HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2018 May 19.

Published in final edited form as:

Science. 2017 May 19; 356(6339): 717–721. doi:10.1126/science.aal3096.

Blocking promiscuous activation at cryptic promoters directs cell type specific gene expression

Jongmin Kim[1,2], **Chenggang Lu**[2,†], **Shrividhya Srinivasan**[2,†], **Stephan Awe**[3], **Alexander Brehm**[3], and **Margaret T. Fuller**[2,4,*]

[1]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA 94305-5329, USA

[2]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305-5329, USA

[3]Institut für Molekularbiologie und Tumorforschung, Philipps-Universität Marburg, Marburg, Germany

[4]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305-5329, USA

Abstract

To selectively express cell-type specific transcripts during development, it is critical to maintain genes required for other lineages in a silent state. Here we show in the *Drosophila* male germ line stem cell lineage that a spermatocyte specific zinc finger protein, Kumgang (Kmg), working with the chromatin remodeler, dMi-2, prevents transcription of genes normally expressed only in somatic lineages. By blocking transcription from normally cryptic promoters, Kmg restricts activation by Aly, a component of the testis-meiotic arrest complex (tMAC), to transcripts for male germ cell differentiation. Our results suggest that as new regions of the genome become open for transcription during terminal differentiation, blocking the action of a promiscuous activator on cryptic promoters is a critical mechanism for specifying precise gene activation.

Highly specialized cell types such as red blood cells, intestinal epithelium, and spermatozoa are produced throughout life from adult stem cells. In such lineages, mitotically dividing precursors commonly stop proliferation and initiate a novel, cell-type specific transcription program that sets up terminal differentiation of the specialized cell type. In the *Drosophila* male germ line, stem cells at the apical tip of the testis self-renew and produce daughter cells that each undergo four rounds of spermatogonial mitotic transit amplifying (TA) divisions, after which the germ cells execute a final round of DNA synthesis (pre-meiotic S phase) and initiate terminal differentiation as spermatocytes (Fig. 1A) (1). Transition to the

*Correspondence to: mtfuller@stanford.edu.
†Equal contribution

Supplementary Materials:
Materials and Methods
Supplementary Text
Supplementary figures S1–S15 and legends
Table S4
Captions for table S1–3 and S5–8
References (*28–54*)

spermatocyte state is accompanied by transcriptional activation of >1,500 genes, many of which are expressed only in male germ cells (2). Expression of two thirds of these depends both on a testis-specific version of the MMB/dREAM complex termed the testis meiotic arrest complex (tMAC) and on testis-specific paralogs of TATA binding protein associated factors (tTAFs) (3–5). Although this is one of the most dramatic changes in gene expression in *Drosophila* (6), it is not yet understood how the testis-specific transcripts are selectively activated during the three day spermatocyte period.

Identification of an early differentiation gene, *kumgang*

To identify the first transcripts upregulated at onset of spermatocyte differentiation (fig. S1A, arrow), we genetically manipulated germ cells to synchronously differentiate from spermatogonia to spermatocytes *in vivo* utilizing $bam^{-/-}$ testes, which contain large numbers of overproliferating spermatogonia (fig. S1B and C) (7–8). Brief restoration of Bam expression under heat shock control in *hs-bam;* $bam^{-/-}$ flies induced synchronous differentiation of $bam^{-/-}$ spermatogonia, resulting in completion of a final mitosis, pre-meiotic DNA synthesis and onset of spermatocyte differentiation by 24 hours after Bam expression, eventually leading to production of functional sperm (Fig. 1B and fig. S1D-O and S2; supplementary text). Comparison of transcripts expressed before vs. 24 hours after heat shock of *hs-bam;* $bam^{-/-}$ testes by microarray identified 27 early transcripts significantly upregulated more than 2-fold in testes from *hs-bam;* $bam^{-/-}$ (Fig. 1C, red) but not from $bam^{-/-}$ flies subjected to the same heat shock regime (fig. S3A). Among these was the early spermatocyte marker *RNA binding protein 4* (*Rbp4*) (Fig. 1C and table S1) (9). Notably at this early time point, the transcript for *CG5204*, now named *kumgang* (*kmg*) based on the Korean name of mythological guardians at the gate of Buddhist temples, had the highest fold increase among all 754 *Drosophila* predicted transcription factors (Fig. 1C and table S1) (10).

Kumgang (*CG5204*) encodes a 747 amino acid protein with six canonical C2H2-type zinc finger domains (fig. S4A) expressed in testes but not in ovary or carcass (fig. S3B). Kmg protein was expressed independently from the tMAC component Always early (Aly) or the tTAF Spermatocyte Arrest (Sa) (fig. S3C), and both *kmg* mRNA and protein were upregulated prior to Topi, another component of tMAC (fig. S3D and E) (3, 11). Immunofluorescence staining of wild-type testes revealed Kmg protein expressed specifically in differentiating spermatocytes (Fig. 1D and fig. S3F-H), where it was nuclear and enriched on the partially condensed bivalent chromosomes (Fig. 1E-E″). Consistent with dramatic upregulation of *kmg* mRNA after the switch from spermatogonia to spermatocyte (Fig. 1C and fig. S3D), expression of Kmg was first detected by immunofluorescence staining after completion of pre-meiotic S phase marked by downregulation of Bam (fig. S3I-I″) (12), coinciding with expression of Rbp4 protein (fig. S3J-J″), and before expression of the tTAF, Sa (fig. S3K- K″).

