A-MYB/TCFL5 regulatory architecture

ensures the production of pachytene piRNAs

in placental mammals

Tiangxiong Yu,[1,†] Adriano Biasini,[2,†] Katharine Cecchini,[2,†] Martin Säflund,[3] Haiwei Mou,[4] Amena Arif,[2,5] Atiyeh Eghbali,[3] Dirk G. de Rooij,[6] Zhiping Weng,[1] Phillip D. Zamore,[2,*] and Deniz M. Özata[3,*]

[1]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA

[2]RNA Therapeutics Institute and Howard Hughes Medical Institute, University of Massachusetts Medical School, 368 Plantation Street, Worcester, MA 01605, USA

[3]Department of Molecular Biosciences, The Wenner-Gren Institute, Stockholm University, S-106 91 Stockholm, Sweden

[4]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

[5]Present address: Beam Therapeutics, 238 Main St, Cambridge, MA 02142, USA

[6]Reproductive Biology Group, Division of Developmental Biology, Department of Biology, Faculty of Science, Utrecht University, Utrecht 3584, the Netherlands

[†]These authors contributed equally to this work

***Corresponding authors**: phillip.zamore@umassmed.edu (P.D.Z) and deniz.ozata@su.se (D.M.O)

Keywords: TCFL5, A-MYB, pachytene piRNAs, spermatogenesis

1

ABSTRACT

In male mice, the transcription factor A-MYB initiates the transcription of pachytene piRNA genes during meiosis. Here, we report that A-MYB activates the transcription factor *Tcfl5* produced in pachytene spermatocytes. Subsequently, A-MYB and TCFL5 reciprocally reinforce their own transcription to establish a positive feedback circuit that triggers pachytene piRNA production. TCFL5 regulates the expression of genes required for piRNA maturation and promotes transcription of evolutionarily young pachytene piRNA genes, whereas A-MYB activates the transcription of older pachytene piRNA genes. Intriguingly, pachytene piRNAs from TCFL5-dependent young loci initiates the production of piRNAs from A-MYB-dependent older loci ensuring the self-propagation of pachytene piRNAs. A-MYB and TCFL5 act via a set of incoherent feedforward loops that drive regulation of gene expression by pachytene piRNAs during spermatogenesis. This regulatory architecture is conserved in rhesus macaque, suggesting that it was present in the last common ancestor of placental mammals.

INTRODUCTION

In human and mouse testes, ~100 genes—first expressed at the pachytene stage of meiosis I—make non-coding transcripts that are processed into 26–30 nt pachytene PIWI-interacting RNAs (pachytene piRNAs) (Li et al., 2013; Özata et al., 2020). Found only in placental mammals, pachytene piRNAs are diverse—comprising ~1 million distinct species—and plentiful—rivaling ribosomal RNA abundance (Gainetdinov et al., 2018; Li et al., 2013; Özata et al., 2020). Pachytene piRNAs bind the PIWI proteins PIWIL1 and PIWIL2, generating RNA-protein complexes that can repress transcripts during male gametogenesis (Aravin et al., 2006; Choi et al., 2021; Girard et al., 2006; Goh et al., 2015; Grivna et al., 2006; Han et al., 2015; Homolka et al., 2015; Lau et al., 2006; Mohn et al., 2015; Wu et al., 2020; Özata et al., 2020). Successful spermiogenesis requires pachytene piRNAs, yet their sequences are poorly conserved and rapidly diverging even among modern humans (Choi et al., 2021; Wu et al., 2020; Özata et al., 2020).

The transcription factor A-MYB binds the promoters of more than half of pachytene piRNA-producing genes ("pachytene piRNA genes") in humans and mice, and *A-Myb* mutant mice make fewer pachytene piRNAs than wild-type (Li et al., 2013; Özata et al., 2020). Although *A-Myb*-mutant germ cells fail to express meiotic genes and arrest early in meiosis I (Bolcun-Filas et al., 2011), human and mouse testes with little or no A-MYB still make some pachytene piRNAs (Özata et al., 2020); and the promoters of half of human and macaque piRNA-producing genes are not bound by A-MYB (Özata et al., 2020), suggesting that one or more additional transcription factors act downstream or in parallel to A-MYB to initiate piRNA production.

