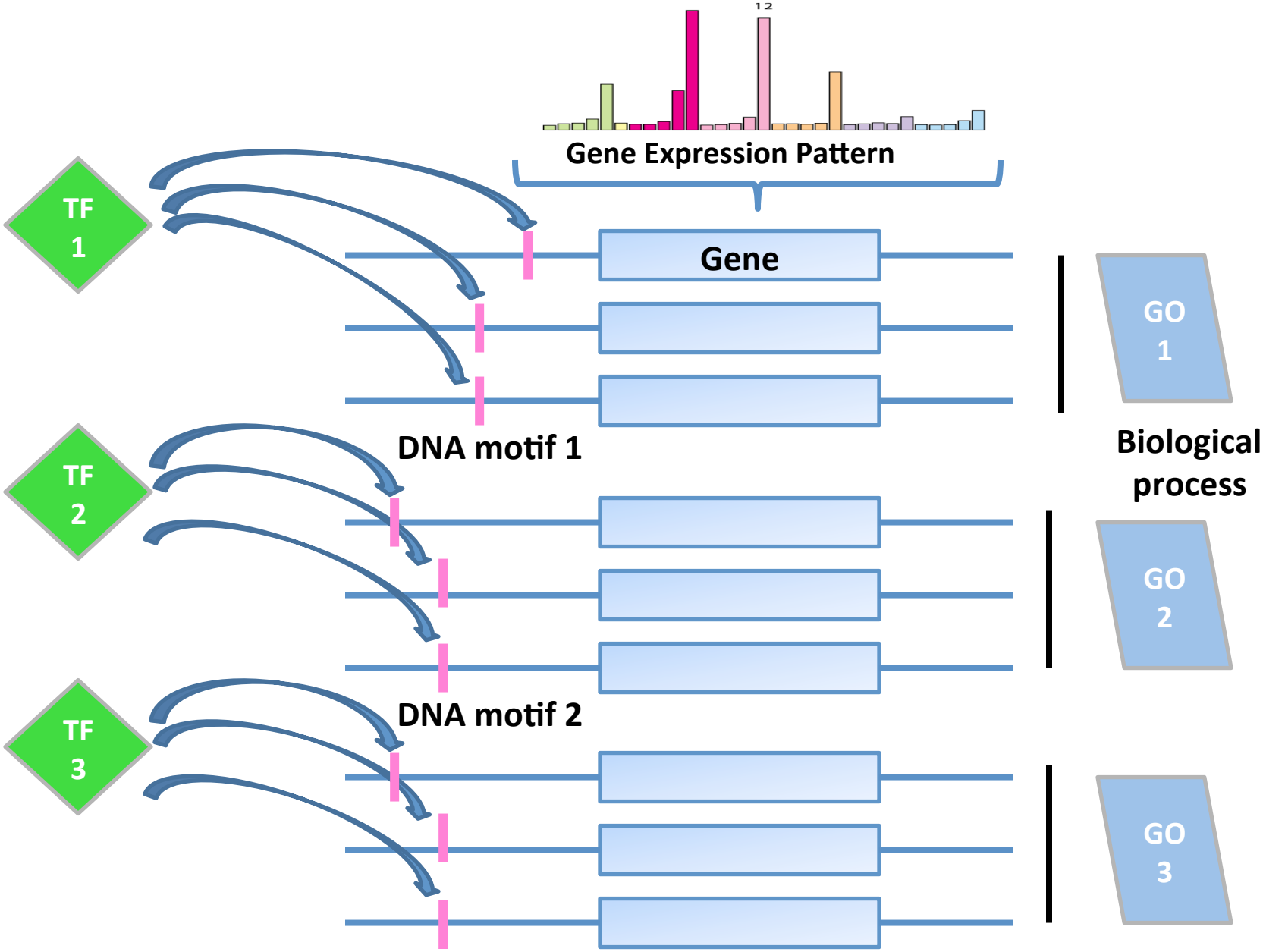
RNAs Encoding Hormone Biosynthetic Enzymes Accumulate Primarily in the Chalazal Endosperm





- 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

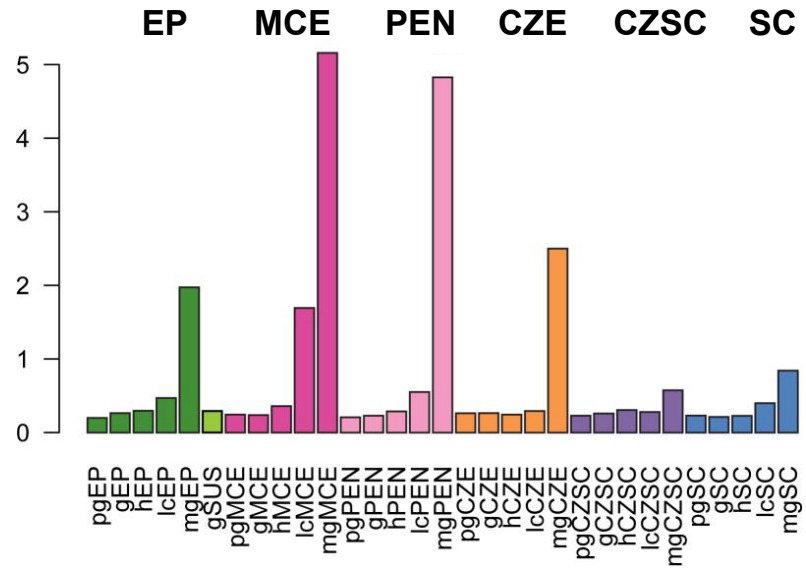
Transcriptional Networks Controlling Coexpressed Gene Sets

