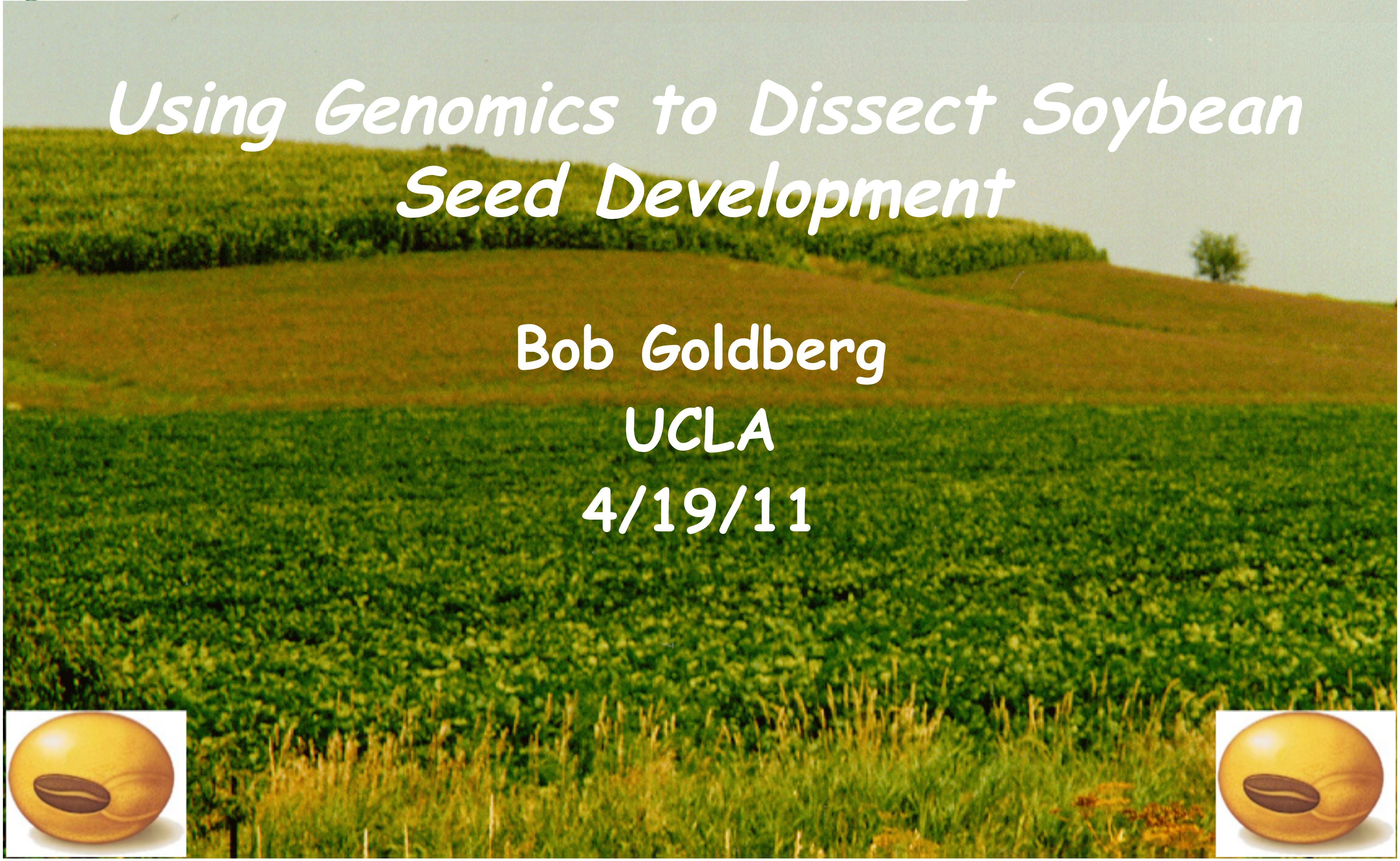
The School of Plant Sciences

The University of Arizona



Using Genomics to Dissect Soybean Seed Development

Bob Goldberg

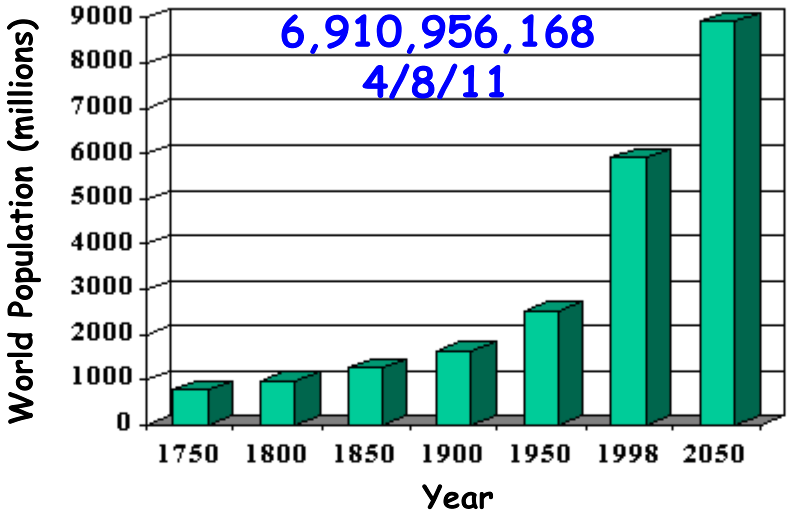
UCLA

4/19/11

“The Bravest are surely *those who have the clearest vision of what is before them, glory and danger alike, and yet notwithstanding go out to meet it”*

Thucydides 400 BC

We Face Major Challenges In Agriculture



OVER THE NEXT 40 YEARS WE WILL NEED TO ~ DOUBLE THE WORLD'S FOOD SUPPLY IN ORDER TO PRODUCE MORE FOOD THAN IN ALL OF HUMAN HISTORY (FAO, October 2009)

AND DO IT ON LESS ARABLE LAND!!!!

CROP YIELDS NEED TO BE INCREASED SIGNIFICANTLY

And…….If We Are Going To Use Plants For Energy Production Crop Yields Will Need To Increase In Order To Grow More On Less Land………

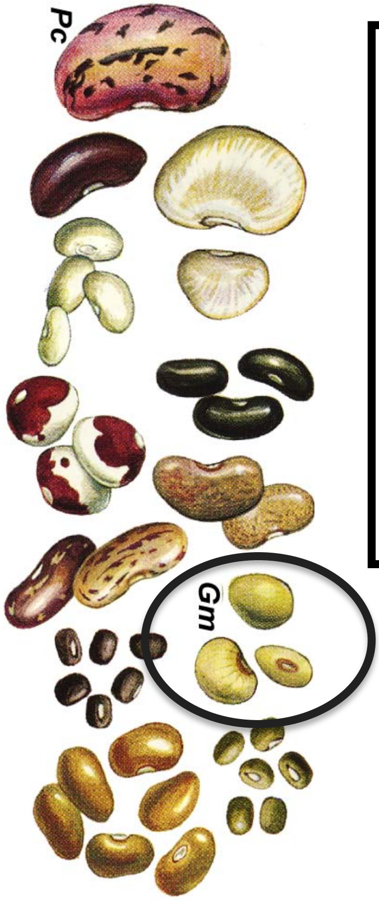
April 6, 2011

Rush to Use Crops as Fuel Raises Food Prices and Hunger Fears

…By Using a Variety of Approaches to Identify Genes and Processes That Will Help Increase Crop Yields and Food Production Significantly in the 21st Century….

Yield (Developmental Traits)

- Seed Number
- Seed Size
- Growth Rate
- Organ Size (More Seeds)
- Plant Architecture
- Flowering Time
- Senescence
- Maturity
- Stature



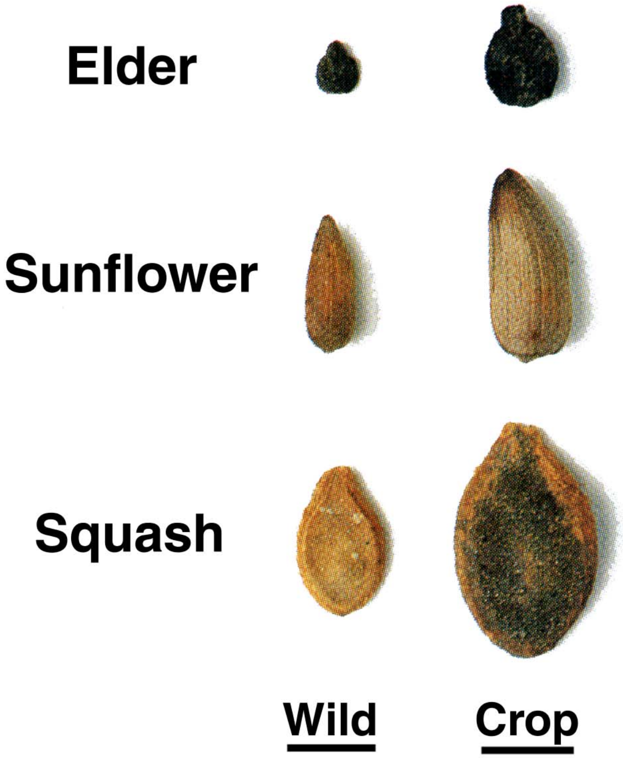
Yield (Stress Traits)

- Nutrient Uptake
- Drought Resistance
- Heat Resistance
- Cold Tolerance
- Salt Tolerance
- Shade Tolerance
- Disease Resistance

…….And by Using Breeding and Genetic Engineering to Introduce These "Yield" Genes Into Crops (One thing we can be sure of-we can't predict what new technology will be the driver 10-25 years out!)

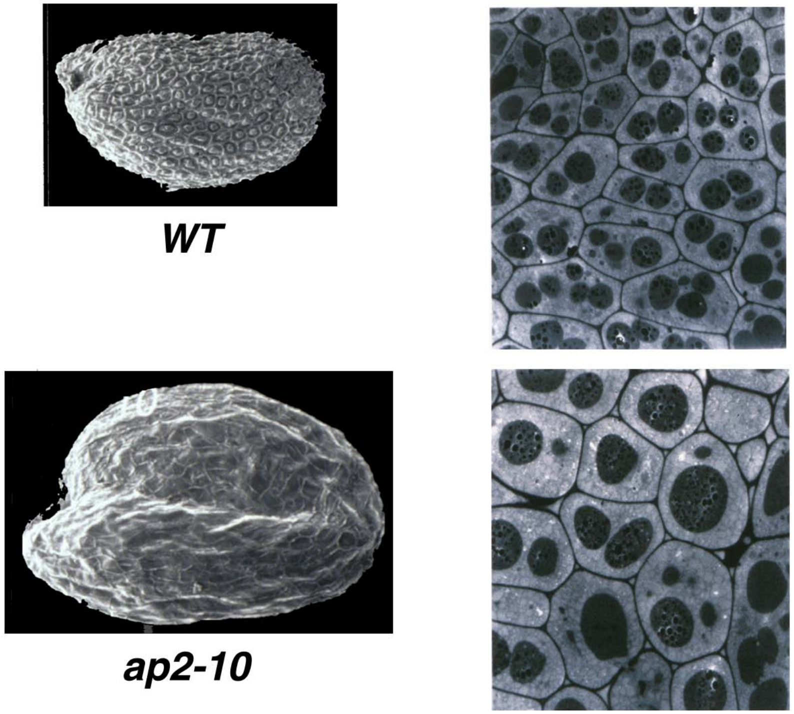
"Manipulating" Plants to Increase Seed Yield Is Not New…………………. Seed Size!

Engineering Bigger Seeds 10,000 Years Ago



Our American Ancestors, 10,000 BC

Engineering Bigger Seeds Today



But Need to Identify the Critical Genes

Jofuku et al., PNAS, 2005

Scientists ALWAYS overestimate how much can be accomplished in a short period of time (1 month to a year) ……… ………….but underestimate how much will be accomplished over the LONG TERM (5-10 years)……..

One thing we can be sure of is - we can't predict what will be the driver of new agriculturally important breakthroughs 15-25 years in the future!

1900: Rediscovery of Mendel's Work

DeVries, Correns and Tschermak independently rediscover Mendel's work.

Three botanists - Hugo DeVries, Carl Correns and Erich von Tschermak - independently rediscovered Mendel's work in the same year, a generation after Mendel published his papers. They helped expand awareness of the Mendelian laws of inheritance in the scientific world.

The three Europeans, unknown to each other, were working on different plant hybrids when they each worked out the laws of inheritance. When they reviewed the literature before publishing their own results, they were startled to find Mendel's old papers spelling out those laws in detail. Each man announced Mendel's discoveries and his own work as confirmation of them.

1909: The Word Gene Coined

Danish botanist Wilhelm Johannsen coined the word gene to describe the Mendelian units of heredity.

He also made the distinction between the outward appearance of an individual (phenotype) and its genetic traits (genotype).

Four years earlier, William Bateson, an early geneticist and a proponent of Mendel's ideas, had used the word *genetics* in a letter; he felt the need for a new term to describe the study of heredity and inherited variations. But the term didn't start spreading until Wilhelm Johannsen suggested that the Mendelian factors of inheritance be called *genes*.

The proposed word traced from the Greek word *genos*, meaning "birth". The word spawned others, like *genome*.

1911: Fruit Flies Illuminate the Chromosome Theory

Using fruit flies as a model organism, Thomas Hunt Morgan and his group at Columbia University showed that genes, strung on chromosomes, are the units of heredity.

Morgan and his students made many important contributions to genetics. His students, who included such important geneticists as Alfred Sturtevant, Hermann Muller and Calvin Bridges, studied the fruit fly *Drosophila melanogaster*. They showed that chromosomes carry genes, discovered genetic linkage - the fact that genes are arrayed on linear chromosomes - and described chromosome recombination.

In 1933, Morgan received the Nobel Prize in Physiology or Medicine for helping establish the chromosome theory of inheritance.

Early Fly drawing

Portrait of Morgan at Work

Courtesy of American Philosophical Society, Curt Stern Papers. Noncommercial, educational use only.

Photo of Morgan's fly room at Columbia University

2000: *Drosophila* and *Arabidopsis* genomes sequenced

Drosophila melanogaster (fruit fly)has been a primary tool for geneticists since the early part of the twentieth century. The sequencing of its genome is the result of a collaborative effort between the *Drosophila* Genome Project Group, led by Gerald Fink at the University of California, Berkeley and researchers from Celera Genomics Corporation led by Craig Venter. The *Drosophila* genome is estimated to have approximately 13,600 genes as compared to 20,000—25,000 genes in humans. The popularity of *Drosophila* as an experimental organism ensures that its genome sequence will be a valuable resource for research in genetics and medicine. Many genes of *Drosophila* have been conserved through evolution and have human counterparts. This means that scientists can perform experiments using flies and apply their findings to human biology.

Arabidopsis thaliana is the first plant to have its genome sequenced. This plant from the mustard family has become the plant biologists' equivalent of the laboratory mouse. Its genome was completed by the collective efforts of an international group of researchers called the *Arabidopsis* Genome Initiative. The *Arabidopsis* genome has an estimated 25,000 genes—apparently even more than humans. Although not a crop plant, *Arabidopsis* was chosen as a model organism because its genome is small and it has relatively little of the noncoding, so-called junk, DNA. It does, however, share very similar biochemistry to crop plants such as rice or barely. The study of its sequence is expected to have widespread applications for agriculture and medicine.

2004: Refined Analysis of Complete Human Genome Sequence

The International Human Gene Sequencing Consortium led in the United States by the National Human Genome Research Institute and the Department of Energy published a description of the finished human gene sequence. The analysis reduced the estimated number of genes (which as recently as the mid-1990's had been ~100,000) from 35,000 to only 20,000-25,000. The fact that the human genome has far fewer genes than was originally thought suggests that humans "get more" out of their genetic information than do other animals. For example, the average human gene is able to produce three different gene products.

The finished sequence contains 2.85 billion nucleotides interrupted by only 341 gaps. It covers 99 percent of the genome with an accuracy of 1 error per 100,000 bases. Researchers confirmed the existence of 19,599 protein-coding genes and identified 2,188 other DNA segments that are thought to be protein-coding genes. Although the genome sequence is described as "finished," it isn't perfect. The small gaps that remain cannot be sequenced by the industrial-scale methods used by the Human Genome Project. Filling in these gaps will have to await a series of small targeted efforts by researchers using other techniques and possibly new technologies. The finished genome sequence can be freely accessed through public databases and may be used by researchers without restrictions.

2008: NextGen Sequencing

2025??

Why Seeds and a Reminder of Why They are Important!

Seeds Are Used in Many Ways as Food, Beverages, Spices, and Fuels!

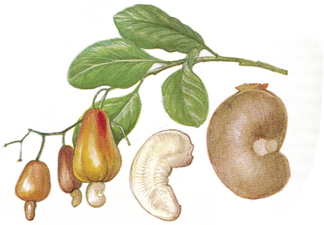
Beans

Peas

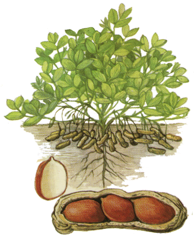
Wheat

Corn

Coconut



Cashew Nuts



Peanuts

Pecans

Cocoa Beans

Coffee Beans

Nutmeg

Mustard

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

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

Most Importantly….. 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

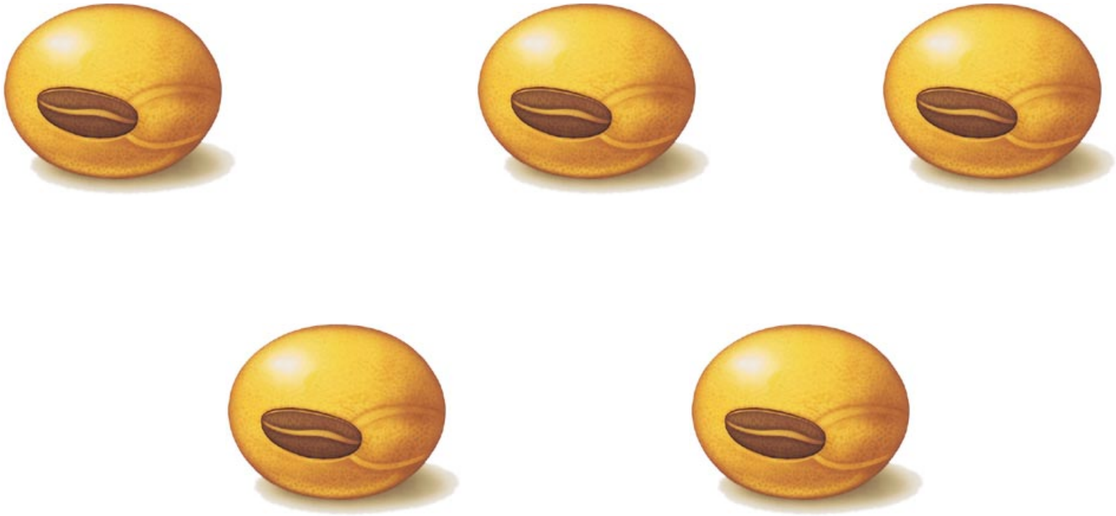
Non-Seed Crops

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

Soy Oil is ~11% of World's Oil/Fat Caloric Intake & Soy Cake is ~61% of World's Meal/Cake Production! (Fao Report, 2004)

Seeds are Important!!!!

How Is a Seed Formed?



A Short Reminder……………………………………

In the Beginning....

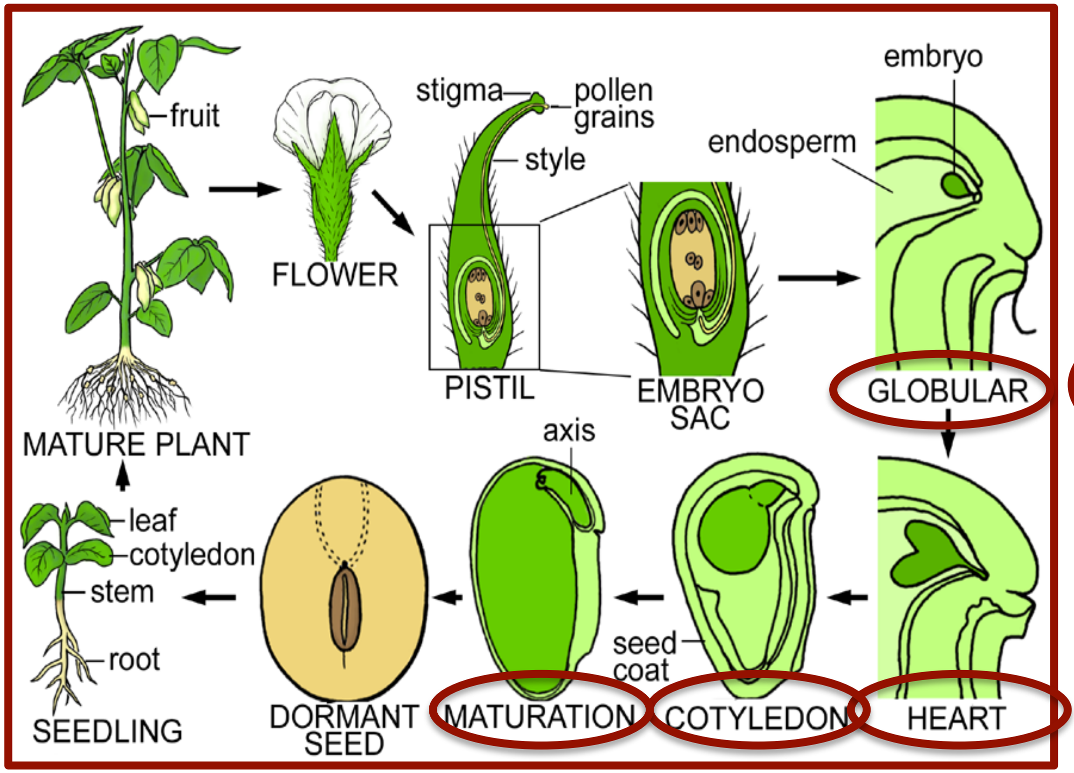
Why Soybean?-A Reminder

- 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
- Major Funding Source

Diversity of Oil Seed Plants

What Are the Questions Focused On In This Talk?

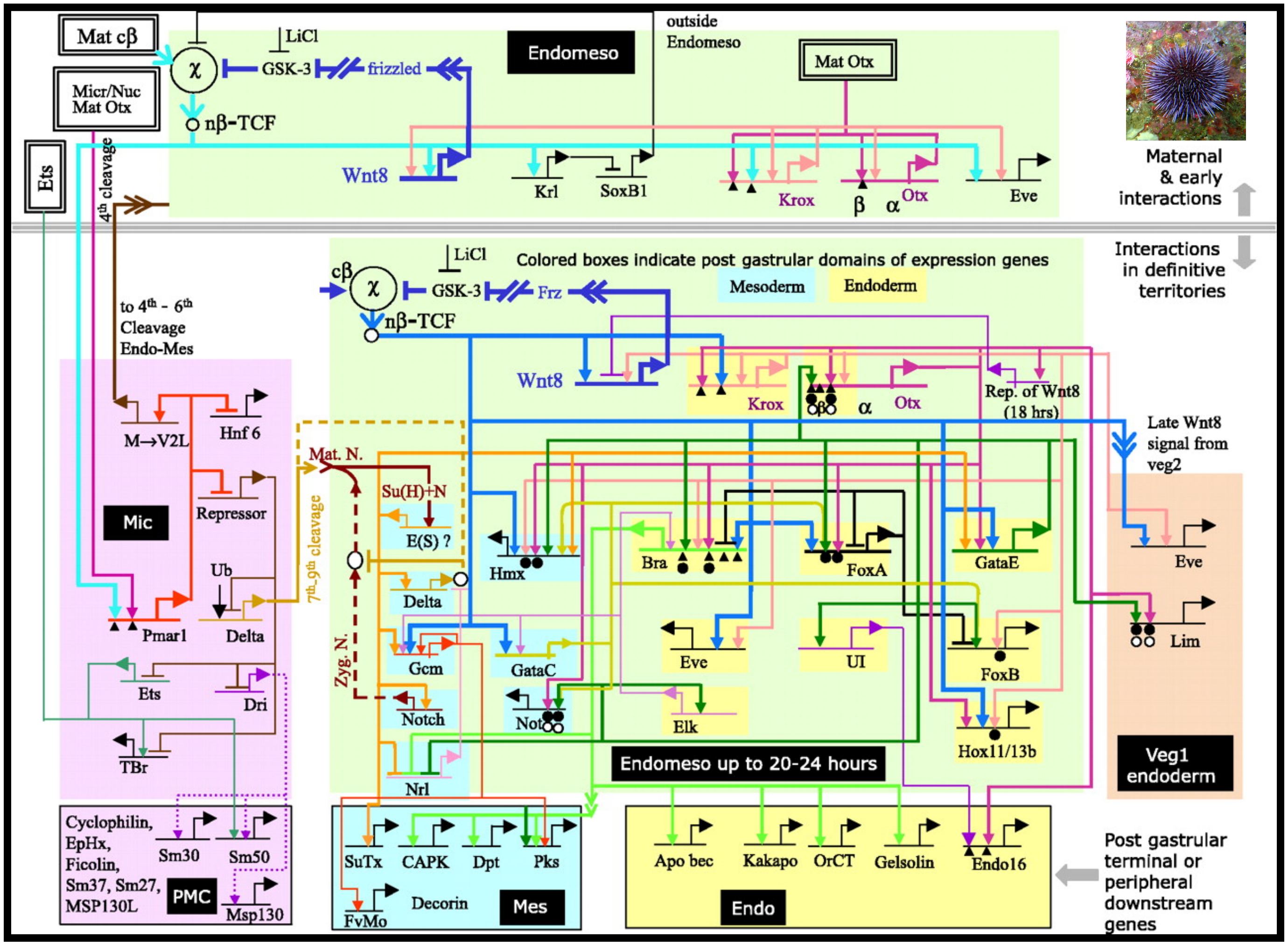
Soybean Seed Development



Seed Stages Studied

- What Is The Spectrum Of Genes And Regulators That Are Active In Specific Seed Compartments Throughout Development?
- How Does Gene Activity Change During Seed Development?
- What Biological Processes Are Specific For Different Seed Compartments/Development?
- What Are The Genes And Epigenetic Processes Required To Make A Seed?

Ultimate Goal……To Uncover Regulatory Genes and Circuits Driving Seed Differentiation and Development Using Genomics



Learn How To Make a Seed!!