Kmg prevents misexpression of genes normally expressed in somatic cells

Function of Kmg in spermatocytes was required for male germ cell differentiation. Reducing function of Kmg in spermatocytes, either by cell-type specific RNAi knock down (KD) (Fig.

2A, B and fig. S4A, B) or in flies trans-heterozygous for a CRISPR-induced *kmg* frameshift mutant and a chromosomal deficiency ($kmg^{\Delta 7}/Df$) (fig. S3A, C, E and F), resulted in accumulation of mature primary spermatocytes arrested just prior to the G2/M transition for meiosis I and lack of spermatid differentiation. A 4.3 kb genomic rescue transgene containing the 2.3kb *kmg* open reading frame (fig. S4A) fully rescued the differentiation defects and sterility of $kmg^{\Delta 7}/Df$ flies (fig. S4D), confirming that the meiotic arrest phenotype was due to loss of function of Kmg. In both *kmg* KD and $kmg^{\Delta 7}/Df$, Kmg protein levels were less than 5% of wild-type (fig. S4G; supplementary text). $kmg^{\Delta 7}/Df$ mutant animals were adult viable and female fertile but male sterile, consistent with the testis-specific expression.

Function of Kmg was required in germ cells for repression of over 400 genes not normally expressed in wild-type spermatocytes. Although the differentiation defects caused by loss of function of *kmg* appeared similar to the meiotic arrest phenotype of testis-specific *tMAC* component mutants by phase contrast microscopy, analysis of gene expression in *kmg* KD testes showed that many Aly (tMAC) dependent spermatid differentiation genes were expressed, although some at lower level than in wild-type. Among the 652 genes with 256-fold lower expression in $aly^{-/-}$ mutant compared to wild-type testes (Fig. 2D, green dots and table S2), only 4 showed 256-fold lower expression in *kmg* KD compared to sibling control (no Gal4 driver) testes (Fig. 2C, green dots). In contrast, transcripts from over 500 genes were strongly upregulated in *kmg* KD testes with almost no detectable expression in testes from sibling control males (Fig. 2C, red dots). Hierarchical clustering identified 440 genes specifically upregulated in *kmg* KD testes (fig. S5A, red line, and table S3) compared to testes from wild-type, $bam^{-/-}$, $aly^{-/-}$ or $sa^{-/-}$ mutant flies. These 440 genes were significantly associated with Gene Ontology terms such as ‘substrate specific channel activity’ or ‘detection of visible light’ that appeared more applicable to non-germ cell types, such as neurons (fig. S5B). Analysis of published transcript expression data for a variety of *Drosophila* tissues (6) revealed that the 440 were normally not expressed or extremely low in wild-type adult testes, but many were expressed in specific differentiated somatic tissues such as eye, brain or gut (fig. S5C). Confirming misexpression of neuronal genes at the protein level, immunofluorescence staining revealed that the neuronal transcription factor Prospero (Pros) (13), normally not detected in male germ cells, was expressed in clones of spermatocytes homozygous mutant for *kmg* induced by Flp-FRT mediated mitotic recombination (Fig. 2E-E‴). The misexpression of Pros was cell autonomous, occurring only in mutant germ cells. Notably, mid stage to mature spermatocytes homozygous mutant for *kmg* misexpressed Pros (Fig. 2E′-E‴, arrow), but mutant early spermatocytes did not (Fig. 2E′-E‴, arrowhead), indicating that the abnormal upregulation of Pros occurred only after spermatocytes had reached a specific stage in their differentiation program.

Kmg functions with dMi-2 to repress misexpression of normally somatic transcripts

A small-scale cell-type specific RNAi screen of chromatin regulators revealed that knock down of *dMi-2* in late TA cells and spermatocytes resulted in meiotic arrest, similar to loss of function of *kmg* (Fig. 3A and table S4). Immunofluorescence analysis of testes from a

protein trap line in which an endogenous allele of dMi-2 was tagged by GFP revealed that dMi-2-GFP, like the untagged endogenous protein (fig. S6A-A‴), was expressed and nuclear in progenitor cells and spermatocytes, as well as in somatic hub and cyst cells (fig. S6B′ and B″). Notably, dMi-2-GFP co-localized to chromatin with Kmg in spermatocytes (fig. S6B-B′ and B‴–C‴) and the level of dMi-2 protein appeared lower and less concentrated on chromatin in nuclei of *kmg*$^{-/-}$ spermatocytes than in neighboring *kmg*$^{+/+}$ or $^{+/-}$ spermatocytes, suggesting that Kmg may at least partially help recruit dMi-2 to chromatin in spermatocytes (Fig. 3B-B‴ and fig. S6D-F). Furthermore, in testis extracts Kmg co-immunoprecipitated with dMi-2 and vice versa, suggesting that Kmg and dMi-2 form a protein complex in spermatocytes (Fig. 3C). Comparison of microarray data revealed that most of the 440 transcripts upregulated in testes upon loss of function of *kmg* were also abnormally upregulated in *dMi-2* KD testes (Fig. 3D), suggesting that Kmg and dMi-2 may function together to repress expression of the same set of normally somatic transcripts in spermatocytes.

