Using Genomics to Dissect Seed Development

BOB GOLDBERG
8/4/08

Today's Headlines

The New York Times

Los Angeles Times

A Global Need for Grain That Farms Can't Fill

Published: March 9, 2008

Economist.com

High Rice Cost Creating Fears of Asia Unrest

By KEITH BRADSHER
Published: March 29, 2008

U.S.News & WORLD REPORT

CNN.com

THE FOOD CHAIN

A Drought in Australia, a Global Shortage of Rice

Across Globe, Empty Bellies Bring Rising Anger

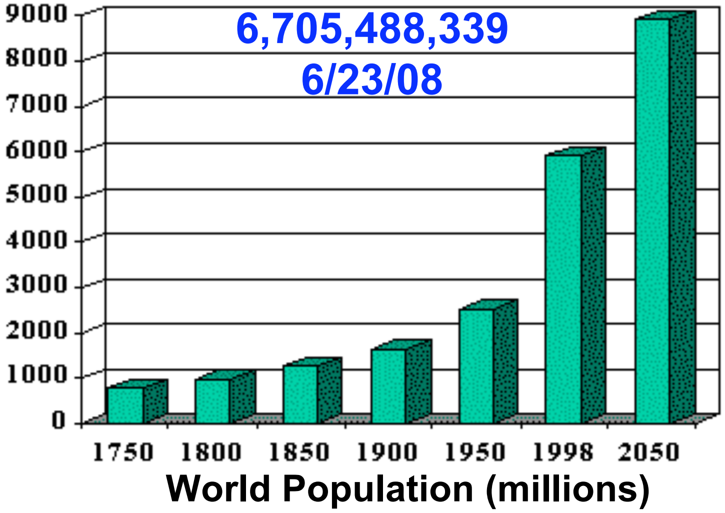
Newsweek

updated 10:42 p.m. EDT, Mon April 14, 2008

Riots, instability spread as food prices skyrocket

The Washington Post

We Face Major Challenges In Agriculture Even Greater Than Those in Today's Headlines



OVER THE NEXT 50 YEARS WE WILL NEED TO PRODUCE MORE FOOD THAN IN THE WHOLE OF HUMAN HISTORY

AND DO IT ON LESS ARABLE LAND!!!!

CROP YIELDS NEED TO BE INCREASED SIGNIFICANTLY

And……..There’s Also A Problem With Using Land For Energy Production………………

A Reminder……….Seeds Are Used in Many Ways as Food, Beverages, Spices, and Fuels!

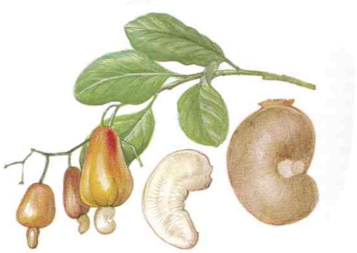
Beans

Peas

Wheat

Corn

Coconut



Cashew Nuts

Peanuts

Pecans

Cocoa Beans

Coffee Beans

Nutmeg

Mustard

Most Importantly….. Our Food is Derived From Fourteen 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*

In Some World Populations 75% of Calories Are Derived From Seeds!

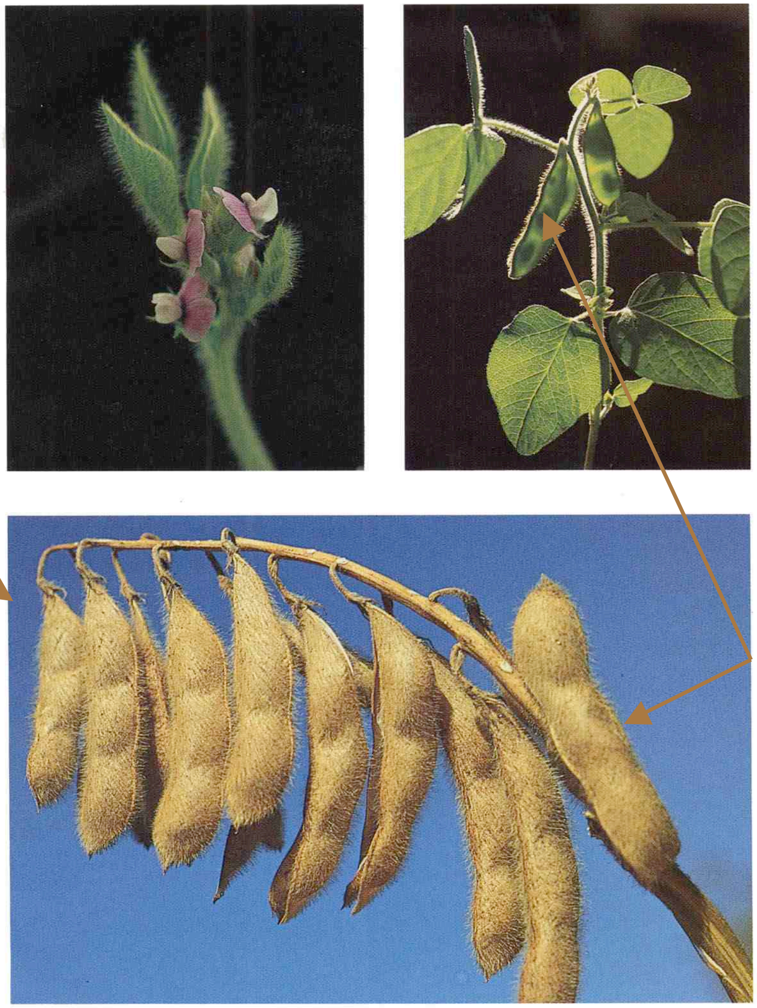
So….How Can Seed Yields Be Improved?
Use a Variety of Approaches To Identify Genes Critical For Fundamental Seed Processes *(<u>Yo!!-It's the Yield That Counts!</u>)*

Seeds Without Fertilization

Hybrids

Reduced Pod Shattering

Architecture Designed For Specific Growth Conditions



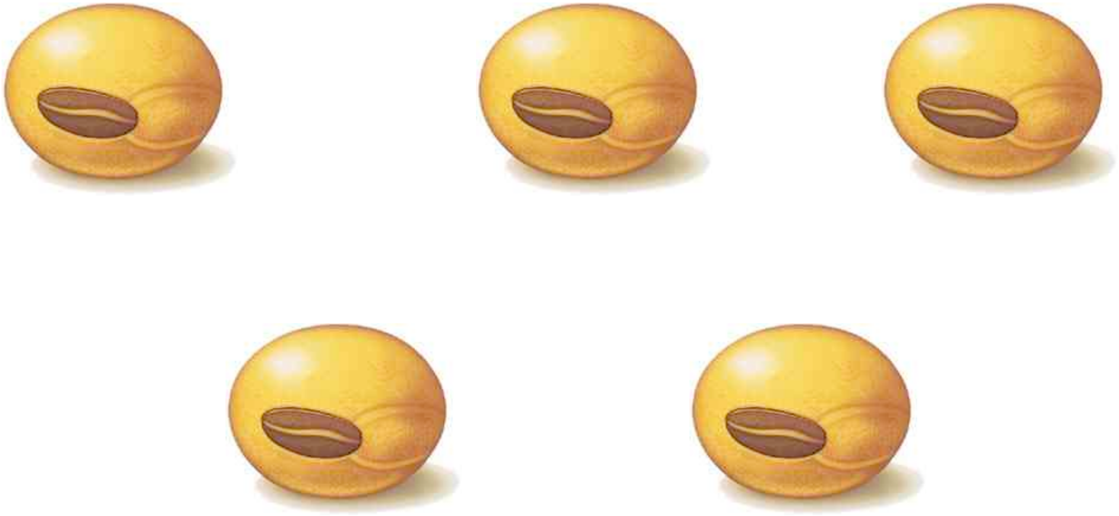
More Seeds

Bigger Seeds

Seeds Optimal For Human/Animal Health & Nutrition

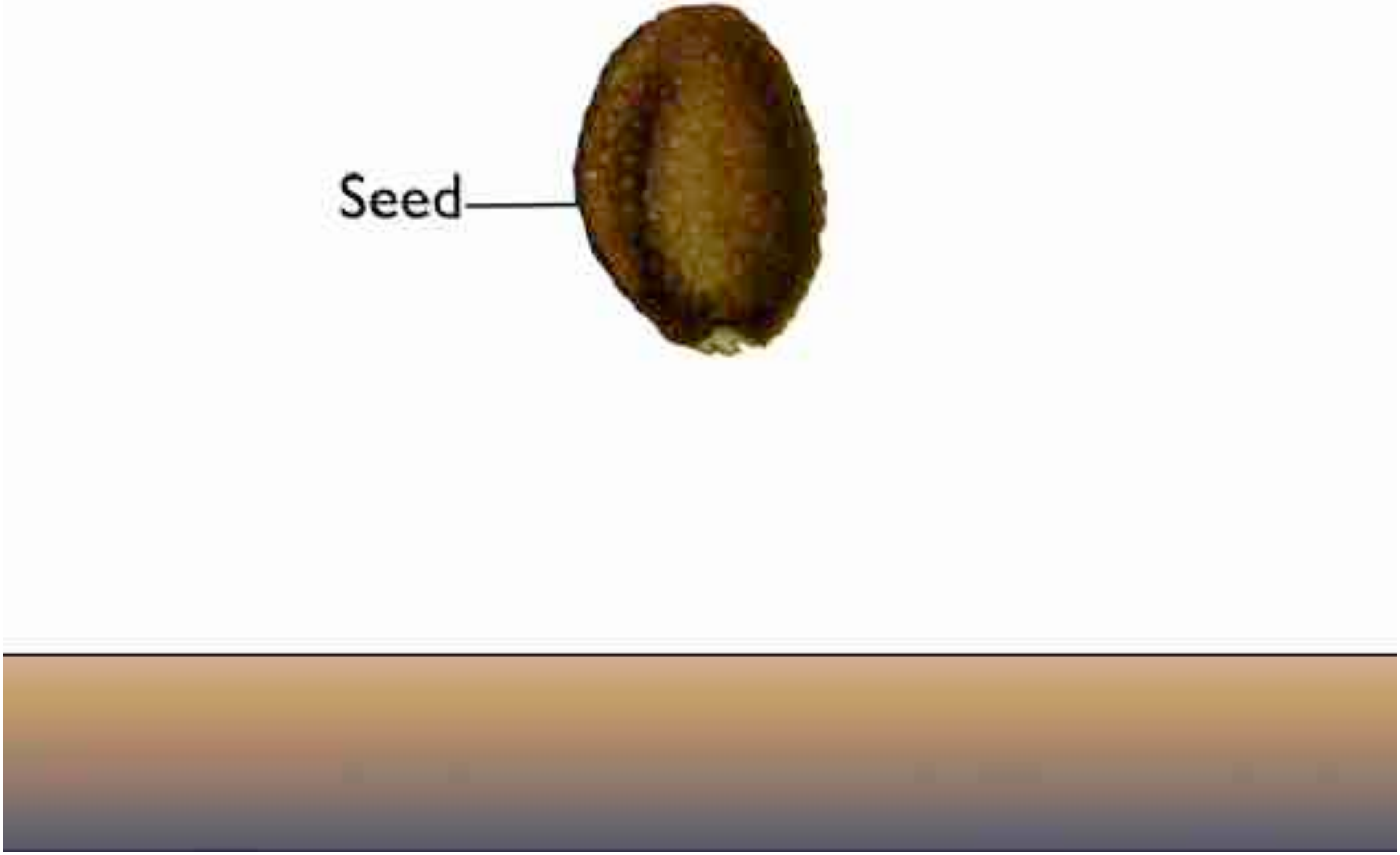
…..and → **Re-Engineer Existing Crops** → *Increase Yields!*

How Is a Seed Formed?

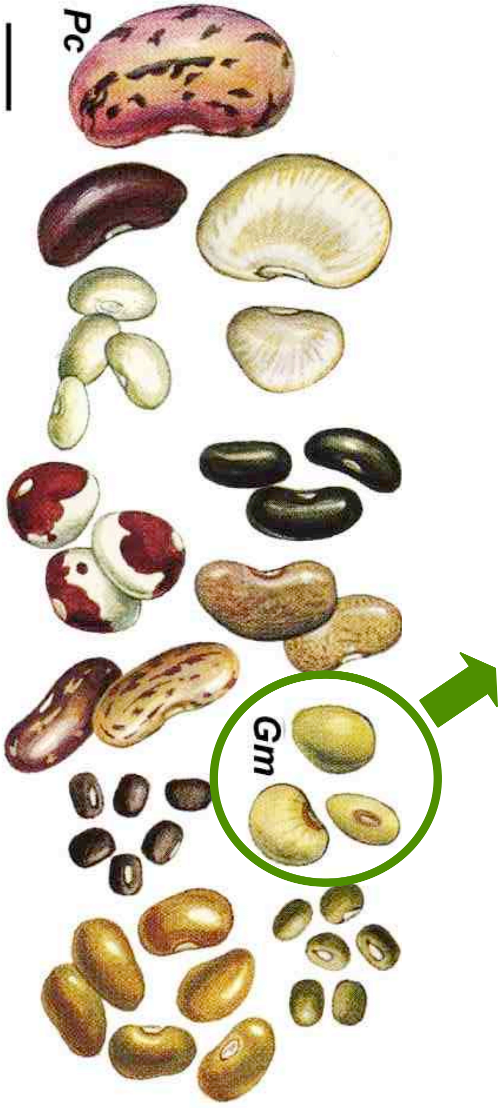


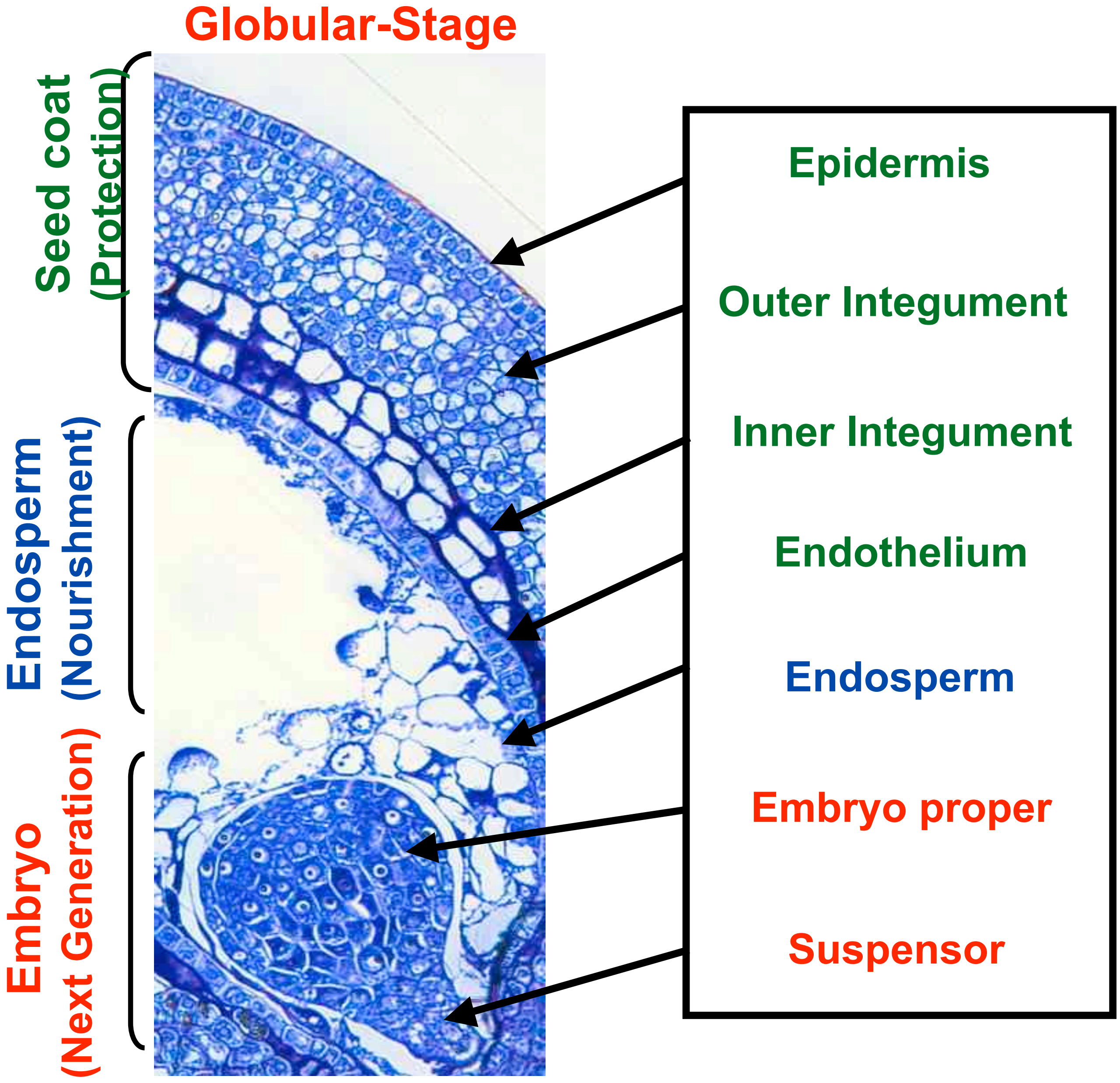
A Reminder……………………………….

In the Beginning....



What Are the Genes Required to Make a Seed?