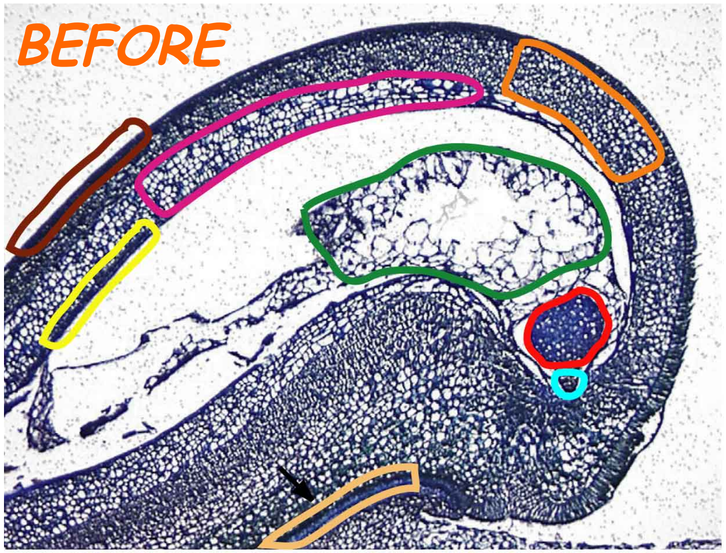
- Big Seeds
- More Seeds
- More Yield
- Increased Food and Fuel

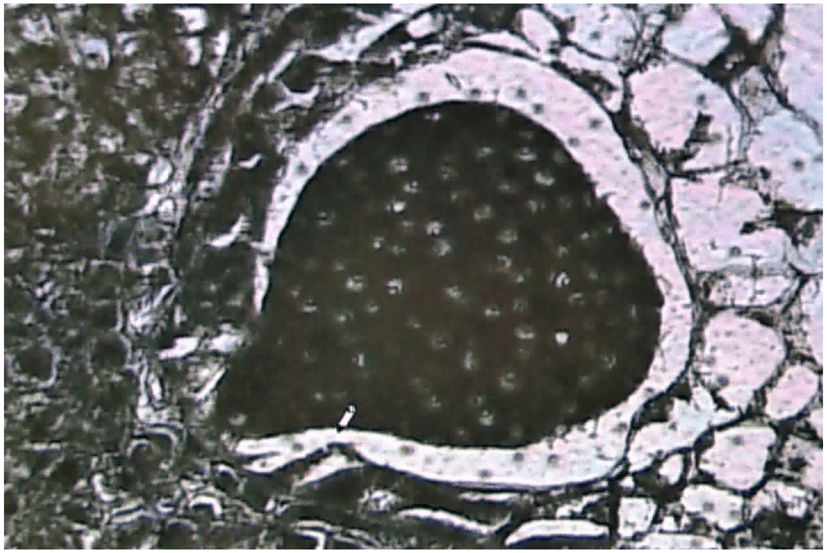
Eric Davidson et al. Science, 2007

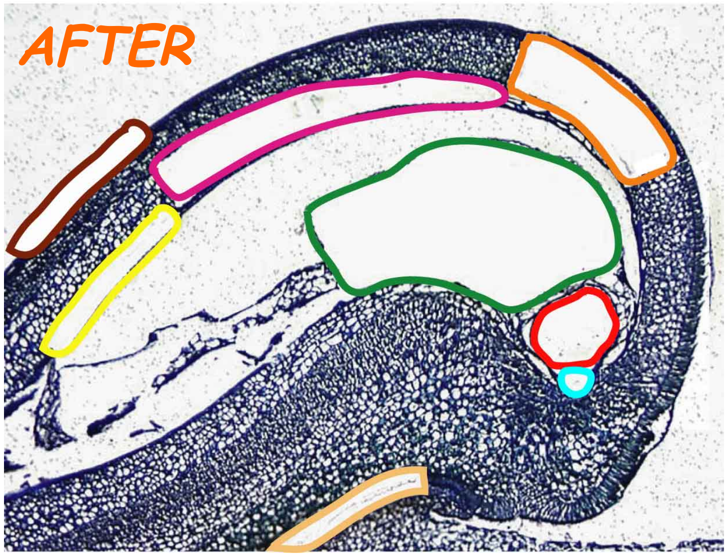
Knowledge of Cell-Specific TF mRNAs and Knock-Down Effects On Embryo Phenotype and TF mRNAs

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

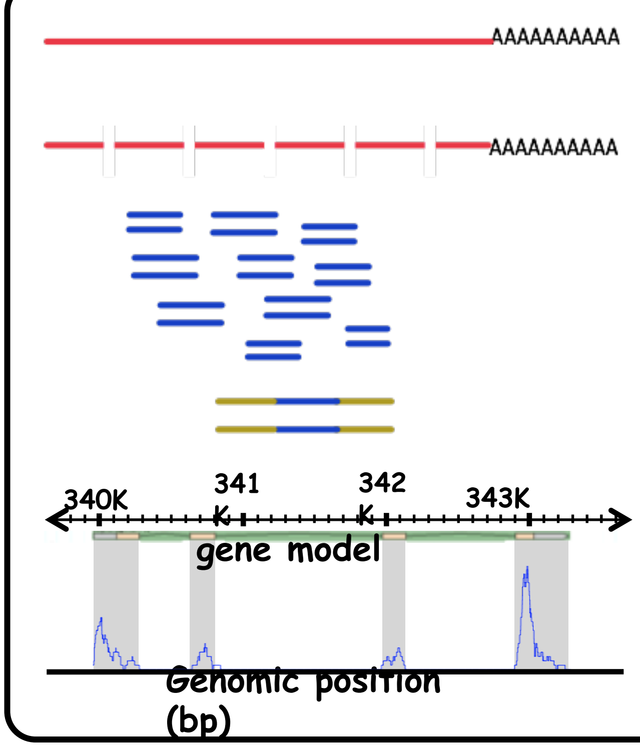
cDNA-Based Array
37,000 Probe Sets
>25,000 Gene Models

2nd Generation GeneChip*

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

SEQUENCING

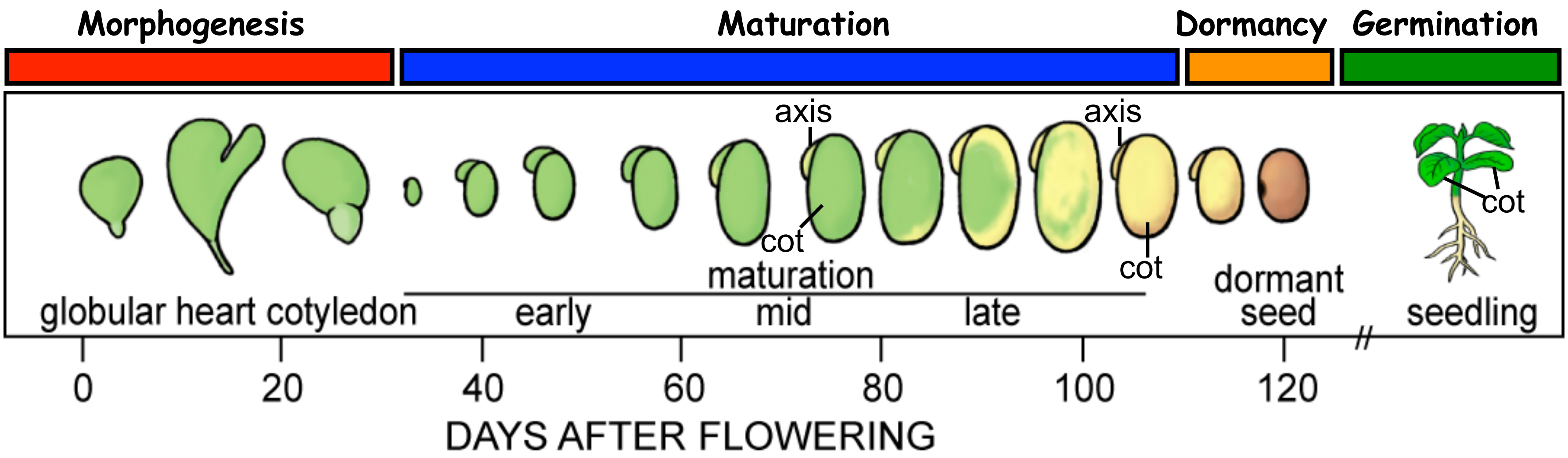
RNASeq



1. mRNA
2. RNA Fragmentation
3. Double-stranded cDNAs
4. Adapter Ligation
5. Sequencing and analysis

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

Soybean Seed Development-A Very Short Overview

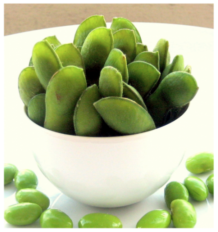


Differentiation of Compartments, Tissues, and Regions

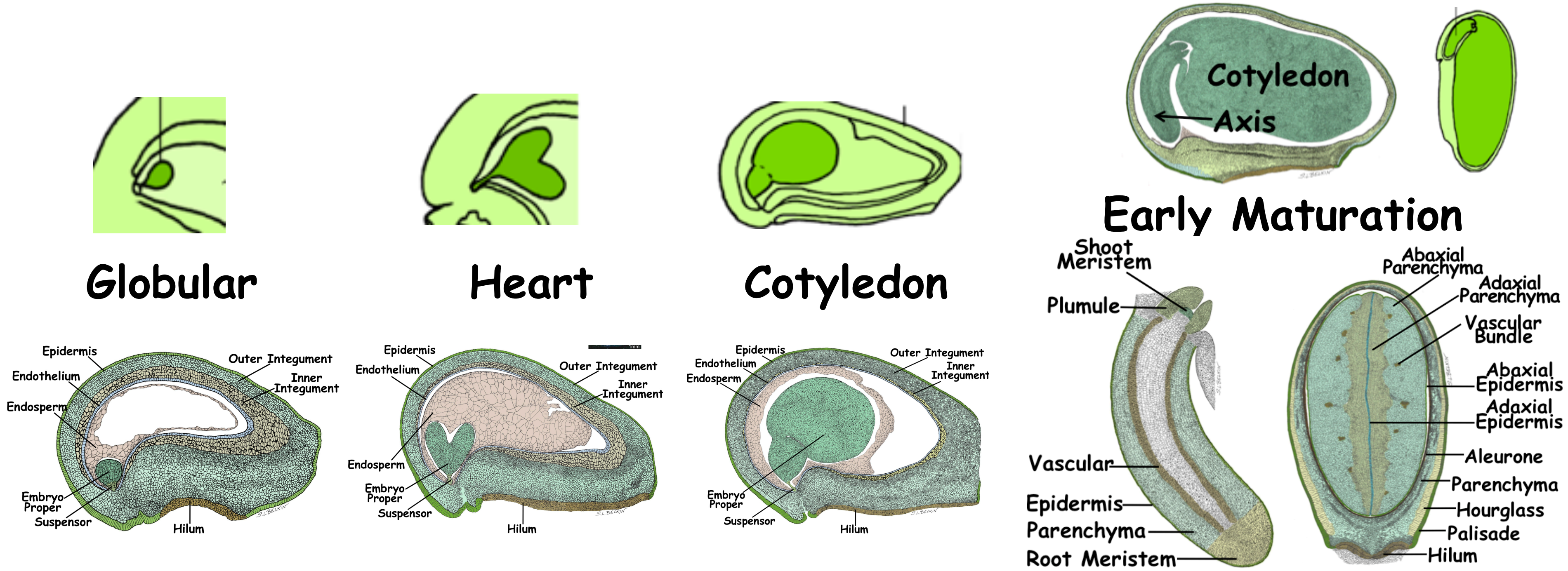
Storage Protein Deposition
Preparation or Dormancy

Mature Seed, Dormancy and Germination

Food Reserve Accumulation



What Developmental Stages and Seed Compartments Studied?



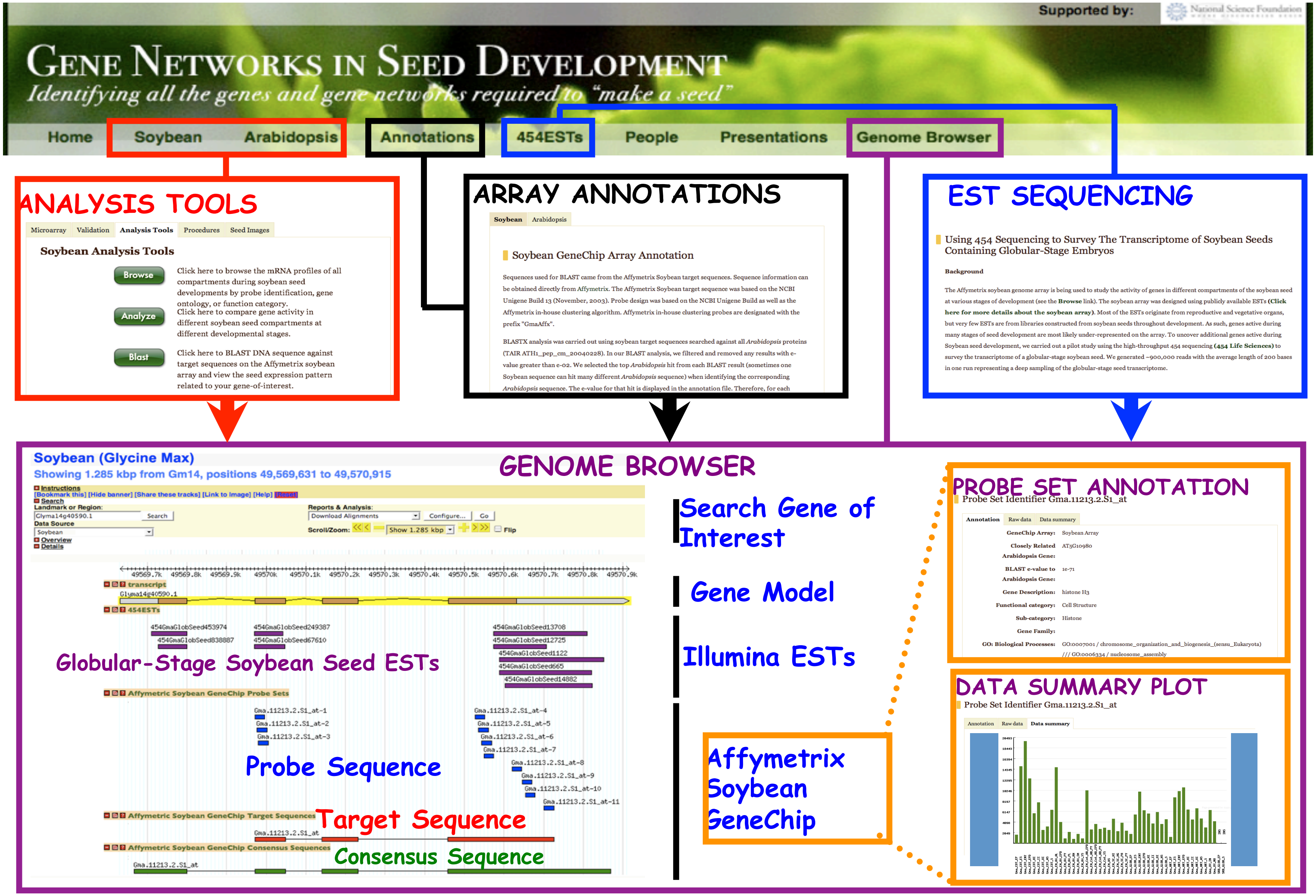
SEED STAGE	GLOB	HRT	COT	EM	TOTAL
# Compartments Studied	8	8	8	16	40
# GeneChip Experiments	26	21	18	34	99
# Genes Validated by qRT-PCR (TF)	42 (28TFs)	36 (20TFs)	55 (55TFs)	21 (21TFs)	154

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



Supported by: National Science Foundation

GENE NETWORKS IN SEED DEVELOPMENT

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

ANALYSIS TOOLS

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 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_cm_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* sequence) 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

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

GeneChip Array: Soybean Array
Closely Related Arabidopsis Gene: AT5G10980
BLAST e-value to Arabidopsis Gene: 1e-71
Gene Description: histone H3
Functional category: Cell Structure
Sub-category: Histone
Gene Family:
GO: Biological Processes: GO:0007001 / chromosome_organization_and_biogenesis_(sensu_Eukaryota) /// GO:0006334 / nucleosome_assembly

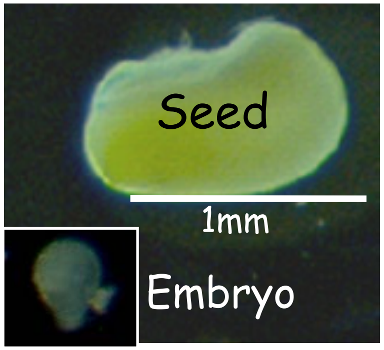
DATA SUMMARY PLOT

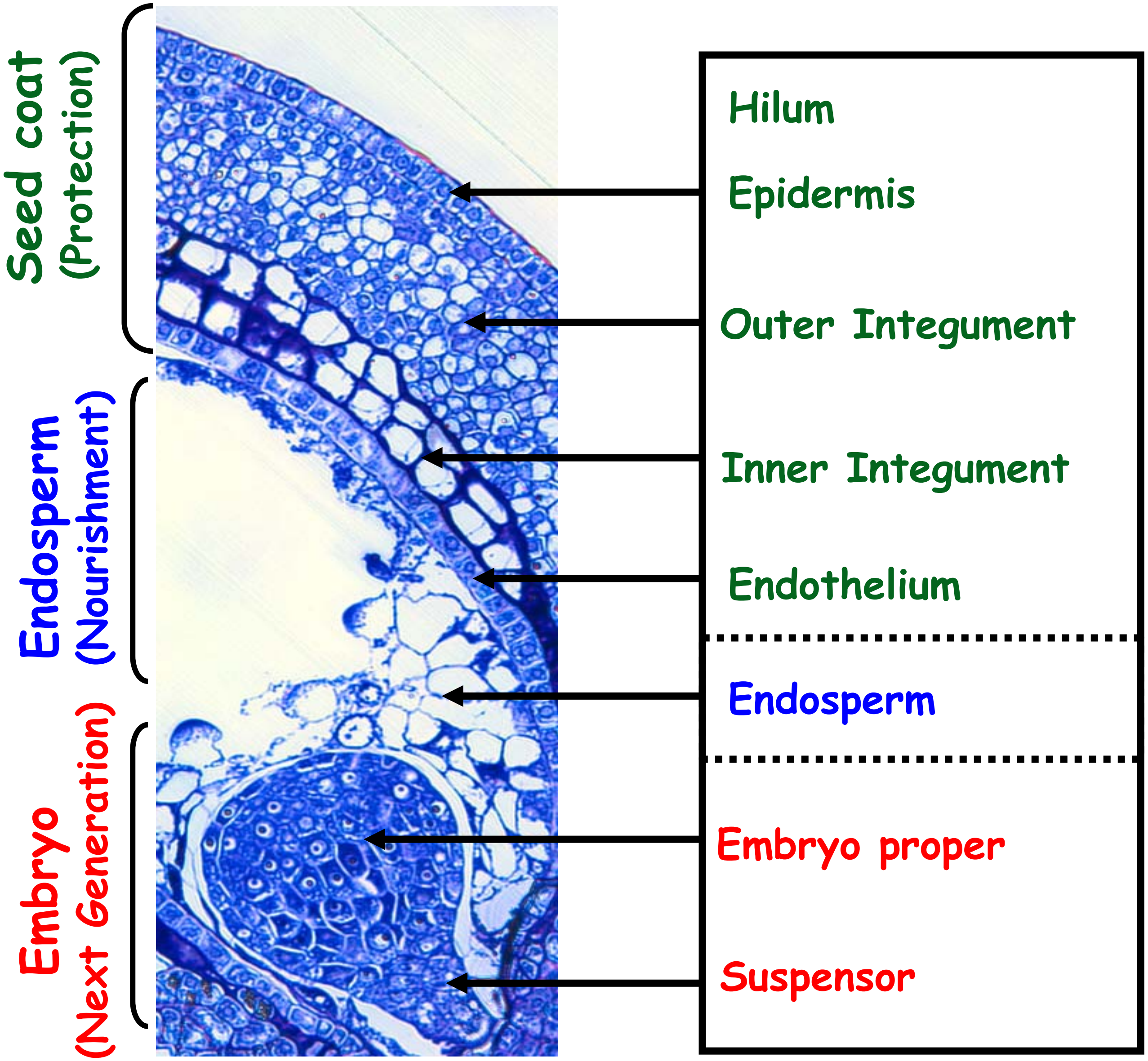
Probe Set Identifier Gma.11213.2.S1_at

What Are The Genes Required To Make a Seed?

An Example-The Globular Stage

Globular Stage Seed After Fertilization





Seed coat (Protection)

Endosperm (Nourishment)

Embryo (Next Generation)

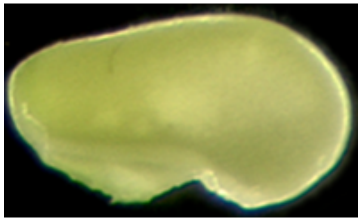
Hilum
Epidermis
Outer Integument
Inner Integument
Endothelium

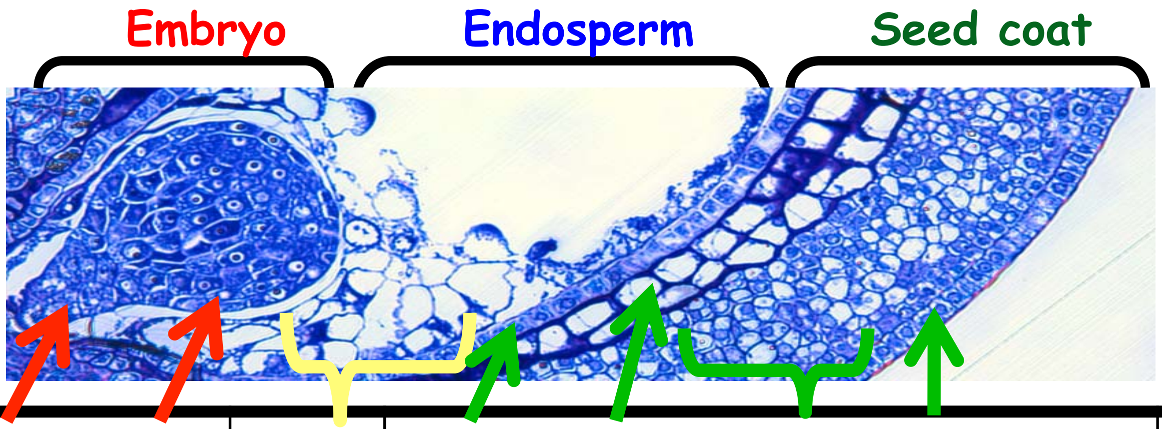
Endosperm

Embryo proper
Suspensor

How Is Gene Activity Regulated in Different Compartments of a Globular-Stage Seed?

Whole Mount



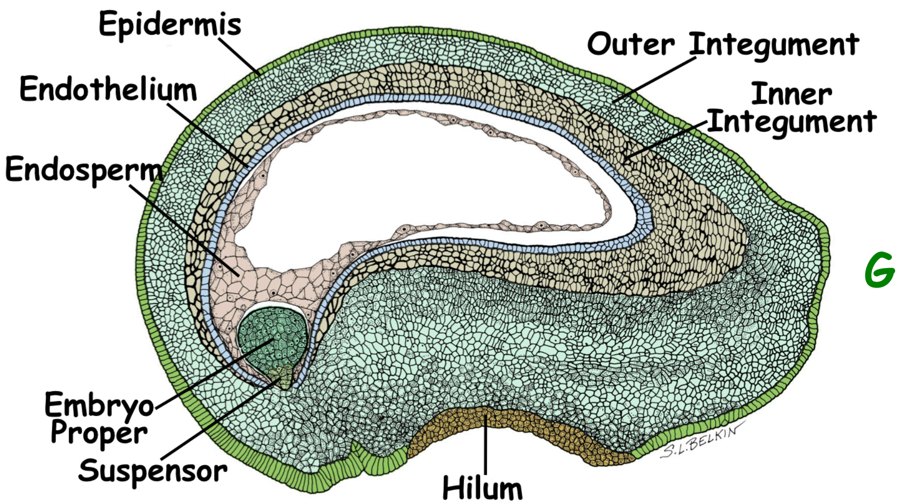


	WM	SUS	EP	ES	ENT	II	OI	EPD	HI	Union
Total mRNAs (TFs)	19,938 (1,373)	10,439 (642)	15,975 (1,040)	13,880 (872)	15,274 (976)	14,767 (952)	16,402 (1,102)	13,451 (865)	16,153 (1,086)	21,989 (1,558)
Unique mRNAs (55 TFs)		62 (2)	70 (14)	115 (13)	47 (4)	44 (5)	79 (7)	29 (2)	66 (8)	
Shared mRNAs (TFs)		← 7,664 (400) →								

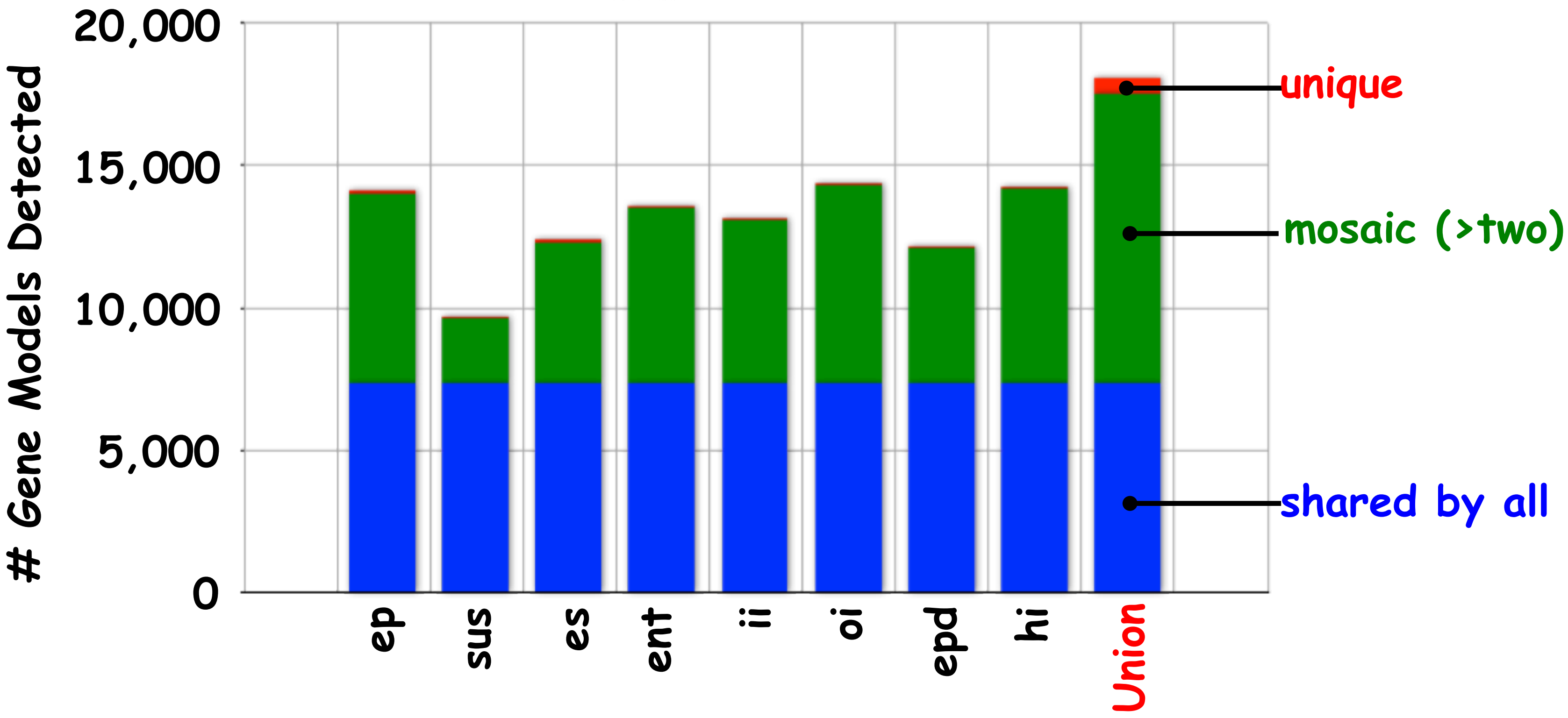
( ) indicates the number of transcription factor transcripts