Chromatin immunoprecipitation followed by sequencing (ChIP-Seq) revealed that Kmg protein localized along the bodies of genes actively transcribed in the testis. ChIP-Seq with anti-Kmg identified 798 genomic regions strongly enriched by immunoprecipitation of Kmg (peaks, q value < 10^{-10}) from wild-type but not from *kmg* KD testes. 698 of the 798 robust Kmg ChIP-Seq peaks overlapped with exonic regions of 680 different genes actively transcribed in testes (Fig. 3E and fig. S7A, C, E, G, I and table S5). The enrichment was often strongest just downstream of the transcription start site (TSS), but with substantial enrichment along the gene body as well (Fig. 3E and fig. S7A, C, E, G).

ChIP-Seq with anti-dMi-2 also showed enrichment along the gene bodies of the same 680 genes bound by Kmg, with a similar bias just downstream of the TSS (Fig. 3E and fig. S7B, D, F, H). The dMi-2 ChIP signal along these genes was partially reduced in *kmg* KD testes, suggesting that Kmg may recruit dMi-2 to the bodies of genes actively transcribed in the testis (Fig. 3E and fig. S7B, D, F, H).

RNA-Seq analysis revealed that the 680 genes bound by Kmg were strongly expressed in testes (Fig. 3E and fig. S7I-K) and most strongly enriched in the GO term categories ‘spermatogenesis’ and ‘male gamete generation’ (fig. S8A). One third of the genes bound by Kmg were robustly activated as spermatogonia differentiate into spermatocytes, and were much more highly expressed in the testes than in other tissues (fig. S8B, red rectangle). The median levels of transcript expression of most of the 680 Kmg bound genes did not change significantly upon loss of Kmg (fig. S7J and K).

Genes that are normally transcribed in somatic cells that became upregulated upon loss of Kmg function in spermatocytes for the most part did not appear to be bound by Kmg. Only 3 of the 440 genes upregulated in *kmg* KD overlapped with the 680 genes with robust Kmg peaks, suggesting that Kmg may prevent misexpression of normally somatic transcripts either indirectly or by acting at a distance.

Kmg and dMi-2 prevent misexpression from cryptic promoters

Inspection of RNA-Seq reads from *kmg* and *dMi-2* KD testes mapped onto the genome showed that about 80% of the transcripts detected as misexpressed in knock down compared to wild-type testes by microarray analysis did not initiate from the promoters used in the somatic tissues in which the genes are normally expressed (Fig. 4A and C, arrows). Metagene analysis (Fig. 4E), as well as visualization of RNA expression centered on the TSSs annotated in the Ensembl database (fig. S9A) showed that most of the 143 genes normally expressed in wild-type heads but not in wild-type testes were misexpressed in *kmg* or *dMi-2* KD testes from a start site different from the annotated TSS utilized in heads. Transcript assembly from our RNA-Seq data using Cufflinks for the 143 genes also showed that the transcripts misexpressed in *kmg* or *dMi-2* KD testes most often initiate from a different TSSs than the transcripts from the same gene assembled from wild-type heads (fig. S9B-D; supplementary text).

Of the 440 genes scored as de-repressed in *kmg* KD testes based on microarray, 346 could be assigned with TSSs in *kmg* KD testes based on visual inspection of our RNA-Seq data mapped onto the genome browser (table S6; supplementary text). Of these, only 67 produced transcripts in *kmg* KD testes that started within 100bp of the TSS annotated in the Ensembl database based on the tissue(s) in which the gene was normally expressed. In contrast, for rest of the 346 genes, the transcripts expressed in *kmg* KD testes started from either a TSS upstream (131/346) or downstream (148/346) of the annotated TSSs. 262 of the 346 genes were misexpressed starting from nearly identical positions in *dMi-2* KD as in *kmg* KD testes (table S6), suggesting that Kmg and dMi-2 function together to prevent misexpression from cryptic promoters.

Kmg prevents promiscuous activity of Aly

Many of the ectopic promoters from which the misexpressed transcripts originated appeared to be bound by Aly, a component of tMAC, in *kmg* KD testes (Fig. 4B, D arrowheads and Fig. 4F, G). ChIP for Aly was performed using anti-HA on testis extracts from flies bearing an Aly-HA genomic transgene able to fully rescue the $aly^{-/-}$ phenotype (fig. S10; see methods). 181 of 346 genes with new TSSs assigned by visual inspection had a region of significant enrichment by ChIP for Aly with its peak summit located within 100bp of the cryptic promoter (Fig 4F and G; supplementary text). Motif analysis by MEME revealed that these regions were enriched for the DNA sequence motif (AGYWGGC) (Fig. 4H and fig. S11). This motif was not significantly enriched in the set of 165 cryptic promoters at which Aly was not detected in *kmg* KD testes (fig. S11B). Enrichment of Aly at the cryptic promoters was much stronger in *kmg* KD compared to wild-type testes (Fig. 4B, D and G), suggesting that in the absence of Kmg, Aly may bind to and activate misexpression from cryptic promoters.

Genetic tests revealed that the misexpression of somatic transcripts in *kmg* KD spermatocytes indeed required function of Aly. The neuronal transcription factor Pros, abnormally upregulated in *kmg* KD or mutant spermatocytes (Fig. 2E-E‴), was no longer misexpressed if the *kmg* KD spermatocytes were also mutant for *aly* (Fig. 5A, B and fig.