Globular-Stage

Seed coat (Protection)

Endosperm (Nourishment)

Embryo (Next Generation)

Epidermis

Outer Integument

Inner Integument

Endothelium

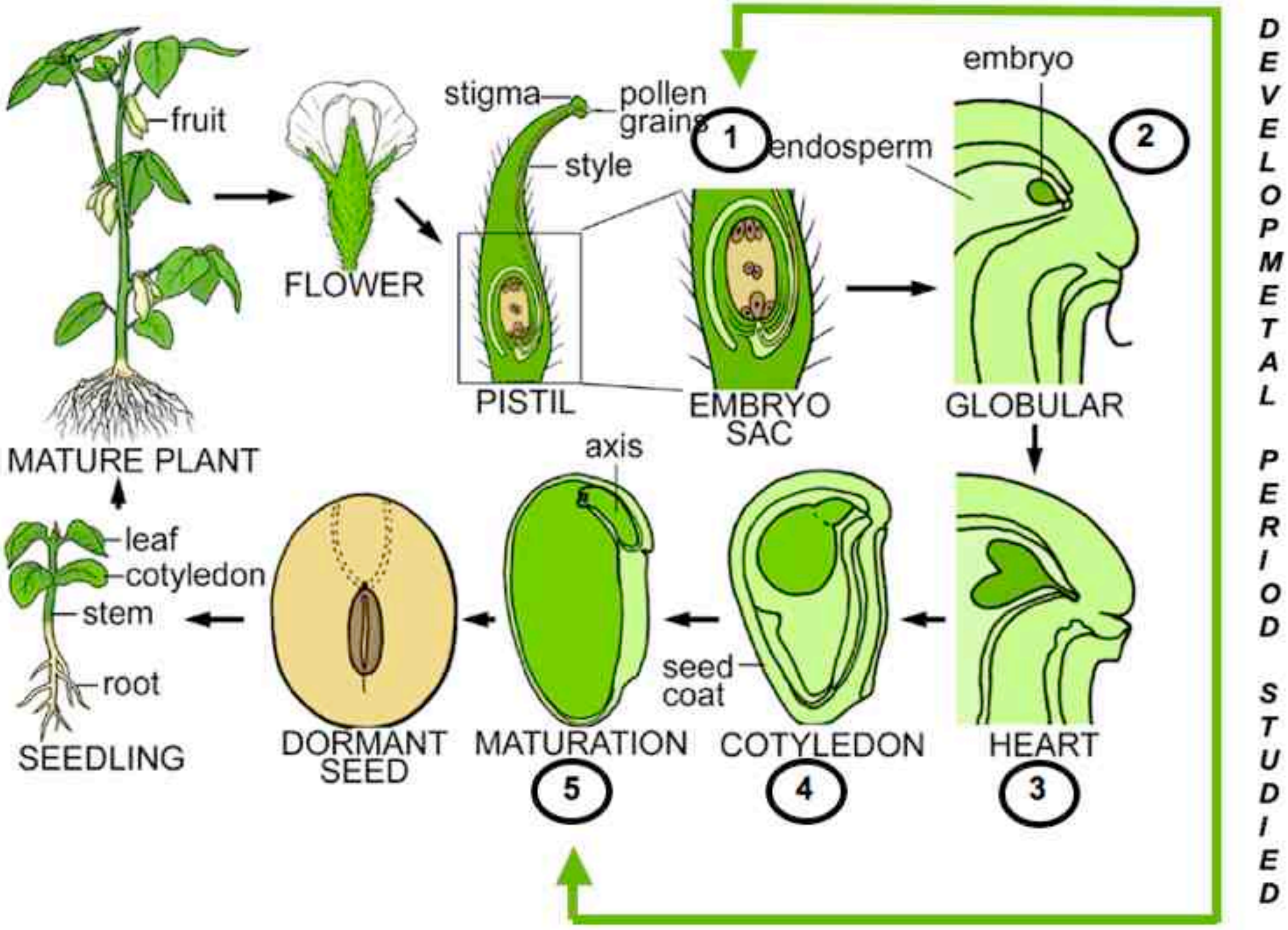
Endosperm

Embryo proper

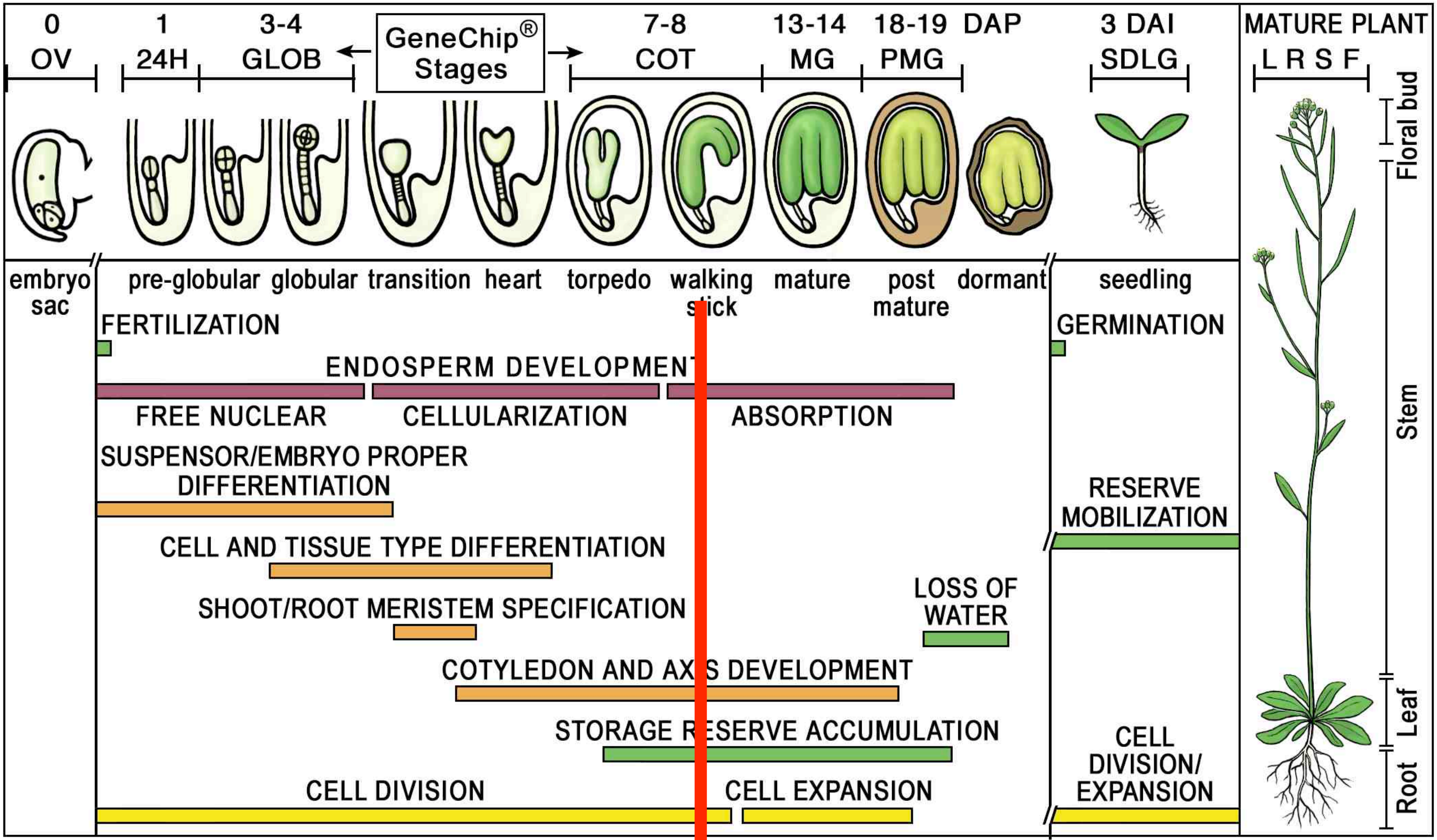
Suspensor

And How Are They Wired in a Plant Genome?

More Specifically………What Are The Genes Required to Program <u>Every</u> Compartment, Tissue, and Cell Type During Seed Development?



Genome-Wide Profiling of mRNAs During Arabidopsis Seed Development & Plant Life Cycle



0 OV

1 24H

3-4 GLOB

GeneChip® Stages

7-8 COT

13-14 MG

18-19 PMG

DAP

3 DAI SDLG

MATURE PLANT

L R S F

embryo sac

pre-globular

globular

transition

heart

torpedo

walking stick

mature

post mature

dormant

seedling

FERTILIZATION

ENDOSPERM DEVELOPMENT

FREE NUCLEAR

CELLULARIZATION

ABSORPTION

SUSPENSOR/EMBRYO PROPER DIFFERENTIATION

CELL AND TISSUE TYPE DIFFERENTIATION

SHOOT/ROOT MERISTEM SPECIFICATION

COTYLEDON AND AXIS DEVELOPMENT

STORAGE RESERVE ACCUMULATION

CELL DIVISION

CELL EXPANSION

LOSS OF WATER

GERMINATION

RESERVE MOBILIZATION

CELL DIVISION/ EXPANSION

Differentiation

Prepare For Dormancy & Germination

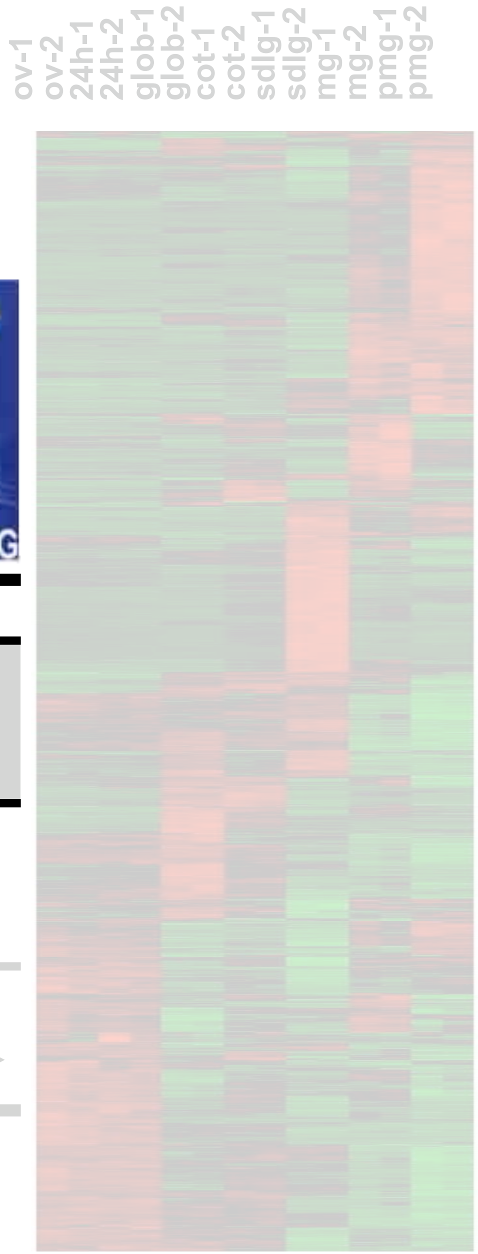
What Are the Genes That Are Important For Programming Specific Stages of Seed Development?

Gene Activity Before, During, And After Arabidopsis Seed Development

SEED DEVELOPMENT

	OV	24H	GLOB	COT	MG	PMG	SDLG
Total mRNAs	12,591	12,421	13,722	13,103	10,875	8,779	13,185
TF mRNAs	999	995	1,089	1,051	851	699	1,016
Unique mRNAs	22	16	100	50	26	31	505
Unique TFs	4	0	17	9	4	6	57
Shared mRNAs (TFs)	6,937 (477)	→	→	→	→	→	→

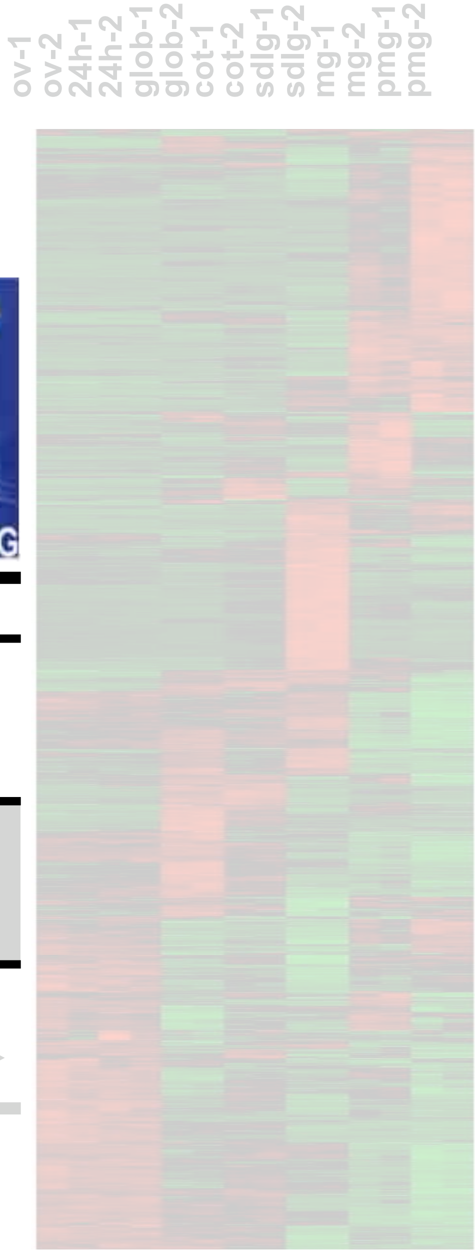
http://estdb.biology.ucla.edu/genechip



Gene Activity Before, During, And After Arabidopsis Seed Development

SEED DEVELOPMENT

	OV	24H	GLOB	COT	MG	PMG	SDLG
Total mRNAs	12,591	12,421	13,722	13,103	10,875	8,779	13,185
TF mRNAs	999	995	1,089	1,051	851	699	1,016
Unique mRNAs	**22**	**16**	**100**	**50**	**26**	**31**	**505**
Unique TFs	**4**	**0**	**17**	**9**	**4**	**6**	**57**
Shared mRNAs (TFs)	6,937 (477)	→	→	→	→	→	→



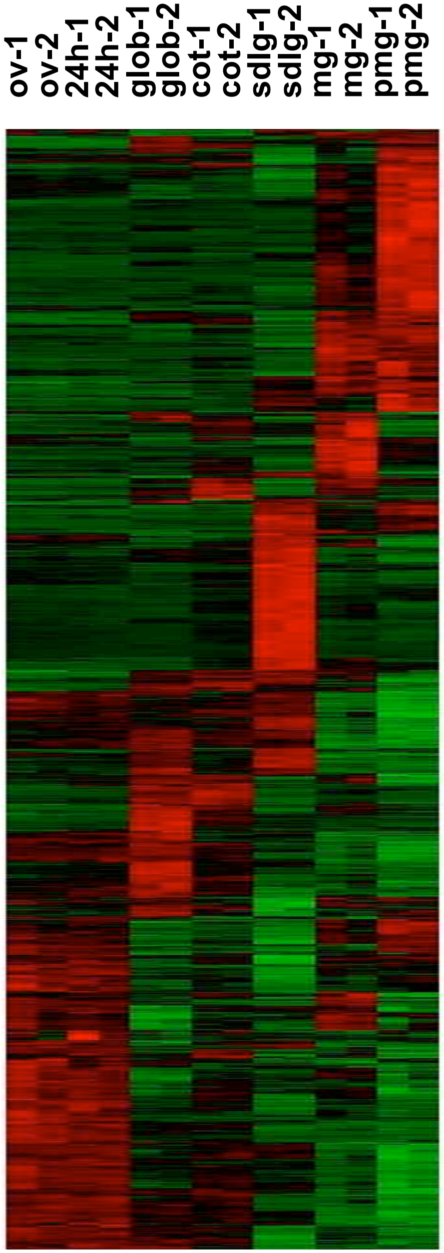
http://estdb.biology.ucla.edu/genechip

Gene Activity Before, During, And After Arabidopsis Seed Development

SEED DEVELOPMENT

	OV	24H	GLOB	COT	MG	PMG	SDLG
Total mRNAs	12,591	12,421	13,722	13,103	10,875	8,779	13,185
TF mRNAs	999	995	1,089	1,051	851	699	1,016
Unique mRNAs	22	16	100	50	26	31	505
Unique TFs	4	0	17	9	4	6	57
Shared mRNAs (TFs)	**6,937 (477)**						

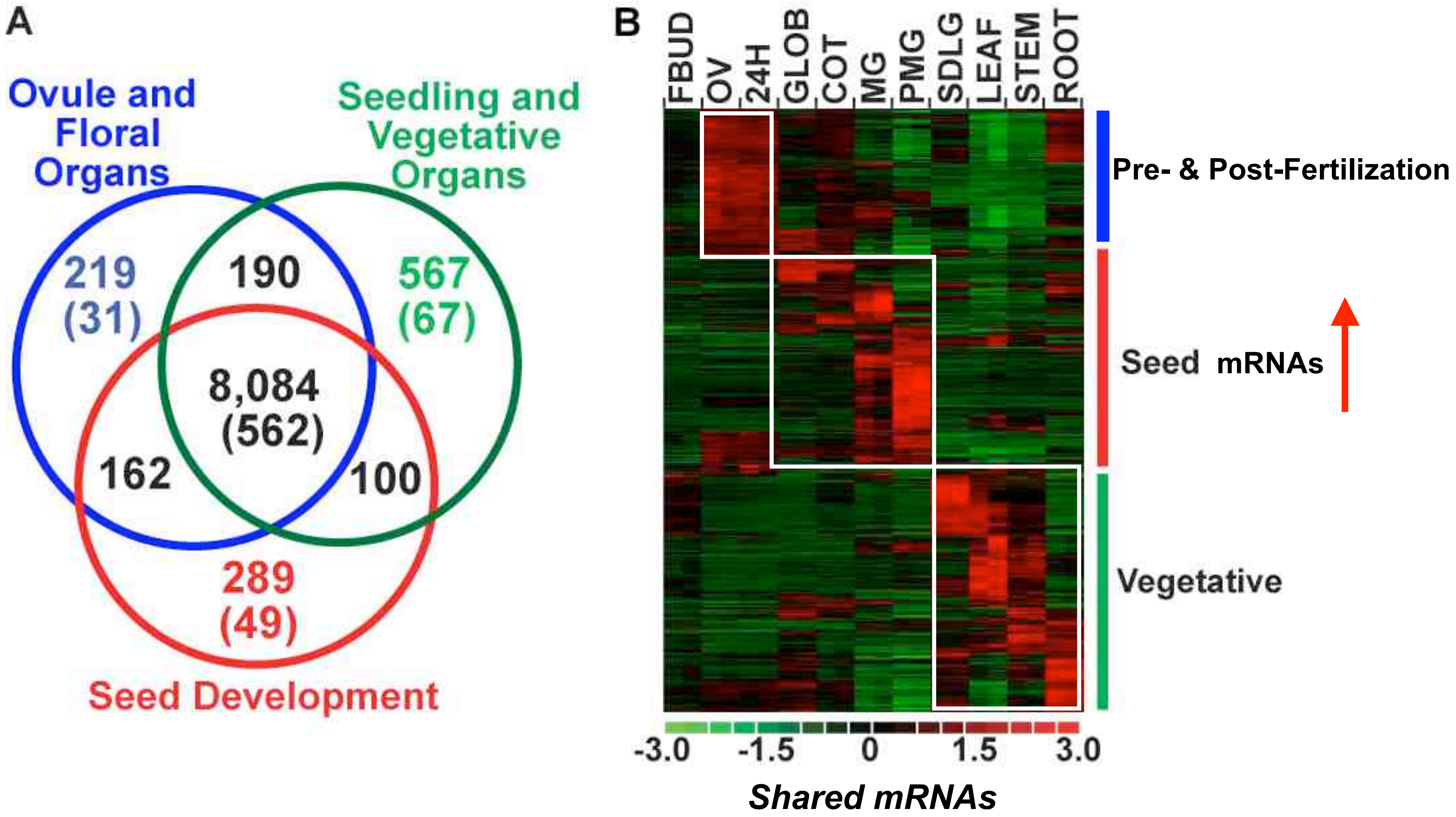
http://estdb.biology.ucla.edu/genechip



Shared mRNAs

Are There Seed-Specific Genes That May Play a Critical Role in Programming Seed Development?

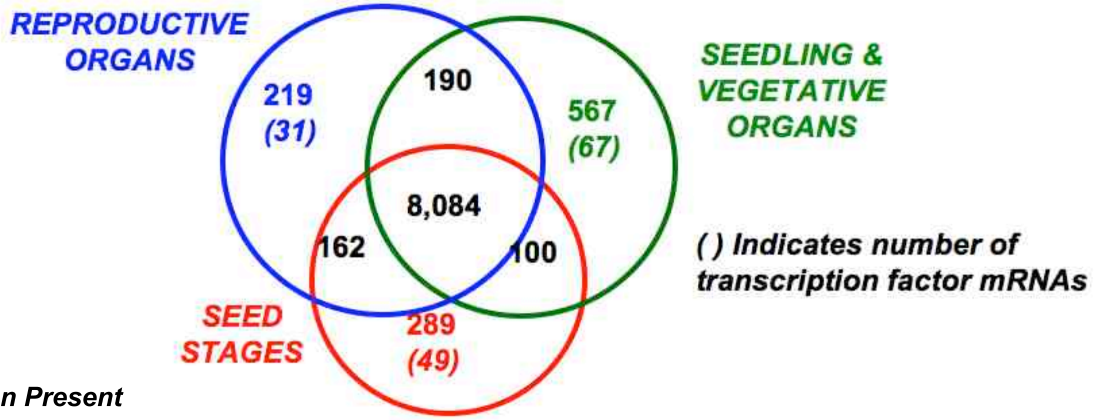
Identification of Seed-Specific mRNAs in the Arabidopsis Life Cycle



…at the GeneChip Level!!