Here, we report that A-MYB initiates transcription of *Tcfl5* early in meiosis I. Subsequently, TCFL5 binds both its own promoter and that of *A-Myb*, establishing a mutually reinforcing positive feedback loop. Whereas A-MYB generally initiates the production of pachytene piRNAs from evolutionarily older piRNA precursor loci,

3

transcription of younger pachytene piRNA genes is more dependent on TCFL5. Like A-MYB, TCFL5 promotes transcription of both piRNA precursors and mRNAs encoding piRNA biogenesis proteins. The A-MYB/TCFL5 positive feedback loop allows each protein to initiate interlocking coherent feed-forward loops dedicated to piRNA production. Together, A-MYB and TCFL5 regulate the transcription of both pachytene piRNAs and their target transcripts, establishing an incoherent feedforward loop during spermatogenesis. This transcriptional architecture is evolutionarily conserved in the old-world monkey rhesus macaque, suggesting it predates the divergence of rodents and primates. Our findings suggest that this transcriptional circuit likely co-evolved with pachytene piRNAs themselves.

RESULTS

A-MYB is expressed before TCFL5 during male meiosis I

Consistent with the testis-specific expression of human *Tcfl5* (Maruyama et al., 1998), two recent studies revealed TCFL5 is first expressed in mouse primary spermatocytes (Galán-Martínez et al., 2022; Xu et al., 2022). We used staged mouse testes (Nebel et al., 1961) to measure the abundance of *Tcfl5* mRNA and protein across spermatogenesis. At 11 days post-partum (dpp), meiosis proceeds no further than the zygotene stage; early pachytene spermatocytes appear at 14 dpp; and mid- and late-pachytene spermatocytes are present by 17 dpp. *A-Myb* mRNA and protein were first detected at 14 dpp. By contrast, *Tcfl5* mRNA was present at 14 dpp, but TCFL5 protein was not detected until after 17 dpp (Fig. 1A, B; Supplemental Fig. S1A).

To directly test whether TCFL5 is expressed in primary spermatocytes, we purified germ cells from *Tcfl5+/em2(flag)Pdz* (henceforth, *Tcfl5+/FLAG*) mice in which the endogenous TCFL5 protein is tagged with 3XFLAG peptide at its amino-terminus (Supplemental Fig. S1B) and performed western blotting using anti-FLAG antibody. FLAG-TCFL5 was not detected in purified spermatogonia; was expressed abundantly in

4

primary spermatocytes; and was present at a lower level in secondary spermatocytes (Fig. 1C; Supplemental Fig. S1C). Reinforcing the view that *Tcfl5* is expressed mostly in primary spermatocytes, we reanalyzed publicly available single-cell RNA-sequencing data (Green et al., 2018): *Tcfl5* mRNA abundance was highest in primary spermatocytes compared to spermatogonia, spermatids, and testicular somatic cells (Fig. 1D). Moreover, consistent with the primary spermatocyte-specific expression of A-MYB (Tang and Goldberg, 2012) (Fig. 1E), immunostaining of TCFL5 in testis sections from staged mouse detected the TCFL5 in the nuclei of primary spermatocytes, but not in those of spermatogonia or sertoli cells (Fig. 1F). In contrast to previous observations (Xu et al., 2022), we detected a single, well-resolved TCFL5 band by Western blotting. A single TCFL5 protein isoform is consistent with publicly available chromatin immunoprecipitation followed by sequencing (ChIP-seq) of histone H3 trimethylated on lysine 4 (H3K4me3) (Maezawa et al., 2018) and transposase-accessible chromatin with sequencing (ATAC-seq) data from purified primary spermatocytes (Maezawa et al., 2018), which detect a single prominent peak of H3K4me3 or ATAC-seq around the transcription start site of *Tcfl5* gene (Supplemental Fig. S1D).

A-Myb expression precedes that of *Tcfl5* (Fig. 2A): dual-color RNA fluorescence in situ hybridization (RNA-FISH) of adult testis sections revealed that *A-Myb* mRNA first appears in the nuclei of leptotene/zygotene cells as one or two brightly staining foci, likely corresponding to its genomic sites of transcription (Supplemental Fig. SA). *Tcfl5* mRNA was not detected in leptotene/zygotene cells, and first appeared in the nuclei and cytoplasm of pachytene cells. Single-color RNA-FISH of multiple seminiferous tubules at different stages confirmed that *A-Myb* mRNA first appears in pre-leptotene, leptotene, and zygotene cells, whereas *Tcfl5* first appears in mid- and late-pachytene cells (Fig. 2B).