- ~ Same Number of mRNAs in each compartment, region, & tissue
- Most mRNAs shared by all seed compartments, regions, & tissues
- There is a small number of compartment-specific transcripts, including transcription factor mRNAs

Most Seed mRNAs Are Shared By Different Compartments & Regions

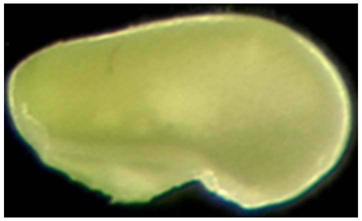


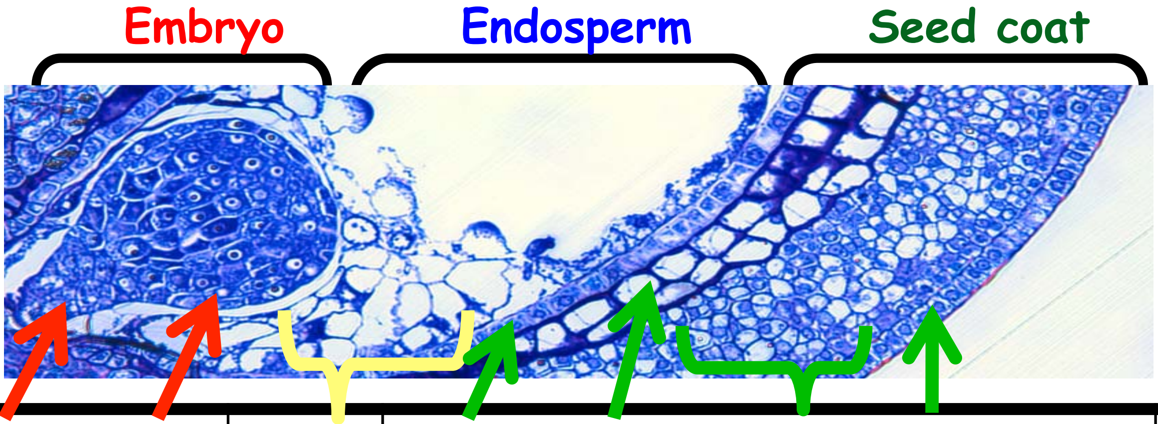
Globular Stage Seed



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

Whole Mount



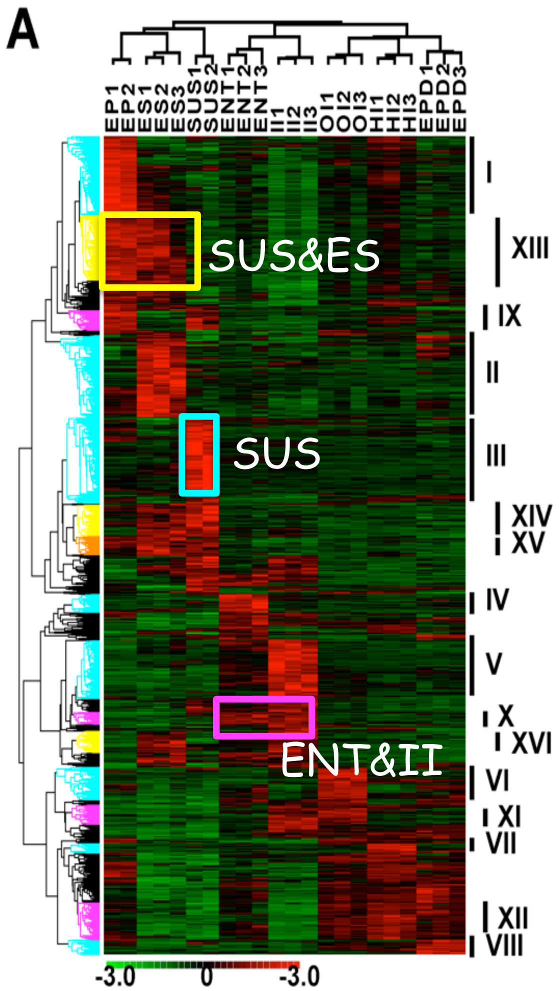


	WM	SUS	EP	ES	ENT	II	OI	EPD	HI	Union
Total mRNAs (TFs)	19,938 (1,373)	10,439 (642)	15,975 (1,040)	13,880 (872)	15,274 (976)	14,767 (952)	16,402 (1,102)	13,451 (865)	16,153 (1,086)	21,989 (1,558)
Unique mRNAs (55 TFs)		62 (2)	70 (14)	115 (13)	47 (4)	44 (5)	79 (7)	29 (2)	66 (8)	
Shared mRNAs (TFs)		← 7,664 (400) 52% quantitatively regulated (ANOVA, $p<0.01$) →								

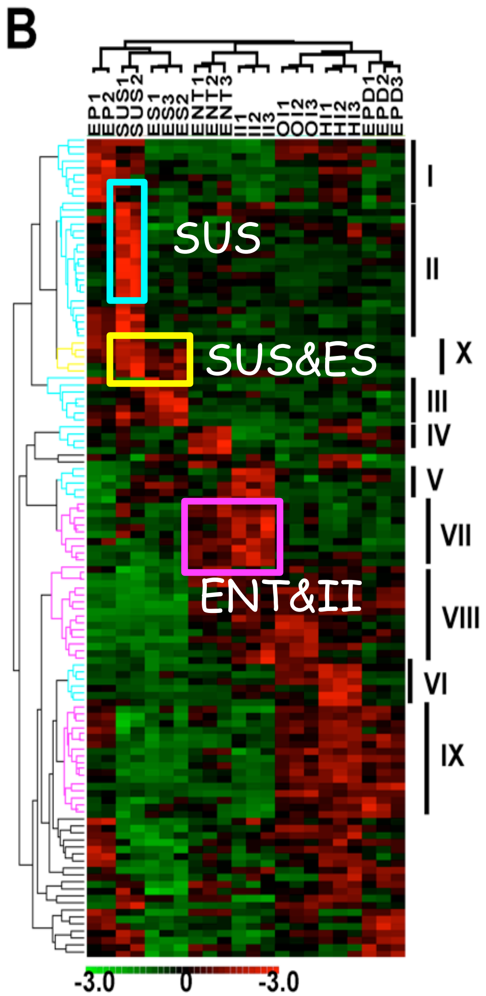
- **Most mRNAs shared by all seed compartments, regions, & tissues**

- **Do shared mRNAs have compartment, region, and tissue-specific patterns?**

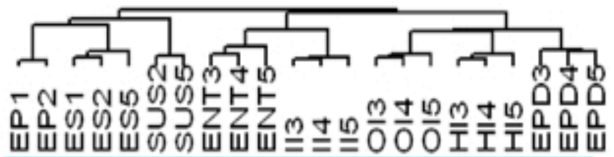
How Are Shared Transcripts Regulated in Globular-Stage Seed Compartments, Regions, and Tissues?



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



Hierarchical Clustering of the Varying TFs mRNAs, p<0.001

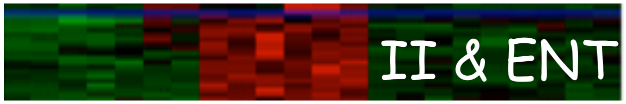


Tissue-Specific Enriched

Sets of mRNAs up-regulated in <u>individual</u> seed tissues

Compartment Enriched

Sets of mRNAs <u>Co-upregulated</u> in multiple seed tissues of the <u>same</u> <u>origins</u> (e.g., Embryo, ENT-II)

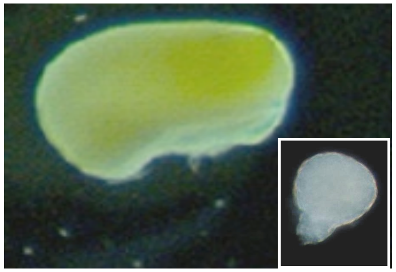


Mosaic

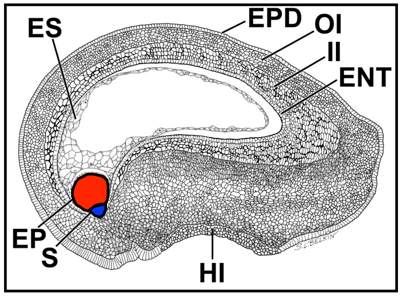
Sets of mRNAs <u>Co-upregulated</u> in multiple seed tissues of the <u>different</u> origins (e.g., EP-ES)

How Many Genes Are Active in the Globular Stage Seed?

Whole Seed

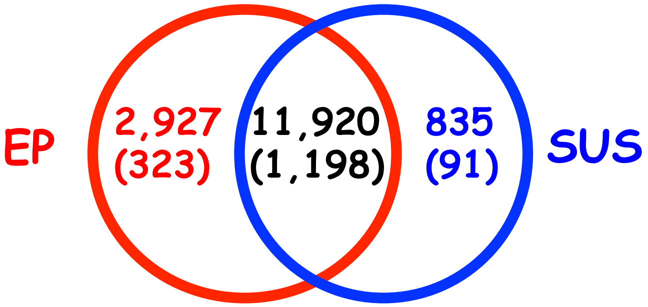


LCM



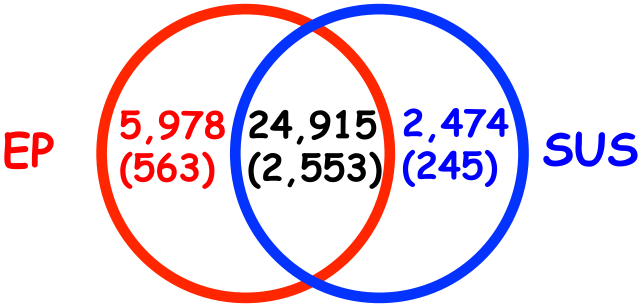
TECHNOLOGY	TISSUE/REGION	GENECHIP	RNA-SEQ	# Mapped Reads
Whole Mount	Whole Seed	17,057 (1,748)	45,821 (4,717)	49.5M
LCM	EP	14,847 (1,521)	30,893 (3,116)	11.6M
	SUS	12,755 (1,289)	27,389 (2,798)	8.4M
	EP + SUS	15,682 (1,612)	33,367 (3,361)	20M

GeneChip



(EP + SUS) = 15,682 (1,612)

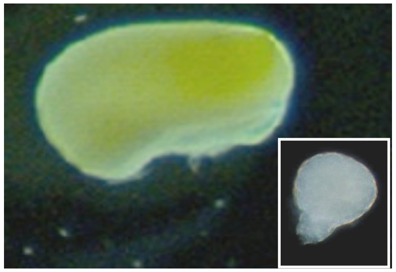
RNA-Seq



(EP + SUS) = 33,367 (3,361)

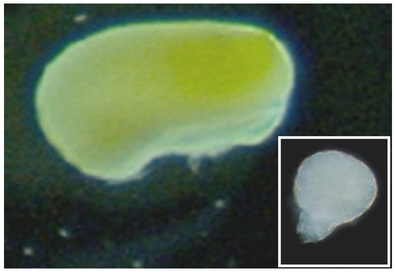
How Many Genes Are Active in the Globular Stage Seed?

Whole Seed GeneChip



17,057
(1,748)

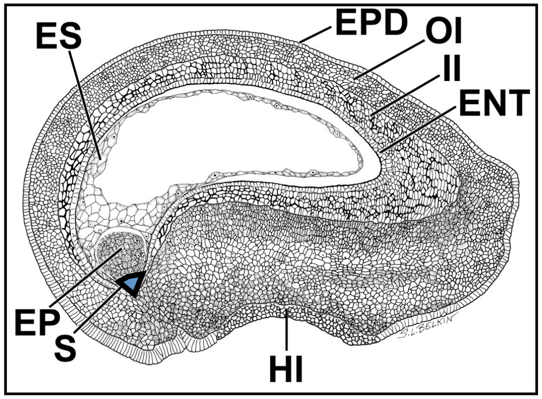
Whole Seed RNASeq



Platform	Illumina GA II
# Reads	89.4M
# Bases	6.8Gb
# Genes Detected	45,821

45,821
(4,717)

LCM GeneChip



18,092
(1,904)

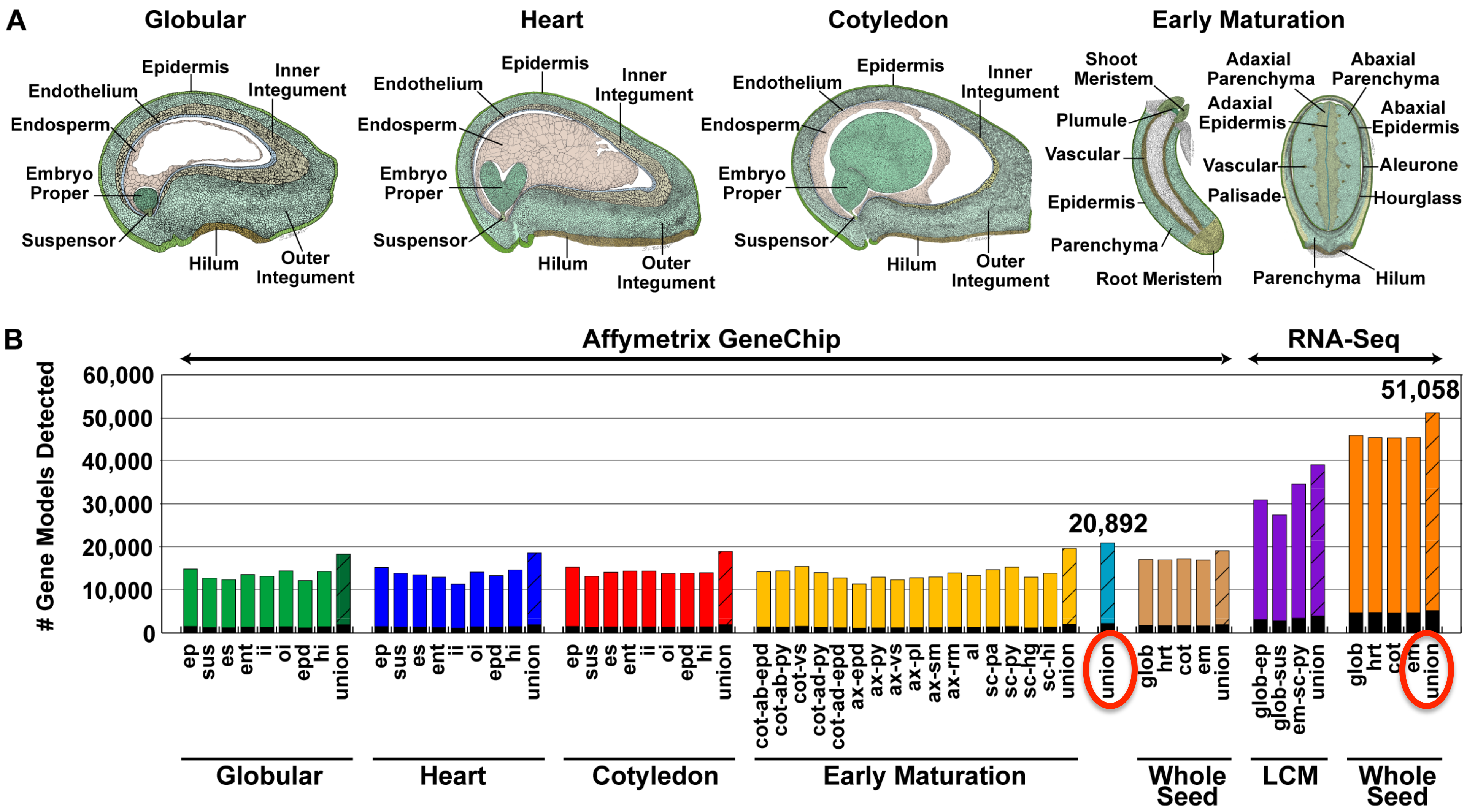
Putative Aldo/Keto Reductase mRNA

Glyma18g52250
(GmaAffx.92851.1.S1_s_at)

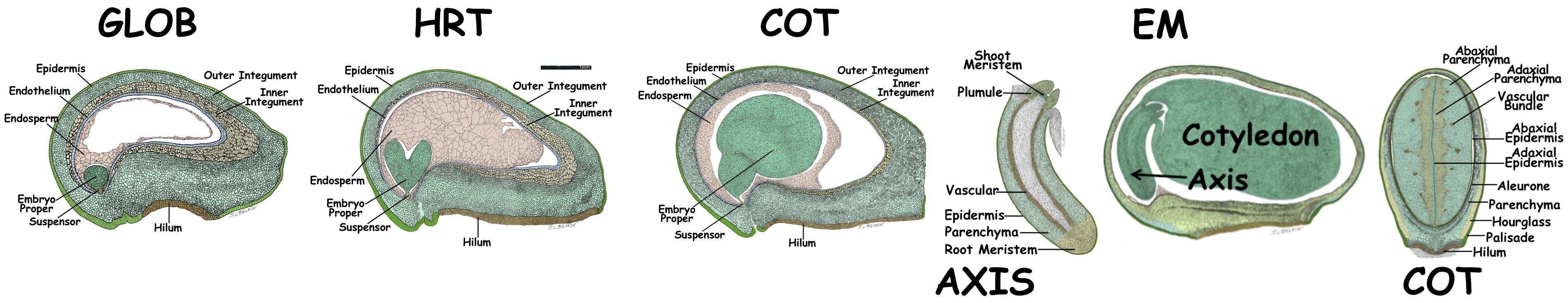
Platform	EP	SUS	ES	ENT	II	OI	EPD	HI	WS
GeneChip	A	5,907	A	A	A	A	A	A	A
qPCR (fold-reduction)	1,328	1	539	ND	ND	33,341	ND	1,474	N/A
RNA-Seq	0	323							12

( ) Transcription Factors

What Are the Genes Active in <u>Every</u> Soybean Compartment, Region, & Tissue Thoughout Development?

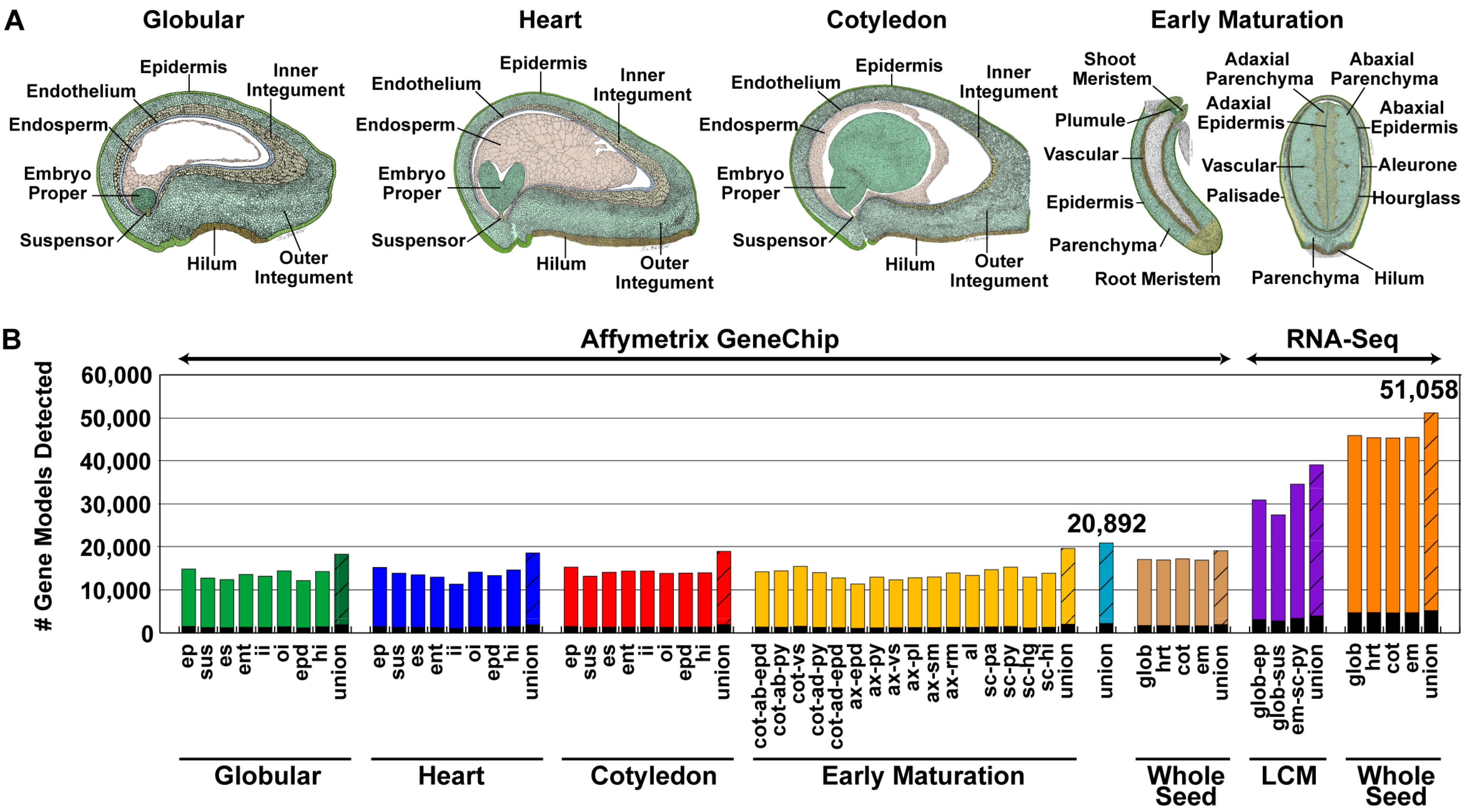


What Are the Genes Active in <u>Every</u> Soybean Compartment, Region, & Tissue Throughout Development?

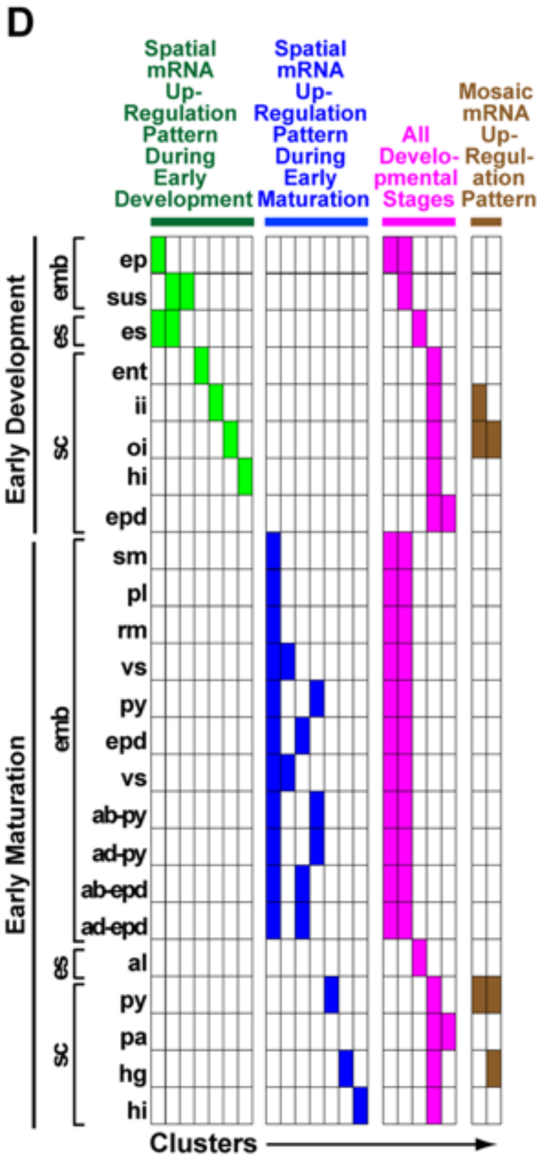


- *~ Same Number of mRNAs/Compartment*
- *Most mRNAs Within a Seed Shared by All Compartments*
- *Most TF mRNAs Within the Seed Shared by All Compartments*
- *Most mRNAs Shared by All Stages of Development*
- *Large Quantitative Changes in Shared mRNA Prevalences*
- *~ Small Number of Compartment & Stage Specific mRNAs*
- *~ 50,000 Genes Required to Make a Seed*

How Are Genes <u>Regulated</u> During Soybean Seed Development?