S12A-F), even though germ cells in *kmg* KD*; aly*$^{-/-}$ testes appear to reach the differentiation stage at which Pros turned on in the *kmg* KD germ cells (fig. S13; supplementary text). Assessment by qRT-PCR revealed that misexpression of 5/5 transcripts in *kmg* KD testes also required function of Aly (fig. S12G). Global transcriptome analysis of *kmg* KD vs. *kmg* KD*; aly*$^{-/-}$ testes by microarray showed that the majority of the 440 genes de-repressed due to loss of function of *kmg* in spermatocytes were no longer abnormally upregulated in *kmg* KD; *aly*$^{-/-}$ testes (Fig. 5C). Even genes without noticeable binding of Aly at their cryptic promoters were suppressed in *kmg* KD; *aly*$^{-/-}$, suggesting Aly may regulate this group of genes indirectly.

Together our ChIP and RNA-Seq data show that Kmg and dMi-2 bind actively transcribed genes (Fig. 3E and fig. S7) (14), but are required to block expression of aberrant transcripts from other genes normally silent in testes. The mammalian ortholog of dMi-2, CHD4 (Mi-2β) has been shown to bind active genes in mouse embryonic stem cells (15) or T lymphocyte precursors (16), but also plays a role in ensuring lineage specific gene expression in other contexts (17–18). We cannot rule out that Kmg and dMi-2 might also act directly at the cryptic promoter sites but that our ChIP conditions did not capture their transient or dynamic binding, as several chromatin remodelers or transcription factors such as the thyroid hormone receptor have been difficult to detect by ChIP (19–20). Kmg and dMi-2 may repress misexpression from cryptic promoters indirectly by activating as yet unidentified repressor proteins. However, it is also possible that Kmg and dMi-2 act at a distance by modulating chromatin structure or confining transcriptional initiation or elongation licensing machinery to normally active genes.

Indeed, changes in the genomic localization of Aly protein in wild-type vs. *kmg* KD testes raised the possibility that Kmg may in part prevent misexpression from cryptic promoters by concentrating Aly at active genes. Of the 1,903 Aly peaks identified by ChIP from wild-type testes, the 248 Aly peaks that overlapped with strong Kmg peaks showed an overall reduction in enrichment by ChIP of Aly from *kmg* KD testes compared to wild-type (Fig. 6A–C, Case #1 and fig. S14A and B). In contrast, the Aly peaks at cryptic promoters were more robust in *kmg* KD testes than in wild-type (Fig. 4G). In general over the genome, 4,129 new Aly peaks were identified by ChIP from *kmg* KD testes that were absent or did not pass the statistical cutoff in wild-type testes (Fig. 6A–C, Case #2 and fig. S14A). More than 30% of the genomic regions with new Aly peaks in *kmg* KD showed elevated levels of RNA expression starting at or near the Aly peak in *kmg* KD but not in wild-type testes (Fig. 6D and fig. S14C), suggesting that misexpression of transcripts from normally silent promoters in *kmg* KD testes is more widespread than initially assessed by microarray. Together, these findings raise the possibility that Kmg may prevent misexpression of aberrant transcript by concentrating Aly to active target genes in wild-type testes, so preventing binding and action of Aly at cryptic promoter sites.

Our results suggest that selective gene activation is not always mediated by a precise transcriptional activator, but can instead be directed by combination of a promiscuous activator and a gene-selective licensing mechanism (fig. S15A). Cryptic promoters may become accessible as chromatin organization is reshaped to allow expression of terminal differentiation transcripts that were tightly repressed in the progenitor state. We posit that

this chromatin organization makes a number of sites accessible for transcription dependent on the testis-specific tMAC complex component Aly. In this context, activity of Kmg and dMi-2 is required to prevent productive transcript formation from unwanted initiation sites, potentially by confining Aly to genes actively transcribed in the testis and limiting the amount of Aly protein acting at cryptic promoters.

The initiation of novel transcripts from cryptic promoters is reminiscent of loss of function of Ikaros, a critical regulator of T and B cell differentiation (21) and a tumor suppressor in the lymphocyte lineage (22–23). Like Kmg, Ikaros is a multiple zinc finger protein associated with Mi-2β, which binds to active genes in T and B cell precursors (16, 24). Notably, in T-cell lineage acute lymphoblastic leukemia (T-ALL) associated with loss of function of Ikaros, cryptic intragenic promoters were activated leading to expression of ligand independent Notch1 protein, contributing to leukemogenesis (25). Thus, in addition to being detrimental for proper differentiation, firing of abnormal transcripts from normally cryptic promoters due to defects in chromatin regulators may contribute to tumorigenesis through generation of novel oncogenic proteins.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank L. Di Stefano, H. White-Cooper, X. Chen, F. Port, S. Bullock, A. Kuo, G. Crabtree, S. Park, S. Kim, J. Wysocka and G. Ramaswami for advice and generously sharing data and reagents, members of the Fuller lab. for discussions and critical reading of the manuscript, the Stanford CSIF for fluorescence microscopy (NCRR grants S10RR017959-01 and 1S10OD010580) and the SFGF for high-throughput sequencing. Genomic data are available under NCBI GEO (GSE89506). J. K. was supported by the Anne T. and Robert M. Bass Stanford Graduate Fellowship and the Bruce and Elizabeth Dunlevie Bio-X Stanford Interdisciplinary Graduate Fellowship. Research support: DFG TRR81 to A.B. and S.A.; Kempkes Stiftung to S.A; NIH 5R01GM061986 and the Reed-Hodgson Professorship in Human Biology to M.T.F.

