Supporting Information

Hsieh et al. 10.1073/pnas.1019273108

SI Methods

Plant Material. Seeds were plated on 0.5× Murashige and Skoog salts (Caisson Laboratories), Gamborg's Vitamins (Sigma), and 2% sucrose, stratified at 4 °C for 2 d, grown in continuous light in a growth chamber for 10 d, and then transplanted to soil and grown in greenhouse conditions (16 h light). *met1-6* (Col-*gl* ecotype) (1) homozygous plants were obtained from a self-pollinated *met1-6* heterozygote that had never been homozygous for *met1-6*. *fie-1* (2) and *dme-2* (3) heterozygous plants are in L*er* and Col-*gl* backgrounds, respectively.

RNA Isolation from Dissected Seed and Endosperm. Stage 12–13 flower buds were emasculated and pollinated 48 h later. Embryo and endosperm were dissected in 0.3 M sorbitol and 5 mM Mes (pH 5.7) on a slide under a dissecting microscope 7–8 d after pollination, which corresponded to the late torpedo to early bent cotyledon stages of embryogenesis. At this stage of development, bilaterally active genes are expressed from both maternal and paternal genomes in embryo and endosperm (4). Embryos were washed to remove contaminating endosperm. Because the *Arabidopsis thaliana* seed coat has considerable tensile strength, it was possible to separate endosperm from the seed coats and extract RNA from endosperm (5). Approximately 7–10 siliques were dissected, and total RNA was extracted using RNAeasy kit (Qiagen) plus in-column DNase digestion. For WT endosperm, two independent sets of total RNA were obtained from each direction of the crosses. For one set, designated as pure endosperm in Dataset S2, RNA was extracted from dissected endosperm devoid of chalazal endosperm to minimize potential seed coat contamination, whereas for the other set, designated as full endosperm, RNA was extracted from all compartments of endosperm.

Laser Capture Microscopy. Laser capture microdissection was done essentially as described (GEO accession no. GSM311287). Briefly, siliques containing embryos at the linear to bent cotyledon stages (6) were fixed according to the methods of Kerk et al. (7). The embryo proper or the entire endosperm, including micropylar, peripheral, and chalazal domains, was dissected and captured from seed sections using a LMD 6000 system (Leica Microsystems), and RNA was isolated using the RNaqueous microkit (Ambion).

Library Construction. Approximately 30–50 ng total RNA from manually dissected embryo or endosperm or 5–10 ng total RNA from laser capture microdissection (LCM) embryo or endosperm were converted to double-stranded cDNAs using the Ovation RNA-seq System (NuGen Technologies) according to the manufacturer's protocol. This system uses the Ribo-SPIA linear amplification process to amplify cDNA directly from total RNA without the need for poly-A RNA selection or ribosomal RNA depletion. About 400–600 ng Ovation-amplified cDNA was used to construct Illumina sequencing libraries following the standard Illumina genomic DNA library protocol. Illumina sequencing was performed at the Vincent J. Coates Genomic Sequencing Laboratory at University of California at Berkeley and the University of California Los Angeles Microarray Core.

Identification of SNPs Between Col and L*er*. To identify SNPs between Columbia (Col), the reference sequenced accession of *A. thaliana*, and Landsberg *erecta* (L*er*), we sequenced L*er* genomic libraries using the Illumina GA2 platform and mapped these sequences onto the TAIR8 Col scaffold. SNPs were identified with the MAQ aligner (8), yielding a L*er* pseudoscaffold with 402,226 SNPs compared with the Col reference (Dataset S1) at an average 8.7-fold coverage per SNP; a minimum consensus quality of 30 and a minimum twofold coverage per SNP was required. Over 200 SNPs were verified by conventional DNA (Sanger method) sequencing (Dataset S3). We also identified 35 SNPs in the chloroplast genome using the same criteria.