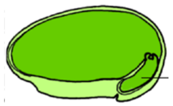
mRNA Accumulation Patterns Throughout Soybean Seed Development – A Summary



1

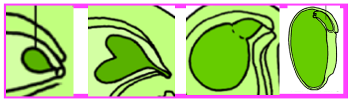
A Spatial pattern of mRNA Up-Regulation During Early Development (GLOB-HRT-COT)

2



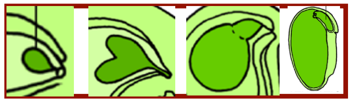
A Spatial pattern of mRNA Up-Regulation During Early Maturation (EM)

3



A Spatial pattern of mRNA Up-Regulation Throughout All Developmental Stages (GLOB to EM)

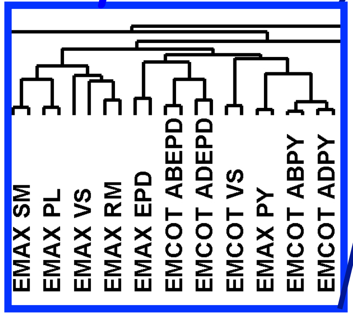
4



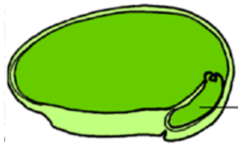
Mosaic mRNA Up-Regulation Pattern

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

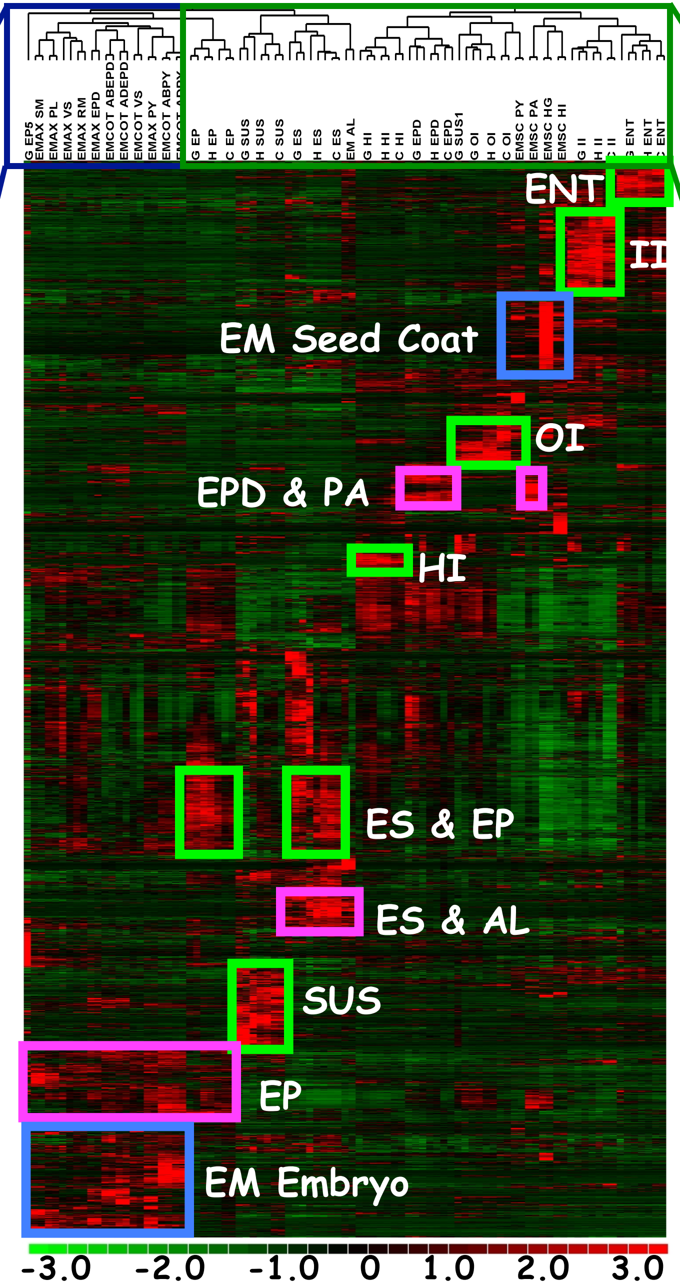
Maturation phase



②

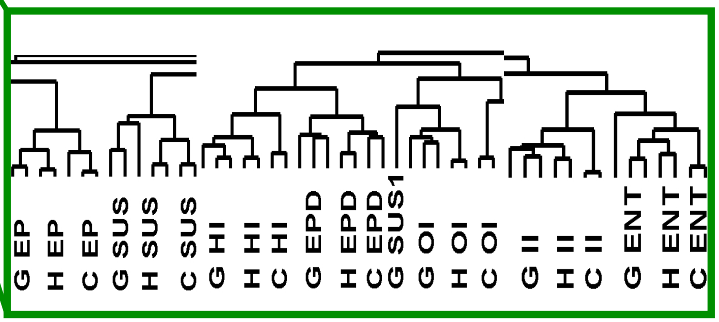


A TEMPORAL pattern of mRNA accumulation is apparent (Early Maturation)



Top 4K Varying mRNAs

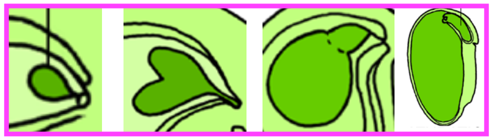
Morphogenesis phase



①

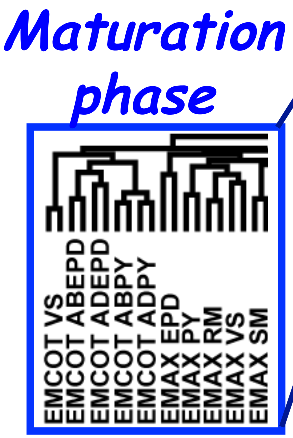
A SPATIAL pattern of mRNA accumulation is apparent (GLOB-HRT-COT)

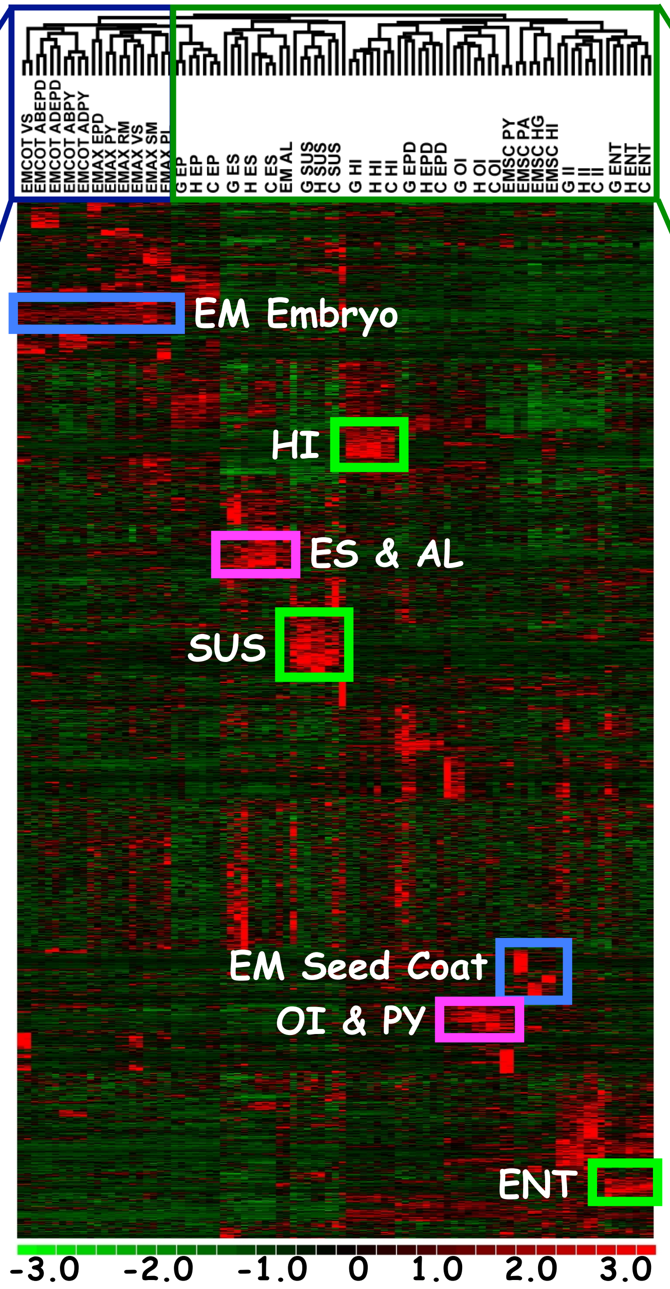
③

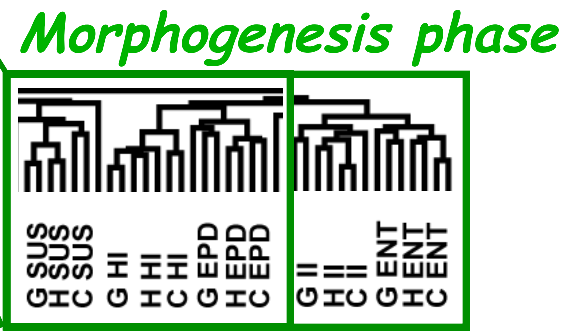


mRNA sets up-regulated throughout development (GLOB to EM)

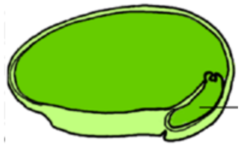
What Are the Temporal and Spatial Transcription Factor mRNA Accumulation Patterns Throughout Seed Development?







②

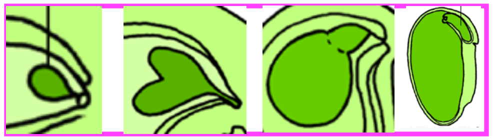


A TEMPORAL pattern of TF mRNA accumulation is apparent (Early Maturation)

①

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

③

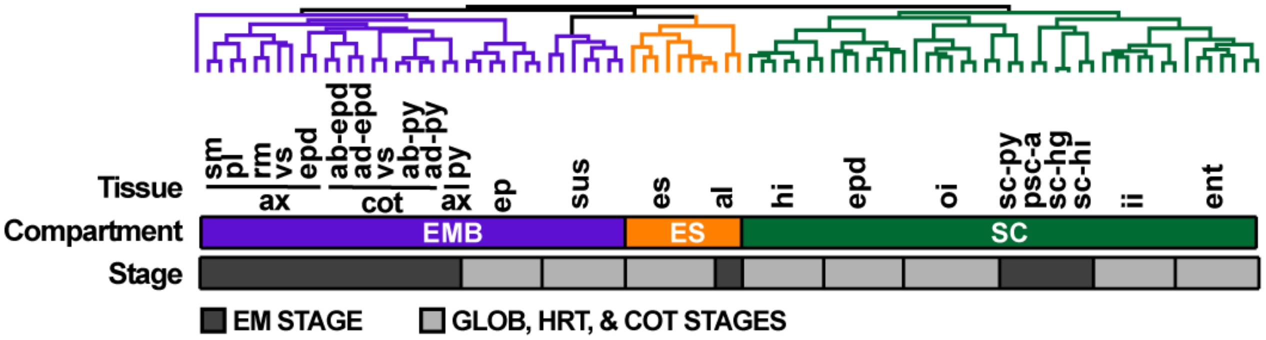


TF mRNA sets up-regulated throughout development (GLOB to EM)

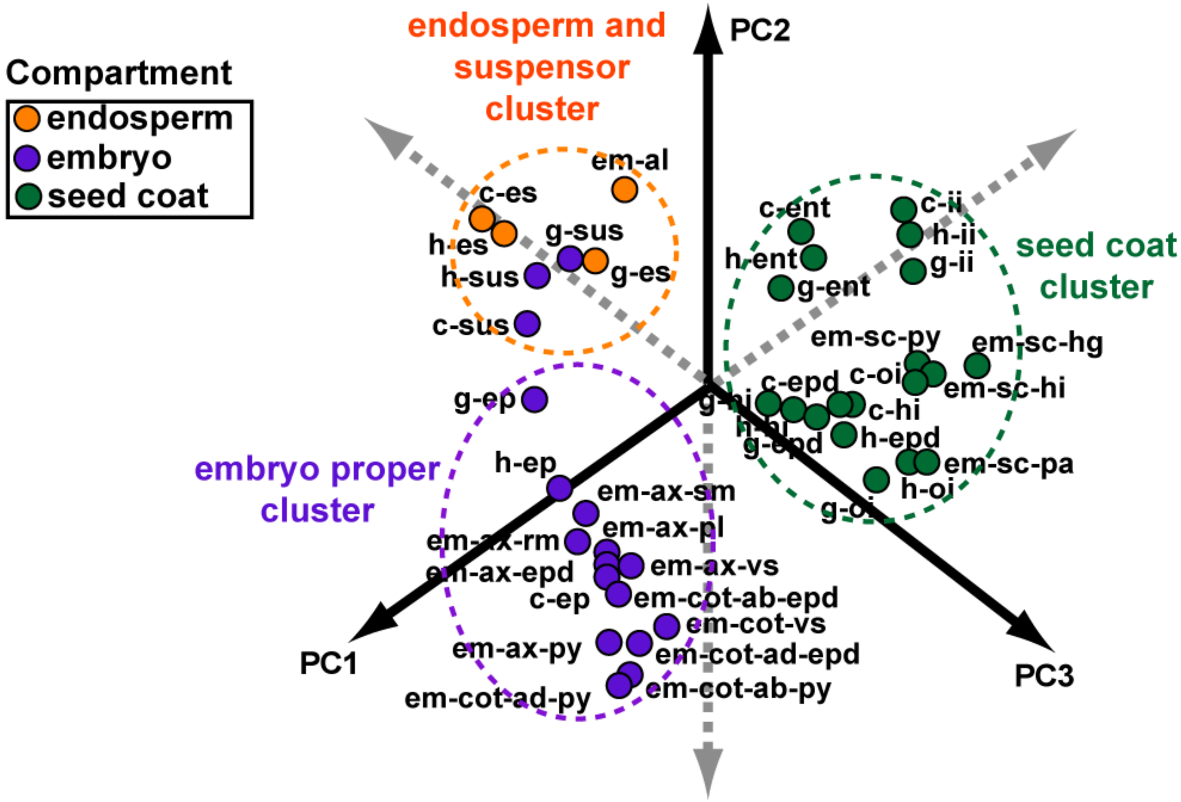
1,930 TF mRNAs Present in at Least One Compartment During Seed Development

What Are the Biological Relationships Between 40 Seed Compartments, Regions, and Tissues Throughout Development?

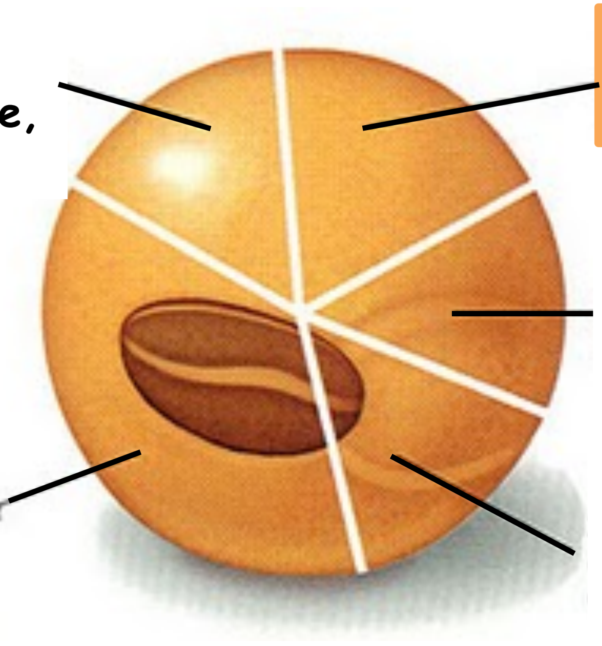
Hierarchical Clustering



Principal Component Analysis



Are Different Seed Compartments Specialized For Specific Metabolic Processes?



15% Soluble Carbohydrates (Sucrose, Raffinose, Stachyose)

20% Oil (Oleate, Linoleate etc)

10% "Other" (Isoflavones, Vitamins, etc)

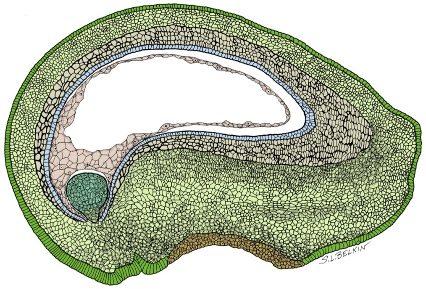
40% Protein (Glycinin, Beta-conglycinin etc)

15% Insoluble Carbohydrates (Dietary fiber)

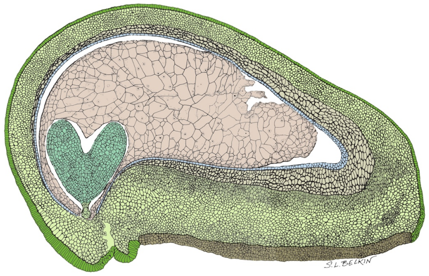
Composition of Soybean Seed

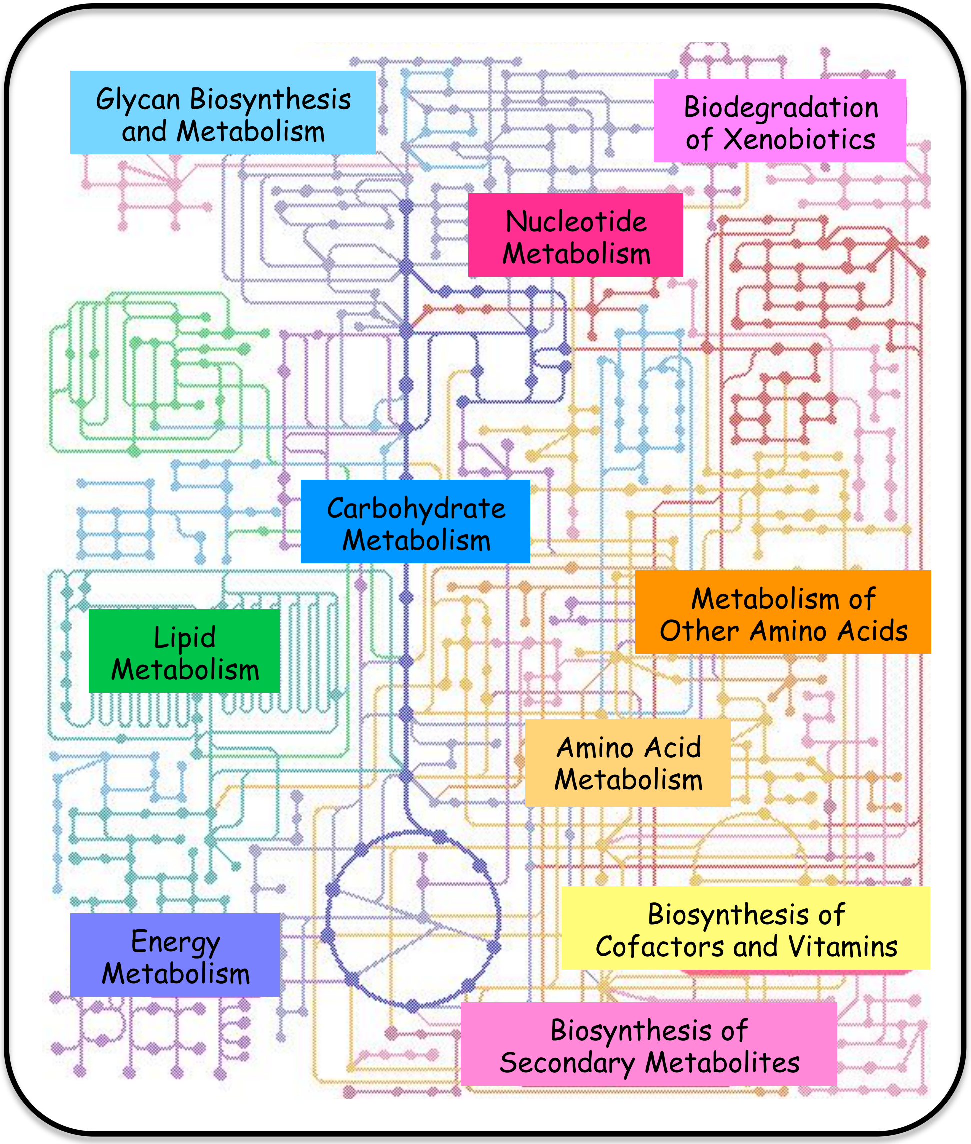
Are Seed Compartments Specialized For Specific Metabolic Processes?

Globular



Heart




Glycan Biosynthesis and Metabolism

Biodegradation of Xenobiotics

Nucleotide Metabolism

Carbohydrate Metabolism

Lipid Metabolism

Metabolism of Other Amino Acids

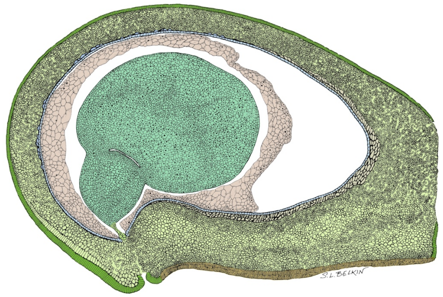
Amino Acid Metabolism

Energy Metabolism

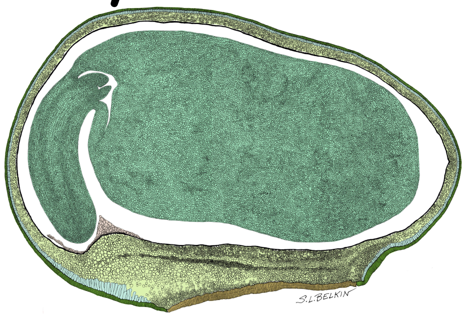
Biosynthesis of Cofactors and Vitamins

Biosynthesis of Secondary Metabolites

Cotyledon

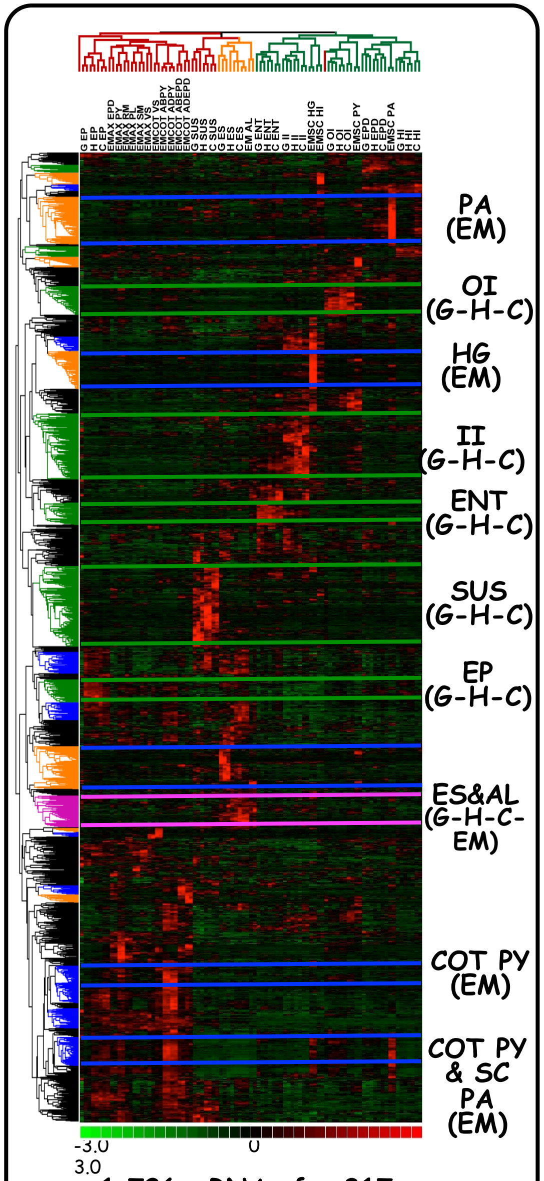


Early Maturation



Examined 325 Metabolic Pathways in 40 Different Seed Compartments

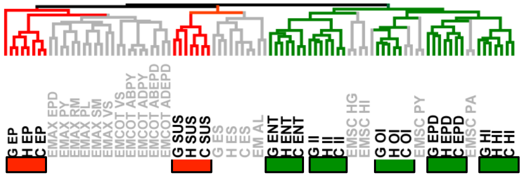
What are the Accumulation Patterns for mRNAs Encoding Enzymes in Metabolic Pathways During Seed Development?

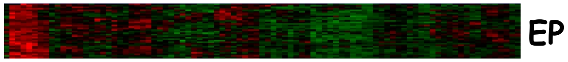


1,726 mRNAs for 317 pathways (ANOVA, p<0.001)

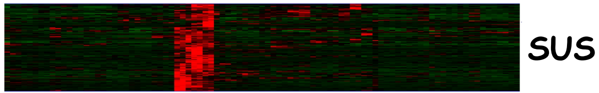
Spatial Pattern of Enzyme mRNA Accumulation During Early Seed Development (GLOB-HRT-COT)

1

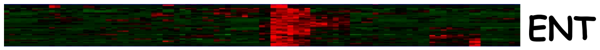




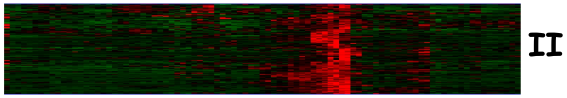
Photosynthesis, Purine Biosynthesis, Sucrose Degradation



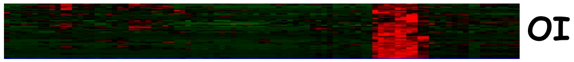
Cutin, Suberin, & Flavonoid Biosynthesis



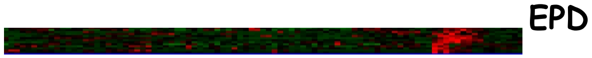
Terpene, Phenylpropanoid, Flavonoid Biosynthesis



Xyloglucan Biosynthesis



Jasmonic Acid & Oleate Biosynthesis

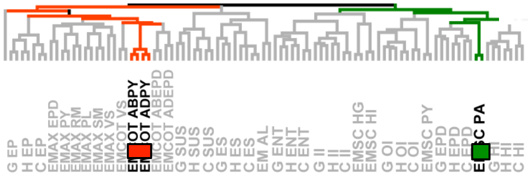


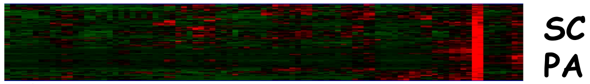
Cutin, ABA Glucose Ester Biosynthesis

Flavonoid & Suberin Biosynthesis

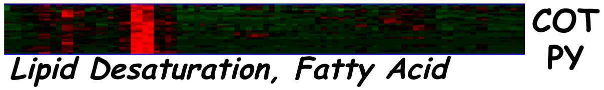
Spatial Pattern of Enzyme mRNA Accumulation Late in Seed Development (Early Maturation)

2





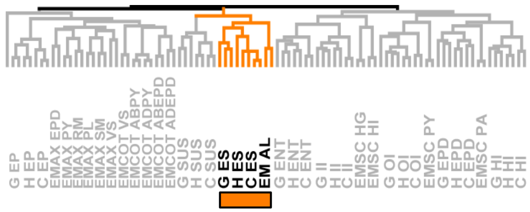
Flavonoid & Suberin Biosynthesis

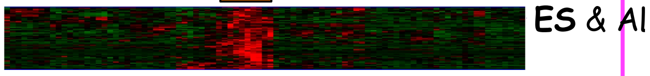


Lipid Desaturation, Fatty Acid Biosynthesis

Enzyme mRNA Sets Up-Regulated Throughout Seed Development (GLOB to EM)

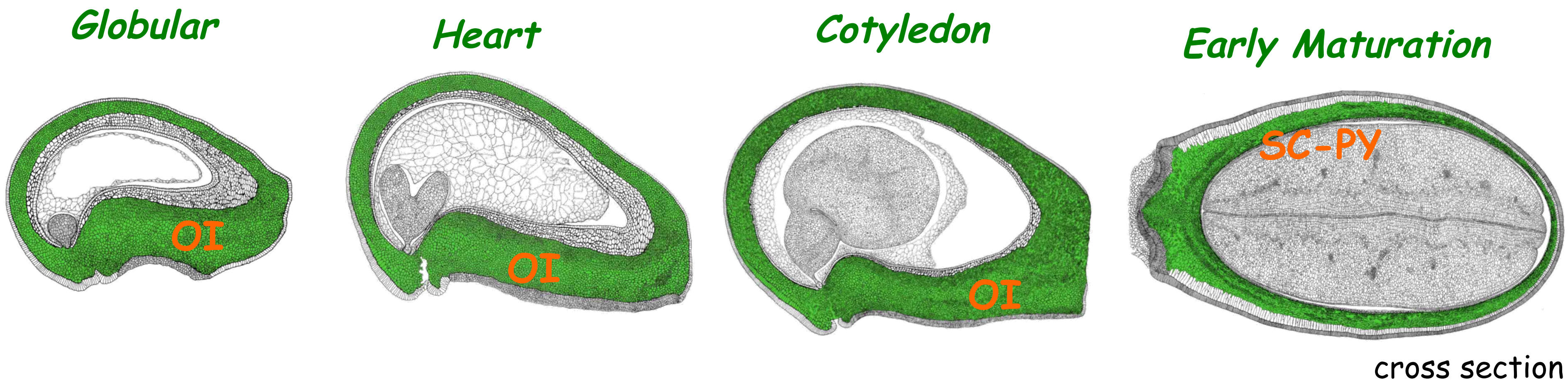
3



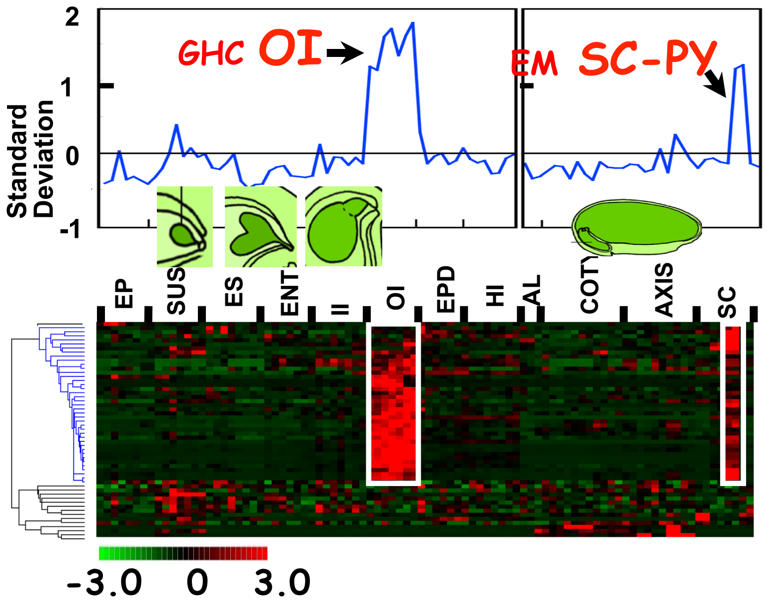


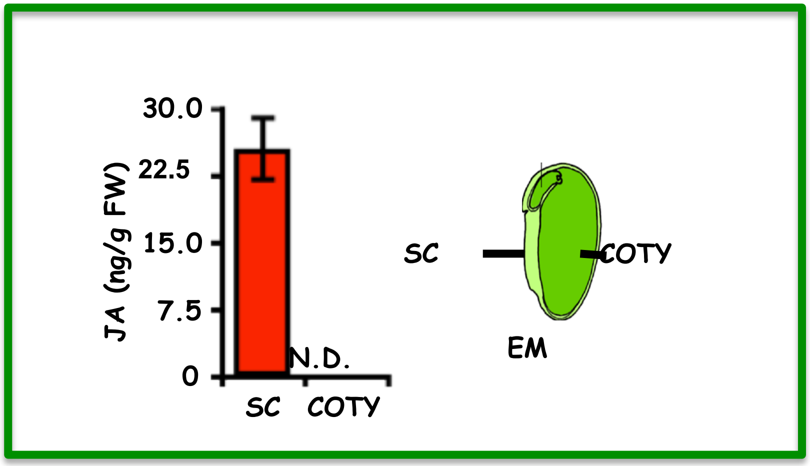
Cellulose Biosynthesis, Purine and Pyrimidine Salvage Pathways