References and Notes

1. Fuller MT. Cold Spring Harbor Laboratory Press. 1993; 1:71.
2. White-Cooper H, Caporilli S. Advances in experimental medicine and biology. 2013; 786:47. [PubMed: 23696351]
3. Beall EL, et al. Genes & development. Apr 15.2007 21:904. [PubMed: 17403774]
4. Hiller MA, Lin TY, Wood C, Fuller MT. Genes & development. Apr 15.2001 15:1021. [PubMed: 11316795]
5. White-Cooper H, Leroy D, MacQueen A, Fuller MT. Development (Cambridge, England). Dec. 2000 127:5463.
6. Chintapalli VR, Wang J, Dow JA. Nature genetics. Jun.2007 39:715. [PubMed: 17534367]
7. Gonczy P, Matunis E, DiNardo S. Development (Cambridge, England). Nov.1997 124:4361.
8. McKearin DM, Spradling AC. Genes & development. Dec.1990 4:2242. [PubMed: 2279698]
9. Baker CC, Gim BS, Fuller MT. Development (Cambridge, England). Oct 1.2015 142:3394.
10. Adryan B, Teichmann SA. Bioinformatics (Oxford, England). Jun 15.2006 22:1532.
11. Perezgasga L, et al. Development (Cambridge, England). Apr.2004 131:1691.
12. Insco ML, Leon A, Tam CH, McKearin DM, Fuller MT. Proceedings of the National Academy of Sciences of the United States of America. Dec 29.2009 106:22311. [PubMed: 20018708]
13. Doe CQ, Chu-LaGraff Q, Wright DM, Scott MP. Cell. May 3.1991 65:451. [PubMed: 1673362]

14. Mathieu EL, et al. Nucleic Acids Res. Jun.2012 40:4879. [PubMed: 22362736]
15. de Dieuleveult M, et al. Nature. Feb 4.2016 530:113. [PubMed: 26814966]
16. Zhang J, et al. Nat Immunol. Jan.2012 13:86.
17. Gomez-Del Arco P, et al. Cell Metab. May 10.2016 23:881. [PubMed: 27166947]
18. O'Shaughnessy-Kirwan A, Signolet J, Costello I, Gharbi S, Hendrich B. Development (Cambridge, England). Aug 1.2015 142:2586.
19. Gelbart ME, Bachman N, Delrow J, Boeke JD, Tsukiyama T. Genes & development. Apr 15.2005 19:942. [PubMed: 15833917]
20. Grontved L, et al. Nat Commun. 2015; 6:7048. [PubMed: 25916672]
21. Georgopoulos K. Nat Rev Immunol. Mar.2002 2:162. [PubMed: 11913067]
22. Mullighan CG, et al. N Engl J Med. Jan 29.2009 360:470. [PubMed: 19129520]
23. Winandy S, Wu P, Georgopoulos K. Cell. Oct 20.1995 83:289. [PubMed: 7585946]
24. Schwickert TA, et al. Nat Immunol. Mar.2014 15:283. [PubMed: 24509509]
25. Gomez-del Arco P, et al. Immunity. Nov 24.2010 33:685. [PubMed: 21093322]
26. Chen X, Lu C, Morillo Prado JR, Eun SH, Fuller MT. Development (Cambridge, England). Jun. 2011 138:2441.
27. Celniker SE, et al. Nature. Jun 18.2009 459:927. [PubMed: 19536255]

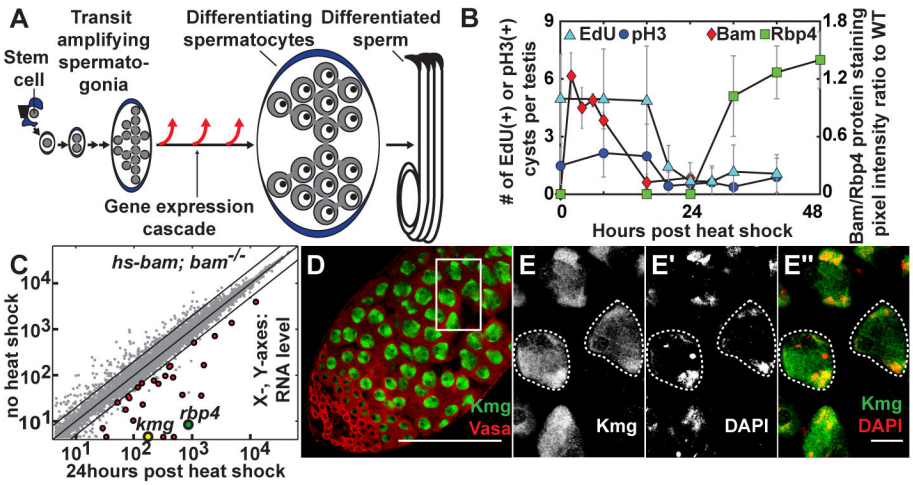

Figure 1. Identification of *kmg* as an early differentiation gene

(A) Male germ line stem cell lineage of *Drosophila melanogaster*.