Identification of Imprinted Genes. Illumina (76-bp) reads from cDNA libraries were aligned to Col and L*er* cDNA scaffolds (TAIR8) using Bowtie (9), accepting up to three mismatches in an alignment. Reads that aligned to either Col or L*er* with fewer mismatches were assigned to that ecotype. Each cDNA model (gene) received Col and L*er* scores, each equal to the number of reads aligned to that gene that was assigned to the respective ecotype. Each gene also received a transcriptional score, equal to the number of reads aligned to the cDNA model (irrespective of ecotype) per 1 kb of sequence per 10 million aligned reads. Because all organelles are derived from the female parent, we used the fraction of reads assigned to the female chloroplast genome as a measure of our method's accuracy. For all libraries, 99.3–99.6% of chloroplast reads were assigned to the female genome.

We calculated the probability that a gene's expression deviates from expectation using Fisher's two-tailed exact test. We used the same test to calculate the probability that a gene's imprinted status is altered by mutation of *MET1*, *FERTILIZATION INDEPENDENT ENDOSPERM (FIE)*, or *DEMETER (DME)*. The scores and *P* values for all genes are listed in Dataset S2. To identify imprinted endosperm genes, we summed the scores from both of our library sets (pure and full). In addition to the *P* value cutoffs, a gene's maternal score had to be at least fourfold greater than the paternal score in both reciprocal crosses to be considered maternally expressed or a gene's paternal score had to be at least 1.5-fold greater than the maternal score in both reciprocal crosses to be considered paternally expressed. For a maternally expressed gene to be scored as biallelic in a mutant background, the gene's mutant maternal scores had to be less than fourfold greater than the paternal scores, the maternal/paternal ratio had to be at least twofold greater in WT, and the *P* value for the difference between mutant and WT had to be below 0.001.

Validation of Gene Imprinting. Primers were designed to amplify cDNA fragments that contain at least one SNP between Col-0 and L*er* (Dataset S3). RNA from F1 endosperm was converted to double-stranded cDNAs by standard random-primed first-strand synthesis (Ambion) or by using the Ovation RNA-seq System (NuGen Technologies) according to the manufacturer's protocol. cDNA was sequenced by the Sanger method. Intensity of the L*er* and Col sequence at the SNP was compared based on its Phred score or peak intensity. For genes that did not have a statistically significant number of reads in *met1* endosperm (Table 2), we carried out traditional Sanger sequence validation.

1. Xiao W, et al. (2003) Imprinting of the MEA Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Dev Cell* 5:891–901.
2. Ohad N, et al. (1999) Mutations in *FIE*, a WD polycomb group gene, allow endosperm development without fertilization. *Plant Cell* 11:407–416.
3. Choi Y, et al. (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *arabidopsis*. *Cell* 110:33–42.
4. Kinoshita T, Yadegari R, Harada JJ, Goldberg RB, Fischer RL (1999) Imprinting of the *MEDEA* polycomb gene in the *Arabidopsis* endosperm. *Plant Cell* 11:1945–1952.

5. Gehring M, et al. (2006) DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* 124:495–506.
6. Hsieh T-F, et al. (2009) Genome-wide demethylation of *Arabidopsis* endosperm. *Science* 324:1451–1454.
7. Kerk NM, Ceserani T, Tausta SL, Sussex IM, Nelson TM (2003) Laser capture microdissection of cells from plant tissues. *Plant Physiol* 132:27–35.
8. Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 18:1851–1858.
9. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10.1186/gb-2009-10-3-r25.