Are Seed Compartments Specialized For Specific Metabolic Processes? The Outer Integument/Seed Coat-Parenchyma Case



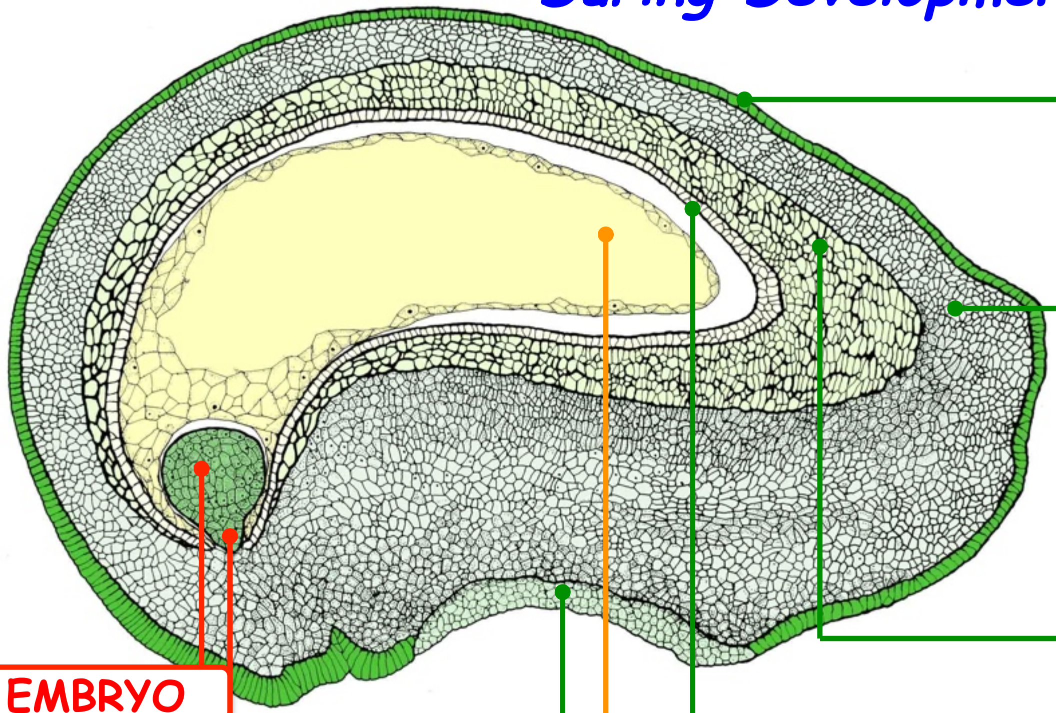
Jasmonic acid biosynthesis pathway mRNAs are up-regulated in the outer integument & seed coat parenchyma tissue during seed development





High detection of JA in the seed coat!

Functional & Regulatory Specialization of of Seed Compartments During Development



SEED COAT

EPIDERMIS
Anthocyanin Biosynthesis
Cellulose Biosynthesis
Cutin Biosynthesis
TF mRNAs: TINY, GLABRA 2, WRKY44

OUTER INTEGUMENT
Photosynthesis
Jasmonic Acid Biosynthesis
Fatty Acid Biosynthesis
Lipid Biosynthesis
TF mRNAs: MYB, GBF6, LIM

INNER INTEGUMENT
Cellulose Biosynthesis
Starch Biosynthesis
Sucrose Biosynthesis
Vitamin E Biosynthesis
TF mRNAs: AP2-EREBP, PCL1, ATAF1

ENDOTHELIUM
Triterpenoid Biosynthesis
Quercetin Biosynthesis
Cutin Biosynthesis
Vitamin E Biosynthesis
TF mRNAs: TT1, GRF5, Zinc Finger

HILUM
Cell Wall Biogenesis
Quercetin Biosynthesis
Regulation of Transcription
TF mRNAs: Dof-type Zinc Finger, DAG2,Ethylene-Responsive

EMBRYO

EMBRYO PROPER
Starch Degradation
Fatty Acid Biosynthesis
Thiamine Biosynthesis
Photosynthesis
TF mRNAs: STF, GRF3, PAP2

SUSPENSOR
Thiamine Biosynthesis
Aromatic Compound Biosynthesis
TF mRNAs:
MYB78, LEC1-like, ABI4

ENDOSPERM

ENDOSPERM
Pathways:
Amino Acid Biosynthesis (Ile, Val, Leu)
Carotenoid Biosynthesis
TF mRNAs: VRN, Zinc Finger, BES1/BZR1

Where Are the Pathways For Soybean "Health" Products Made in the Seed?

AXIS

PLUMULE
Storage Protein (β-conglycinin)

SHOOT MERISTEM
Storage Protein (β-conglycinin)

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

EPIDERMIS
Vitamin B1

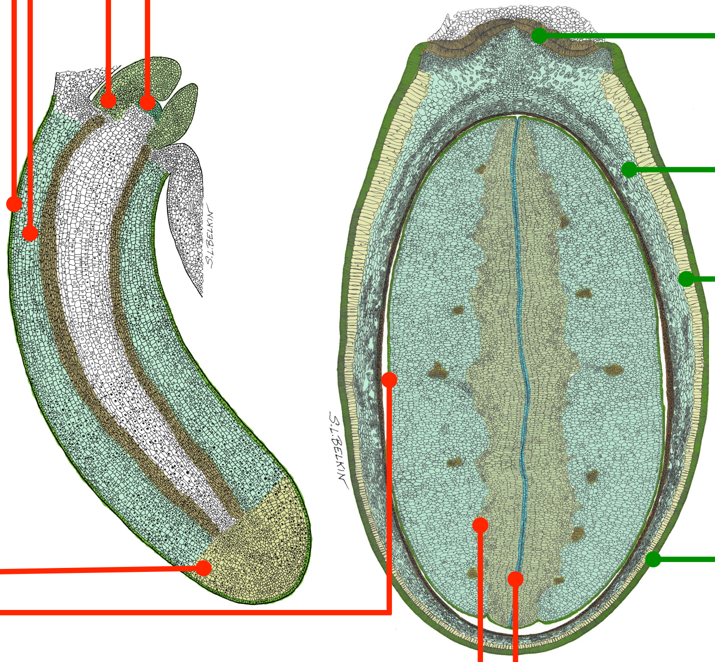
ROOT MERISTEM
Isoflavones (Genistein, Daidzein)

COTYLEDON

PARENCHYMA
Vitamin C, Vitamin E
Fatty acids (Linoleate, Oleate)
Amino acids (Arg, Gln, Leu, Lys)
Storage Proteins (β-conglycinin, Glycinin)

ADAXIAL EPIDERMIS
Storage Proteins (β-conglycinin, Glycinin)

ABAXIAL EPIDERMIS
Amino acids (Arg, Gln, Leu)



SEED COAT

HILUM
Amino acid (Phe)

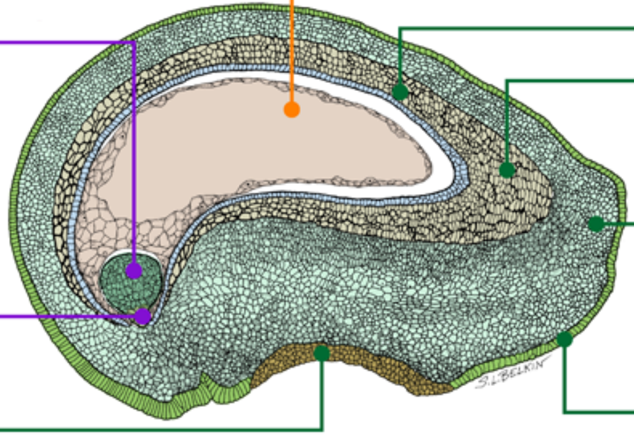
PARENCHYMA
Sucrose, Stachyose

HOUR GLASS
Vitamin E
Sucrose
Amino acids (Leu, Lys, Ser)

PALISADE
Isoflavones (Genistein, Daidzein)
Linoleate
Sucrose
Amino acid (Phe)

*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

Regulatory Specialization of Seed Compartments During Development



EMBRYO

EMBRYO PROPER
YABBY5 (ghc), STM(ghc), AtGRF3 (gh), WOX2 (g), SRS5 (g), ARF4 (g), AS1(hc), HSFB4 (hc), YABBY1 (hc), AtGRF5 (hc), AN3 (h), LEC1-like (c), KNAT7 (c)

SUSPENSOR
ANAC033 (ghc), AtHB-1 (ghc), PLT1 (ghc), ARF10 (ghc), LEC1-like (g), ANAC073 (h), SHY2 (c)

SEED COAT

HILUM
LRP1 (ghc), CPC (ghc), BEL1 (gh), ANT (gh), KNAT4 (gh), LRL1 (gh), AP3 (gc), AIL5 (gh), HSFB4 (g), IAA14 (g), SHP1 (h), SHI (c)

ENDOSPERM

ENDOSPERM
VRN2 (ghc), DAG2 (g), FLP (hc), AIL7 (hc), GL2 (h), FIE (h)

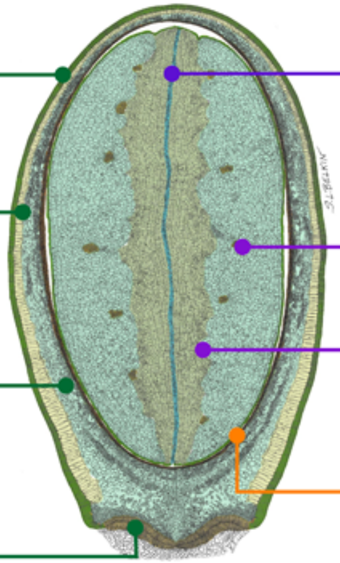
SEED COAT

ENDOTHELIUM
ARR4 (ghc), TT1 (gc), JKD (g), SEP3 (g), BEL1 (h)

INNER INTEGUMENT
TCP20 (ghc), BLH1 (gh), AtGRF7 (g), GAI (h), SEP3 (h)

OUTER INTEGUMENT
AP2 (ghc), GBF6 (gh), LBD38 (gh), RL5(g), SPL9 (g), BHLH093 (c)

EPIDERMIS
GL2 (ghc), TT2 (ghc), WLIM1 (ghc), CPC (gh), TCP2 (gh), WRKY2 (hc), AP2 (h), TCP3 (h)



SEED COAT

PALISADE
WRKY2, GL2, AGL20, AP2

HOURGLASS
AtERF9, AP2 domain containing transcription factor

PARENCHYMA
RAP2.1, WRKY40, ANAC102, AP3, LBD 38, SEP1, AtHB-1

HILUM
DAG2, BEL1

EMBRYO COTYLEDON

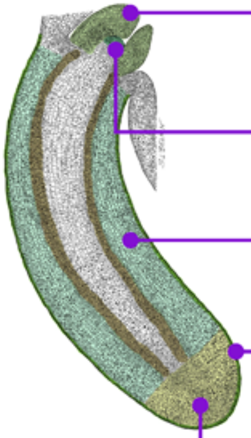
EPIDERMIS
MYB106, ARR16, ZFP4, ARR17

VASCULAR
MYB59, WRKY13, PAP2, AS1

PARENCHYMA
SPL8, ANAC036, ATHB-1, AtIDD4

ENDOSPERM

ALEURONE
AtHB22/MEE68, RAP2.1, WRKY23, KNAT3



EMBRYO AXIS

PLUMULE
AFO, SRS5

SHOOT MERISTEM
STM, AN3, SEP1

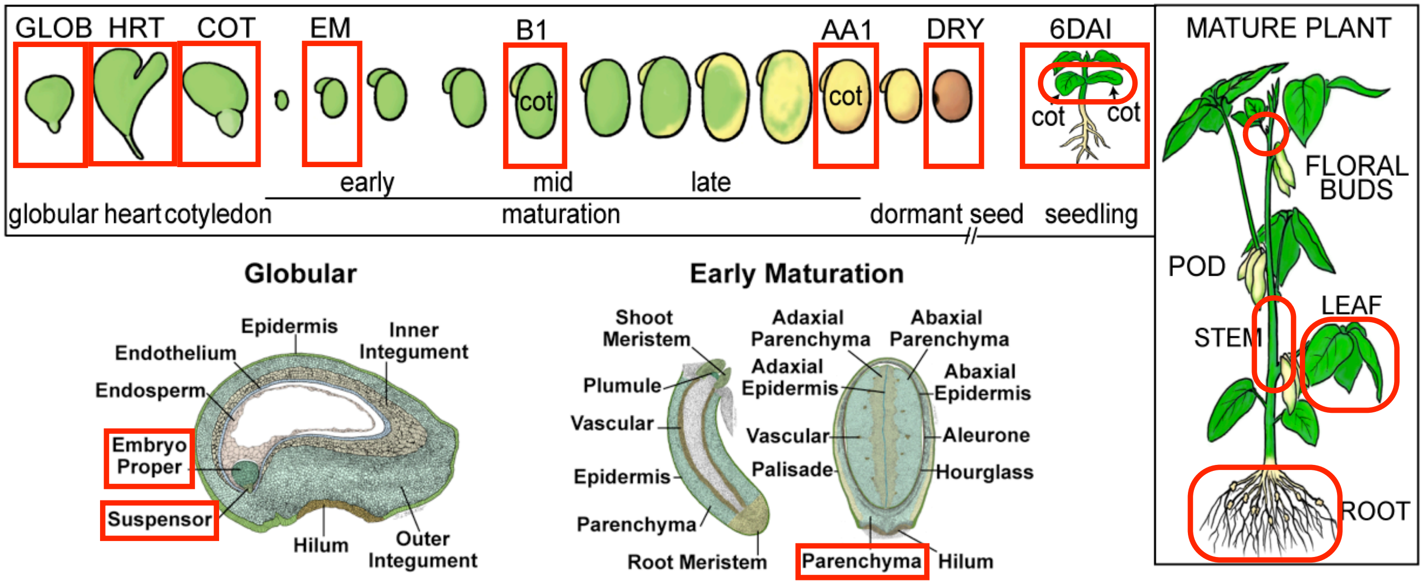
PARENCHYMA
ATS, AtHB13

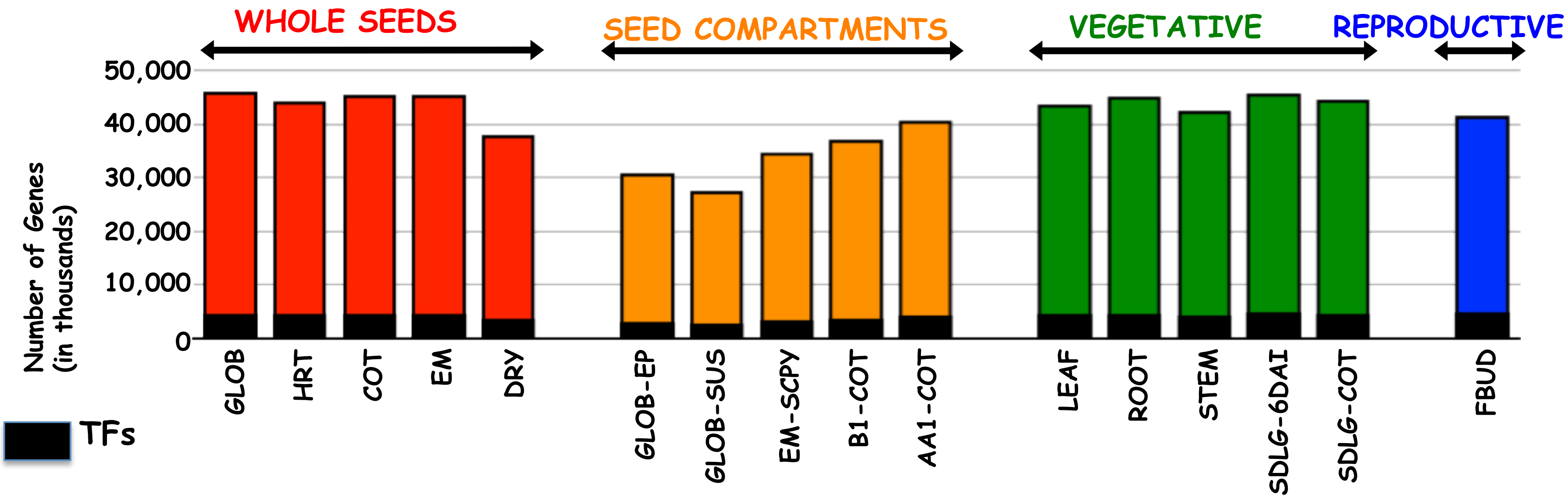
EPIDERMIS
SHY2, AIL6

ROOT MERISTEM
PLT1, SPT

( ) developmental stage

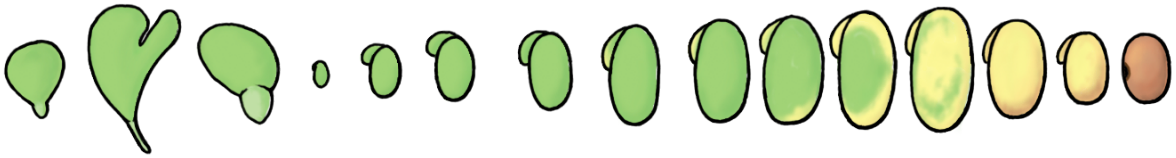
How Many Genes Are Active Throughout the Soybean Life Cycle?

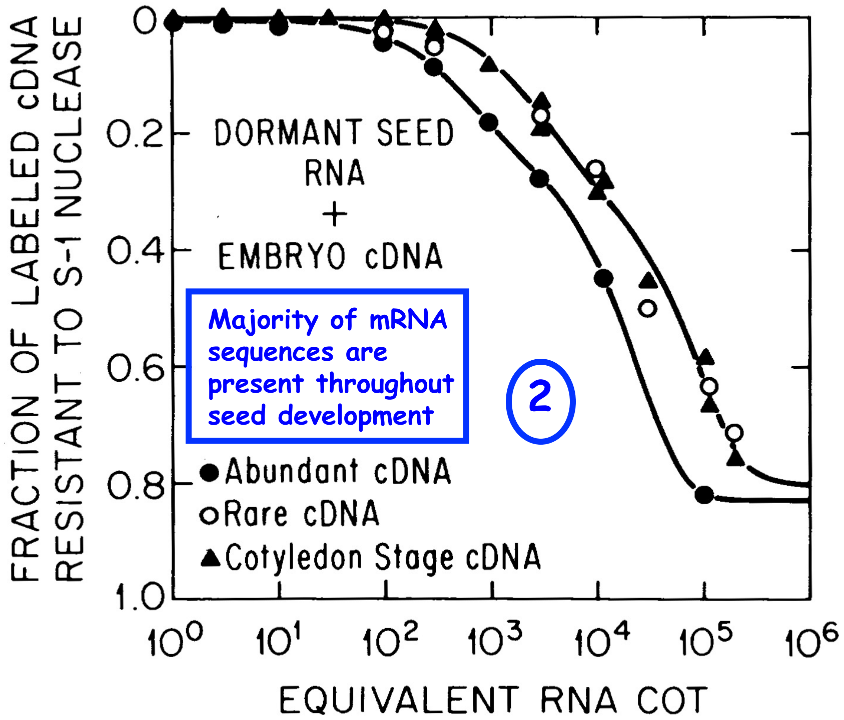




There are 55,343 Genes Active Throughout The Soybean Life Cycle!

A Blast From the Past..........





Majority of mRNA sequences are present throughout seed development

②

①

Gene Numbers in Seed

Soybean RNASeq 2011	Soybean Rot Curve 1981
51,000	32,000

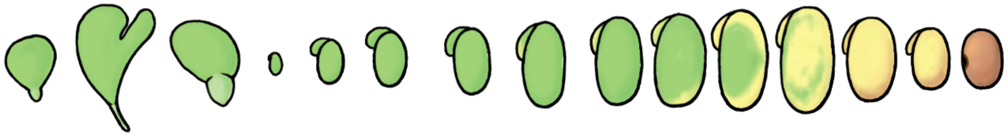
DEVELOPMENTAL BIOLOGY 83, 201-217 (1981)

Abundance, Diversity, and Regulation of mRNA Sequence Sets in Soybean Embryogenesis

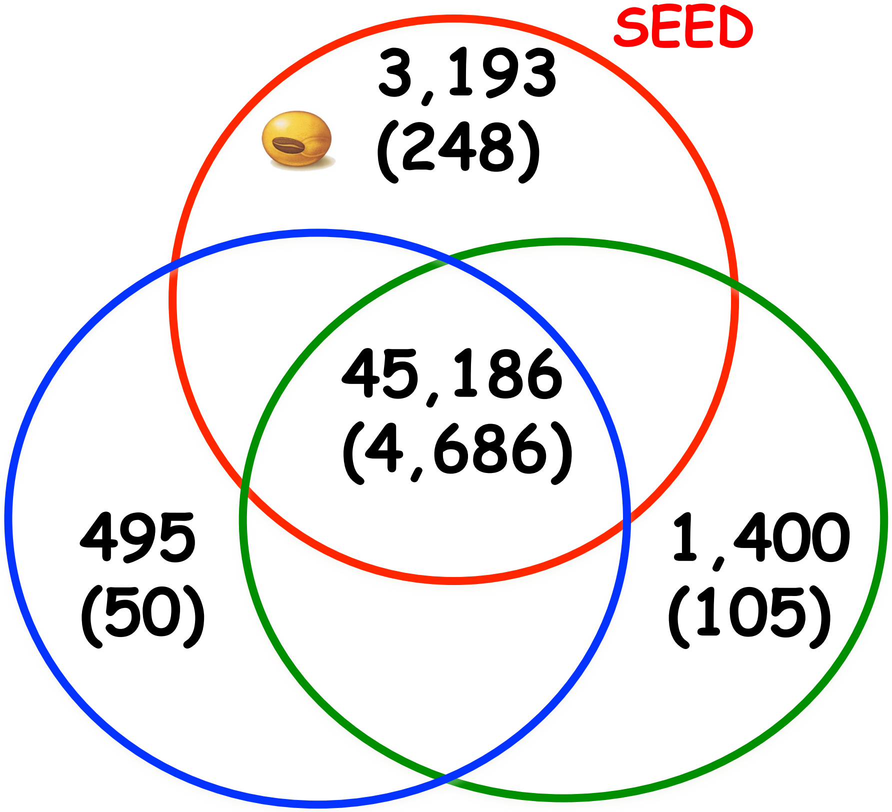
Robert B. Goldberg,*,[1] Gisela Hoschek,* Susan H. Tam,* Gary S. Ditta,*,[2] and R. W. Breidenbach†

*Department of Biology, University of California, Los Angeles, California 90024 and †Plant Growth Laboratory,
University of California, Davis, California 95616

Are There mRNAs Specific to Seed Development?