(B) Number of EdU and pH3 positive cysts per testis and relative expression of Bam and Rbp4-YFP proteins in testes from *hs-bam; bam*$^{-/-}$ flies after transient induction of Bam by heat shock. Error bars: standard deviation (n > 15 testes).

(C) Comparison of transcript levels based on microarray in testes from no heat shock vs. 24 hour post heat shock (PHS) *hs-bam; bam*$^{-/-}$ flies (n = 2) (Red dots) Transcripts significantly upregulated (fold > 2, $p < 0.05$) at 24hr.

(D–E″) Immunofluorescence images of wild-type testis stained for (D) Kmg (green) and Vasa (red). (E- E″) high magnification of box in (D). (E″, merge) Kmg (green), DAPI (red). Dotted outline: spermatocyte nuclei.

Scale bars: 100μm in D; 10μm in E- E″.

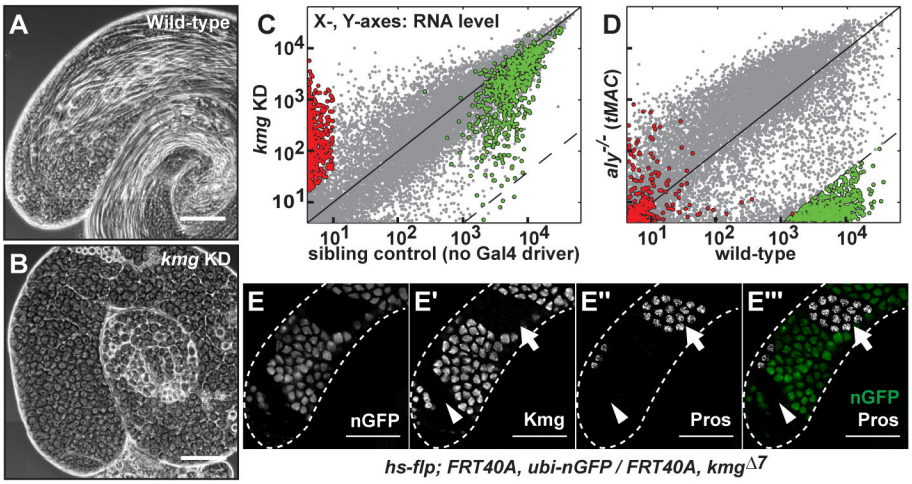

Figure 2. Kmg is required for spermatid differentiation and preventing misexpression of somatic transcripts

(A and B) Phase-contrast images of (A) wild-type and (B) *kmg* KD testes.

(C and D) Comparison of gene expression by microarray between (C) *kmg* KD vs. sibling control (no Gal4 driver) (n = 2) and (D) $aly^{-/-}$ vs. wild-type testes (26) (n = 3). (Red) 555 transcripts significantly upregulated in *kmg* KD testes. (Green) 652 most Aly-dependent transcripts (see methods for statistical cut-offs).

(E-E‴) Immunofluorescence images of (E) GFP visualized by native fluorescence, (E′) Kmg, (E″) Prospero (Pros) in testis bearing germ line clones homozygous mutant for *kmg* marked by absence of GFP. (E‴, merge) GFP (green) and Pros (white). (Arrowhead) young $kmg^{-/-}$ spermatocytes; (Arrow) mature $kmg^{-/-}$ spermatocytes. White dotted lines: testis outline.

Scale bars: 100μm.

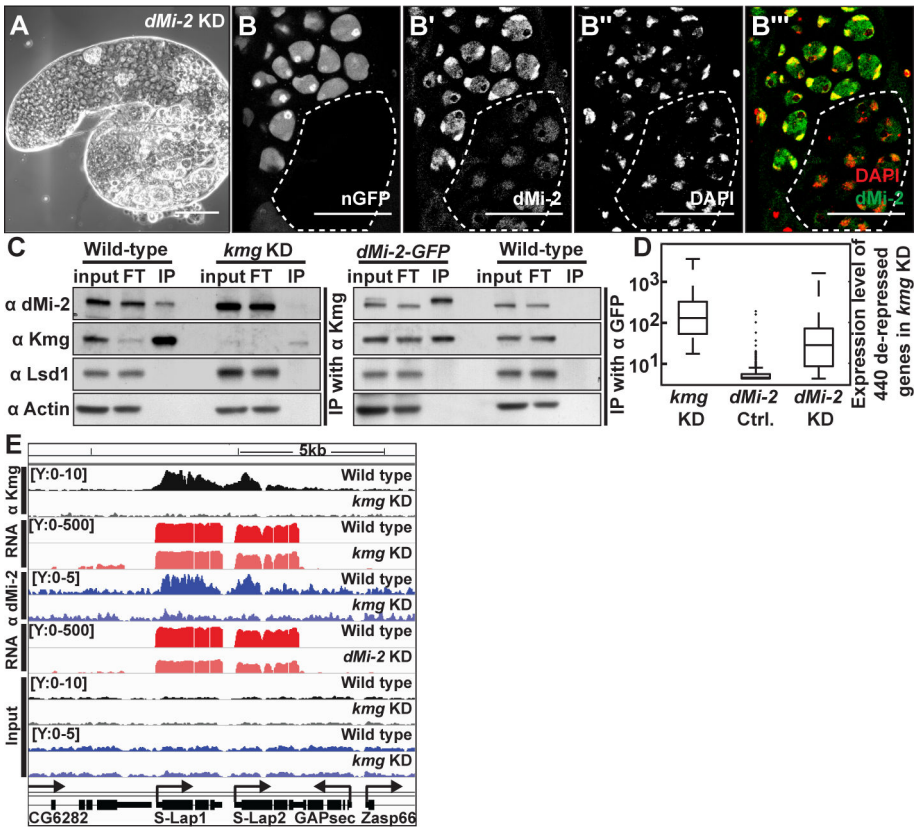

Figure 3. Kmg recruits dMi-2 to actively transcribed genes in spermatocytes

(A) Phase-contrast image of a *dMi-2* KD testis.