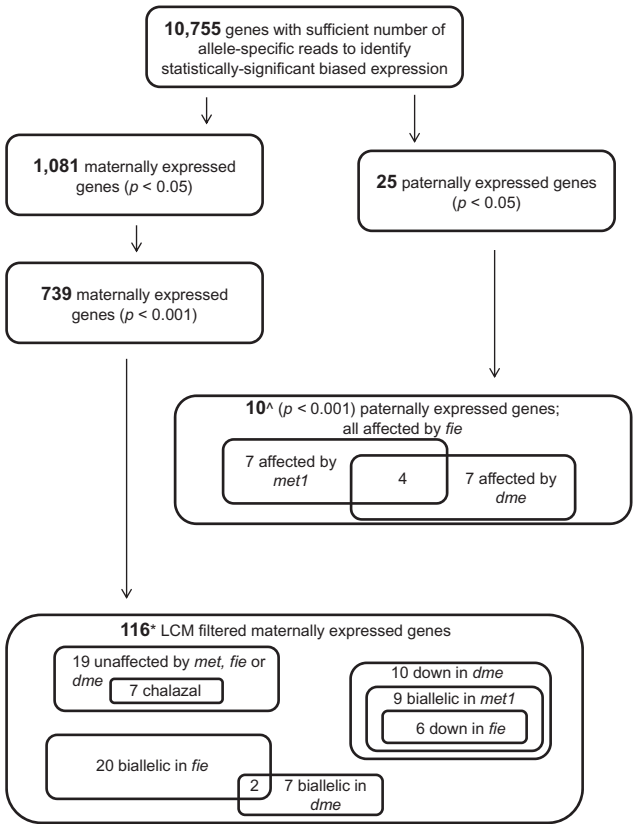

Fig. S1. Flowchart diagram of imprinted gene identification. The terminal diagrams for male- and female-expressed genes are modified Venn diagrams in which numbers within a box indicate the entire contents of the box to maintain consistency with values stated in the manuscript. For example, there are 20 maternally expressed genes with biallelic expression in *fie* and 7 with biallelic expression in *dme*; the two sets have two genes in common. ^Includes *SUVH7* identified by Sanger sequencing. *Includes genes that were close to the LCM cutoff, but were affected by *met1, fie*, or *dme*, and *SUVH8* and *JMJ15* identified by Sanger sequencing.