**Union of Whole Seed & Seed Compartment RNASeq Sets = 52,685 mRNAs (5,294)*



SEED

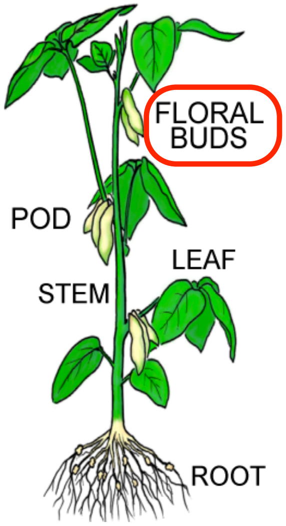
3,193
(248)

45,186
(4,686)

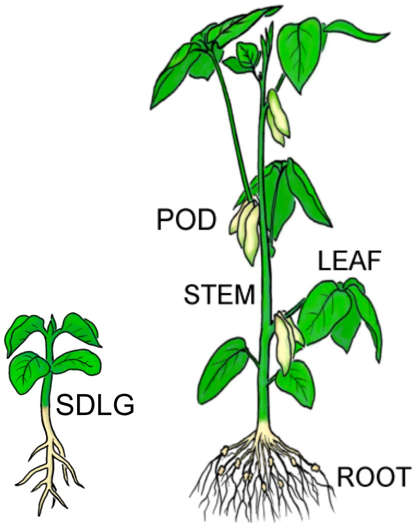
495
(50)

1,400
(105)

REPRODUCTIVE

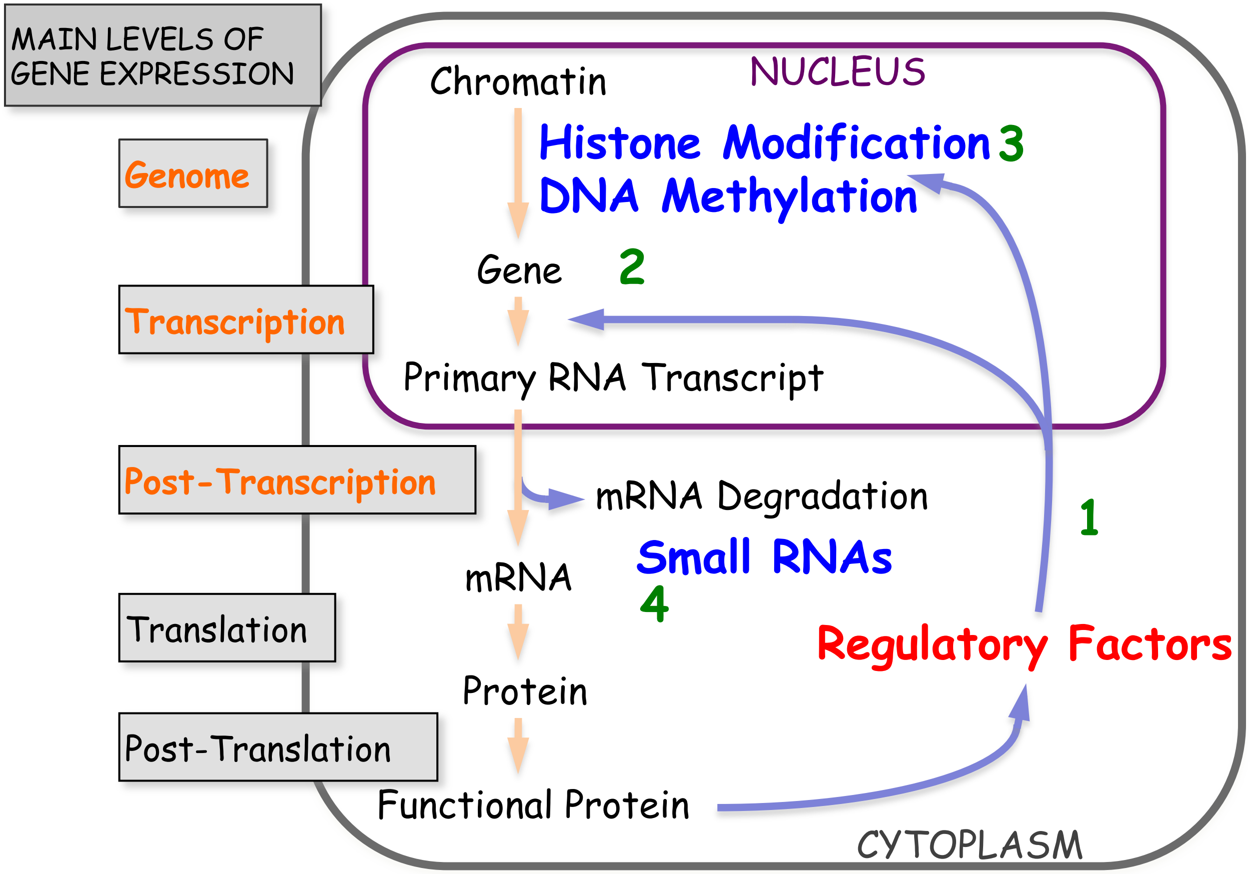


VEGETATIVE



*Union of all LCM & WM (glob-dry)

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



MAIN LEVELS OF GENE EXPRESSION

Genome

Transcription

Post-Transcription

Translation

Post-Translation

NUCLEUS

Chromatin

Histone Modification 3

DNA Methylation

Gene 2

Primary RNA Transcript

mRNA Degradation

Small RNAs

4

mRNA

1

Regulatory Factors

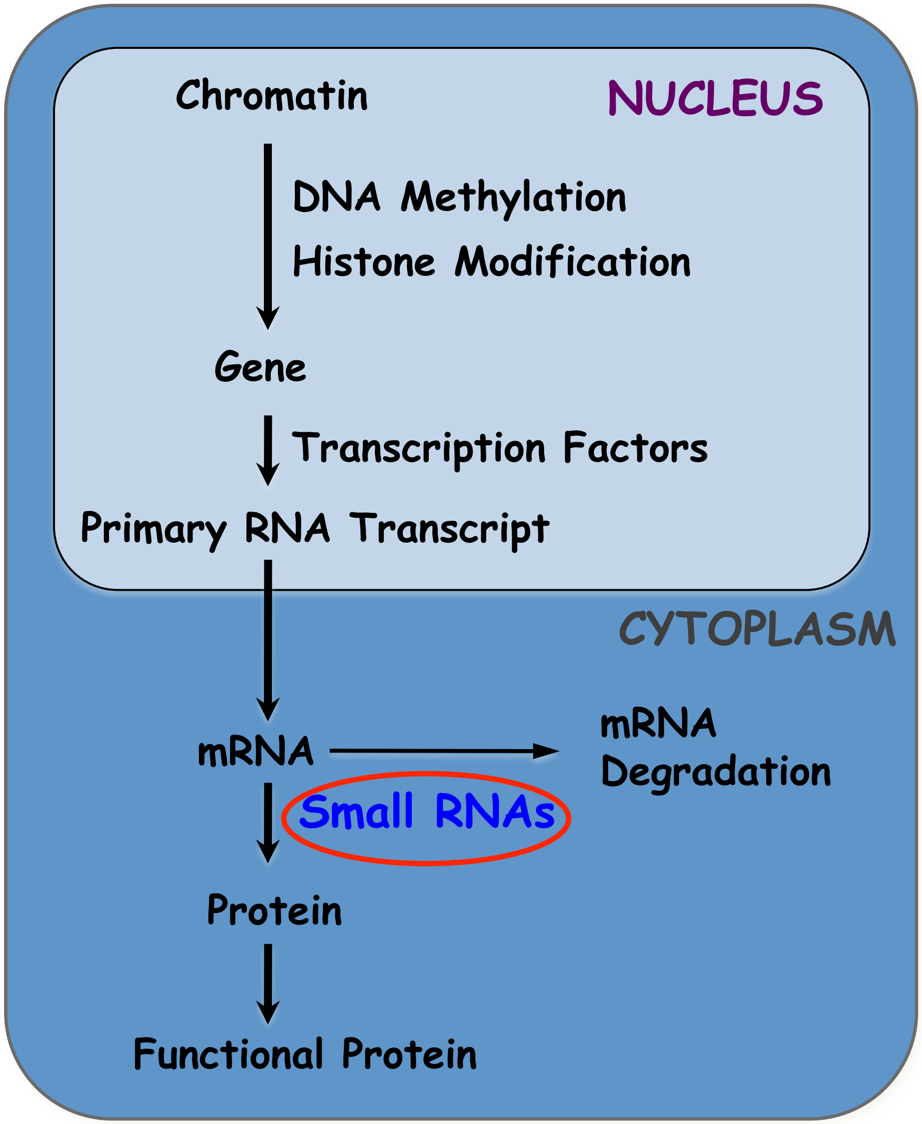
Protein

Functional Protein

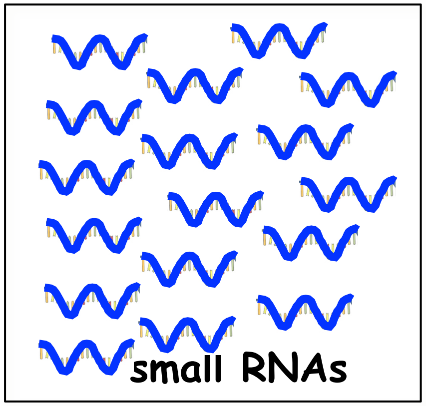
CYTOPLASM

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

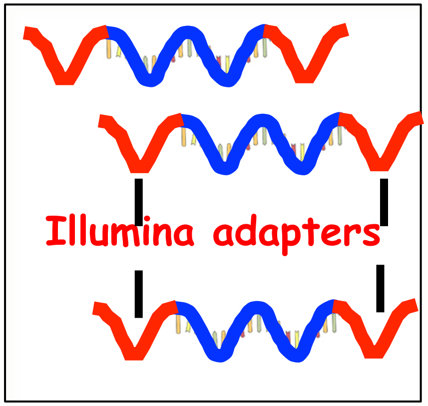
Are Their Compartment-Specific MicroRNAs & Do They Regulate Specific mRNAs?



Enriched Small RNA Isolation



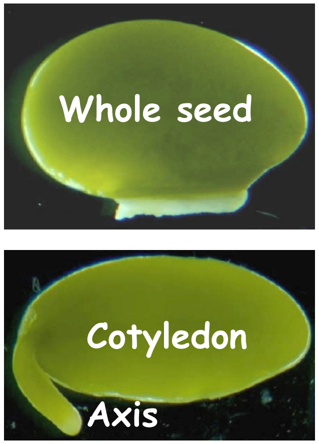
cDNA Library Construction



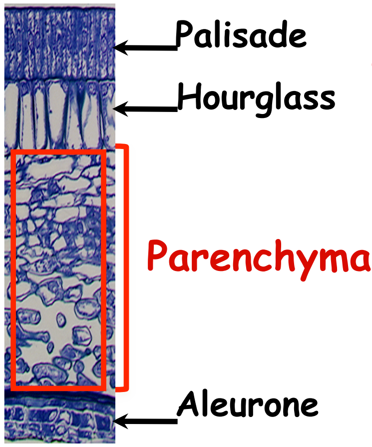
Illumina Sequencing

Using LCM to Sequence and Identify Seed Compartment-Specific MicroRNAs?

Early Maturation Stage Seed



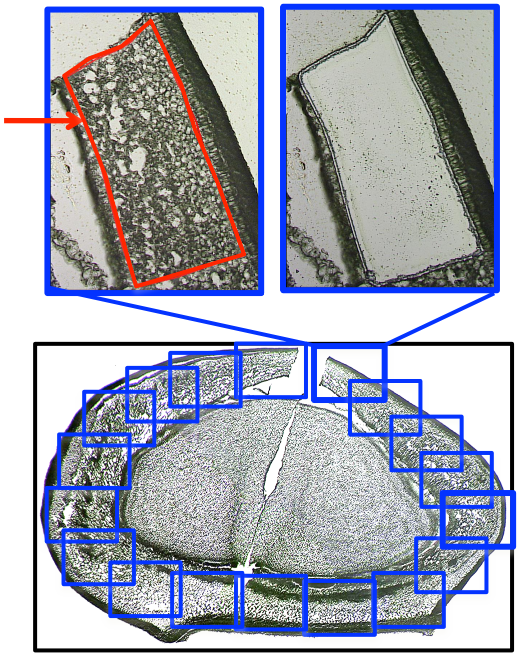
Early Maturation Seed Coat Layers



A LCM Section of Seed Coat Parenchyma

Before LCM

After LCM

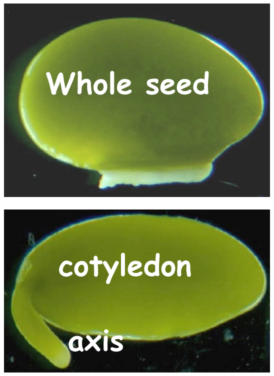


Longitudinal Section

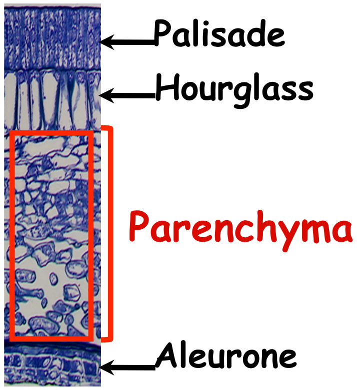
# Longitudinal sections (10μm)	23
Avg. # LCM sections per longitudinal section	20
Total # LCM sections captured	460
Amount of Enriched Small RNAs obtained (ng)	~126

Do MicroRNAs Regulate Compartment-Specific mRNAs?

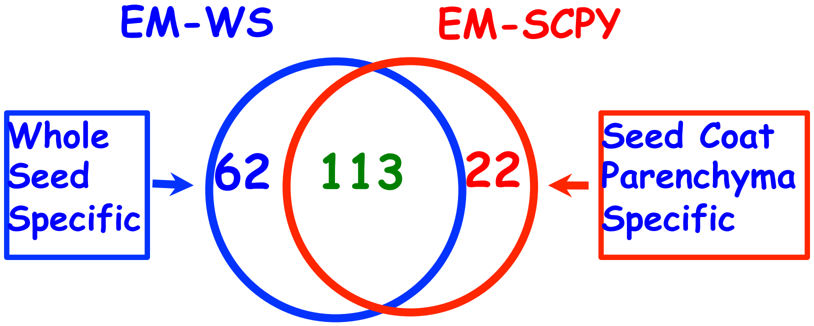
Early Maturation
Whole Seed
(EM-WS)

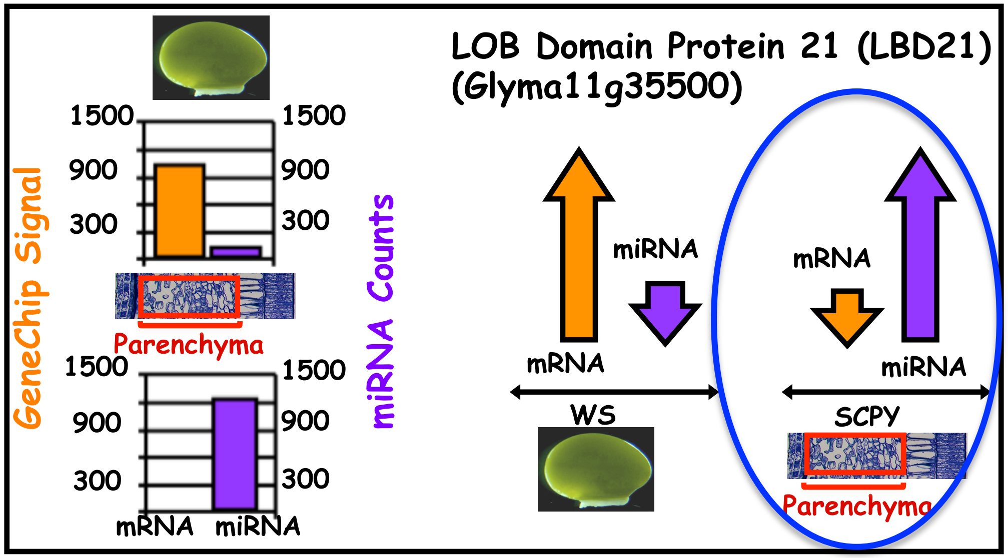


Early Maturation
Seed Coat Parenchyma
(EM-SCPY)



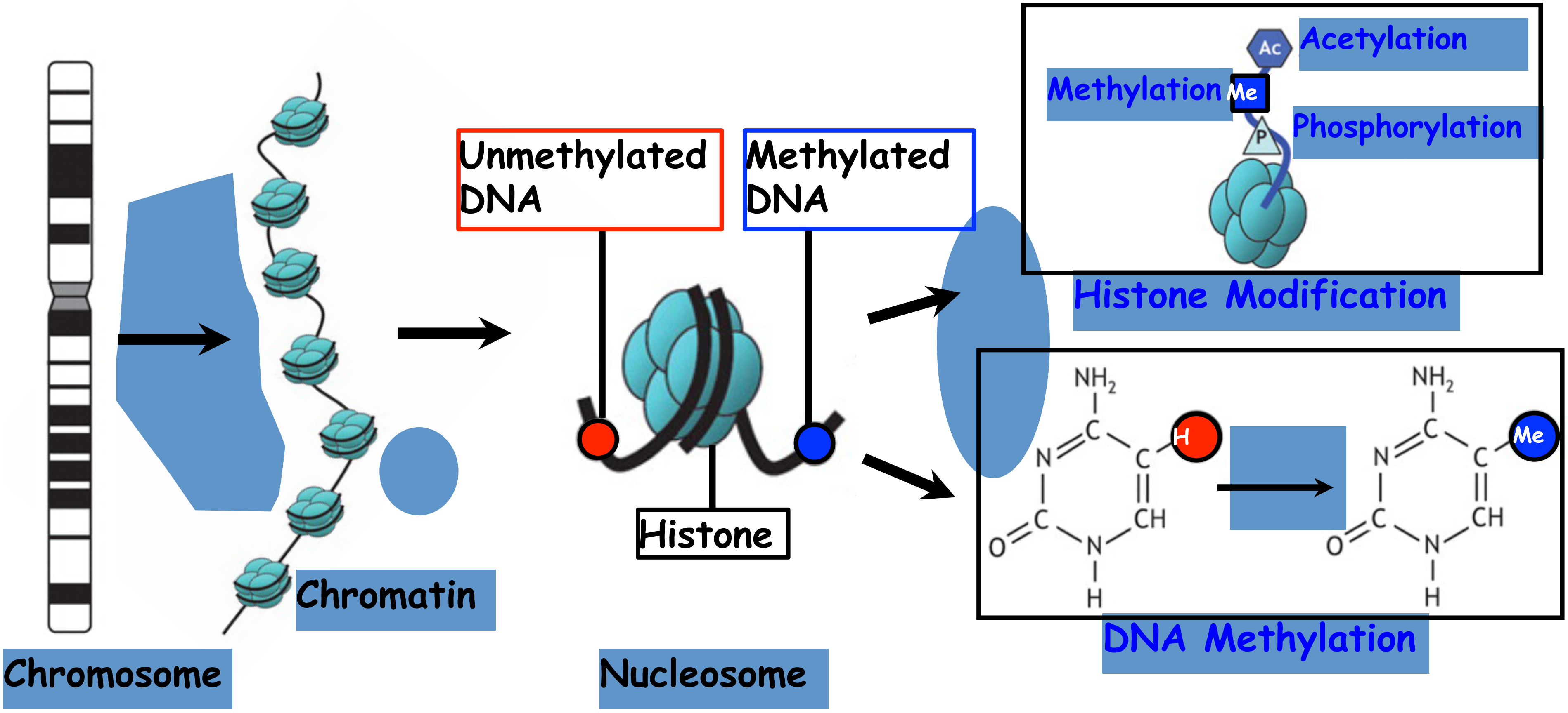
	EM-WS	EM-SCPY
# Reads	28M (2.1Gb)	28M (2.1Gb)
# Cleaned Reads (18-24 nt) (e.g. remove tRNA, rRNA)	15M	7.7M
# Cleaned Reads With Homology to Known miRNAs	493,034	124,590
# Known miRNA Families Identified	175	135





LOB Domain Protein 21 (LBD21)
(Glyma11g35500)

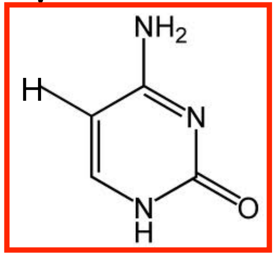
What Role Does Epigenetics Have in Establishing Compartment-Specific Gene Expression Patterns?



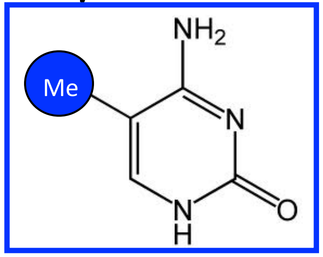
"Basic Principles of Genetics," http://cnx.org/content/m26565/1.1/

How Is DNA Methylation Studied Using Bisulfite Sequencing (BS-Seq)?

Unmethylated Cytosine

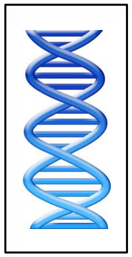


Methylated Cytosine



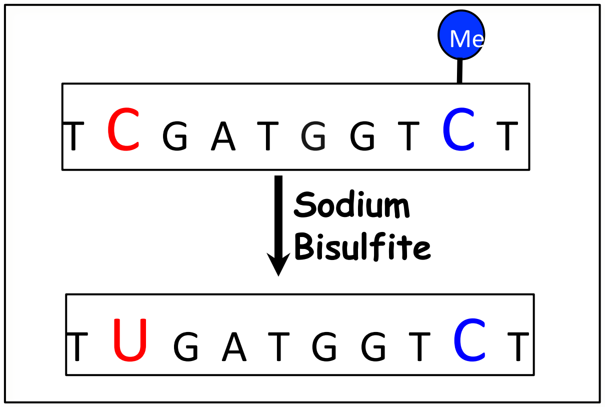
→

DNA Isolation



→

Bisulfite Conversion



Unmethylated Cytosines Converted to Uracil

PCR Amplification

T T G A T C G T C T

↓

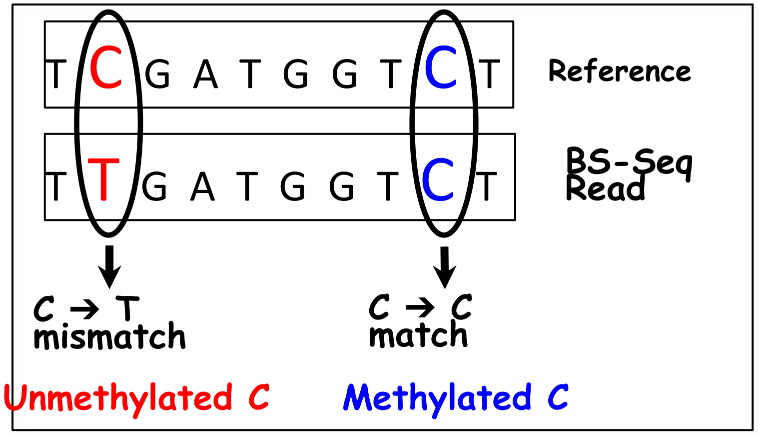
Uracil (U) Replaced by Thymine (T) During PCR Amplification

→

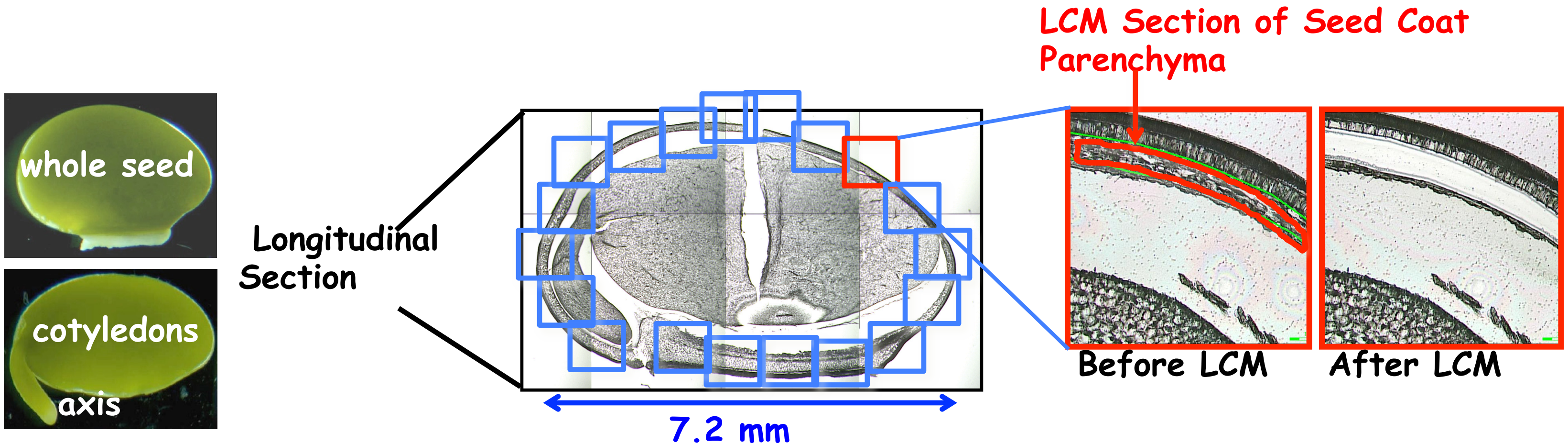
Illumina Sequencing

→

Align Sequences to Soybean Genome



How Much DNA Was Isolated From LCM Captured Early Maturation Stage Seed Coat Parenchyma?



Early Maturation Stage

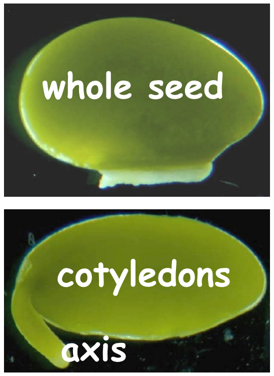
# Longitudinal sections (10μm)	190
Avg. # LCM section per longitudinal section	20
Total # of LCM sections captured	3,800
Amount of DNA obtained (ng)	~445

~445 ng DNA obtained from early-maturation seed coat parenchyma layer for bisulfite sequencing

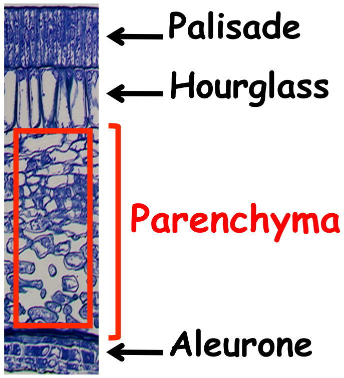
How Many BS-Seq Sequences Do We Have To Date?