(B-B‴) Immunofluorescence images of (B) GFP visualized by native fluorescence, (B′) dMi-2, (B″) DAPI, (B‴, merge) DAPI (red), dMi-2 (green) in testis bearing germ line clones homozygous mutant for *kmg* (dotted line), marked by absence of GFP.

(C) Western blot of dMi-2, Kmg, Lysine Specific Demethylase 1 (Lsd1), and Actin after immunoprecipitation of (left) Kmg from wild-type and *kmg* KD or (right) GFP from *dMi-2-GFP* and wild-type testis lysates. Input: (Kmg IP) 15% or (dMi-2-GFP IP) 7.5% of lysate, FT: flow through after immunoprecipitation (equal proportion as input), IP: complete eluate after immunoprecipitation. Note: two bands in input lane from dMi-2-GFP testes, which express both dMi-2-GFP and untagged dMi-2.

(D) Box plot of expression of 440 de-repressed somatic transcripts in *kmg* KD, *dMi-2* sibling control (no Gal4 driver) and *dMi-2* KD testes. (whiskers) the most extreme data points excluding outliers.

(E) Genome browser screenshot showing ChIP-Seq results for Kmg and dMi-2. Y-axes: normalized read counts based on total one million mapped reads per sample. Y-axes for RNA-Seq reads are in log scale.

Scale bars: 100μm in A; 50μm in B-B‴.

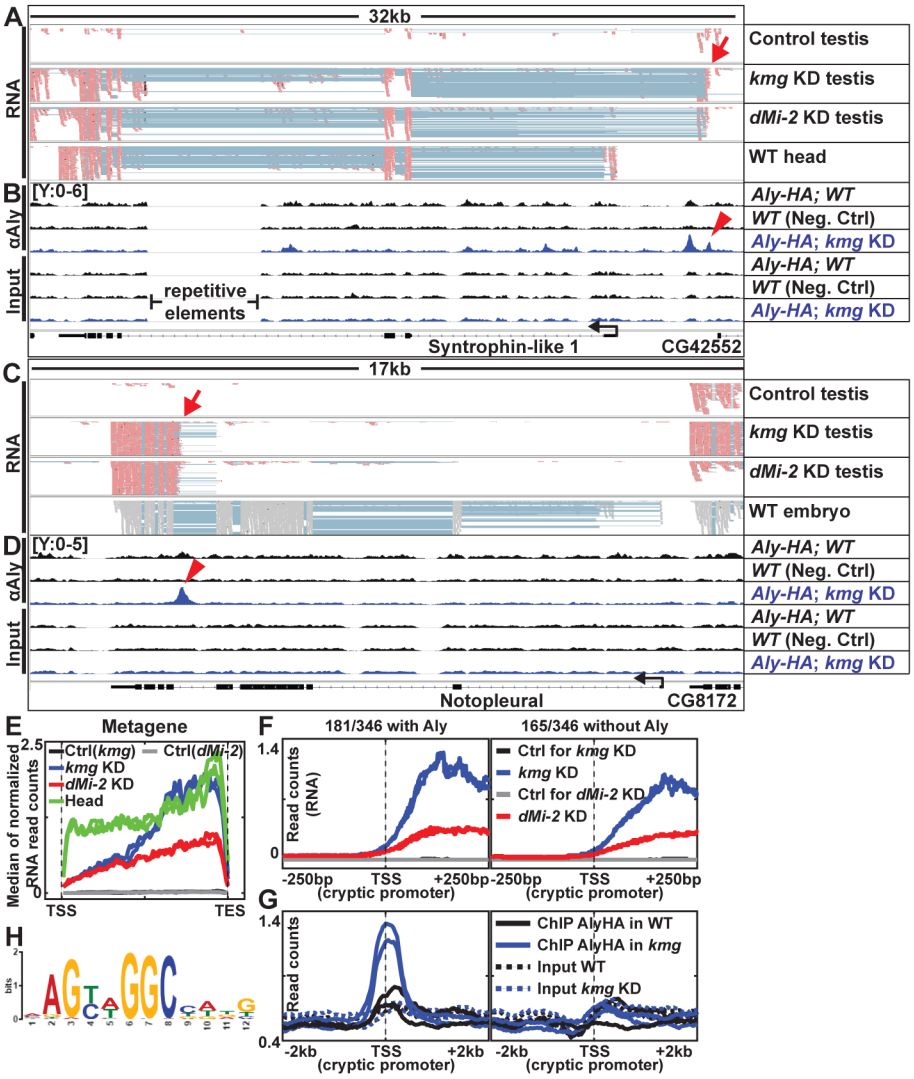

Figure 4. Kmg and dMi-2 prevent misexpression from cryptic promoters

(A and C) Genome browser screenshots of RNA-Seq reads mapped to Crick (−) strand (pastel red) or mapped without strand information (gray) (embryo data from modENCODE) (27). Pastel blue lines: a single read that skipped an intron and mapped to two adjacent exons. Red arrows: cryptic promoter sites.