( ) Indicates number of transcription factor mRNAs

Identification of "Seed-Specific" Transcription Factor mRNAs



		ABI3-VP1	
(G,C)	*	**LEC2**	*John Harada*
(G,C,M,PM)	*	**FUS3**	
(C,M)		**B3 Family TF**	
		AP2/EREBP	
(PM)	*	**AP2 Domain Protein**	
(C,M,PM)		**Aintegumenta-Like 7**	
		ARF	
(G)	*	**ARF21**	
		ARR-B	
(G,C)		**ARR19**	
(C)	*	**ARR21 / ARR13**	
(24H,G,C,M,PM)	*	**ARR22**	
		AS2	
(C,M)		**LOB Domain Protein 18**	
(G)		**LOB Domain Protein 35**	
		AUX-IAA	
(24H,G,C,M)		**IAA31**	
		Basic Leucine Zipper	
(C,M,PM)	*	**AtbZIP67; DPBF2**	
(C)	*	**AtbZIP72**	
(G,C)		**AtbZIP15**	
		CCAAT-Box	
(G,C)	*	**LEC1**	*John Harada*
(G,C,M)	*	**L1L**	
(G)		**HAP5A-Like**	
		Heat-Shock	
(G,C)	*	**HSF1-Like**	
(PM)	*	**Heat Shock TF**	

Stages When Present

		Homeobox
(G,C,M)		**Homeodomain Protein**
(G,C)	*	**ATML1-Like**
(G,C)	*	**Homeodomain Protein**
		MADS-Box
(M,PM)	*	**MADS-Box TF**
(M,PM)	*	**MADS-Box TF**
(G,C)		**AGL33**
(G)		**AGL35**
(G)		**AGL36**
(G)		**AGL45**
(G)		**AGL57**
(G)		**AGL91**
		MYB
(M,PM)		**AtMYB107**
(C)	*	**Myb-Related Protein**
(PM)	*	**AtMYB67; AtY53**
(G,C)	*	**Myb-Related Protein**

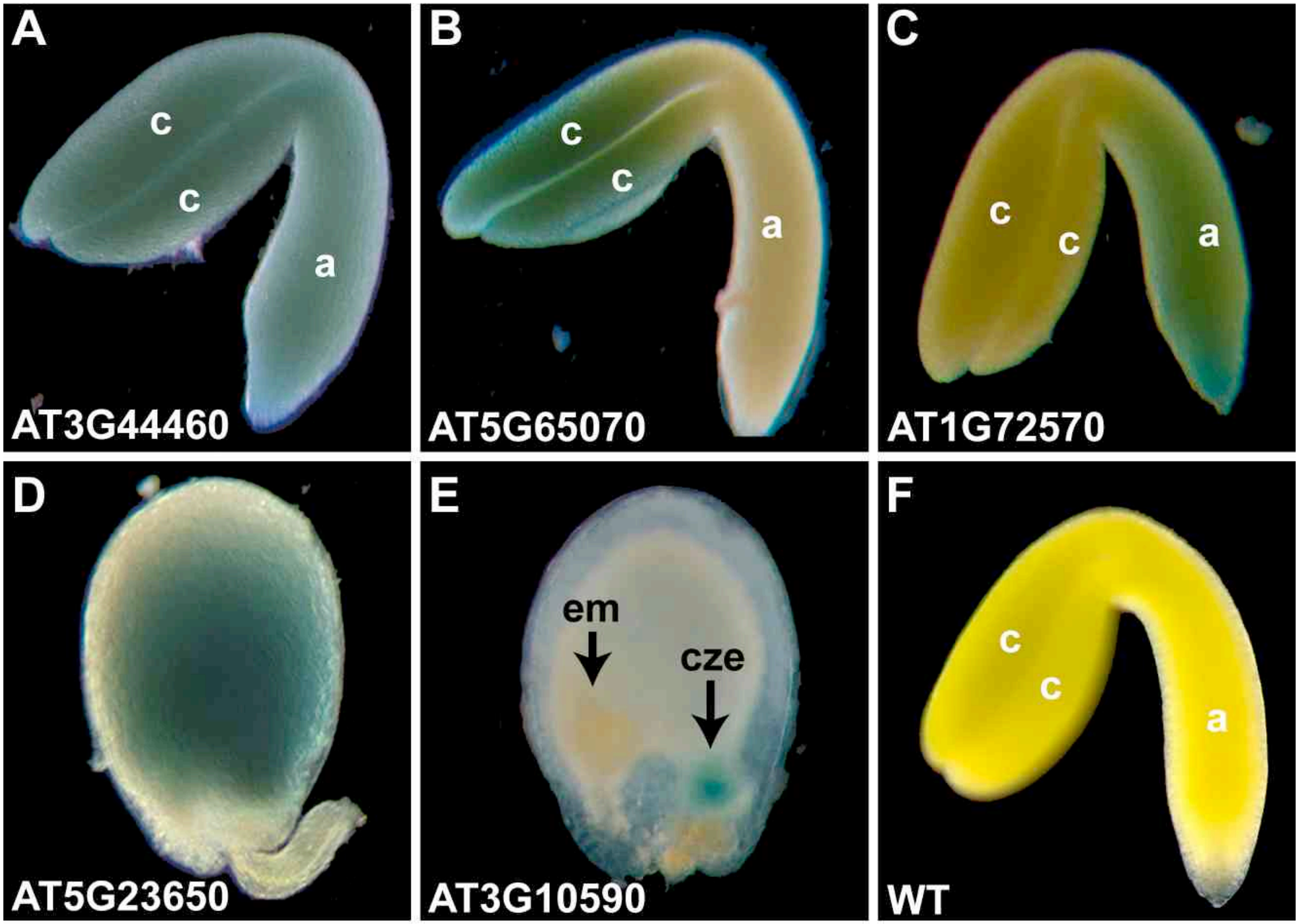
(G,C)		**Myb-Related Protein**
		NAC Domain
(C)		**No Apical Meristem (NAM)**
		Polycomb Group
(M)	*	**MEDEA**
(PM)		**CCHC-Type Family Protein**
		WRKY
(G)	*	**WRKY10; MINI3**
		Zinc Finger
(C,M,PM)	*	**PEI1**
(G)		**C2H2-Type Zinc Finger**
(G,C)		**C2H2-Type Zinc Finger**
(G,C)		**C3HC4-Type RING Finger**
(G)		**C3HC4-Type RING Finger**
(G,C)		**Dof-Type Zinc Finger**
(24H,G)		**RABBIT EARS (RBE)**
		Unclassified
(C)		**SAP; Sterile Apetala**

** Mutations do not affect development (15)*

** Mutations disrupt development (7)*

Seed Stages: 24H, 24Hr Post-Pollination; G, Globular; C, Cotyledon; M, Mature Green; PM, Post-mature Green

Seed-Specific Transcription Factor Genes Are Transcribed in Different Seed Regions

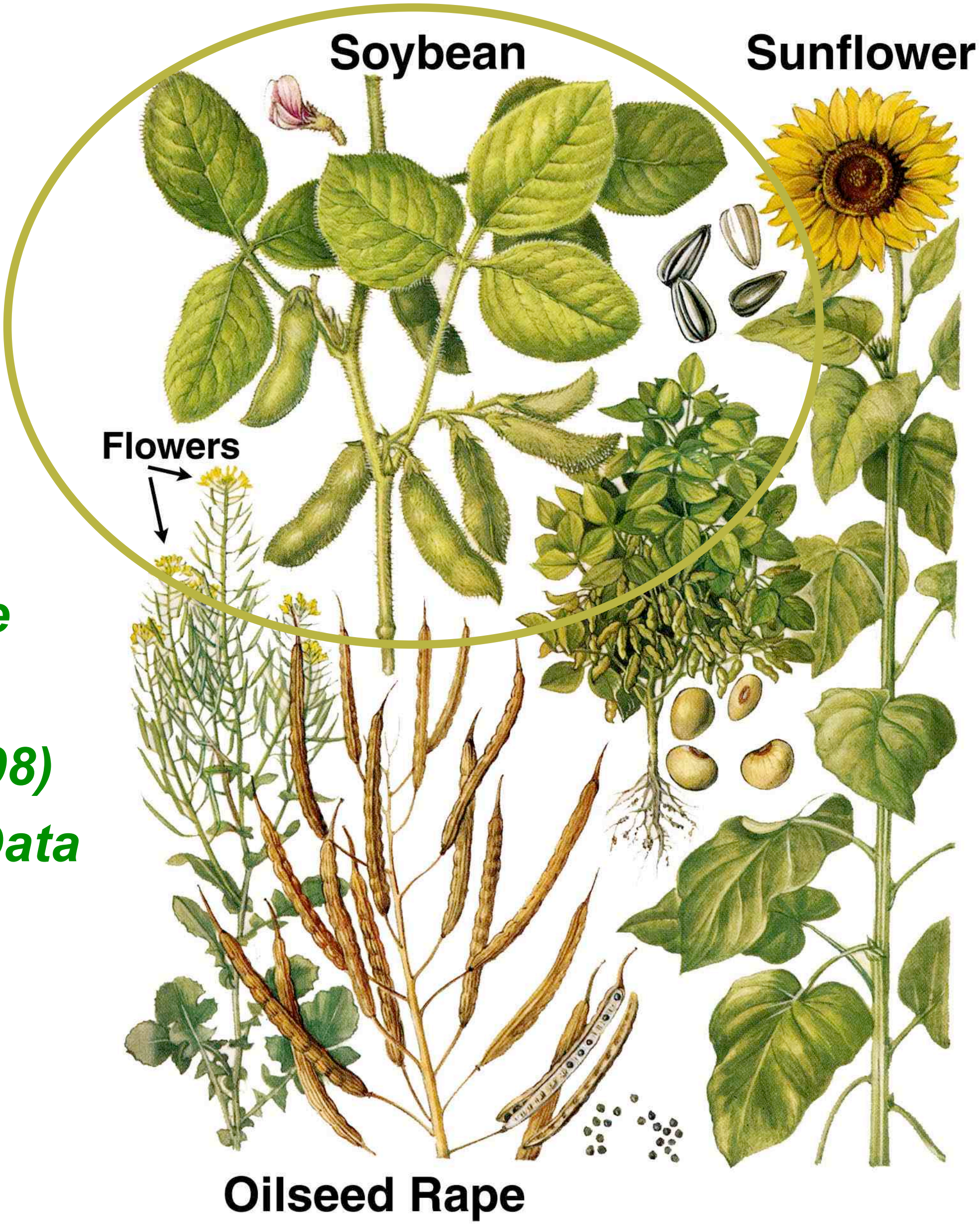


Major Unanswered Question: What Processes Do These TF Genes Regulate?

“Making A Globular Stage Soybean Seed”

All Tissues, Compartments, and Regions?

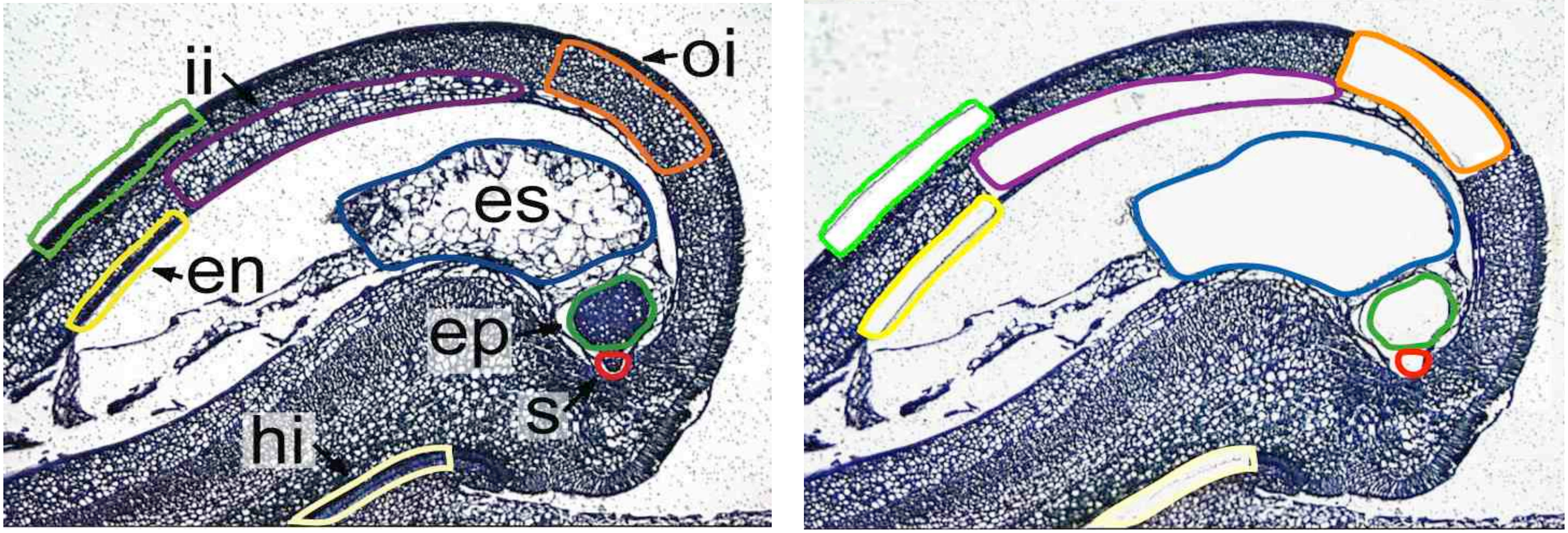
Diversity of Oil Seed Plants

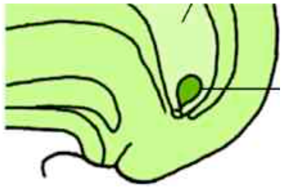


Why Soybean?

- *Second Major US Crop*
- *Major Food Source*
- *Important Biofuel Source*
- *Excellent Model Plant*
- *Genome Sequenced (2008)*
- *Seed Gene Expression Data*
- *Major Funding Source*

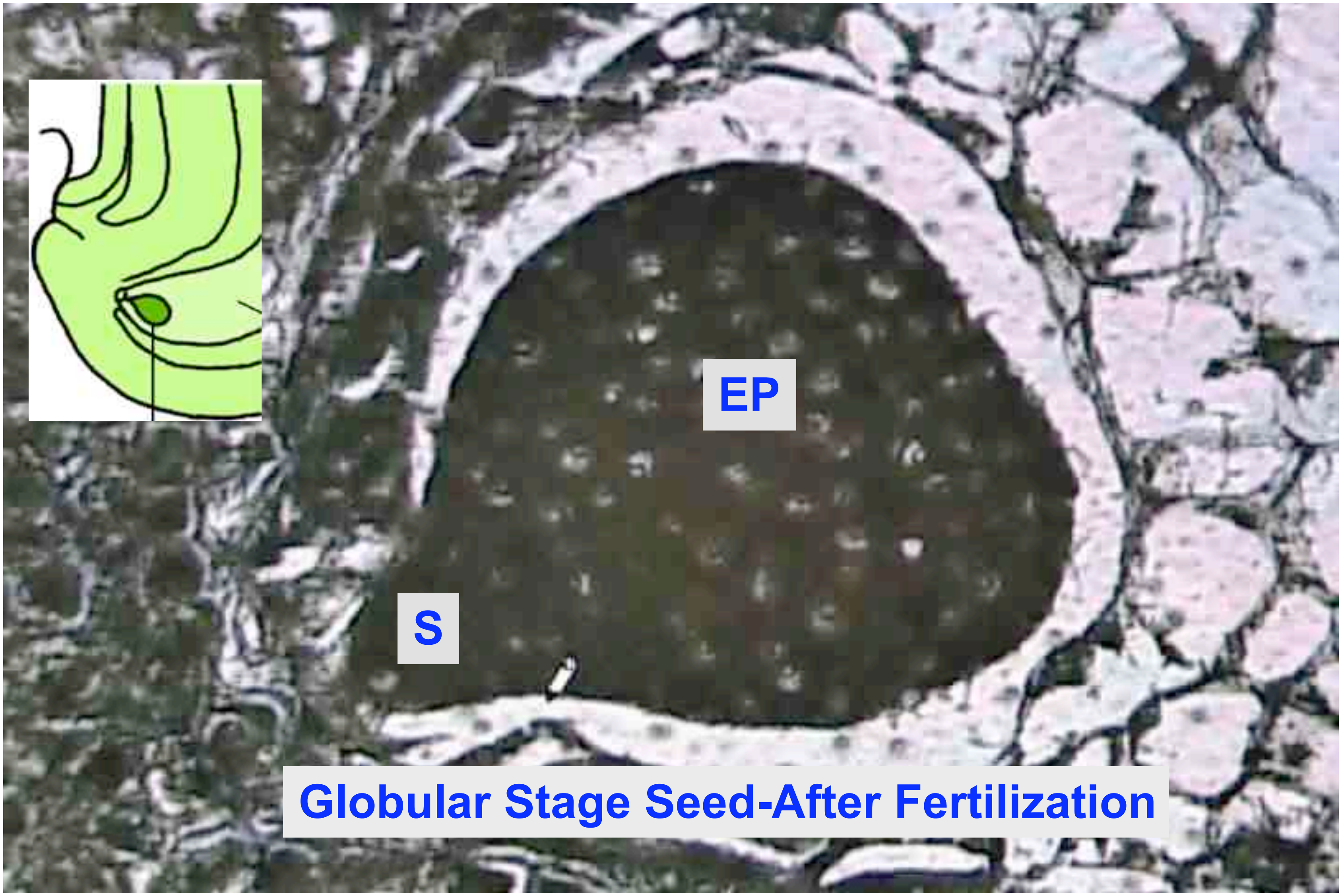
How Can We Profile Gene Activity in All Seed Compartments, Regions, & Tissues?





Combine Laser Capture Microdissection (LCM) Technologies With Genomics Approaches.