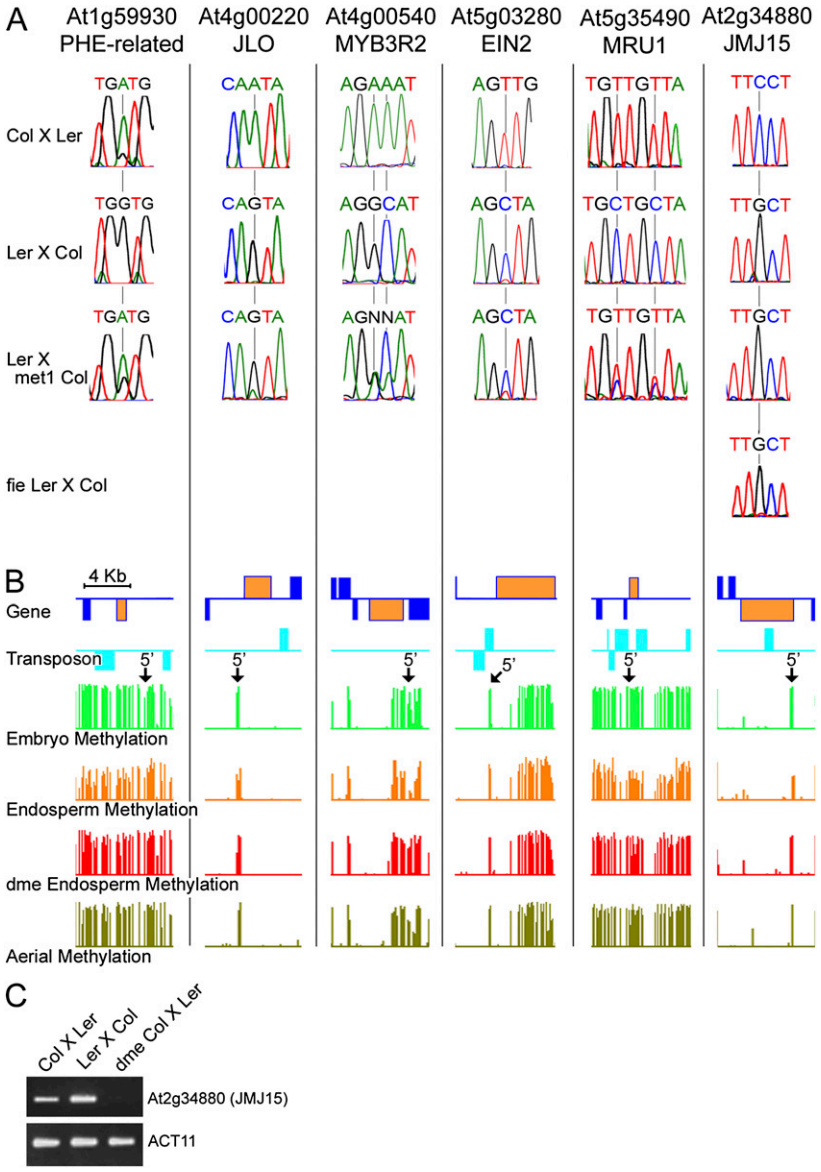

Fig. S2. Maternally expressed imprinted genes. (*A*) RT-PCR sequencing chromatographs at selected SNP regions depicting allele-specific expression in reciprocal crosses between L*er* and Col ecotypes, and expression in female L*er* crossed to male *met1-6* Col-*gl*. (*B*) CG methylation profiles in WT embryo, endosperm, aerial tissue, and *dme* endosperm for genes shown in *A* are displayed. Genes and transposable elements oriented 5′ to 3′ and 3′ to 5′ are shown above and below the line, respectively. Gene models indicated in yellow represent the imprinted genes as shown in *A*. Arrows indicate the 5′ end of imprinted genes where CG demethylation is apparent. (*C*) Expression analysis by semiquantitative RT-PCR in WT reciprocal crosses between L*er* and Col ecotypes and female *dme-2* Col-*gl* crossed to male WT L*er*.