Globular Stage
Whole Seed

Early Maturation
Whole Seed



Early Maturation
Seed Coat Parenchyma

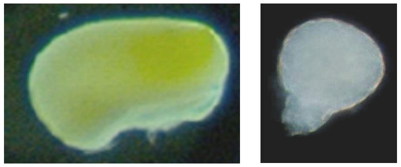


Library	Globular Stage Whole Seeds	Early Maturation Stage Whole Seeds	Early Maturation Seed Coat Parenchyma
# Reads	102M (8.4Gb)	77M (6.3Gb)	154M (12.1Gb)
# Unique Reads (i.e. non-clonal reads)	58M (4.4Gb)	48M (3.7Gb)	81M (6.1Gb)
# Aligned Unique Reads	37M (3.4Gb)	26M (2.5Gb)	24M (2.2Gb)
Coverage of Each Strand	1.7 fold	1.25 fold	1.1 fold

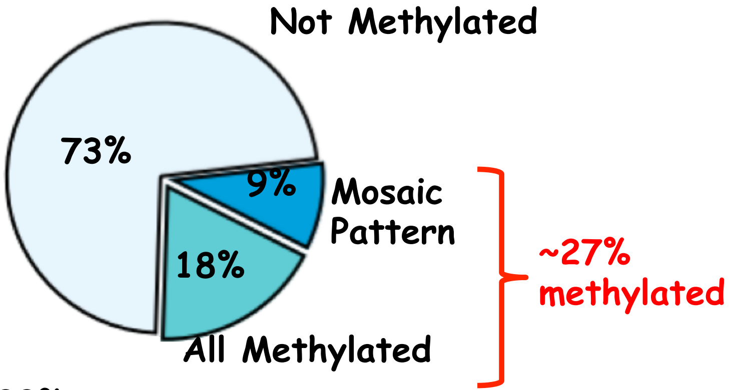
Total Sequences ~ 27X Soybean Genome

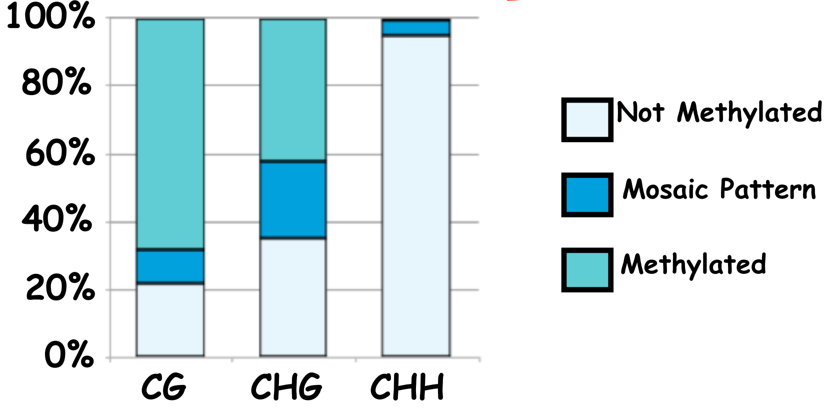
BS-DNA Sequencing of Globular- and Early Maturation-Stage Whole Seed

Globular Stage

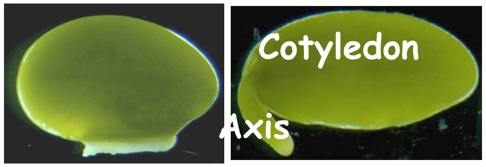


Uniquely Mapped Reads (Gigabases)	Coverage Per Strand
37M (3.4Gb)	1.7 fold

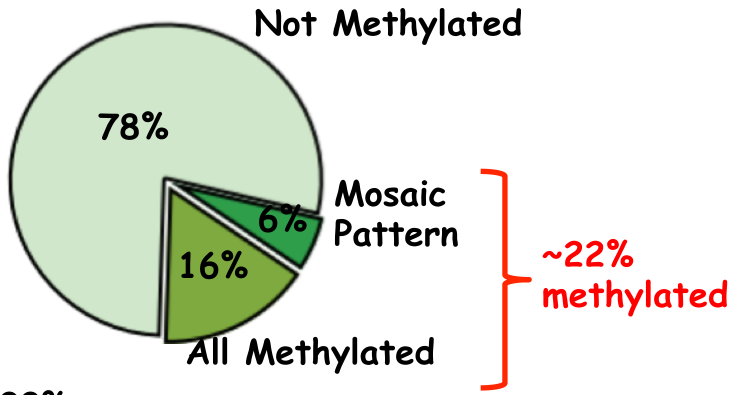


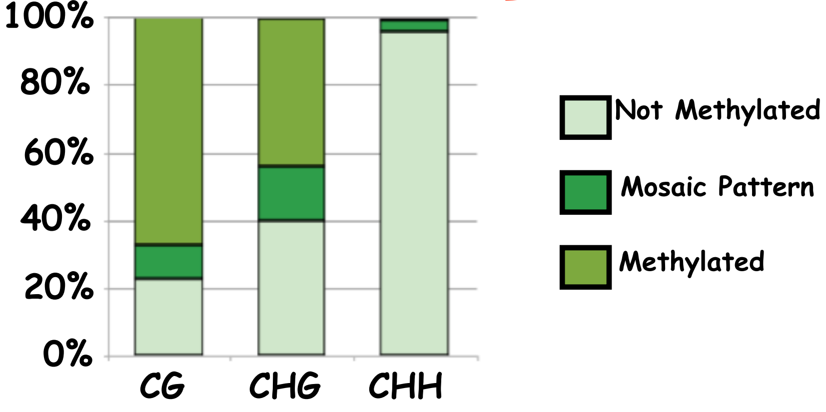


Early Maturation Whole Seed

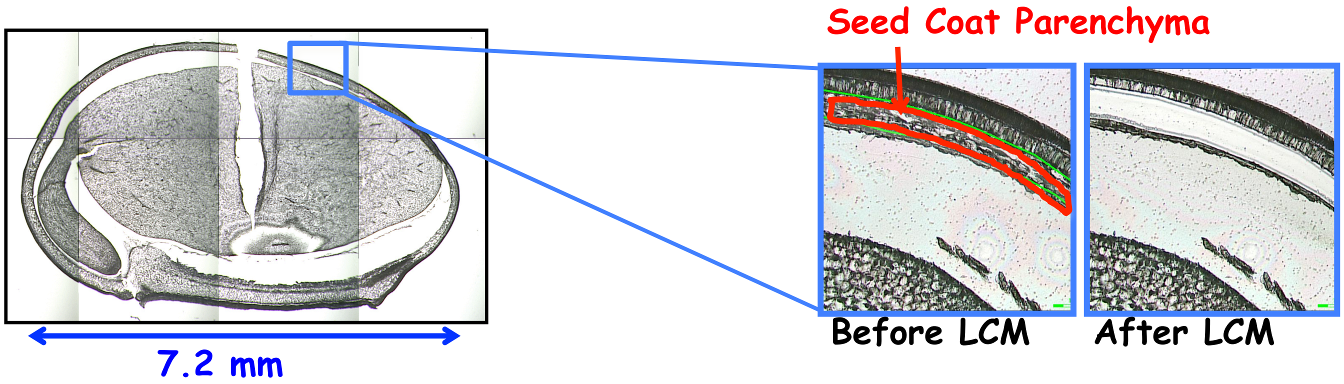


Uniquely Mapped Reads (Gigabases)	Coverage Per Strand
26M (2.5Gb)	1.25 fold

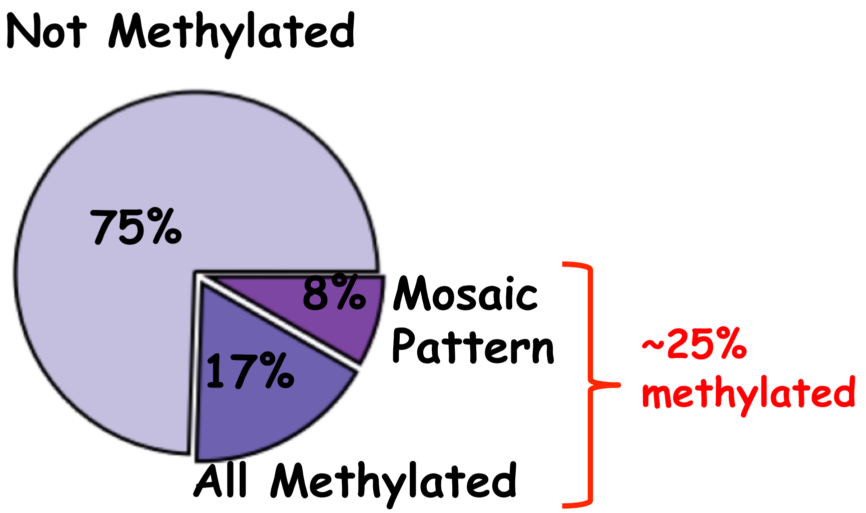


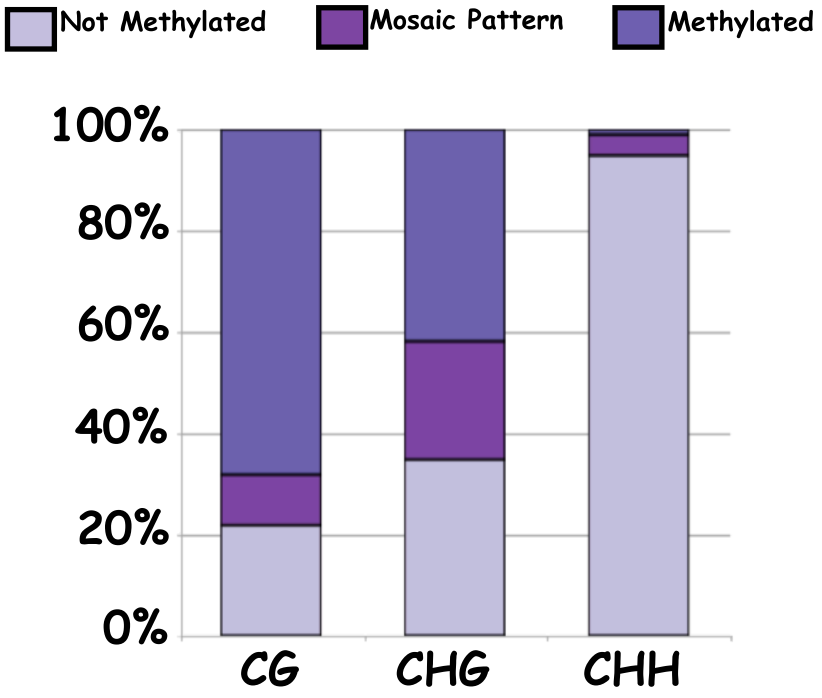


BS-DNA Sequencing of Early Maturation Seed Coat Parenchyma Layer (EM SCPY)

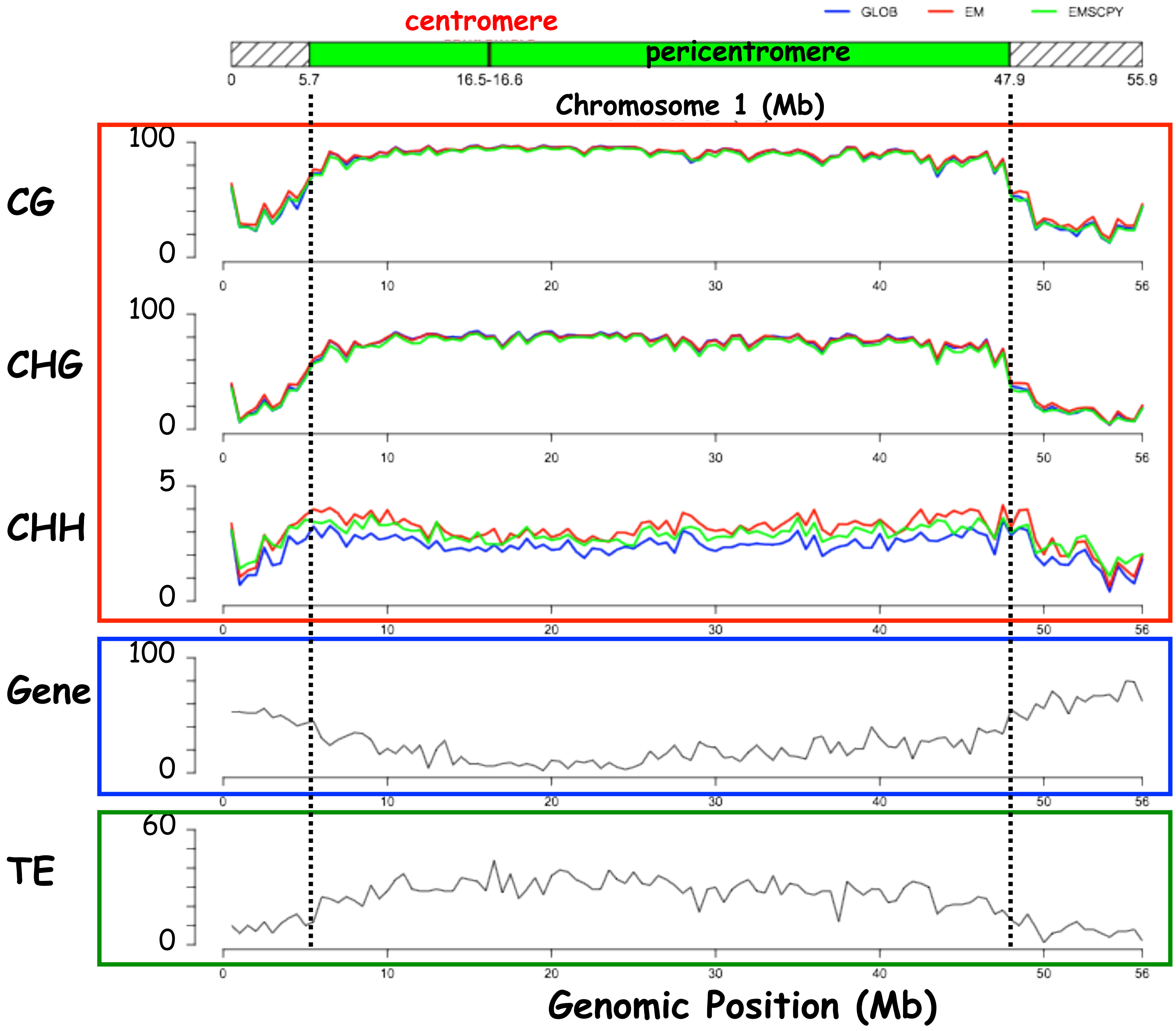


Aligned Unique Reads (Gigabases)	Coverage per Strand
24M (2.2Gb)	1.1 fold





What is The Methylation Pattern Along the Genome For Whole Seeds and Seed Compartment Layers?

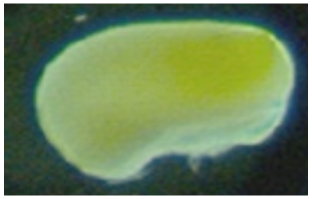


Methylation Level is Higher Around the Centromere and Pericentromere Regions!

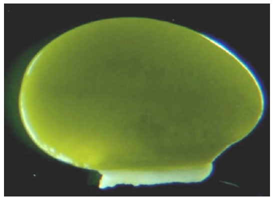
Methylation Level is Inversely Related to Gene Density!

Methylation Level is Related to TE Density!

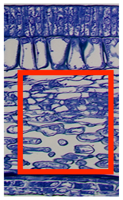
CG Methylation Distribution Across 46,367 High Confidence Genes



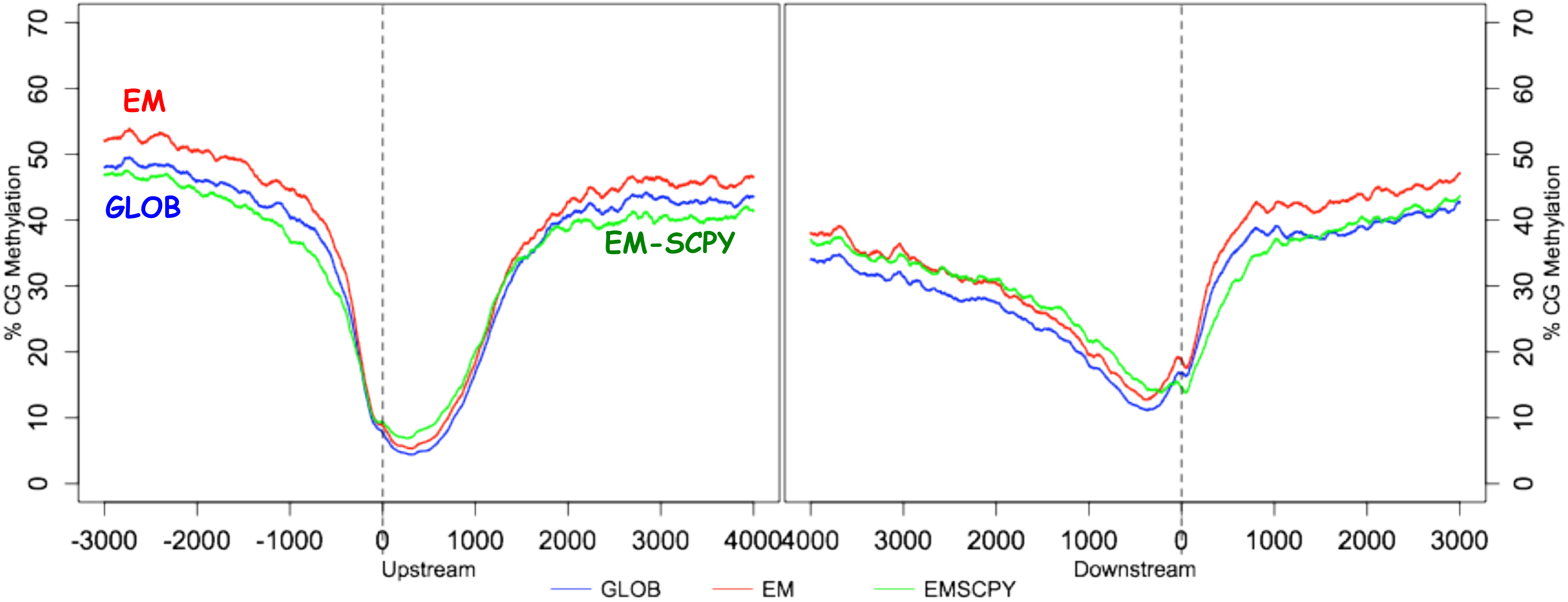
Globular Seed
(GLOB-WS)

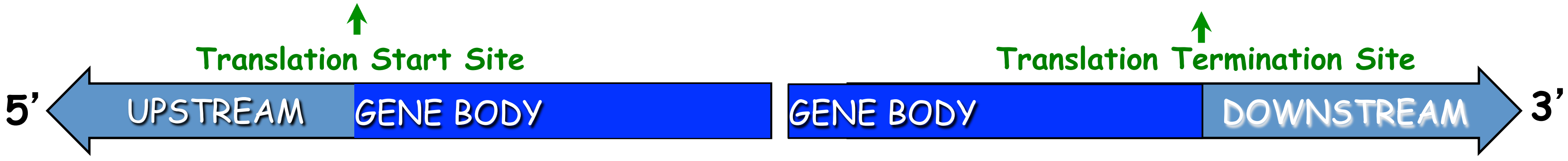


Early Maturation Seed
(EM-WS)



Early Maturation Seed
Coat Parenchyma
(EM-SCPY)

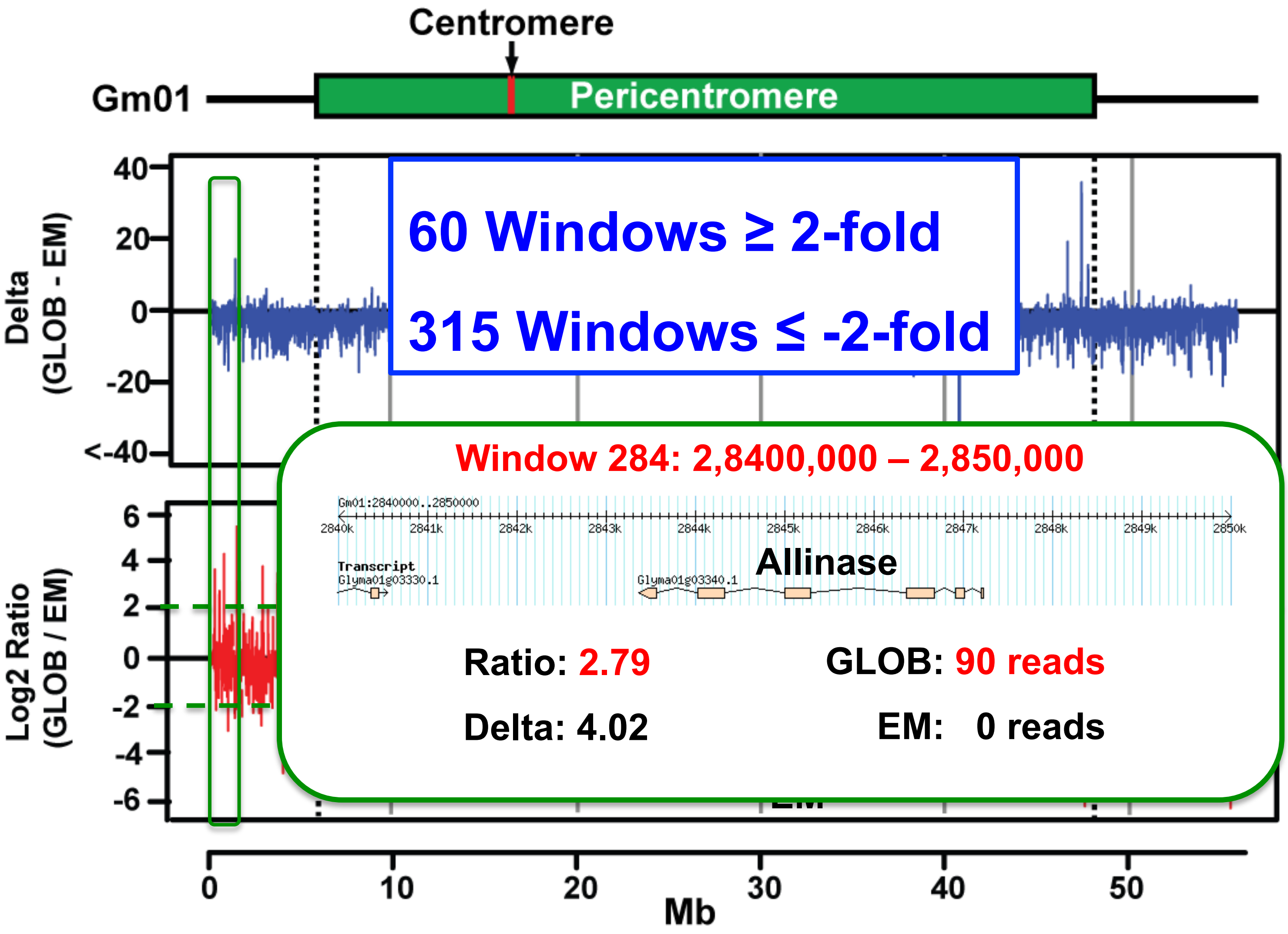




Translation Start Site

Translation Termination Site

Are There Differences in Methylation Level Between the Globular Stage and Early Maturation Stage Seed at the Genome Level ?



60 Windows ≥ 2-fold

315 Windows ≤ -2-fold

Window 284: 2,8400,000 – 2,850,000

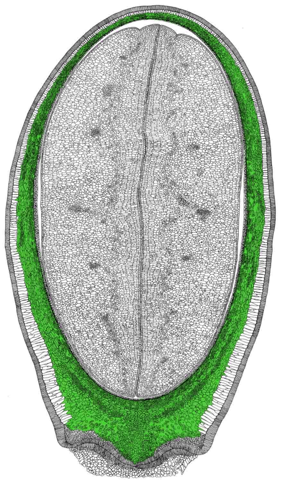
Ratio: 2.79

Delta: 4.02

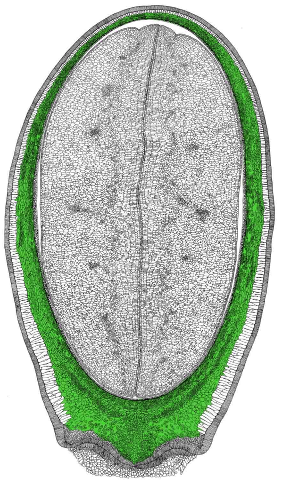
GLOB: 90 reads

EM: 0 reads

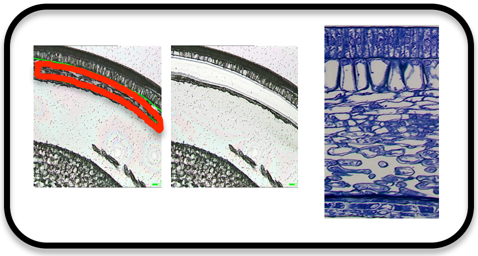
10 kb Windows

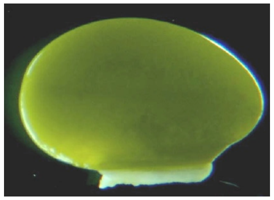


Genes In Different Seed Compartments Can Have Distinct Methylation Patterns

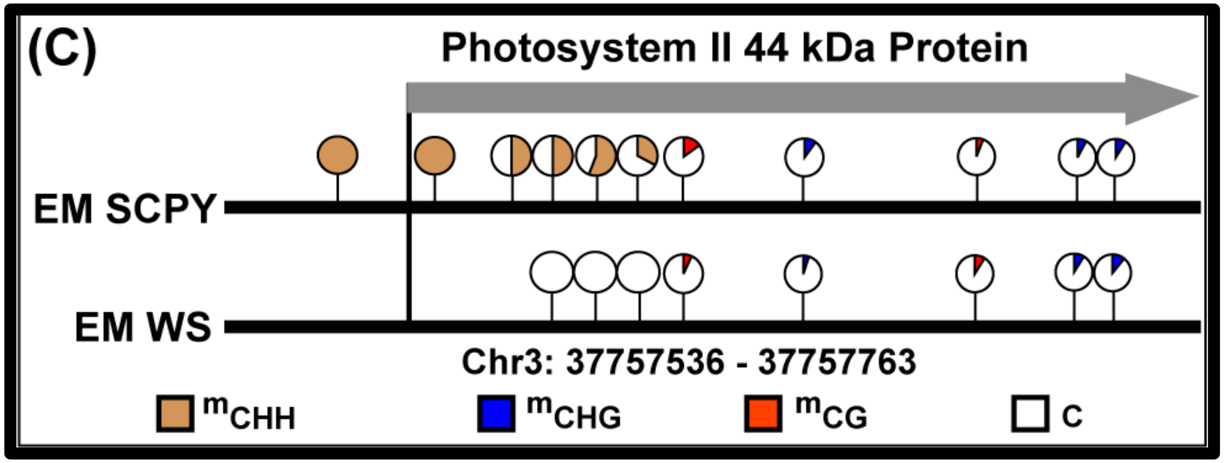


EM SCPY



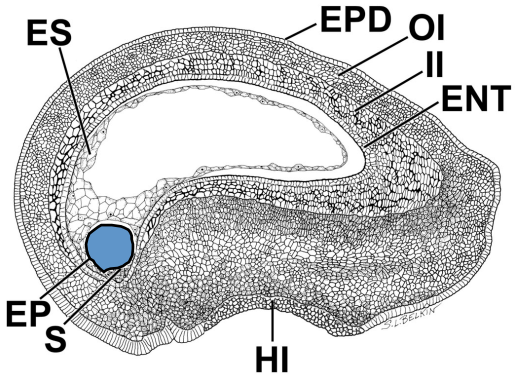


EM Whole Seed (WS)

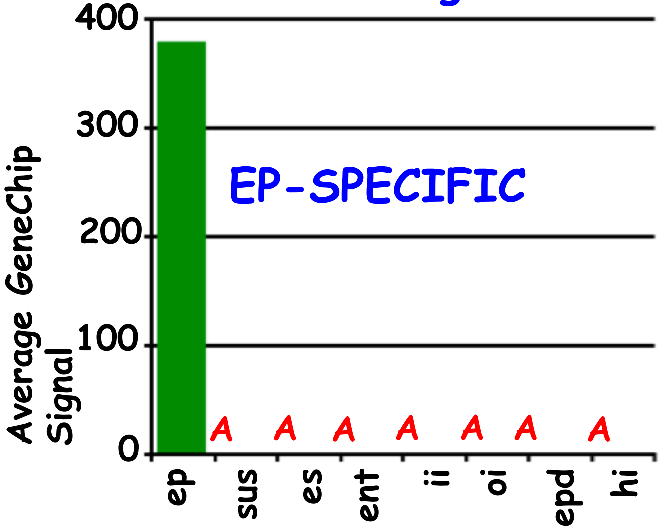


What Are The Functions of Compartment-Specific Transcription Factor mRNAs?