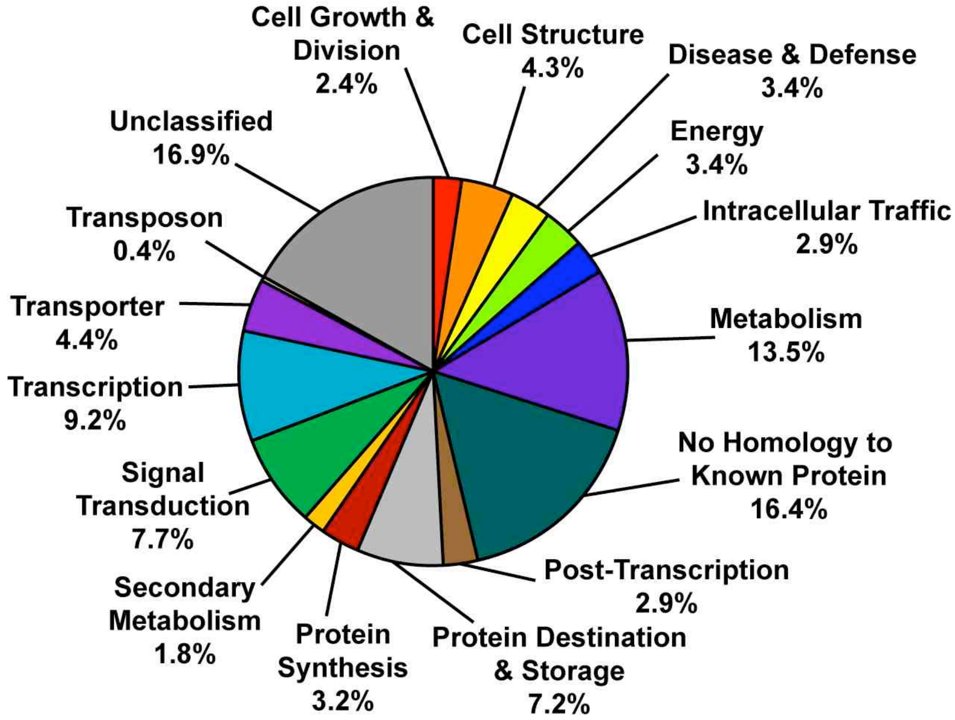
Using Laser Capture Microdissection (LCM) & Soybean GeneChips to Investigate Gene Activity In Seeds



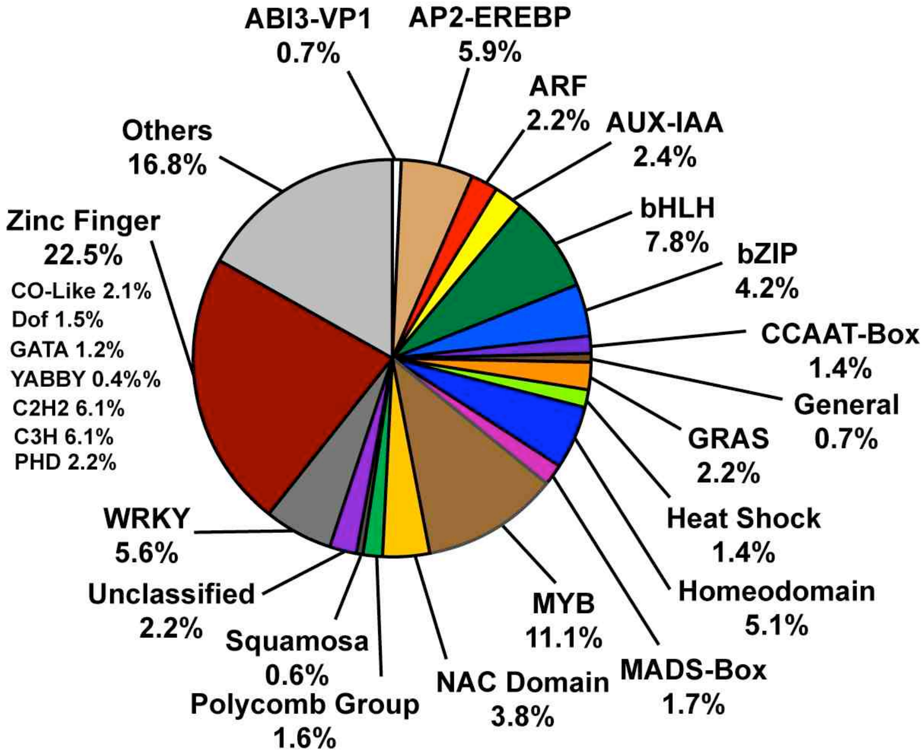
Globular Stage Seed-After Fertilization

Spectrum of Gene Sequences Represented on the Soybean Affymetrix EST GeneChip (2007)

Functional Categories



Transcription Factors



~2,800 TF Transcripts

Contains Probe Sets Representing 38,000 Soybean Transcripts
(~30,000 Clusters/~23,000 Predicted cDNAs) Derived From ~85 cDNA Libraries
From Plant Regions and Multiple Developmental Stages
(Not a Whole Genome Chip)

GENE NETWORKS IN SEED DEVELOPMENT

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

Click here to learn about the Seed Gene project.

Click here to browse the gene expression profiles of different compartments in Soybean and Arabidopsis seed at different developmental stages.

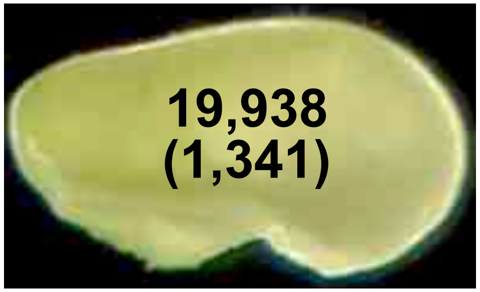
Click here to compare gene activity in different Soybean and Arabidopsis seed compartments.

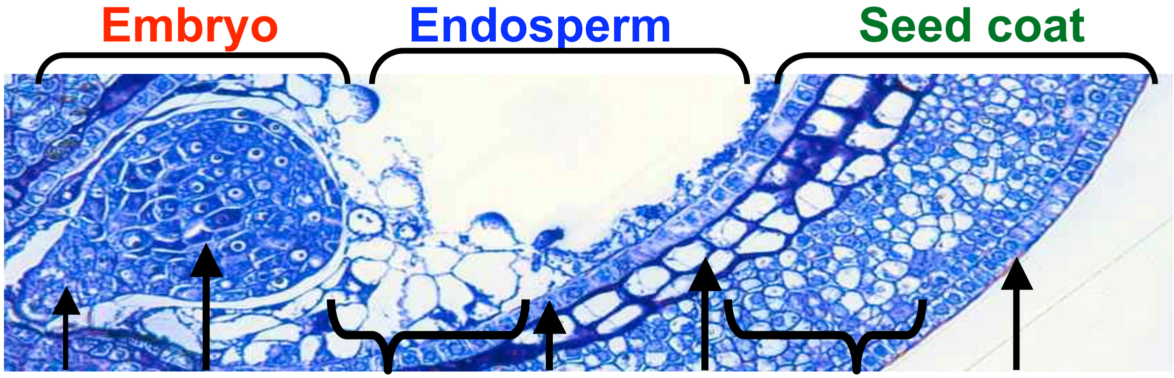
Click here to BLAST your sequence against target sequences on the GeneChip arrays and view the seed expression pattern related to your sequence.

http://estdb.biology.ucla.edu/seed

What Are the Genes Active in an Entire Globular-Stage Seed Immediately After Fertilization?

Whole Mount

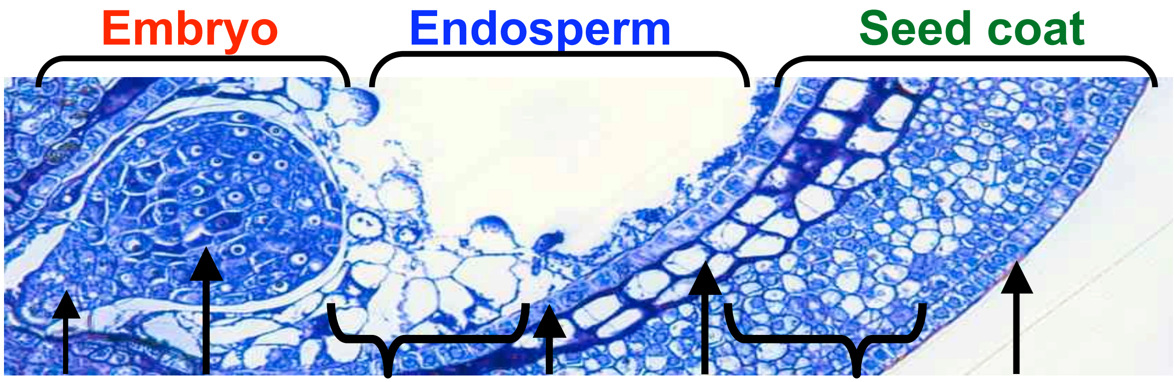




	S	EP	ES	EN	II	OI	EPD	HI
Total mRNAs*	14,177	16,998	13,880	15,274	14,767	16,402	13,451	16,153
TF mRNAs	909	1,100	848	950	927	1,073	837	1,057
Unique mRNAs	74	96	98	39	37	36	23	49
Unique TF mRNAs	5	17	11	3	4	2	2	6
Shared mRNAs	9025 →							

* The present call in globular stage is defined as “present” at least in two biological replicates. One factor ANOVA p>0.05

What Are the Genes Active in an Entire Globular-Stage Seed Immediately After Fertilization



	S	EP	ES	EN	II	OI	EPD	HI
Total mRNAs*	14,177	16,998	13,880	15,274	14,767	16,402	13,451	16,153
TF mRNAs	909	1,100	848	950	927	1,073	837	1,057
Unique mRNAs	74	96	98	39	37	36	23	49
Unique TF mRNAs	5	17	11	3	4	2	2	6
Shared mRNAs	**9025**							

*** The present call in globular stage is defined as "present" at least in two biological replicates**

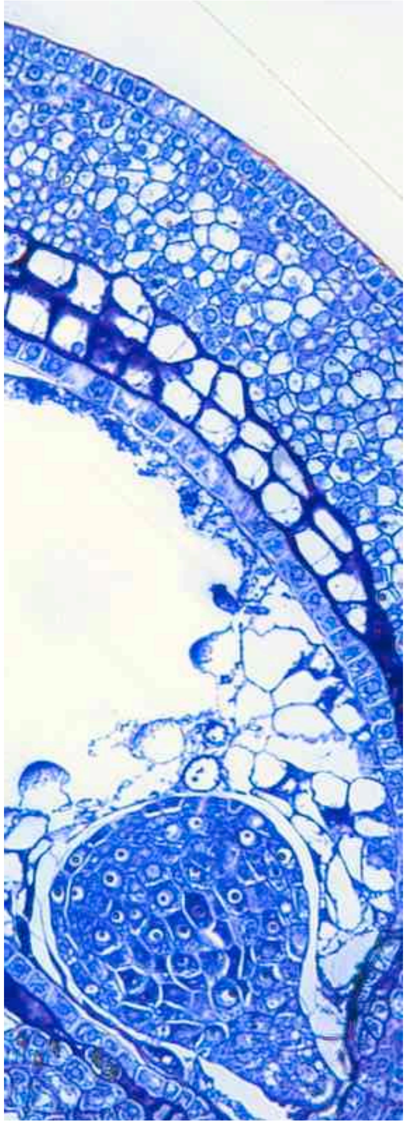
Note: Unique genes are specific within the seed at the level of the GeneChip


Conclude

- *Most Genes Shared By Different Seed Compartments*
- *There Are Small Sets of Compartment-Specific Genes Including Transcription Factor Genes*

A Repeating Theme!!!!


Globular-Stage Seed Compartments Have a Unique Set of Transcription Factor Genes



Outer Integument: 36 (2)

1 bZip
1 Zinc Finger

Inner Integument: 37 (4)

1 General (TFIIA-L)
1 Homeodomain
1 MADS, MYB

Endosperm: 98 (11)

3 Zinc Finger (Dof 2)
2 WRKY
1 ARF (ARF8), bHLH, JUMONJI
1 MADS-Box (PISTILLATA)
1 NAC Domain (NAM)
1 Polycomb Group (TRX1)

Embryo Proper: 96 (17)

4 MYB (MYB 44)
3 Homeodomain (STM, WOX2)
2 IAA (IAA8), bHLH, Zinc Finger
1 bZip (OBF4), WRKY, G2-like
1 GRAS (Scarecrow-like)

Hilum: 49 (6)

3 bHLH (PIF4)
1 Zinc Finger (IDD11),
1 MYB (MYB 111)
1 Aux/IAA (SOLITARY ROOT)

Epidermis: 23 (2)

1 AP2/EREBP
1 bZip

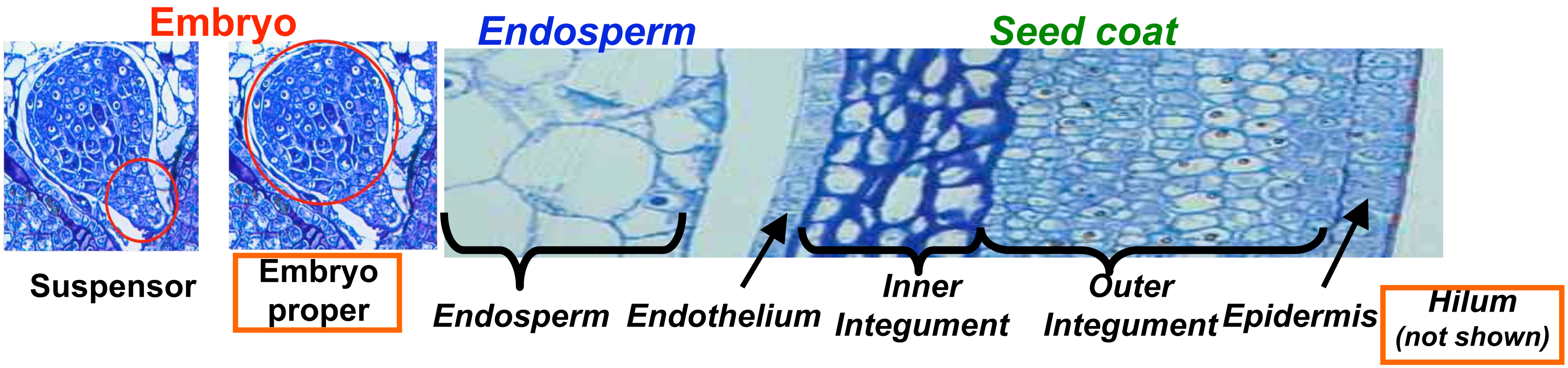
Endothelium: 39 (3)

2 Zinc Finger
1 WRKY

Suspensor: 74 (5)

2 WRKY
1 ARF (ARF16)
1 NAC Domain
1 Zinc Finger

Quantitative RT-PCR and Transcriptional Validation of Globular-Stage "Seed Region-Specific" mRNAs (one example)



GmSTM-Like (Shoot Meristemless-like) Transcription Factor (GmaAffx.63050.1.S1_at)

GeneChip	A	P	A	A	A	A	A	A
qRT-PCR	ND	30.0±0.6	ND	ND	ND	ND	ND	39.5±1.1
Fold Reduction	-	1	-	-	-	-	-	↓700

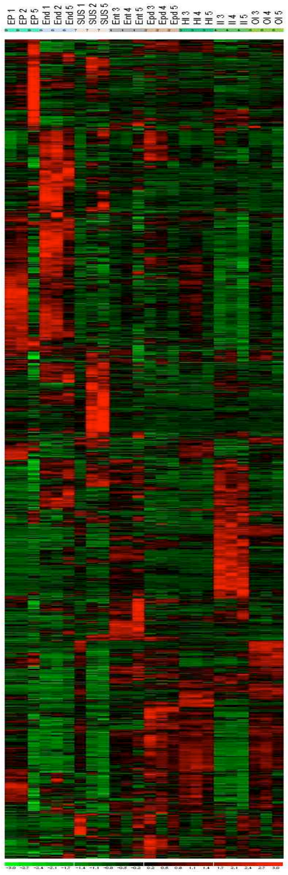
A=Absent = Not Detected at GeneChip Level

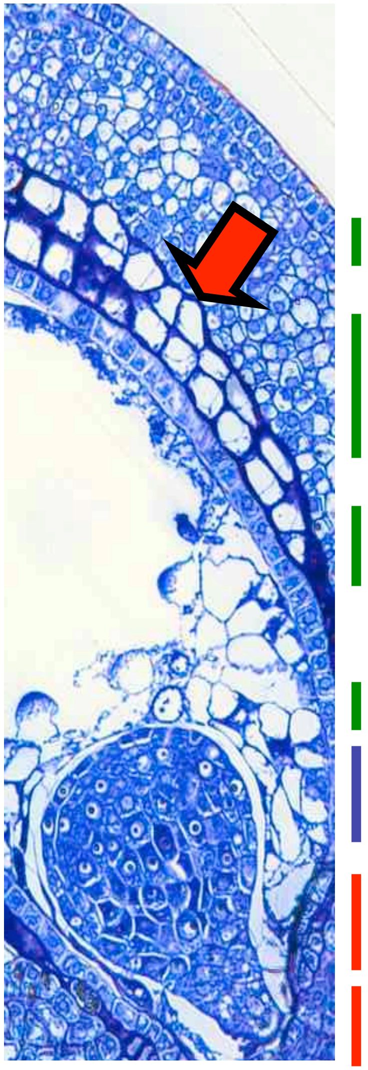
ND = Not Detected at Level of qRT-PCR

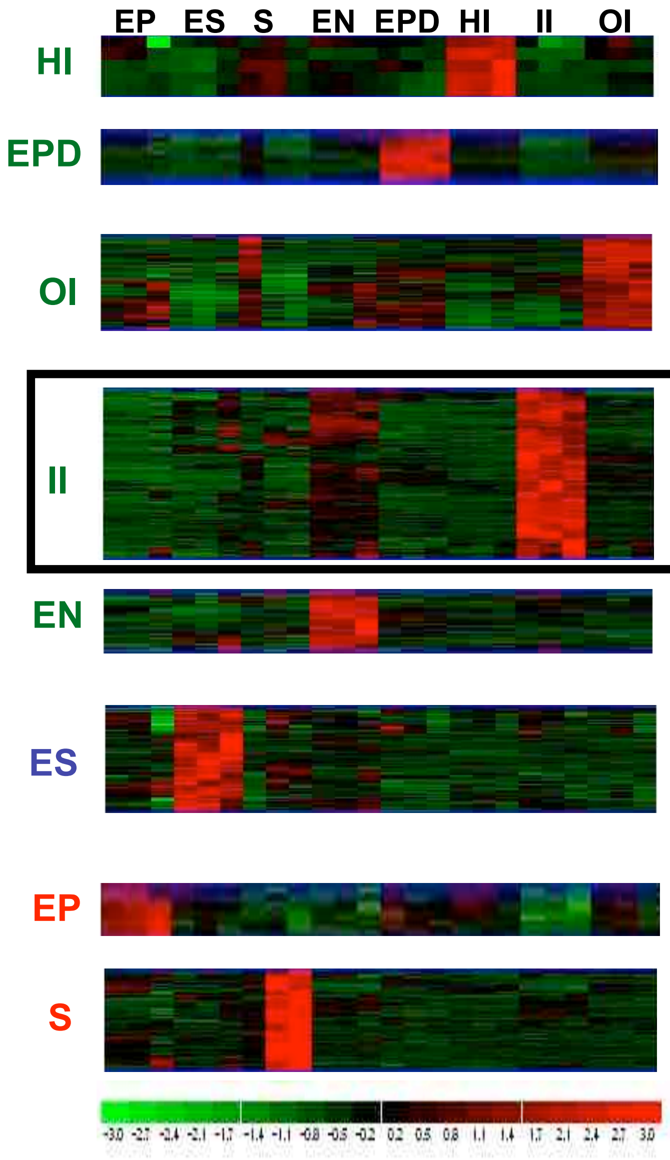
GmSTM-like promoter/GUS Gene Activity in Soybean Seeds

Quantitative Regulation of mRNAs Shared by Soybean Globular-Stage Seed Compartments