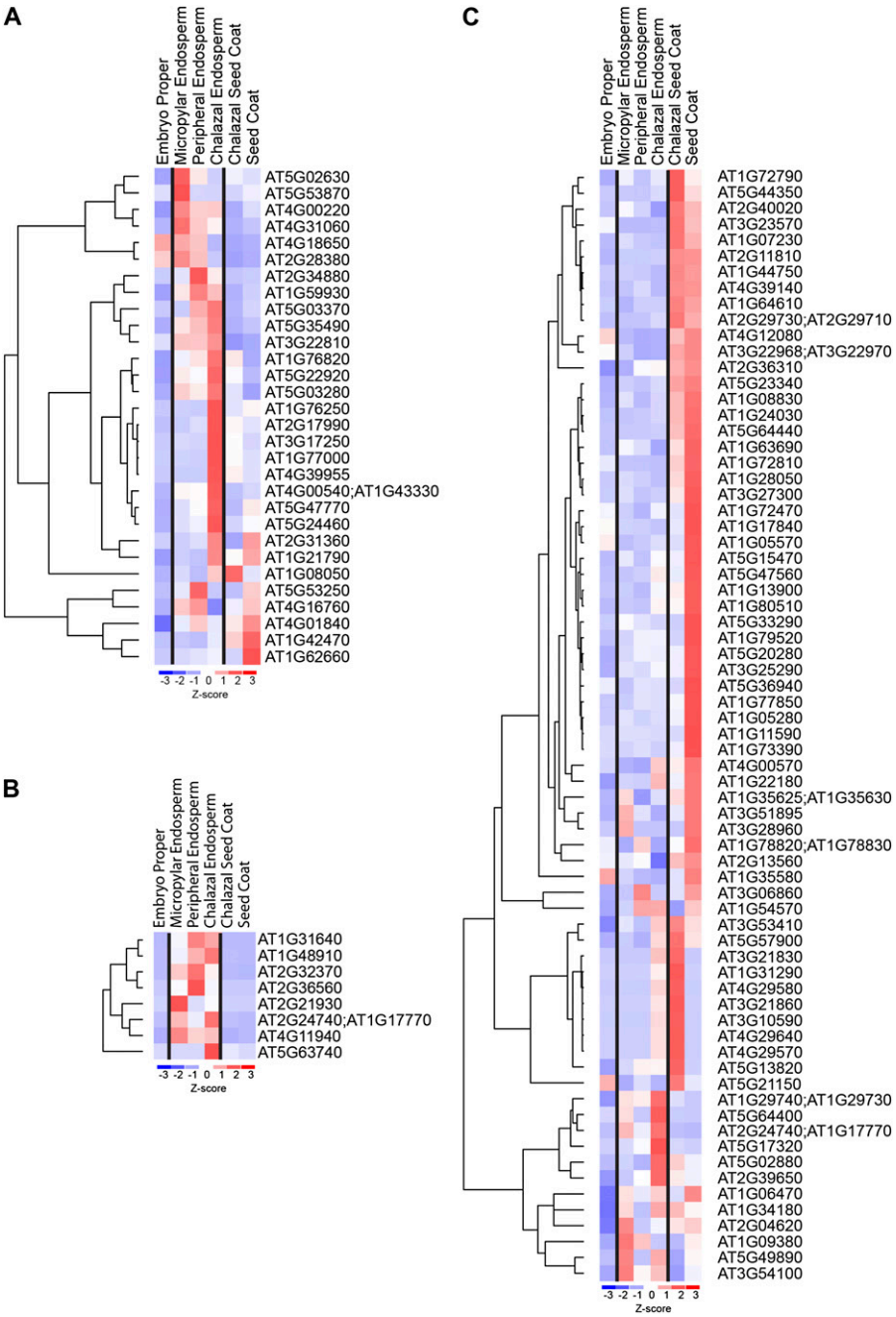

Fig. S3. Unsupervised hierarchical clustering of expression patterns from LCM microarray data (GEO accession no. GSE12404). Hierarchical clustering was carried out using dChip 2008 software (1). Heat maps are based on signal intensities from the Arabidopsis LCM dataset using the average signal value of biological replicates. Signals for preglobular, globular, heart, linear cotyledon, and mature green stages were averaged for each seed compartment. Coloration is based on the number of SDs away from mean signal of a gene in all compartments (*Z* score). Color intensity indicates deviation away from the mean signal for a particular gene: red, higher; blue, lower. Primary red and primary blue are +3 and −3 SDs, respectively. Heat maps have had vertical bars added to differentiate compartments from one another. (*A*) Maternally expressed genes that are regulated by *MET1*, *DME*, and/or *FIE*. Genes not represented on the ATH1 microarray are *AT2G17690*, *AT1G69900*, *AT5G42235*, *AT1G77960*, and *AT4G18150*. (*B*) Paternally expressed genes. Genes not represented on the ATH1 microarray are *AT1G57800* and *AT1G60410*. (C) Maternally expressed genes not regulated by *MET1*, *DME*, or *FIE*. Genes not represented on the ATH1 microarray are *AT1G54575*, *AT1G61090*, *AT1G61097*, *AT3G05700*, *AT3G54740*, *AT4G15080*, *AT4G16180*, *AT4G18320*, *AT5G02970*, *AT5G22200*, and *AT5G29860*.

1. Li C, Wong WH (2001) Model-based analyses of oligonucleotide arrays: Model validation, design issues and standard error application. *Genome Biol* 2:RESEARCH0032.

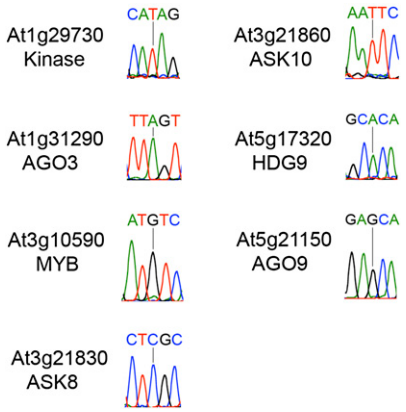

Fig. S4. Expression of genes in LCM endosperm tissue. Female Col-*gl* was crossed with male L*er*, and F1 endosperm was obtained through laser capture microdissection (LCM). RNA was isolated and used to generate RT-PCR sequencing chromatographs at selected SNP regions measuring allele-specific expression.