A-MYB initiates a reciprocal positive feedback loop between TCFL5 and A-MYB

Tcfl5 mRNA was not detectable by RNA-FISH in *A-Myb*[repro9/repro9] (henceforth, *A-Myb*[−/−]) mutant testis (Supplemental Fig. S2B). We used a pair of single-guide RNAs (sgRNAs) to generate a *Tcfl5* knockout by deleting *Tcfl5* exons 2 and 3 (2,274 bp) (Fig. 2C, D; Supplemental Fig. S2C). *A-Myb* mRNA was readily detected in *Tcfl5*[em1Pdz/em1Pdz] (henceforth, *Tcfl5*[em1/em1]) (Supplemental Fig. S2B). A-MYB protein was less abundant in *Tcfl5*[em1/em1] than in C57BL/6 control testes (Fig. 2E; Supplemental Fig. S2D). By contrast, no TCFL5 protein was detected in either *A-Myb*[−/−] or *Tcfl5*[em1/em1] mutants (Fig. 2E; Supplemental Fig. S2D). These data suggest that A-MYB initiates *Tcfl5* expression, while TCFL5 acts to increase the steady-state abundance of A-MYB.

To determine whether A-MYB regulates *Tcfl5* directly or indirectly, we performed ChIP-seq of adult mouse testes. A prominent A-MYB peak was located at the first nucleotide of the *Tcfl5* transcription start site (Fig. 2F). TCFL5 ChIP-seq also revealed a significant TCFL5 peak within the promoters of both *A-Myb* and *Tcfl5* itself (Fig. 2F). A-MYB and TCFL5 "Cleavage Under Targets and Release Using Nuclease" (CUT&RUN; Skene and Henikoff (2017)) of FACS-purified primary spermatocytes confirmed these observations (Supplemental Fig. S2E). We conclude that A-MYB promotes its own transcription and directly initiates transcription of *Tcfl5*; TCFL5 responds by reinforcing its own transcription and that of A-MYB via positive transcriptional feedback.

TCFL5 regulates the expression of piRNA pathway genes

At the onset of the pachynema in mice and humans, A-MYB activates transcription of genes producing pachytene piRNA precursors and genes encoding piRNA biogenesis proteins, thereby generating a feedforward loop that drives pachytene piRNA production (Li et al., 2013; Özata et al., 2020). Given the expression pattern of TCFL5 and A-MYB during mouse spermatogenesis and the previously identified requirement for *A-Myb* in

pachytene piRNA production, we sought to test if TCFL5 also plays a role in pachytene piRNA production.

In FACS-purified primary spermatocytes, the promoters of 14 piRNA biogenesis genes featured a TCFL5 peak near their transcription start sites (Chi-square test $\rho = <2.7 \times 10^{-9}$; median distance = 0 bp) (Fig. 3A; Supplemental Fig. S3A Fig. S3A; Supplemental Table S1A). The steady-state mRNA abundance of four of the TCFL5-bound piRNA biogenesis genes—*Henmt1*, *Mael, Pld6*, and *Tdrd6*—was reduced ≥2-fold in *Tcfl5^{em1/em1}* and *Tcfl5^{+/em1}* testes (Supplemental Fig. S3A, B; Supplemental Table S1A). Among the remaining ten piRNA biogenesis genes—*Piwil1*, *Piwil2*, *Ddx39*, *Tdrd12*, *Tdrd5*, *Tdrd9*, *Tdrd1*, *Mov10l1*, *Fkbp6*, and *Ddx4*, the mRNA abundance of five decreased modestly in *Tcfl5^{+/em1}* (Supplemental Fig. S3A; Supplemental Table S1A). The promoters of the nine genes were all bound by A-MYB, suggesting that their transcription is either driven entirely by A-MYB or that A-MYB largely compensates in *Tcfl5* mutant mice. We conclude that in parallel to A-MYB, TCFL5 also establishes a feedforward loop that powers pachytene piRNA production.

TCFL5 activates evolutionarily younger pachytene piRNA genes

Our data suggest that TCFL5 activates transcription of almost half of pachytene piRNA genes during male meiosis I. CUT&RUN of FACS-purified primary spermatocytes identified TCFL5 peaks within 500 bp of the transcription start sites of 47 of the 100 mouse pachytene piRNA genes (median distance = 0 bp), whereas 91 were bound by A-MYB (median distance = 0 bp) (Fig. 3A; Supplemental Table S1B).