Top 2000 Varying mRNAs







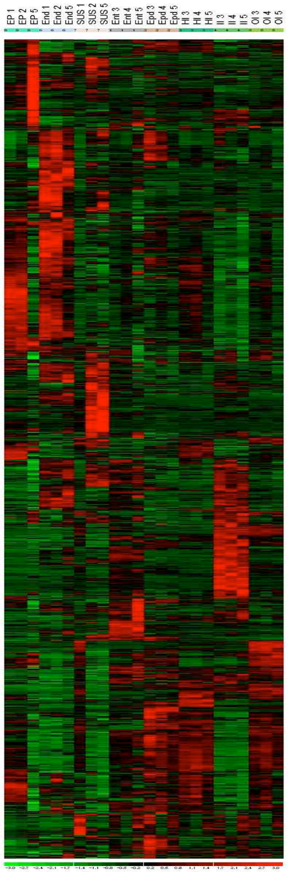
> 2-fold ($p<0.05$)

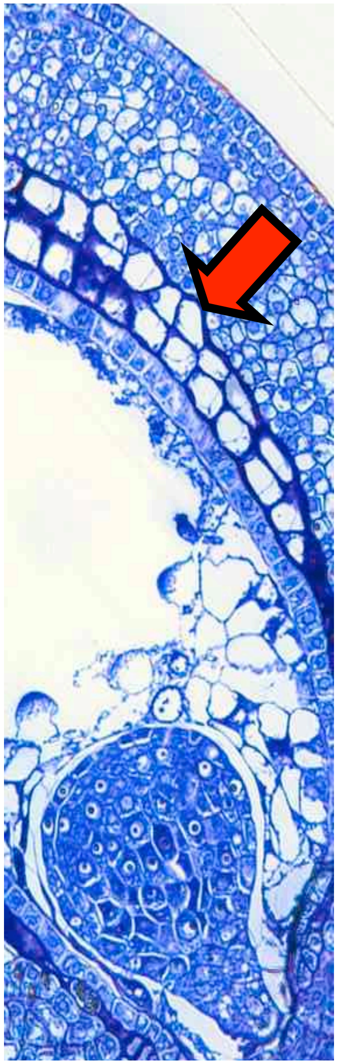
HI	4/5	Protein Storage Transcription
EPD	12/12	Signaling Energy Metabolism
OI	28/62	Energy Metabolism Transporter
II	86/120	Intracellular Traffic Transporter Metabolism
EN	17/31	Metabolism Cell Structure Protein Storage
ES	23/73	Energy Protein Synthesis Protein Storage
EP	4/13	Cell Growth Protein Storage Post-Transcription
S	1/66	Energy Metabolism Protein Storage

FDR < 5%

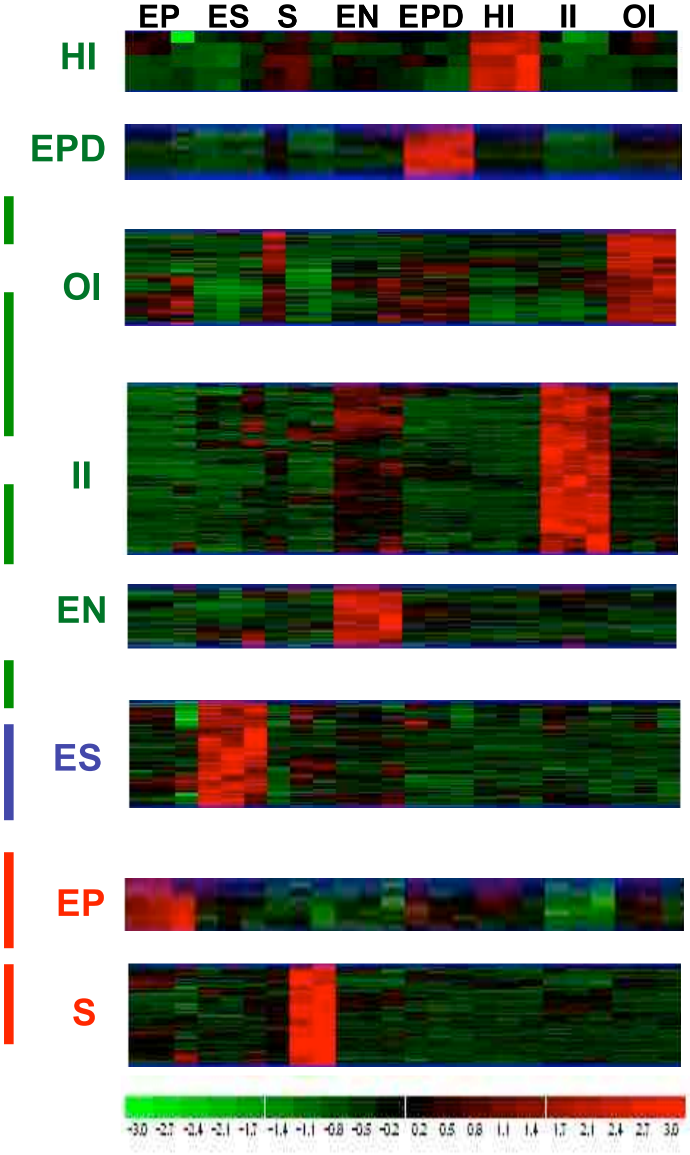
Quantitative Regulation of mRNAs Shared by Soybean Globular-Stage Seed Compartments

Top 2000 Varying mRNAs





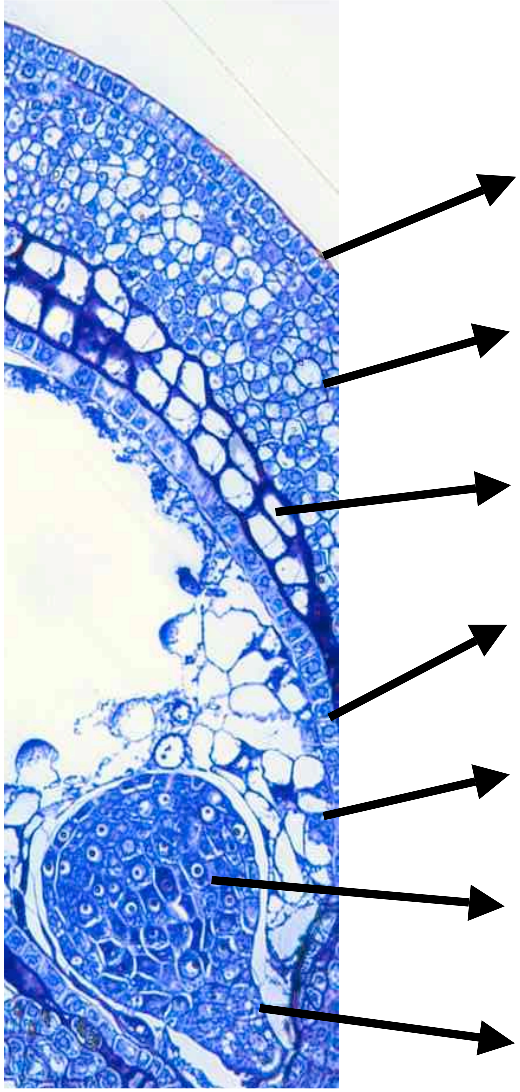
> 2-fold (p<0.05)



4/5	Protein Storage Transcription
12/12	Signaling Energy Metabolism
28/62	Energy Metabolism Transporter
86/120	Vacuolar ATPase STP1 - CAT7 SIS - GATL7 - SAH7
17/31	Metabolism Cell Structure Protein Storage
23/73	Energy Protein Synthesis Protein Storage
4/13	Cell Growth Protein Storage Post-Transcription
1/66	Energy Metabolism Protein Storage

FDR < 5%

How Many Genes Are Active in a Globular-Stage Soybean Seed?



Hilum
16,153 (1,057)

Epidermis
13,451 (837)

Outer Integument
16,402 (1,073)

Inner Integument
14,767 (927)

Endothelium
15,274 (950)

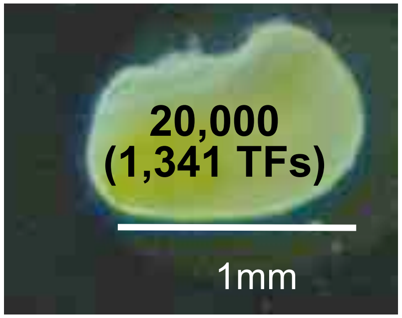
Endosperm
13,880 (848)

Embryo Proper
16,998 (1,100)

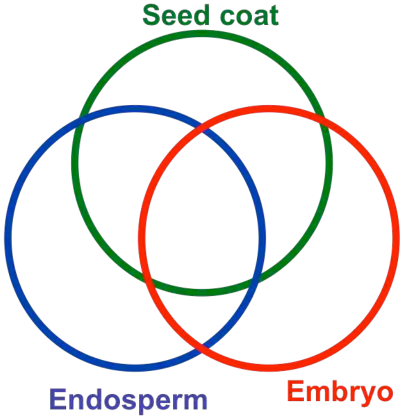
Suspensor
14,177 (909)

At the Level of GeneChip Detection!

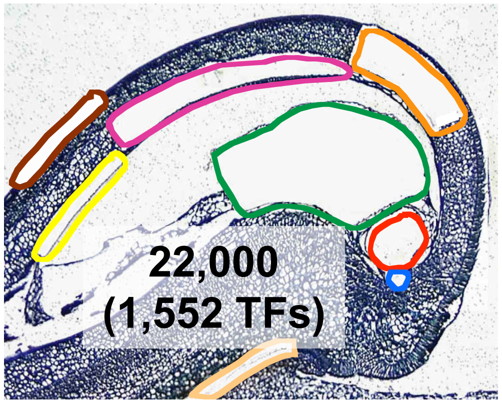
Whole Mount



Union of Seed Compartments



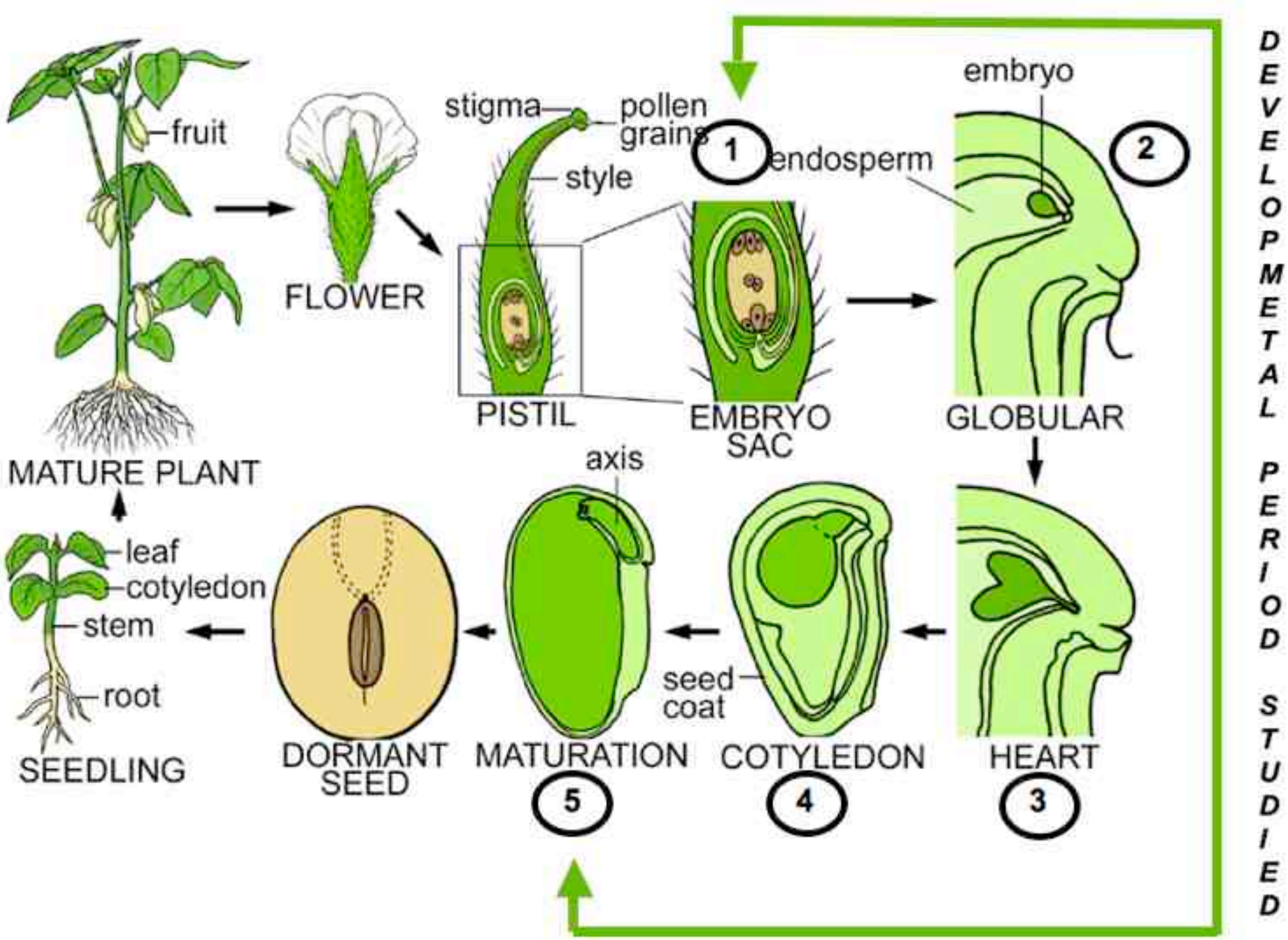
Union of LCM mRNA Sets



<u>Note</u>: Minimum Numbers & Overlap of Gene Activity!!

(Transcription Factors)

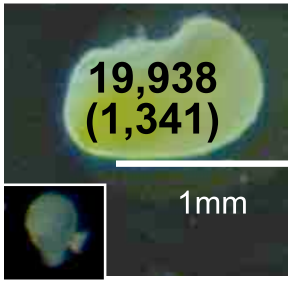
What Are The Genes Required to Program Every Compartment, Tissue, and Cell Type During Soybean Seed Development?



***Note*: Numbers Refer to Specific Seed Stages Studied**

How Many Genes Are Required to Program Soybean Seed Development?

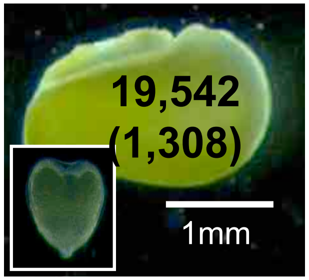
Globular



19,938
(1,341)

+

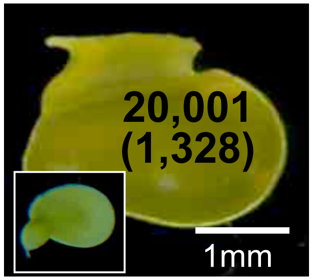
Heart



19,542
(1,308)

+

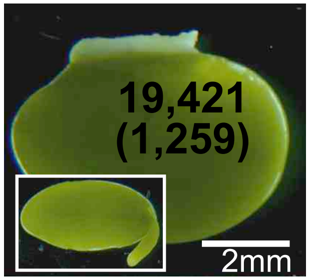
Cotyledon



20,001
(1,328)

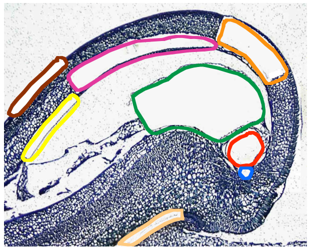
+

Early Maturation



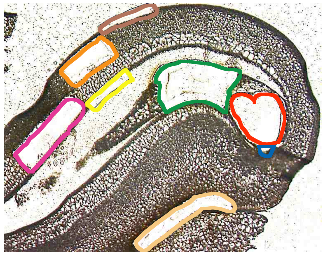
19,421
(1,259)

= 23,033
(1,582)



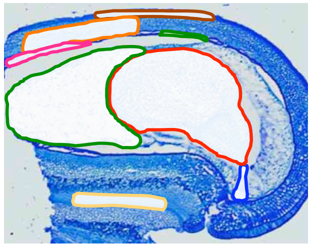
21,989
(1,552)

+



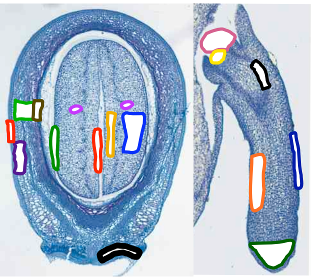
22,377
(1,538)

+



22,903
(1,595)

+

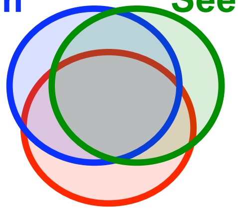


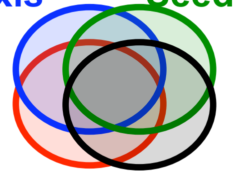
24,003
(1,685)

= 26,359
(1,894)

Minimum Numbers!