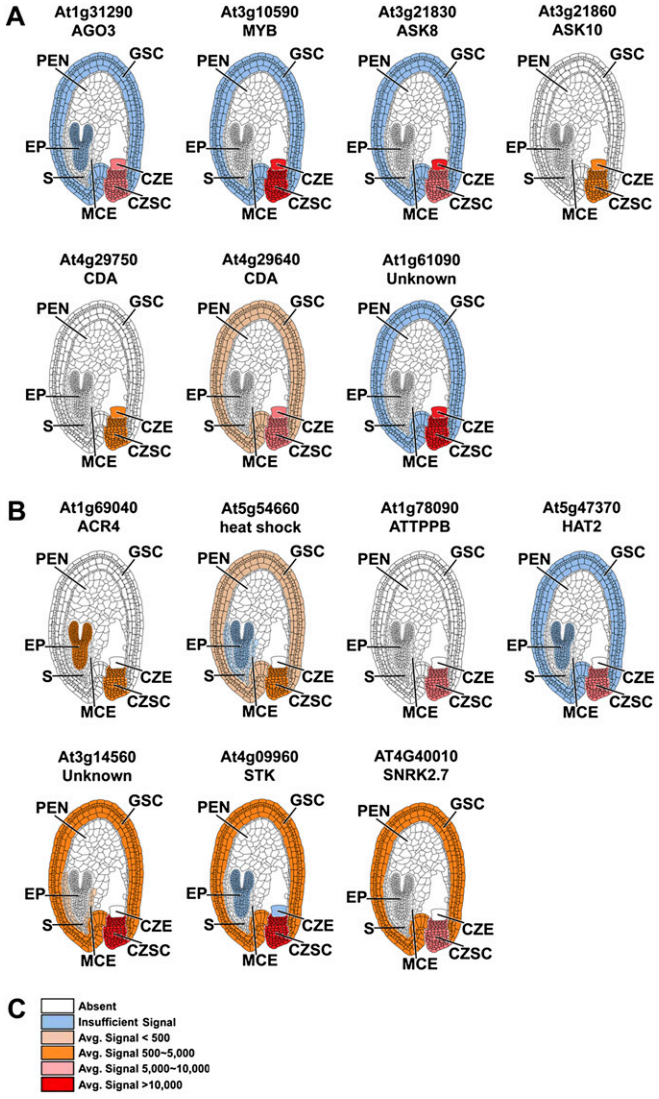

Fig. S5. Seed compartment-specific expression patterns from publicly available LCM microarray data (http://seedgenenetwork.net/arabidopsis). (*A*) Seven genes maternally expressed in the chalazal seed coat (CZSC) and chalazal endosperm (CZE). (*B*) Seven genes expressed in the CZSC. (*C*) Color-coded bars represent relative hybridization signals shown in *A* and *B*. Additional abbreviations for seed compartments: EP, embryo proper; GSC, general seed coat; MCE, micropylar endosperm; PEN, peripheral endosperm; S, suspensor. Abbreviations for gene names: CDA, cytidine deaminase; AGO3, ARGONAUTE 3*; ASK, Arabidopsis skip1-like; ACT, act repeat 4 amino acid binding; ATTPPB, A. thaliana trehalose-6-phosphate phosphatase; HAT2, homeodomain A. thaliana transcription factor; STK, seed stick transcription factor; SNRK2.7, SNF1-related protein kinase 2.7.*

Table S1. Previously described imprinted genes

Number	Annotation	CxL M/P	LxC M/P	*met1* M/P	*fie* M/P	*dme* M/P	Endo exp*	Emb exp*	*met1* exp	*fie* exp	*dme* exp
AT1G02580	MEA	NA	NA	NA	NA	NA	15/0	1/0	6	169	544
AT1G65330	PHE1	NA	NA	NA	NA	NA	0/0	0/0	0	9,737	4,795
AT2G32370 #/F/D/M	HDG3	43/126	42/216	2/1	527/595	208/123	108/256	8/0	3	1,798	304
AT2G35670 $	FIS2	9/0	4/0	0/0	21/2	3/0	9/13	1/1	3	65	3
AT3G03260	HDG8	17/0	4/2	0/0	757/618	583/177	2/5	0/0	0	547	243
AT3G19350	MPC	0/0	0/0	0/0	2/0	2/0	0/0	0/0	0	11	13
AT4G00540 #/m/d	MYB3R2	100/0	75/2	2/7	118/20	36/2	18/28	0/1	8	44	7
AT4G25530	FWA/HDG6	3/0	2/0	1/3	333/1	3/0	6/8	1/3	5	657	6
AT5G17320 #	HDG9	45/2	208/7	2/0	409/15	59/4	75/203	13/1	10	314	44
AT5G54650	FH5	340/53	186/160	37/20	547/210	885/140	45/8	29/11	29	171	164
AT5G62110		2/5	1/2	0/0	295/166	184/99	1/3	0/0	0	185	91