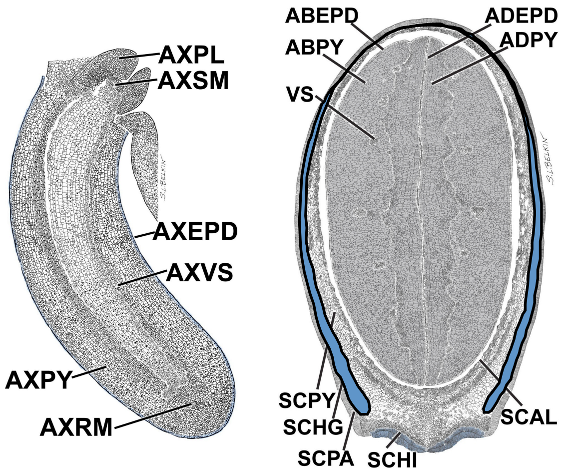
GmSpeechless TF
(Glyma04g41710)



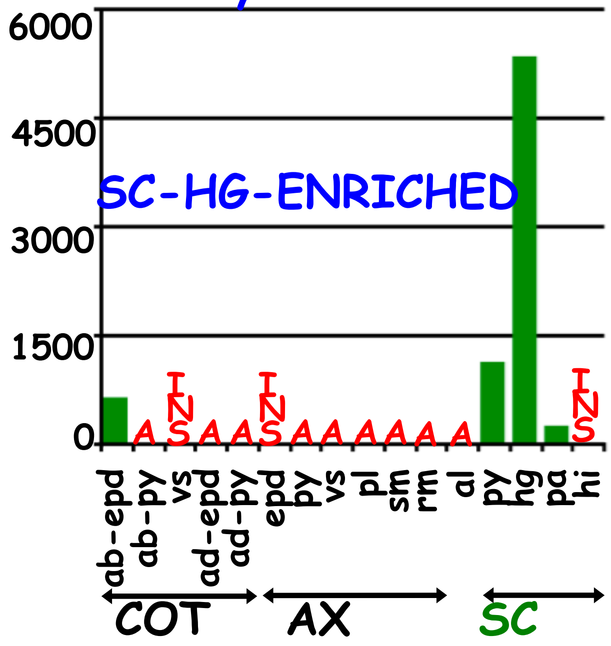
Globular Stage



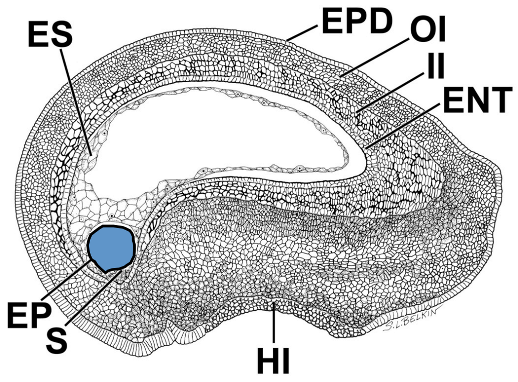
GRAS TF
(Glyma01g38360)



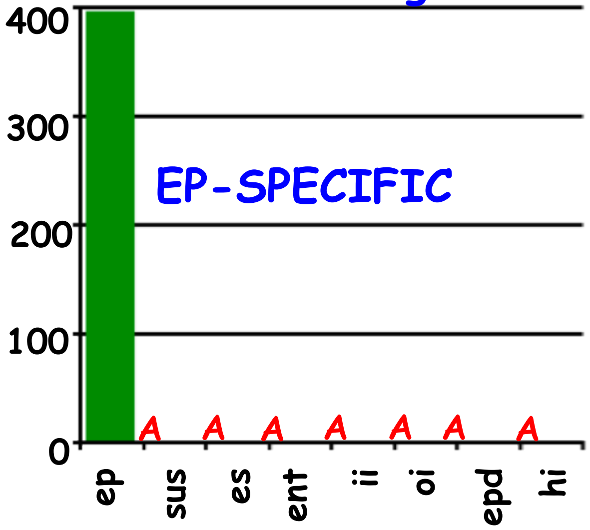
Early Maturation



SCARECROW TF
(Glyma09g40620/Glyma18g45529)



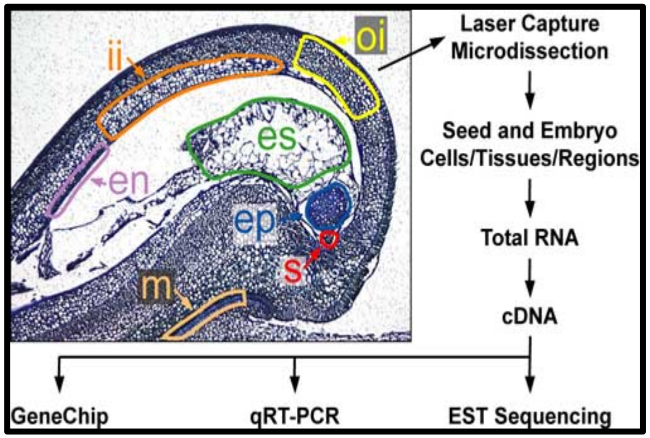
Globular Stage



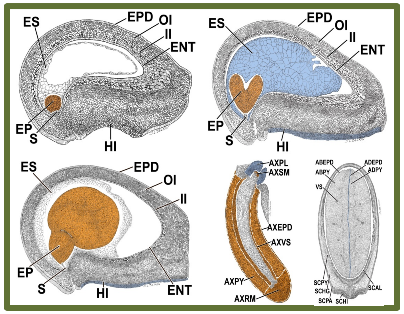
RNAi Gene Suppression Strategy

UCLA

Analysis



Identify Seed Specific TFs



Validation (qRT-PCR)

Gene Nomination

Select Promoter

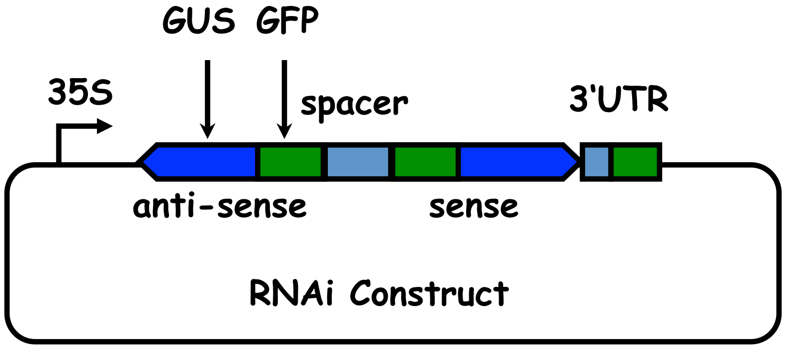
Develop RNAi Constructs

Transgenic Soybean Pipeline

Generate Transgenic Plants and Characterize RNAi Phenotypes

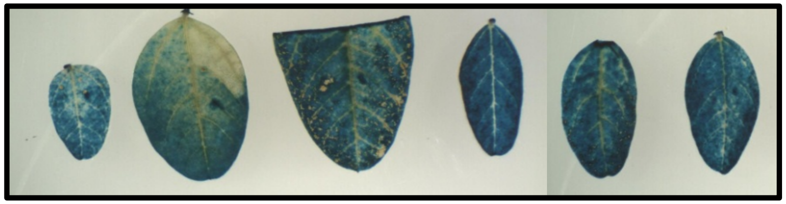
Monsanto

Using RNAi to Knock-Down Seed Compartment-Specific Transcription Factor mRNAs



35S-GUS

35S-GUS



35S-GUS; 35S-RNAi (GUS)

35S-RNAi (GUS) suppresses the 35S-Gus reporter in stably transformed shoot leaves.

Sampling and Quality Control

Transformation Pipeline

R0 – single copy plants

R1 seed population

RNA isolation for qRT-PCR

Fix tissue for histology

Phenotyping

R1 homozygous


Seed Chipping (Genotyping)

R2 seed population

Harvest RNA isolation for qRT-PCR

Fix tissue for histology

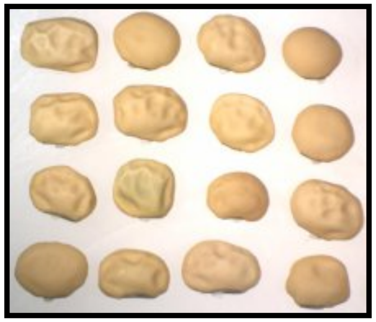
Phenotyping

Project Metrics	
Constructs	
Promoters:GUS Reporters	23
RNAi Suppression Constructs	70
Transgenic Material	
Transgenic Promoter:GUS	230
Transgenic RNAi Plants	1,716
Tissue Sampling	
Tissue samples for qRT-PCR	20,592
Samples for histology	3,432

Knocking Out Seed Transcription Factor mRNAs

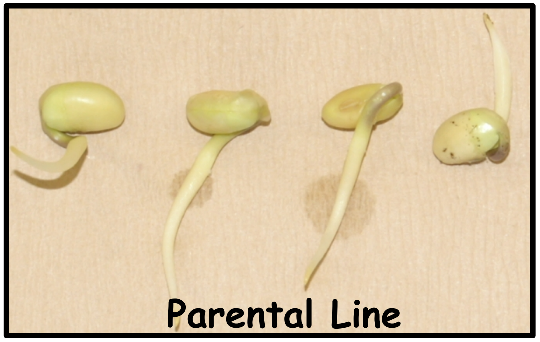
GmFIE
(Glyma10g02690)

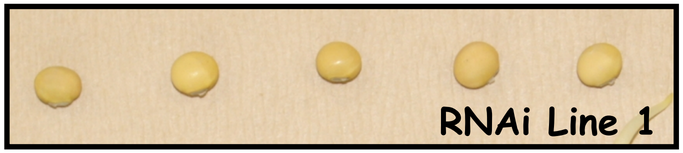

Parental Line

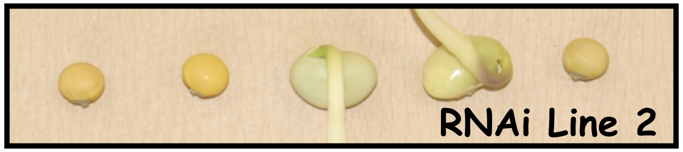

RNAi Line 1


RNAi Line 2

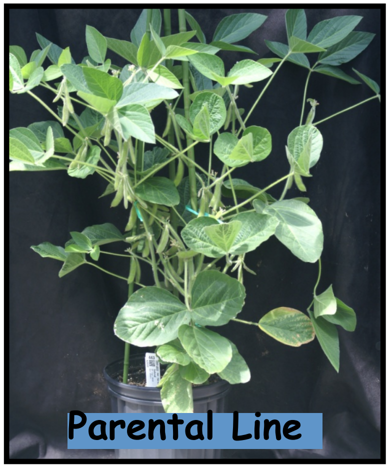
GRAS TF
(Glyma01g38360)
SC-HG-ENRICHED


Parental Line

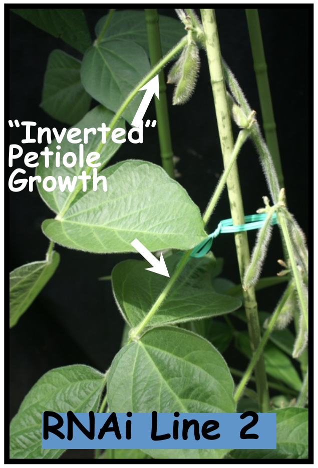

RNAi Line 1


RNAi Line 2

SCARECROW TF
(Glyma09g40620/Glyma18g45520)

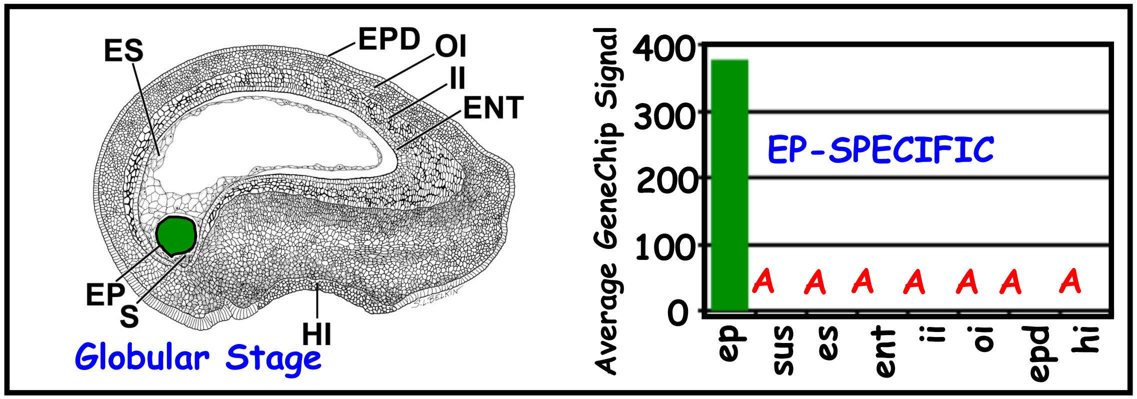

Parental Line

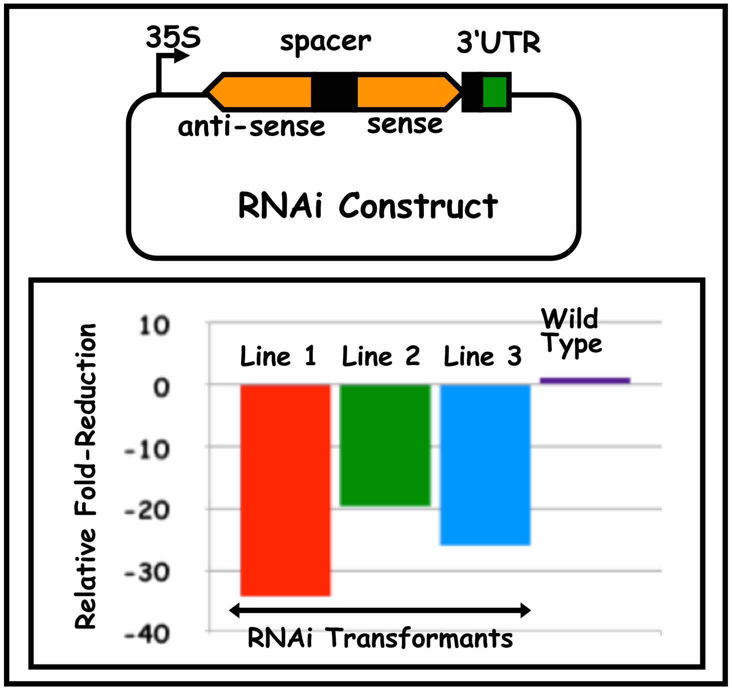

RNAi Line 1

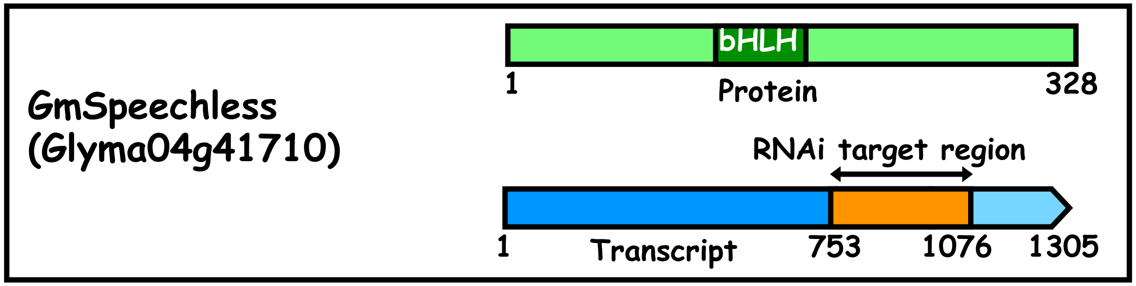


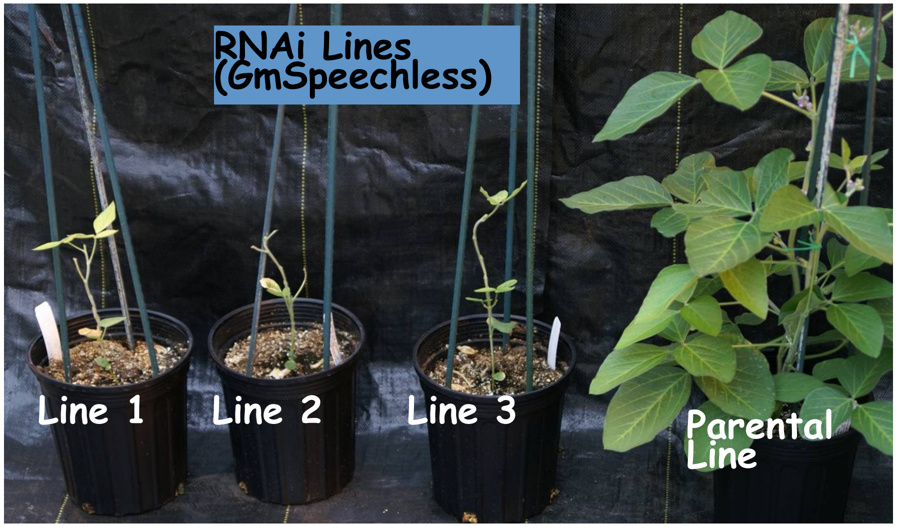
RNAi Line 2

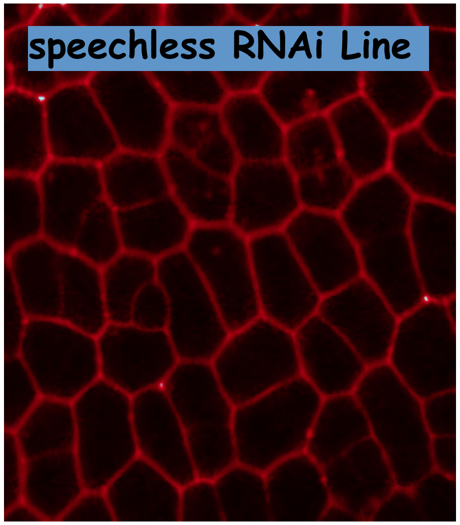
Soybean Speechless-Like TF mRNA Prevents Stomata Formation on Developing Cotyledons

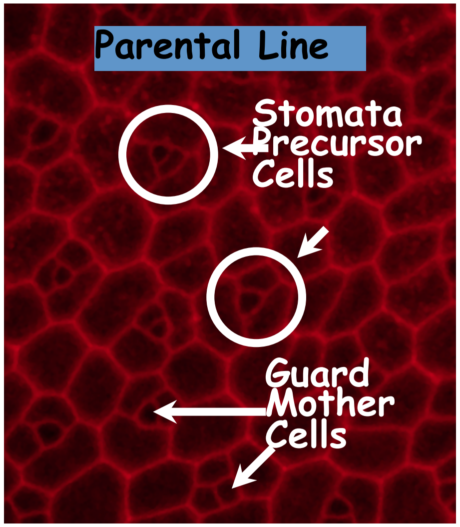




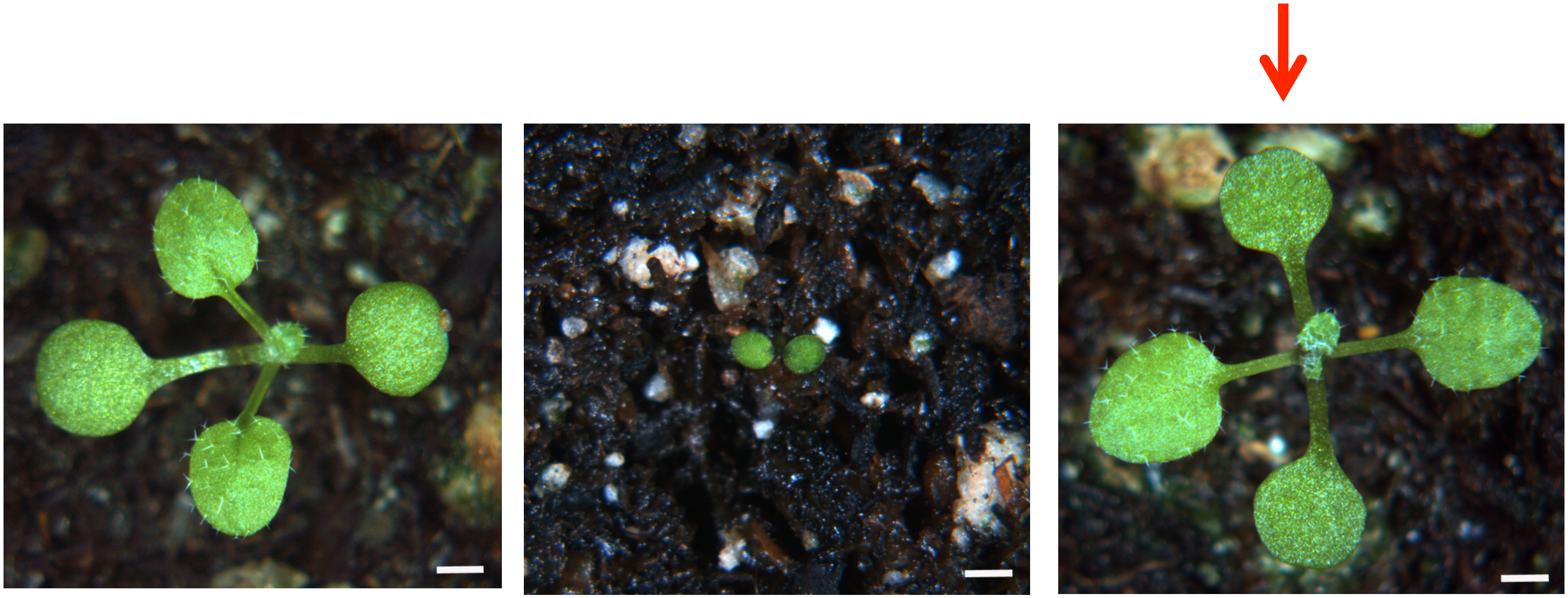








Will the GmSpeechless cDNA rescue the Arabidopsis Speechless Mutant?

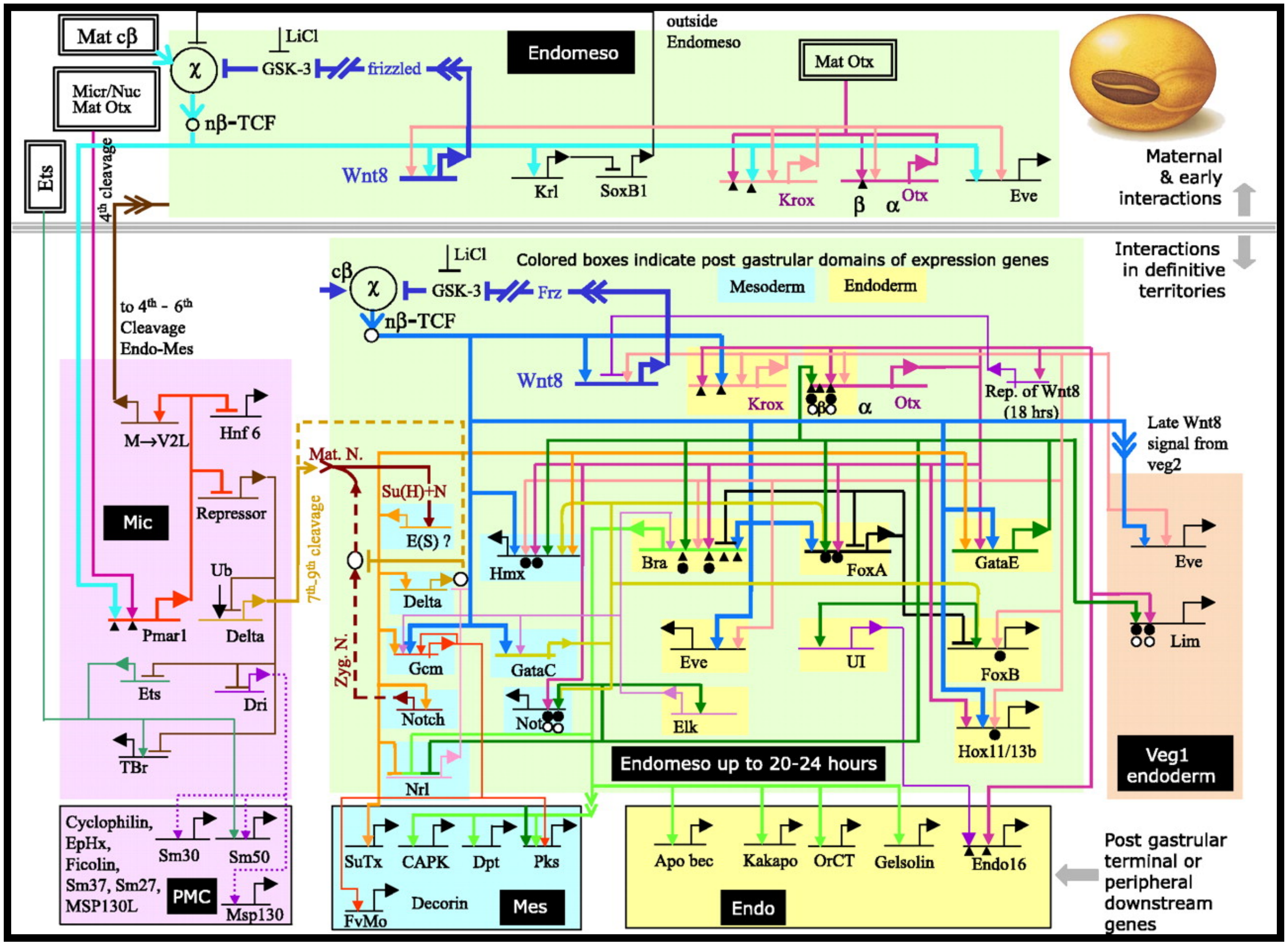

wild type seedling

spch3 mutant

AtSPCH::GmSPCH cDNA
(spch3 background)

15 days after germination

What Are the Regulatory Circuits Controlling Soybean Seed Development?



Come Back in Five Years!

Eric Davidson et al. Science, 2007

The End…………..

…….or is it the Beginning?

Current Lab Members

Bob Goldberg
Brandon Le
Min Chen
Jungim Hur
Kelli Henry
Weihong Yang
John Danzer
Jer-Young Lin
Ann Amores
Jennifer Kwan

Former Lab Members

Anhthu Bui
Javier Wagmaister
Shundai Li
Xingjun Wang
Harry Hahn
Tomo Kawashima
Chen Cheng

Monsanto Collaborators

Dave Somers

UC Davis Collaborators

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

UC Berkeley Collaborators

Bob Fischer
Tzung-Fu Hsieh

UCLA Collaborators

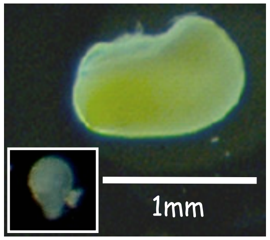
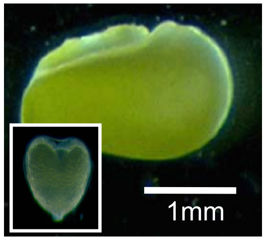
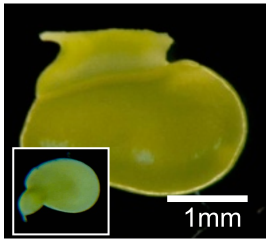
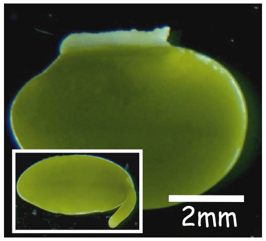
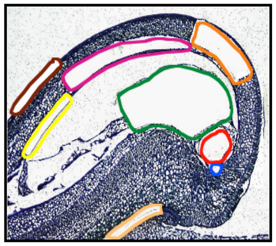
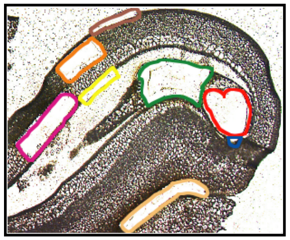
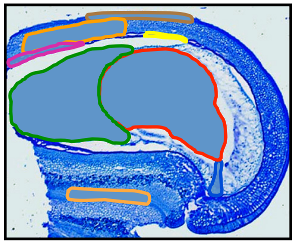
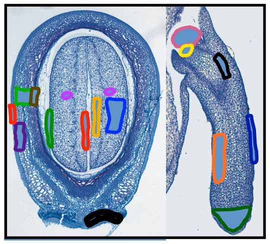
Matteo Pelligrini

National Science Foundation
WHERE DISCOVERIES BEGIN

Funded By NSF Plant Genome Grants To Bob Goldberg & John Harada

What are the Genes Active in Every Tissue, Cell Type, and Compartment During All of Soybean Seed Development?

Gene Models

	Globular		Heart		Cotyledon		Early Maturation		UNION
WHOLE SEED		+		+		+			
GeneChip	17,057 (1,748)		16,938 (1,715)		17,213 (1,725)		16,906 (1,680)		19,098 (1,968)
RNASeq	45,821 (4,717)		45,331 (4,743)		45,227 (4,680)		45,396 (4,712)		51,058 (5,199)
LCM SEED		+		+		+		=	In Progress
GeneChip	18,092 (1,904)		18,564 (1,929)		18,939 (1,985)		19,621 (2,047)		20,874 (2,213)
RNASeq	33,367* (3,361)		In Progress		In Progress		34,521** (3,404)		39,013 (3,943)

Most genes are shared in different seed compartments throughout seed development

(Transcription Factors)

* EP & S - only

** SCPY-only