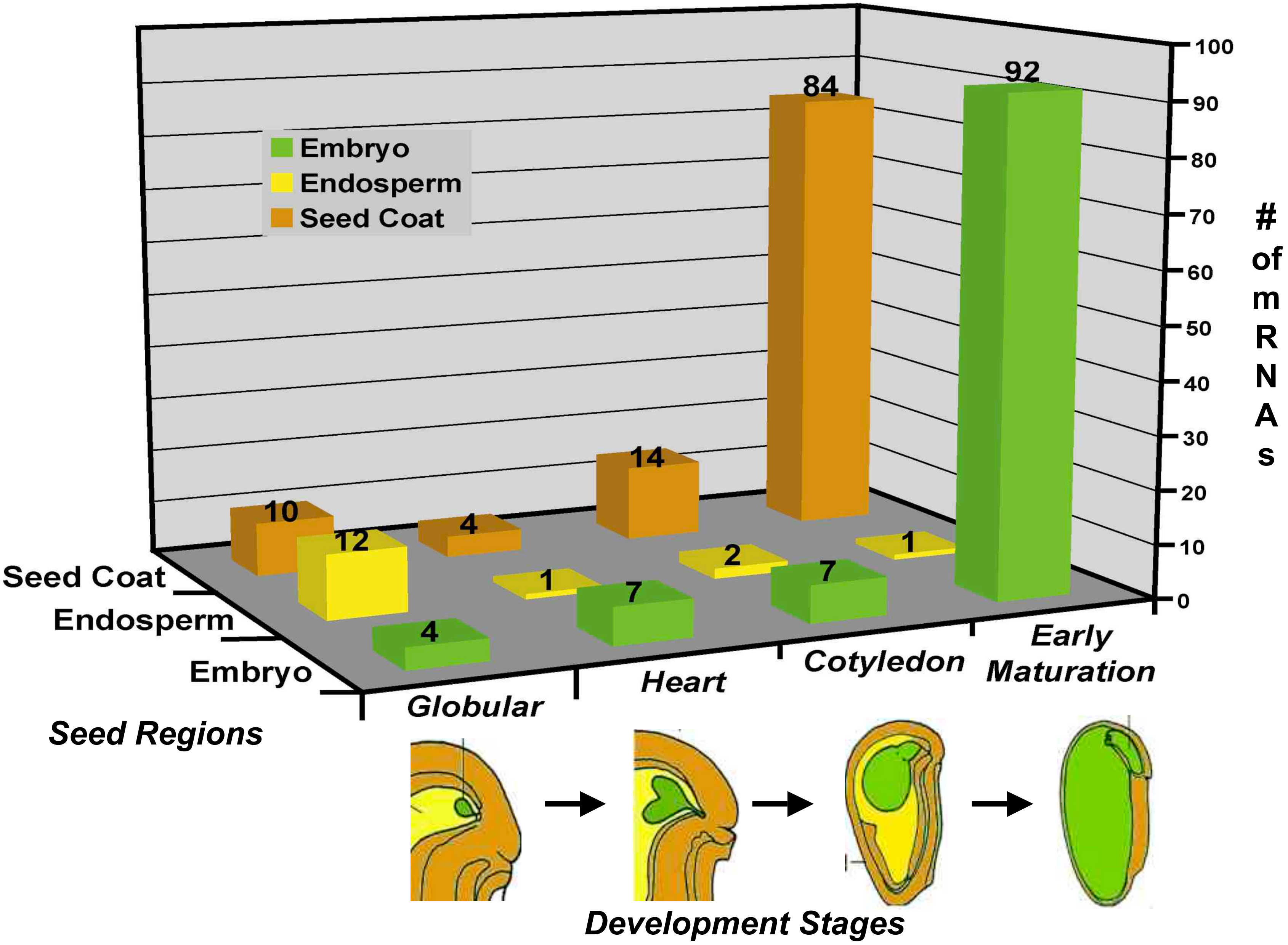
(Transcription Factors)



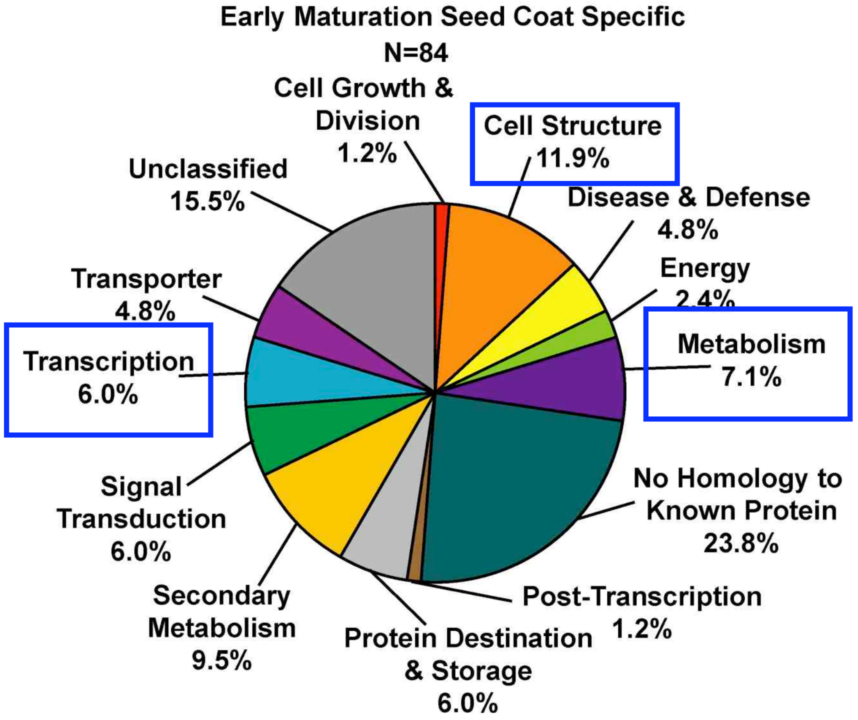


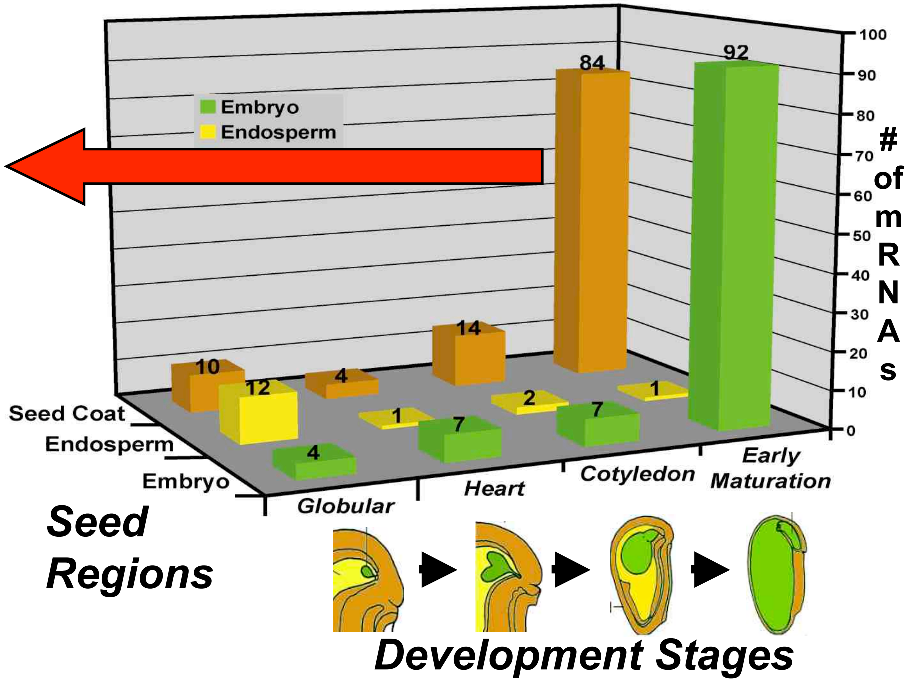
40 Compartments & Tissues Profiled-More than 3.7 Million Data Points!!

Each Soybean Seed Compartment Has a Small Set of Unique mRNAs at Different Developmental Stages



Each Soybean Seed Compartment Has a Small Set of mRNAs at Specific Developmental Stages (e.g., Early Maturation-Stage Seed Coat)





Cell Structure
Cellulose Synthase
(CESA8, CESA4, IRX3)
Pectinacetylesterase Family
Glucan 1,3-beta-glucosidase

Secondary Metabolism
BANYULS (DFR)
TRANSPARENT TESTA 10
Isoflavone Reductase Homolog
2OG-Fe(II) Oxygenase

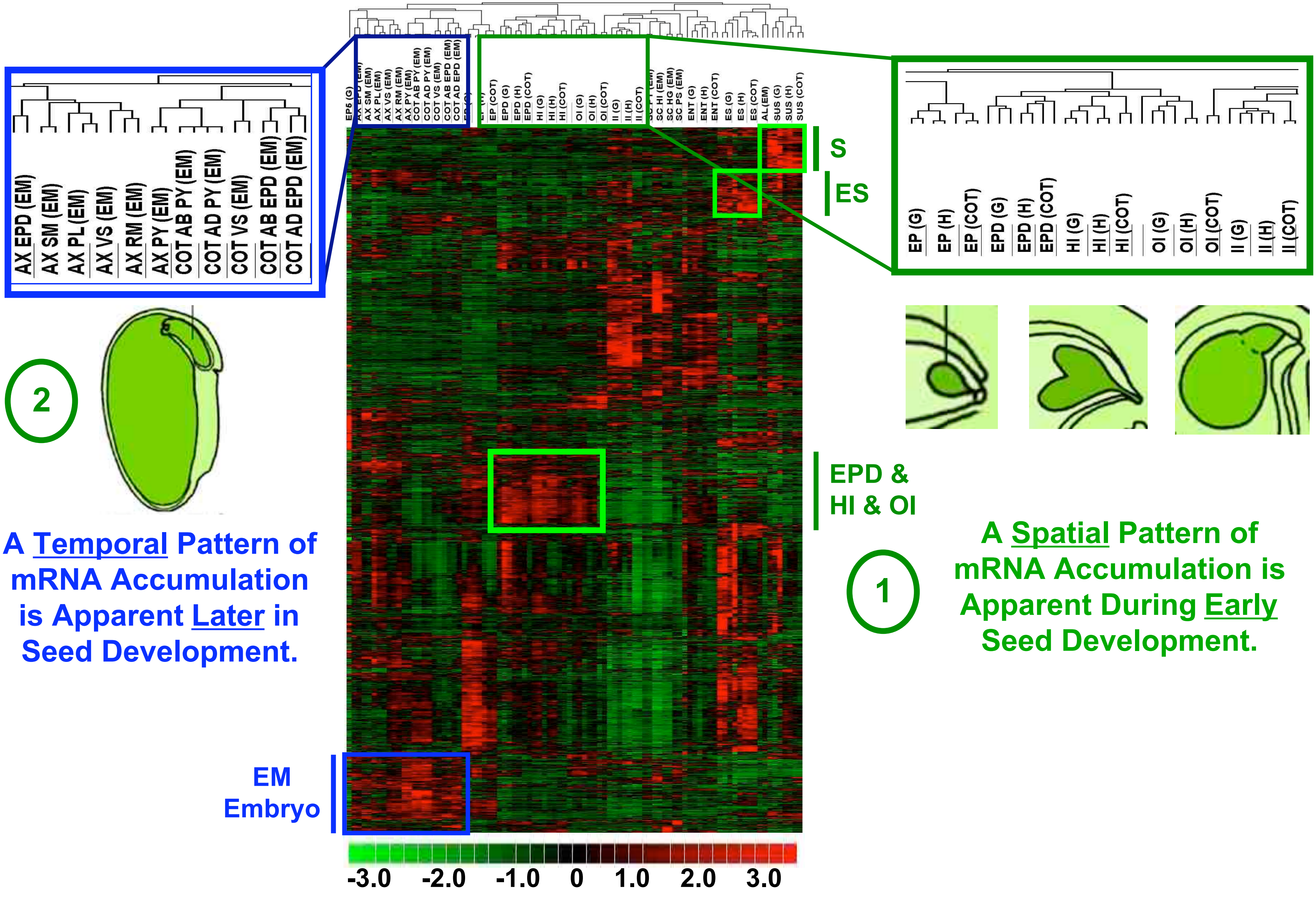
Transcription
SCARECROW-LIKE 3
C2H2 Zinc Finger
B-Box Type Zinc Finger


Cell Wall Enzymes


Brown Color Enzymes


Gene Regulators

Do Shared mRNAs Have Specific Temporal and Spatial Patterns of mRNA Accumulation Throughout Seed Development?



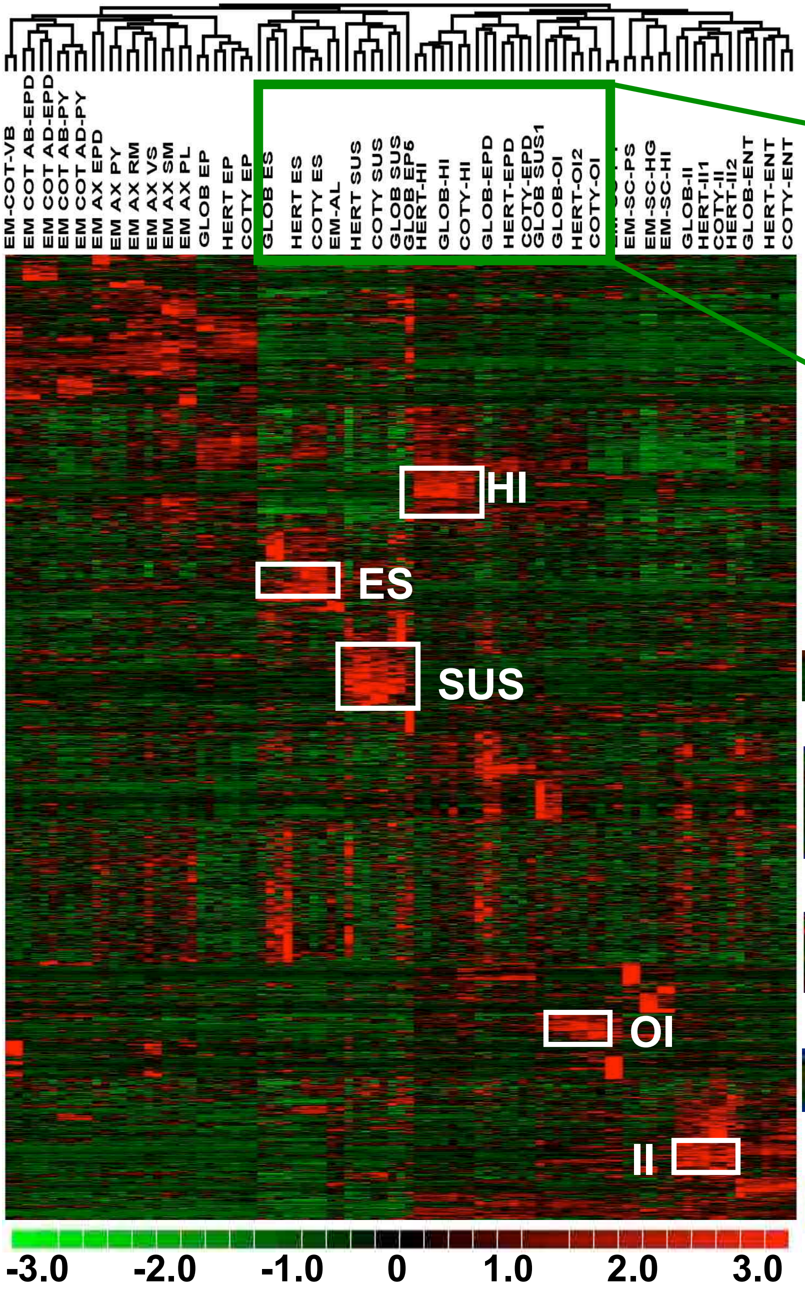
②

A Temporal Pattern of mRNA Accumulation is Apparent Later in Seed Development.

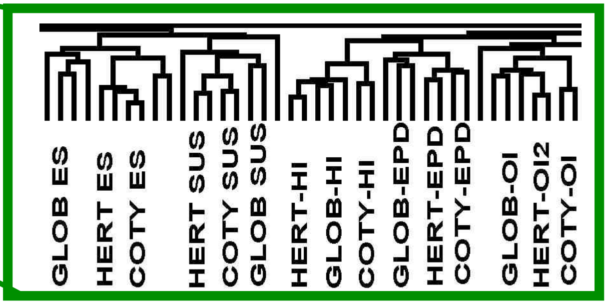
①

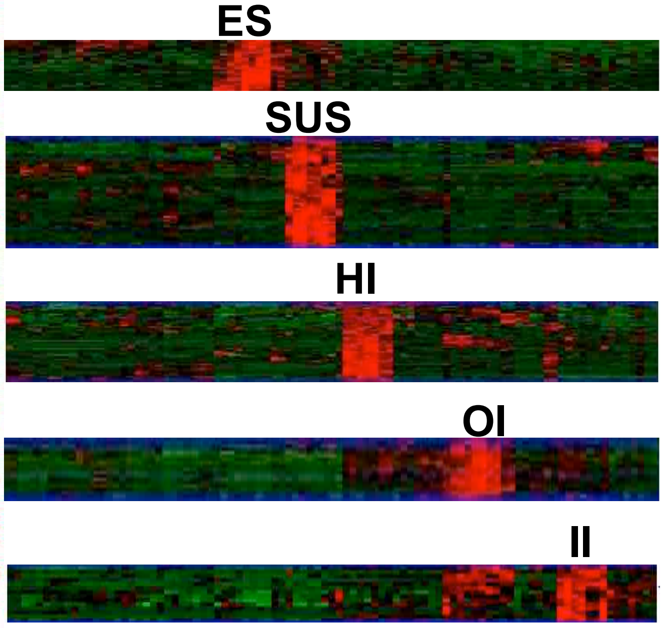
A Spatial Pattern of mRNA Accumulation is Apparent During Early Seed Development.

Are There Spatial Patterns of Transcription Factor mRNA Accumulation During Early Soybean Seed Development?



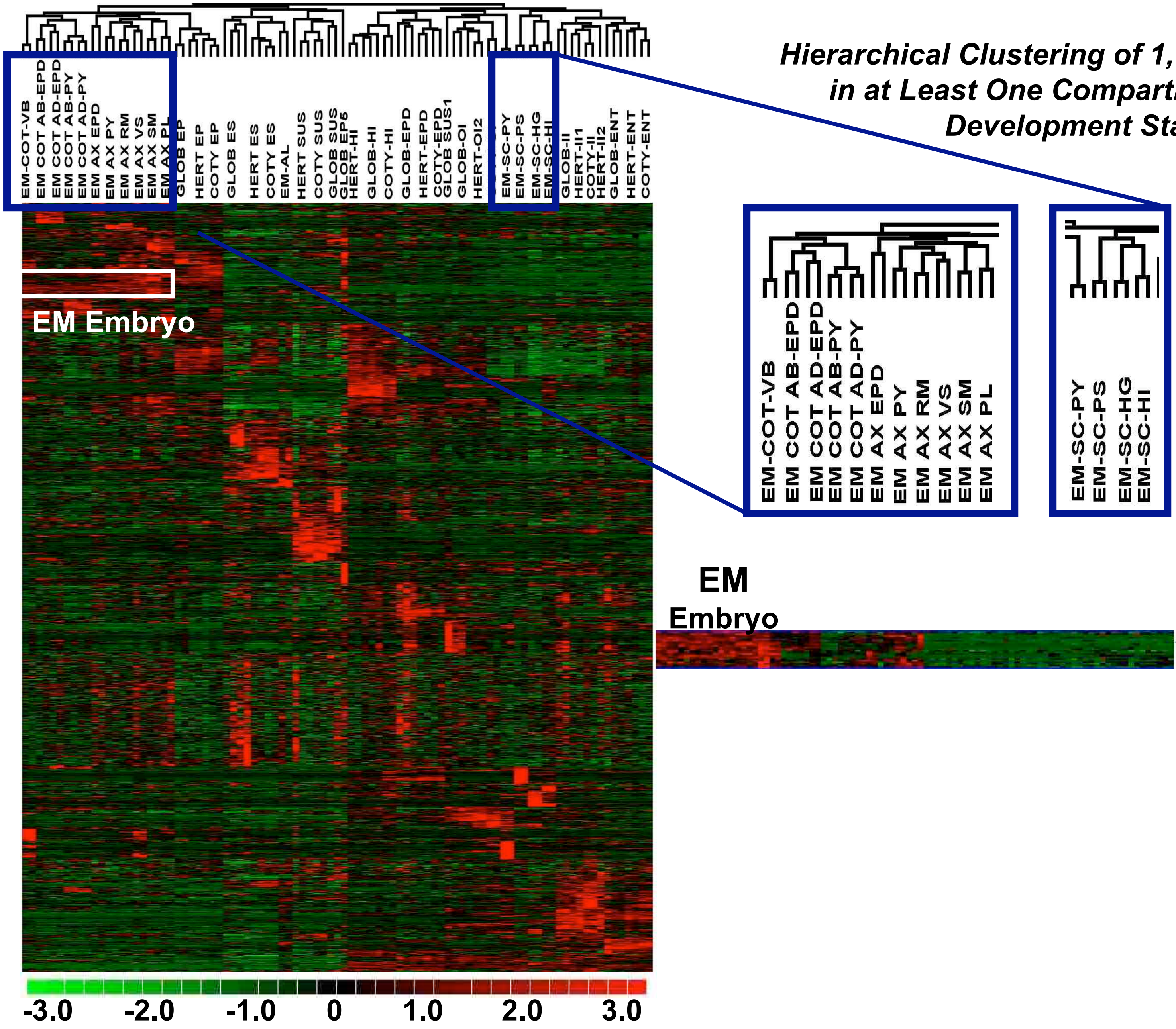
Hierarchical Clustering of 1,932 TFs Present in at Least One Compartment Across Development Stages



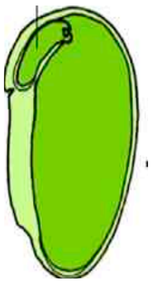


	Over-represented (2-fold, $p<0.05$)
22/42	**ARF (1)**
51/58	**WRKY (15)** **CCAAT-Box (4)**
39/53	**bHLH (10)** **Squamosa (2)**
15/18	**Unclassified (3)**
9/9	**No Over-representation**

Are There Temporal Patterns of Transcription Factor mRNA Accumulation During Late Soybean Seed Development?