Total maternal (M) and paternal (P) reads are shown for the indicated genotypes, as well as transcriptional scores (number of reads per kb of sequence per 10 million aligned reads) for endosperm (Endo exp), embryo (Emb exp), and the indicated mutant genotypes.
*Transcriptional scores derived from manually-dissected and LCM tissue are shown before and after the slash (/), respectively. # = scored as imprinted with $p < 0.001$; $ = scored as imprinted with $p < 0.05$; F = biallelic in *fie* endosperm; D = biallelic in *dme* endosperm; M = downregulated in *met1* endosperm; m = biallelic in *met1* endosperm; d = downregulated in *dme* endosperm; NA = no SNPs between Col and L*er*.

Table S2. Maternally expressed genes that are biallelic in *fie* endosperm

Number	Annotation	CxL M/P	LxC M/P	*met1* M/P	*fie* M/P	*dme* M/P	Endo exp*	Emb exp*	*met1* exp	*fie* exp	*dme* exp
AT1G08050	Zinc finger	24/1	42/4	1/0	35/34	25/4	81/36	11/8	66	476	136
AT1G21790		66/3	57/6	25/3	144/38	222/16	500/94	148/60	640	2,282	1,753
AT1G69900 D		61/4	64/10	5/6	1,501/567	767/322	75/54	19/0	38	3,897	1,380
AT1G76250		140/3	36/1	11/0	36/31	168/6	44/29	2/0	70	90	139
AT1G76820	GTP binding	45/4	50/12	3/0	72/38	155/15	71/19	15/1	18	300	337
AT1G77000	SKP2B	279/64	264/60	69/6	1,602/874	728/201	150/119	24/4	166	1,926	605
AT2G17990		311/32	185/45	41/6	584/261	456/57	430/263	315/345	240	2,751	1,185
AT2G31360	ADS2	369/5	174/28	180/13	243/107	570/41	2,448/507	1,134/1,056	6,124	6,208	5,705
AT3G17250	PP2C-related	33/8	64/8	9/0	147/76	90/11	220/255	150/45	156	1,073	311
AT3G22810	PI binding	263/48	328/32	40/10	162/185	116/20	138/100	38/35	85	208	50
AT4G01840	TPK5	96/3	68/5	15/1	74/40	77/1	30/12	34/26	22	46	31
AT4G16760	ACX1	898/92	645/86	235/9	1,641/654	4,928/678	76/38	104/107	81	396	761
AT5G02630		31/0	26/0	5/0	91/26	100/21	6/0	1/0	9	51	38
AT5G03370	Acylphosphatase	168/13	83/16	13/1	130/46	119/23	42/21	45/74	15	103	59
AT5G22920	Zinc finger	64/7	55/10	2/0	57/31	51/7	62/48	81/10	17	144	87
AT5G24460		342/79	278/43	41/8	804/307	1,087/192	194/181	268/150	169	770	708
AT5G42235		2,226/160	4,438/949	798/13	4,325/1,649	6,478/457	1,907/457	147/2	1,544	5,202	4,732
AT5G47770 D	FPS1	45/6	111/15	18/1	382/155	202/70	66/60	47/66	59	754	298
AT5G53250	AGP22	58/2	28/0	4/0	21/53	70/11	24/6	0/0	12	91	63
AT5G53870	Plastocyanin-like	241/21	222/13	50/44	3,781/2,611	4,519/30	38/9	7/0	43	1,154	1,004

A list of maternally expressed imprinted genes that are biallelically expressed in *fie* endosperm. Total maternal (M) and paternal (P) reads are shown for the indicated genotypes, as well as transcriptional scores (number of reads per kb of sequence per 10 million aligned reads) for endosperm (Endo exp), embryo (Emb exp), and the indicated mutant genotypes.
*Transcriptional scores derived from manually-dissected and LCM tissue are shown before and after the slash (/), respectively. D = biallelic in *dme* endosperm; ADS = ARABIDOPSIS DESATURASE; PP2C = PROTEIN PHOSPHATASE 2C; PI = phosphoinositide; TPK5 = tandem-pore K+ channel; ACX1 = ACYL-COA OXIDASE 1; FPS1 = FARNESYL DIPHOSPHATE SYNTHASE 1; AGP22 = ARABINOGALACTAN PROTEIN 22.