(B and D) Aly-HA ChIP and input read density in wild-type (with and without Aly-HA transgene) and *kmg* KD testes for the same regions. Red arrowheads: peak of enrichment by ChIP for Aly coinciding with potential cryptic promoter sites.

(E) Median RNA expression profile of 143 genes expressed in wild-type heads but not in wild-type testes and misexpressed in *kmg* KD testes, plotted as a metagene excluding introns. TSS: Transcription start site. TES: Transcription end site.

(F and G) Median profile of (F) RNA-Seq and (G) Aly ChIP (solid lines) and corresponding input (dotted lines) signals centered at cryptic promoters.

(H) Position weight matrix representation of sequence motif most significantly enriched in a collection of 181 cryptic promoters (±150bp around the TSS) with peaks of Aly binding in *kmg* KD testes.

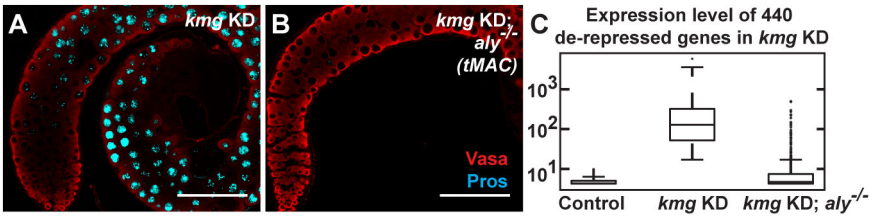

Figure 5. Aly is required for misexpression of aberrant transcripts in *kmg* KD testes

(A and B) Immunofluorescence images of (A) *kmg* KD and (B) *kmg* KD; $aly^{-/-}$ testes stained for Vasa (red) and Pros (cyan).

(C) Box plot of expression of 440 somatic transcripts de-repressed in *kmg* KD testes, showing transcript levels in *kmg* sibling control (no Gal4 driver), *kmg* KD, and *kmg* KD; $aly^{-/-}$testes. (whiskers) the most extreme data points excluding outliers.

Scale bars: 100μm.

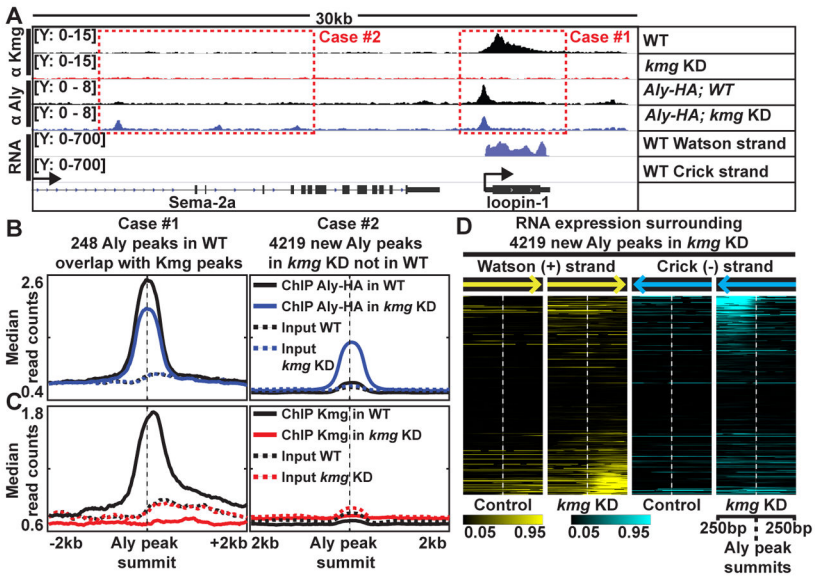

Figure 6. Kmg prevents promiscuous activity of Aly

(A) Genome browser screenshot showing ChIP-Seq results for Kmg and Aly in control and *kmg* KD testes and RNA-Seq results from wild-type testes. Y-axes: normalized read counts based on one million mapped reads per sample.

(B and C) Median profile of (B) Aly and (C) Kmg ChIP (solid lines) and corresponding input (dotted lines) signals centered around (Case #1, left) Aly peaks identified in wild-type testes ($q < 10^{-10}$ and identified in both replicates) that overlap with Kmg peaks and (Case #2, right) new Aly peaks identified in *kmg* KD testes ($q < 10^{-10}$ and identified in both replicates) that were not detected in wild-type testes. For the 248 Aly peaks in the Case #1 - because these peaks were present at the promoters of genes with known strand information, read counts were plotted according to the direction of transcription from 5′ to 3′.

(D) Heatmap representation of normalized RNA-seq read counts centered around 4,219 new Aly peaks that appeared in *kmg* KD (Case #2). Darkest (black) color: read count value at the 5th percentile. Brightest (yellow for reads mapped to Watson (+), cyan for Crick (−) strands) colors: values at the 95th percentile among all values in *kmg* KD.