Hierarchical Clustering of 1,932 TFs Present in at Least One Compartment Across Development Stages

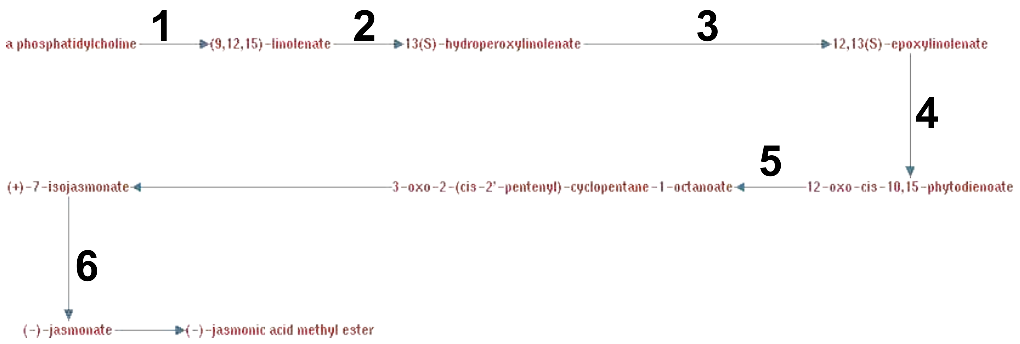


Over-represented (2-fold, p<0.05)

bZIP (2)
CCAAT-Box (3)

Do Soybean Seed Compartments Express Specific Metabolic Pathway Genes?

- Overview -

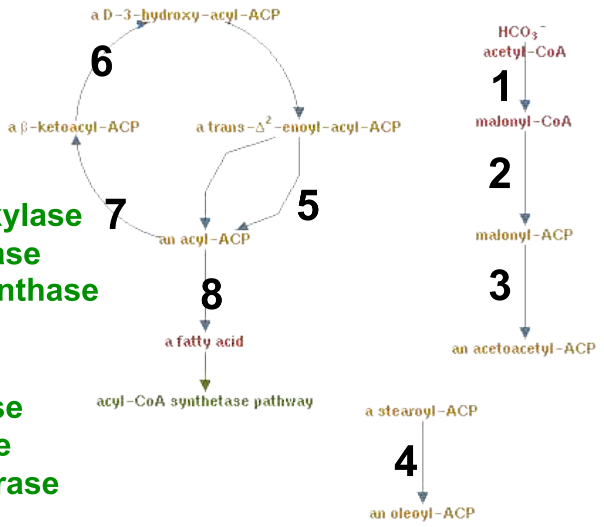


Jasmonic Acid Biosynthesis Pathway

1. Phospholipae A
2. Lipoxygenase
3. Allene oxide synthase
4. Allene oxide cyclase
5. 12-oxophytodienoate reductase
6. JA carboxyl methyltransferase

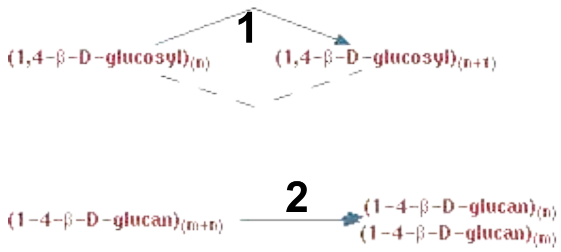
Fatty Acid Biosynthesis Pathway:

1. Acetyl-CoA carboxylase
2. S-malonyltransferase
3. β-ketoacyl-ACP synthase
4. Acyl-cleasaturase
5. Enoyl reductase
6. 3-oxoaxyl reductase
7. 3-oxoaxyl synthase
8. Acyl-ACP thioesterase



Cellulose Biosynthesis Pathway

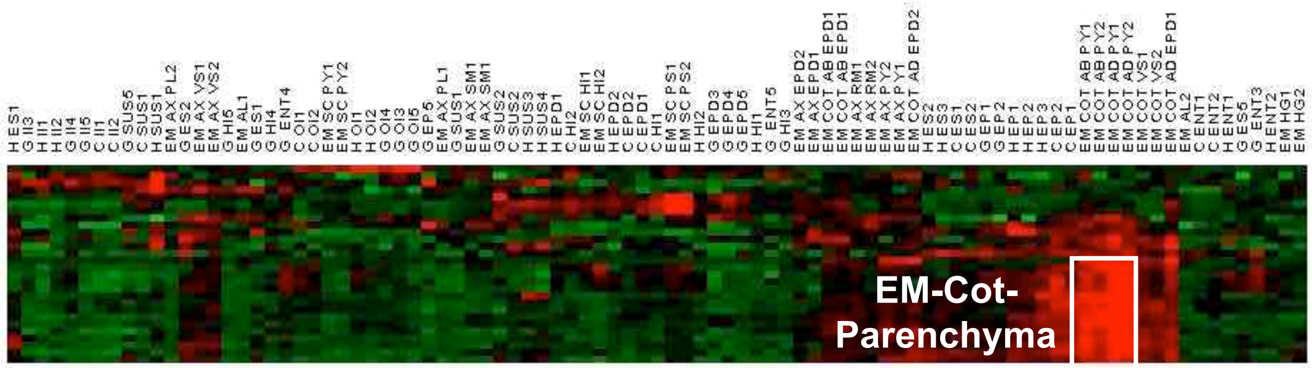
1. Cellulose synthase
2. Cellulase (Glucan-glucanohydrolase)



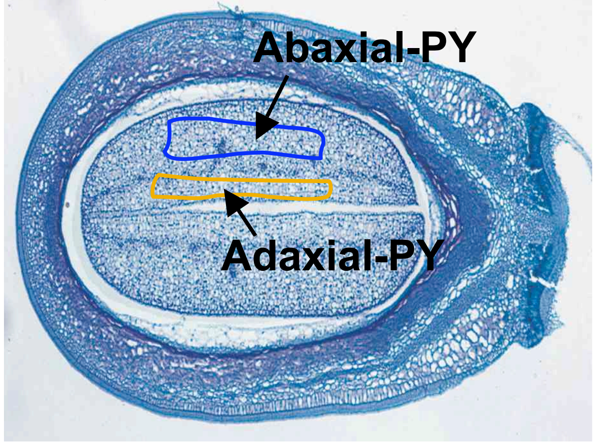
Fatty Acid Biosynthesis mRNAs?

Transcripts Encoding <u>Fatty Acid Biosynthesis</u> Pathway Enzymes Are Prevalent in Early Maturation-Stage Cotyledon Adaxial and Abaxial Parenchyma Cells

Transcripts Encoding FA Biosynthesis Enzymes



Early Maturation



<u>mRNA Upregulated >2-fold (number of mRNAs)</u>

Acetyl-CoA carboxylase (4)

S-malonyltransferase (1)

Acyl-deasaturase (1)

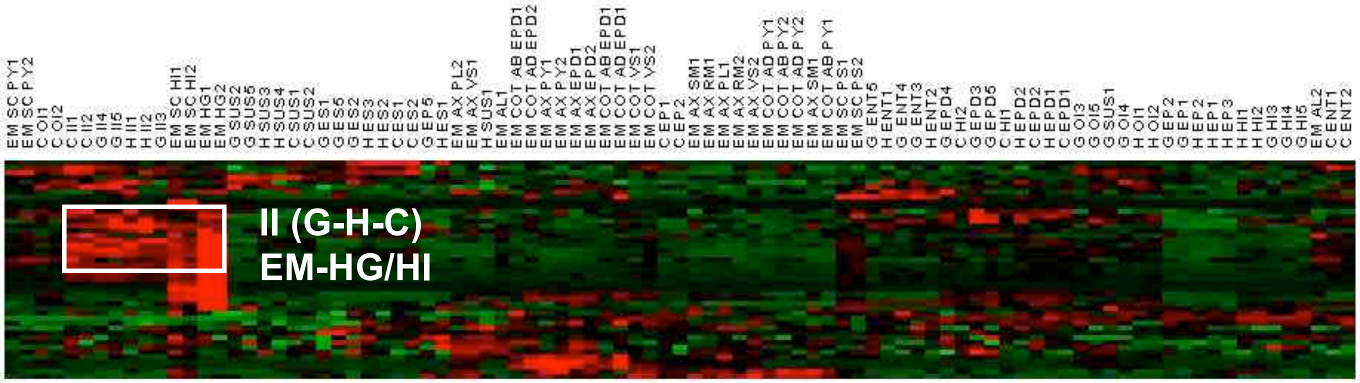
Enoyl reductase (2)

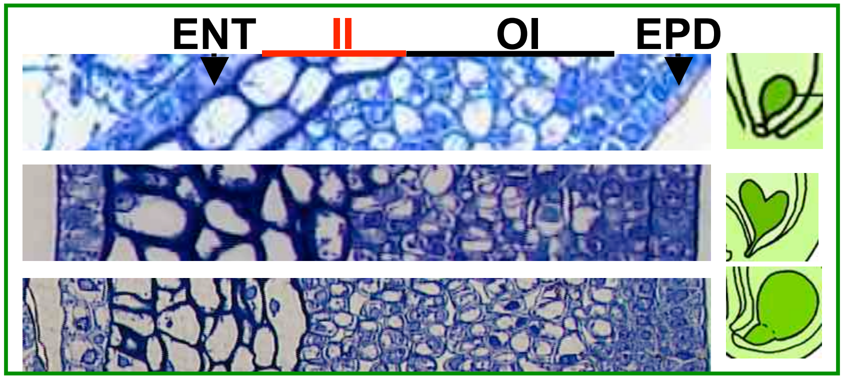
3-oxoaxyl reductase (1)

Cellulose Biosynthesis Enzymes?

Transcripts Encoding <u>Cellulose Biosynthesis</u> Pathway Enzymes Are Prevalent in Seed Coat Inner Integument, EM Hourglass, and Hilum Cells

Transcripts Encoding Cellulose Biosynthesis Enzymes





<u>mRNA Upregulated >2-fold (number of mRNAs)</u>

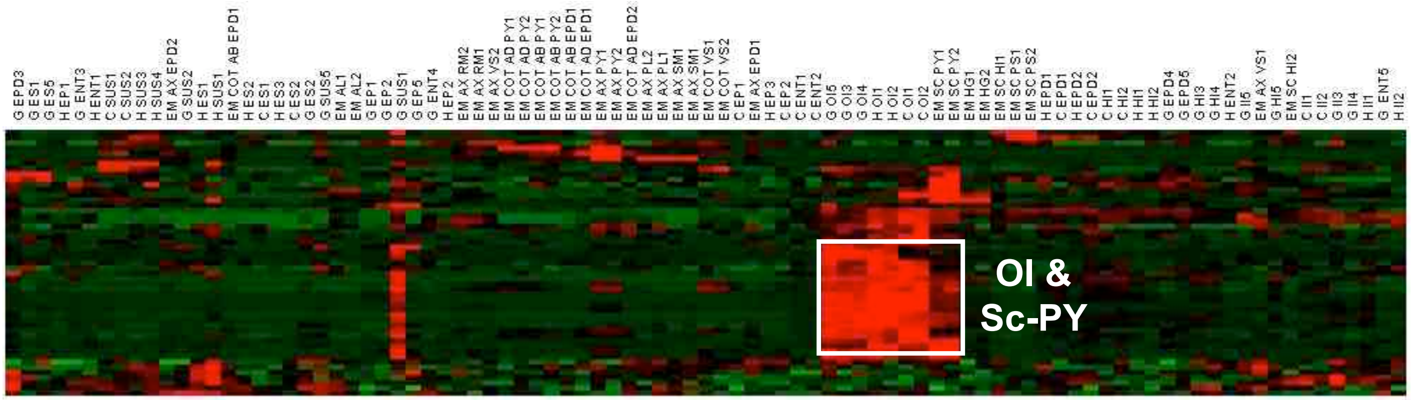
<u>Cellulose synthase:</u>
CesA1 (1),
CesA3 (5),
CesA9 (1)

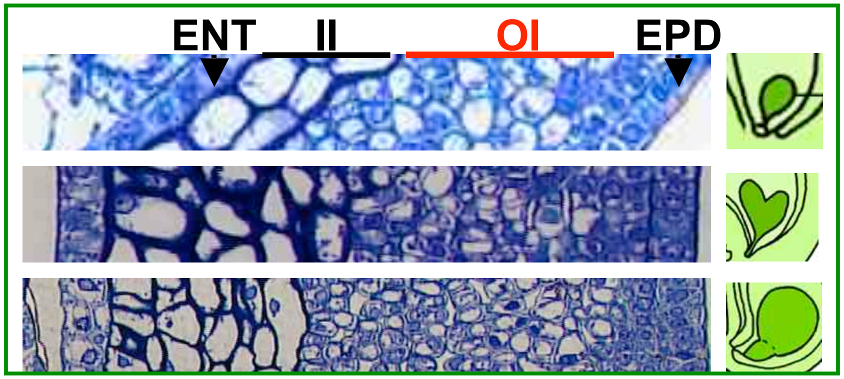
<u>Cellulase:</u>
1,4-β-D-glucan 4-glucanohydrolase (3)

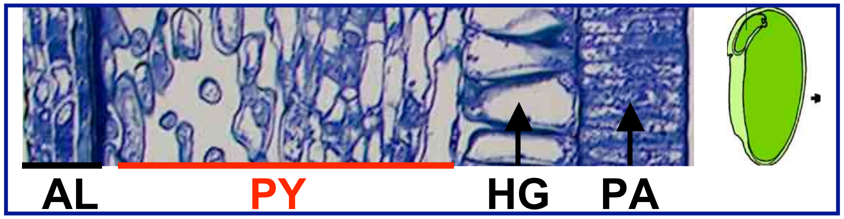
Jasmonic Acid Pathway mRNAs?

Transcripts Encoding Jasmonic Acid Biosynthesis Pathway Enzymes Are Prevalent in the Outer Integument and Seed Coat-Parenchyma

Transcripts Encoding JA Biosynthesis Enzymes







<u>mRNA Upregulated >2-fold (number of mRNAs)</u>

<u>Lipoxygenase:</u>
LOX1 (14),
LOX2 (2),
LOX similar to At3g22400 (1)

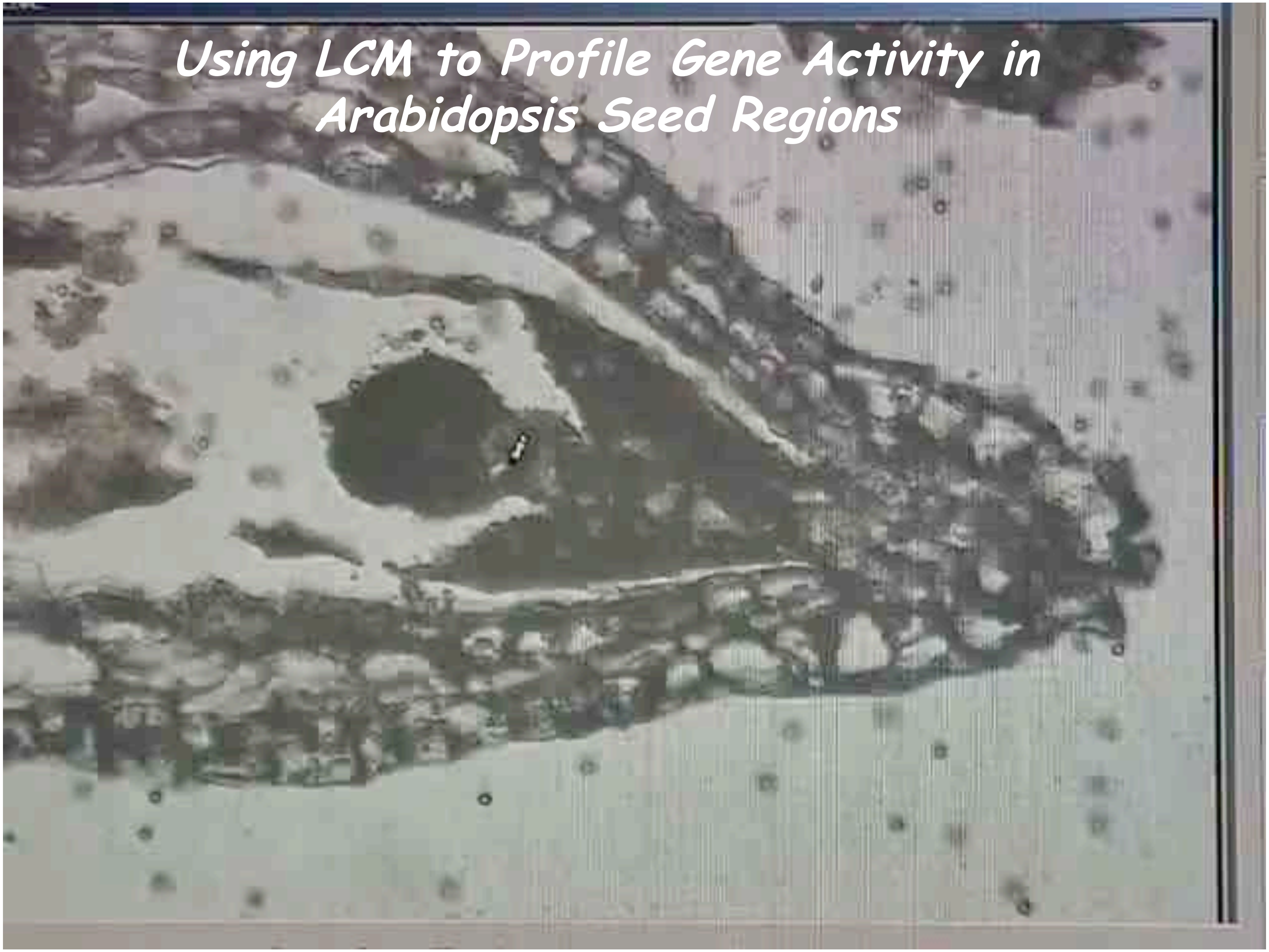
<u>Allene oxide synthase:</u>
AOS (3)

<u>Allene oxide cyclase</u> (2)


- ***How Do Soybean Seed Gene Activity Patterns Compare With Those in Other Plants?***