Dataset S1. SNPs that distinguish L*er* and Col-0

Dataset S1

Dataset S2. Statistical assignments and maternal, paternal, and transcription scores for all genes

Dataset S2

end_M_*P* < 0.05 = maternally expressed endosperm imprinted genes, *P* < 0.05; end_M_*P* < 0.001 = maternally expressed endosperm imprinted genes, *P* < 0.001; end_M_*P* < 0.001_LCM-filtered = maternally expressed endosperm imprinted genes, *P* < 0.001, filtered by endosperm LCM expression data (details in the text); end_M_met1 = maternally expressed endosperm genes, the imprinting of which is significantly affected by a paternal met1 mutation; end_M_fie = maternally expressed endosperm genes, the imprinting of which is significantly affected by a maternal fie mutation; end_M_dme = maternally expressed endosperm genes, the imprinting of which is significantly affected by a maternal dme mutation; end_M_no_mutant = maternally expressed endosperm genes, the imprinting of which is not significantly affected by met1, fie, or dme; chalazal = gene expressed only in chalazal endosperm and seed coat (details in the text); end_P_*P* < 0.05 = paternally expressed endosperm imprinted genes, *P* < 0.05; end_P_*P* < 0.001 = paternally expressed endosperm imprinted genes, *P* < 0.001; emb_M_*P* < 0.001 = maternally expressed embryo imprinted genes, *P* < 0.001 (details in the text); emb_P_*P* < 0.001 = paternally expressed embryo imprinted genes, *P* < 0.001 (details in the text); CxL_p_M = Col × L*er* pure endosperm (*Methods*) maternal score; CxL_p_*P* = Col × L*er* pure endosperm paternal score; LxC_p_*P* = L*er* × Col pure endosperm paternal score; LxC_p_M = L*er* × Col pure endosperm maternal score; M/*P* = Log 2 of the maternal score divided by the paternal score; Fisher:: twotailed = Fisher's two-tailed *P* value (*Methods*); CxL_f = Col × L*er* full endosperm (*Methods*); CxL_a = sum of Col × L*er* pure and full endosperm scores (used for analysis presented in the text); dme_M = dme endosperm maternal score; dme_*P* = dme endosperm paternal score; fie_M = fie endosperm maternal score; fie_*P* = fie endosperm paternal score; met1_M = met1 endosperm maternal score; met1_*P* = met1 endosperm paternal score; CxL_p_n = Col × L*er* pure endosperm expression score (number of reads per kilobase of sequence per 10 million aligned reads); LxC_p_n = L*er* × Col pure endosperm expression score; end_*P* = pure endosperm average expression score; CxL_f_n = Col × L*er* full endosperm expression score; LxC_f_n = L*er* × Col full endosperm expression score; end_f = full endosperm average expression score; end_a = average of end_p and end_f scores (reported in the text); LCM_end_n = LCM endosperm expression score; dme_n = dme endosperm expression score; fie_n = fie endosperm expression score; met1_n = met1 endosperm expression score; CxL_emb_M = Col × L*er* embryo maternal score; CxL_emb_*P* = Col × L*er* embryo paternal score; LxC_emb_*P* = L*er* × Col embryo paternal score; LxC_emb_M = L*er* × Col embryo maternal score; CxL_emb_n = Col × L*er* embryo expression score; LxC_emb_n = L*er* × Col embryo expression score; emb_a = average embryo expression score (reported in the text); LCM_emb_n = LCM embryo expression score.

Dataset S3. Validated SNPs that distinguish Let and Col-0

Dataset S3