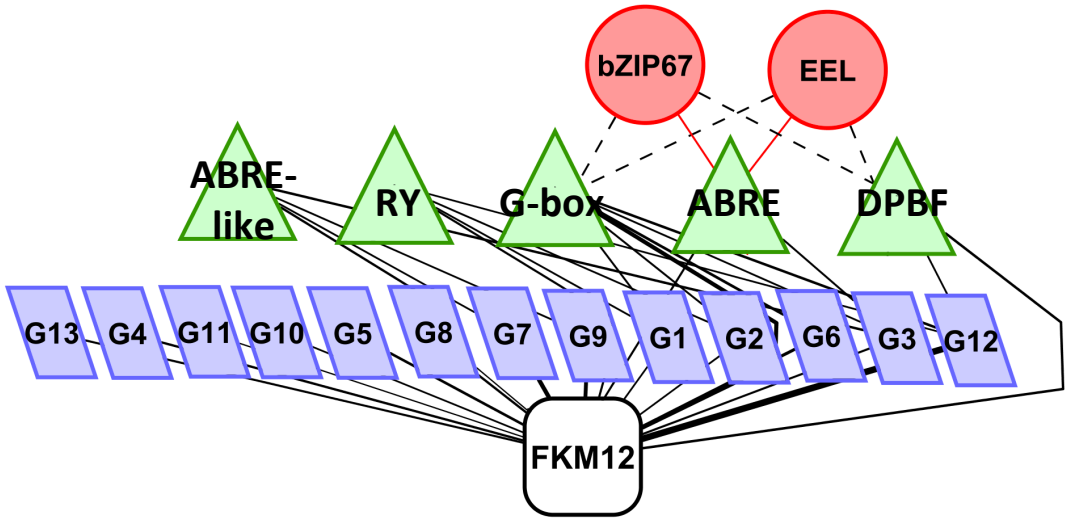


ChipEnrich (with Siobhan Brady)

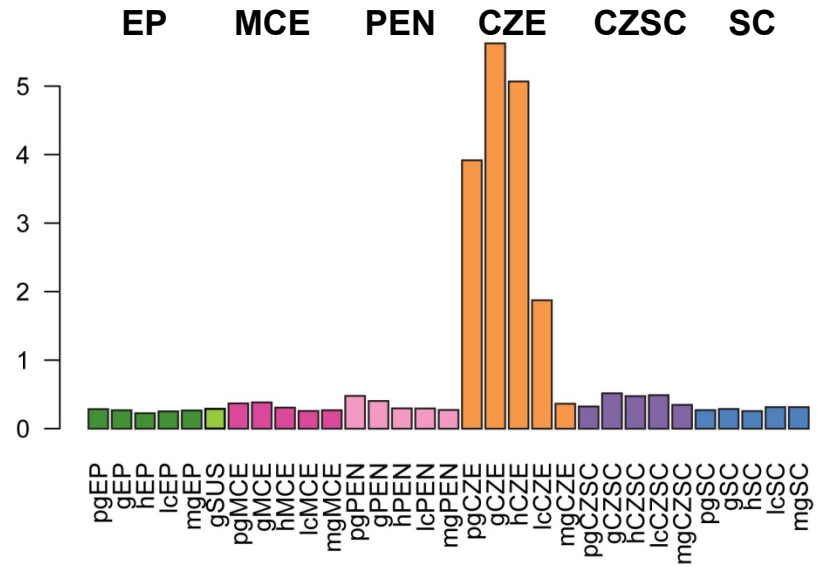
Coexpressed Gene Sets Define Transcriptional Modules

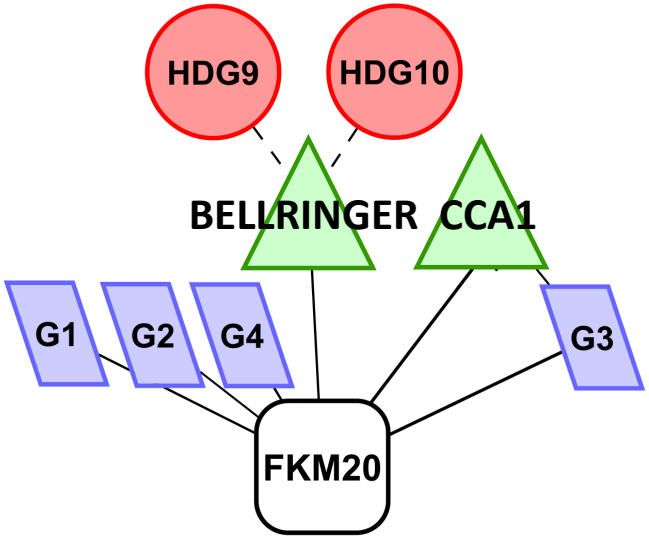
Maturation module





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

Anhthu Bui
Javier Wagmaister
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 Pelligrini

National Science Foundation
WHERE DISCOVERIES BEGIN

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