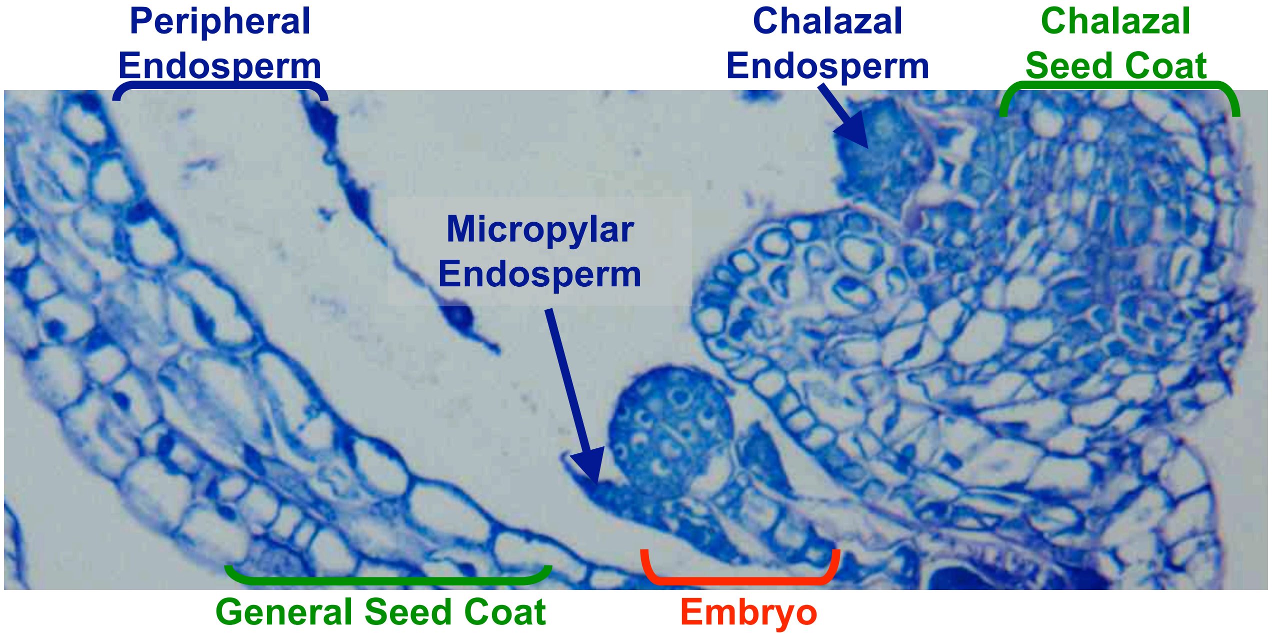
- ***What Are the "Basal" Gene Regulators Required to Make Seeds?***


Using LCM to Profile Gene Activity in Arabidopsis Seed Regions



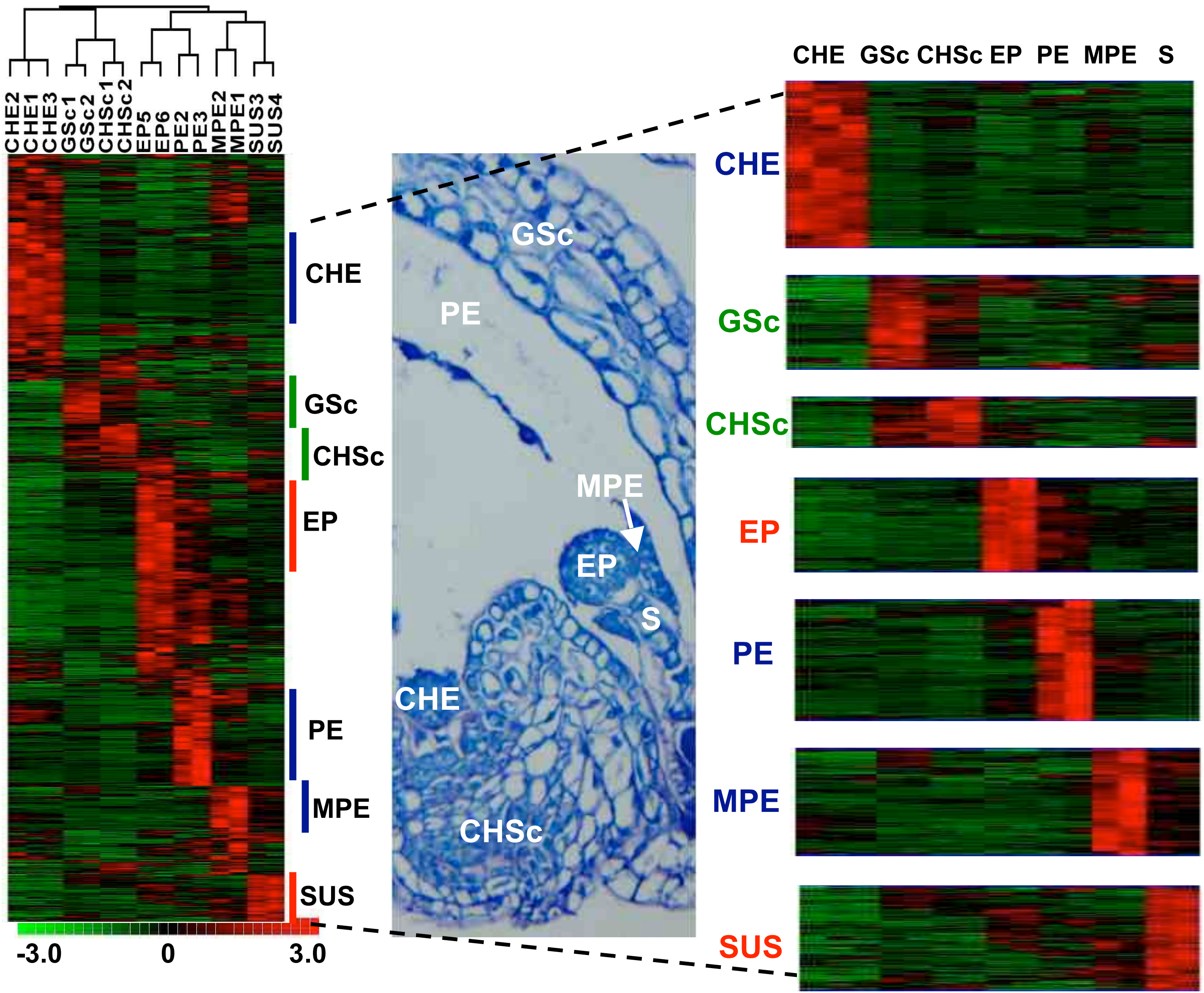
Gene Activity in an Entire Globular-Stage Arabidopsis Seed Immediately After Fertilization

*Unamplified



	GSC	PE	MPE	EP	S	CHE	CHSC
mRNA Present BR1	11,402	11,445	7,844	10,532	11,273	7,709	12,440
mRNA Present BR2	12,316	9,972	10,658	10,674	12,965	7,002	11,818
mRNA Present BR3	-					9,793	-
Total mRNAs	10,513	9,136	7,216	9,130	10,692	7,776	10,794
Unique mRNA	108	54	14	56	95	133	152
Unique TFs	16	3	0	17	8	10	26
Shared	4,028						

Quantitative Regulation of Transcripts Shared by Arabidopsis Globular-Stage Seed Compartments



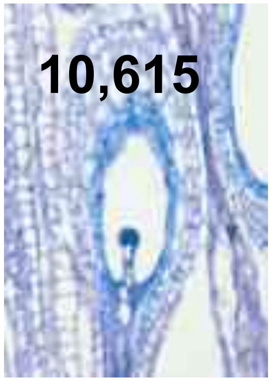
Top 2000 Varying mRNAs

		Over-represented (2-fold, p<0.05)
CHE	204/248	Intracellular Traffic (16) Post-Transcription (11)
GSc	45/112	Energy (4) Metabolism (17)
CHSc	40/83	Energy (1)
EP	114/119	Cell Growth & Division (9) Post-Transcription (13) Protein Synthesis (46)
PE	132/167	Energy (24) Protein Synthesis (28)
MPE	59/114	Intracellular Traffic (2) Protein Destination & Storage (9)
SUS	59/124	Energy (5) Metabolism (20) Protein Destination & Storage (8)

How Many Genes Are Active in Arabidopsis During Early Seed Development?

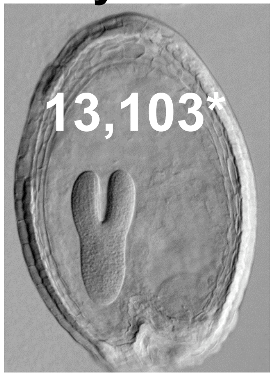
Pre-Globular **Globular** **Linear Cotyledon**

Whole Seed

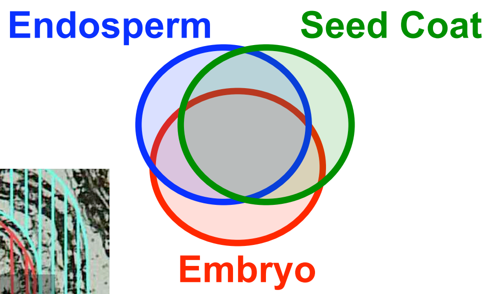


+

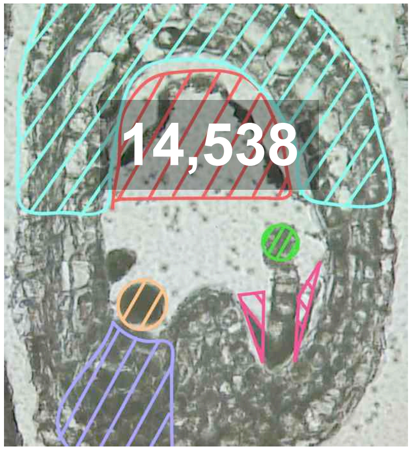
+



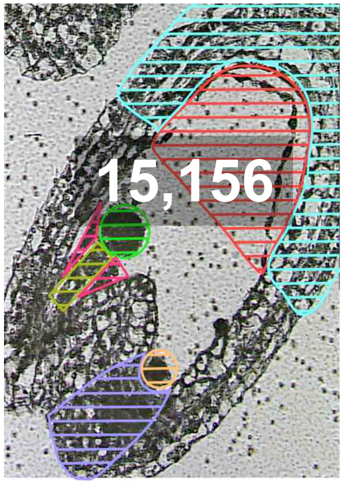
= 15,011^



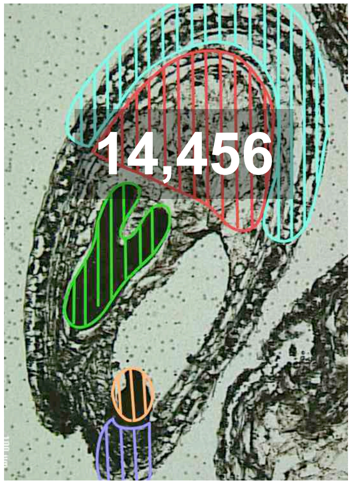
LCM Seed



+



+

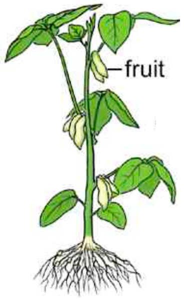


= 16,660^

***Unamplified**

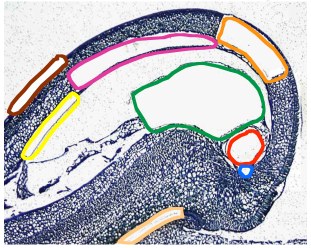
^ <u>Note</u>: Minimum Numbers (82% Whole Genome Chip & Detection Limit)

How Many Genes Are Required to Program Seed Development?
- Comparison of Soybean and Arabidopsis -



Soybean

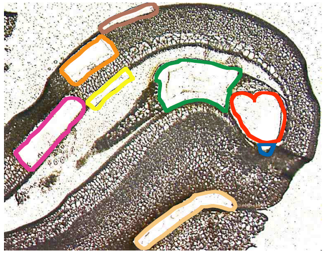
Globular-Stage



21,989
(1,552)

+

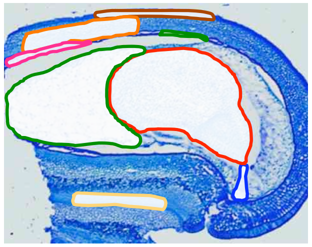
Heart-Stage



22,377
(1,538)

+

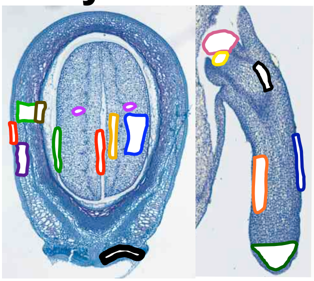
Cotyledon-Stage



22,903
(1,595)

+

Early Maturation

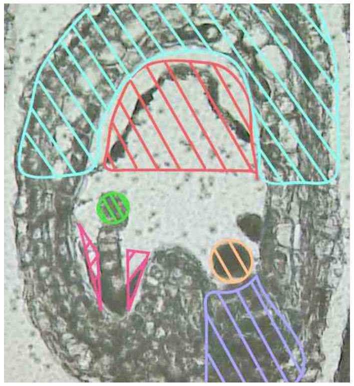


24,003
(1,685)

= 26,359
(1,894)

Arabidopsis

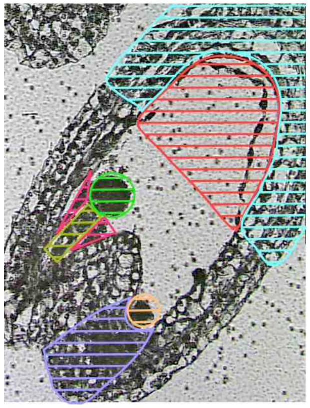
Pre-Globular-Stage



14,538
(1,199)

+

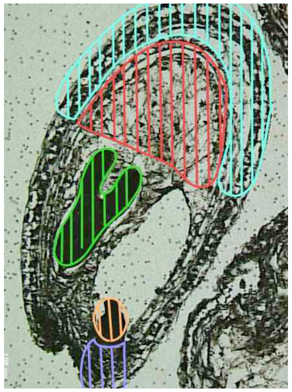
Globular-Stage



15,156
(1,259)

+

Cotyledon-Stage



14,456
(1,210)

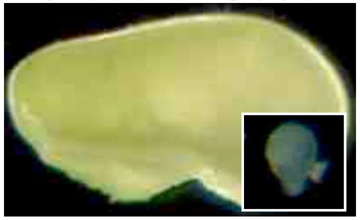
= 16,660
(1,450)

Note: Minimum Numbers

(Transcription Factors)

Are There Additional Genes Active in Soybean Seeds?

Globular Seed



454 Life Science Sequencing

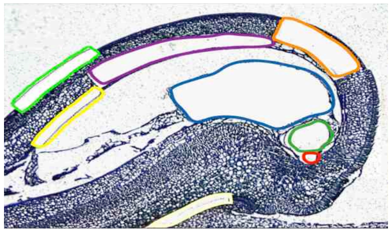
403,511 ESTs

40,674
(Unique EST Clusters)

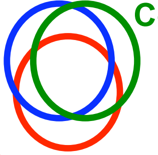
27,901
(Hit Affy Array)

27,901 mRNAs
(2,255 TFs)

Globular LCM



Affymetrix cDNA Array



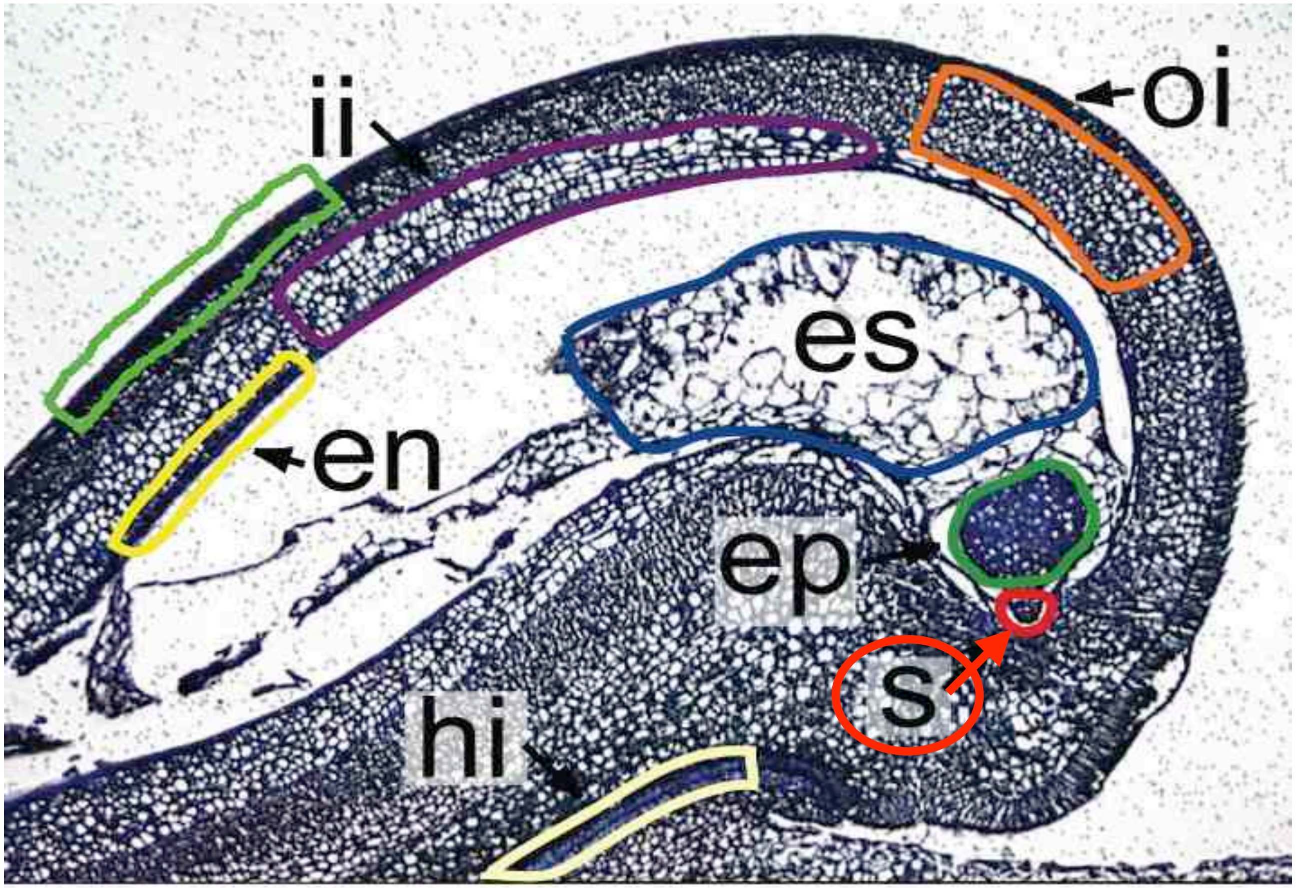
19,938
(Probe Sets)

19,938 mRNAs
(1,341 TFs)

21,989
(Probe Sets)

21,989 mRNAs
(1,552 TFs)

What Are the Regulatory Networks That Activate Genes in Different Seed Compartments?



For Example…. in the Suspensor after Fertilization?

……….To Be Continued by Tomo Kawashima

The End………………..or Is It the Beginning?

A Giant Seed!

GOLDBERG LAB

Current Lab Members

Anhthu Bui
Brandon Le
Chen Cheng
Min Chen
Tomo Kawashima
Jungim Hur
Kelli Henry
Bekah Charney
Daisy Robinton

Former Lab Members

Javier Wagmaister
Xinjun Wang
Shundai Li

Monsanto Collaborators

Dave Somers
John Danzer

UC Davis Collaborators

John Harada
Julie Pellitier
Ryan Kirkbride
Mark Belmonte
Sandra Stone

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