Pachytene piRNA genes with evolutionary conservation of genomic location (synteny) have first exons longer than 10 kbp and a low level of CG dinucleotides, and their promoters and gene bodies are bound by the transcription elongation factor BTBD18 and marked by histone lysine acylation, including acetylation, butyrylation, and crotonylation (Yu et al., 2021; Zhou et al., 2017). BTBD18 encodes a transcriptional

7

elongation factor that has been proposed to facilitate the production of piRNA precursors from evolutionarily older pachytene piRNA genes (Yu et al., 2021), and precursor transcript and mature piRNA abundance from these loci declines significantly in *Btbd18* mutant testes (Zhou et al., 2017). Although TCFL5 binds the *Btbd18* promoter (distance from transcription start site to the nearest TCFL5 peak = 0 bp), *Btbd18* mRNA increased in both *Tcfl5*$^{em1/em1}$ mutant testes (mean increase = 4.5 ± 0.5 fold) and *Tcfl5*$^{+/em1}$ primary spermatocytes (mean increase = 7 ± 3 fold) compared to wild-type (Fig. 3B). We do not yet know whether TCFL5 serves to repress BTBD18 expression. Increased BTBD18 levels in mutants may help explain why loss of TCFL5 has little effect on the transcription of many evolutionary older pachytene piRNA genes.

A-MYB occupancy was greatest at the promoters of pachytene piRNA genes with low CG dinucleotide content (Spearman's $\rho = -0.35$; Supplemental Fig. S4A), long first exons ($\rho = 0.28$), high histone lysine acylation ($\rho = 0.38$–0.63) and BTBD18 occupancy ($\rho = 0.44$; Supplemental Fig. S4A). High A-MYB and low TCFL5 promoter occupancy was also associated with pachytene piRNA genes producing abundant piRNAs (Supplemental Fig. S4B). By contrast, TCFL5 occupancy was correlated with high CG dinucleotide promoter sequence (Spearman's $\rho = 0.67$; Supplemental Fig. S4A), a feature more likely to be found in evolutionarily younger pachytene piRNA genes, which typically produce fewer piRNAs (Yu et al., 2021).

Given that pachytene piRNA genes are marked by histone lysine acylation, it is striking that TCFL5 binds the promoters of seven genes encoding histone acyl transferases (reviewed in ref. Sabari et al. (2017)). Of these seven genes, four were also bound by A-MYB (Supplemental Fig. S4C). Histone acyl transferases use acetyl-, butyryl-, or crotonyl-CoA as acyl donors. Histone crotonylation, which promotes gene expression, is regulated by the intracellular concentration of crotonyl-CoA (Sabari et al., 2015). *Acss2* encodes a short chain acyl-CoA synthetase (Luong et al., 2000) that uses

ATP to synthesize acetyl- and crotonyl-CoA. TCFL5 binds the *Acss2* promoter (Supplemental Fig. S4D).

Our data suggest that pachytene piRNA genes with high A-MYB occupancy are more likely to have conserved synteny than those with high TCFL5 and low A-MYB occupancy (Table S1B). Mouse pachytene piRNA genes were generally present at syntenic sites in other placental mammals (eutheria) when their promoters had high A-MYB occupancy (median = 3.8 RPM, 95% confidence interval [CI] = 2.8–4.7) compared to loci found only in murine rodents (median = 2.6 RPM, 95% CI = 2.1–3.2) or only in mouse (median = 2.3 RPM, 95% CI = 1.6–3.2) (Supplemental Fig. S4E; Table S1B). Conversely, TCFL5 promoter occupancy was higher for murine (median = 1.0 RPM, 95% CI = 0.6–2.3) and mouse pachytene piRNA genes (median = 1.3 RPM, 95% CI = 0.4–2.1) than at eutherian loci (median = 0.8 RPM, 95% CI = 0.5–1.1). Finally, the promoters of 20 of the 21 eutherian pachytene piRNA genes were bound by A-MYB, but only eight were occupied by TCFL5 (Fig. 3A). Our finding that A-MYB activates transcription of older pachytene piRNA genes, whereas TCFL5 tends to activate younger loci, suggests that acquisition of a TCFL5-binding site may facilitate the evolution of new pachytene piRNA genes.

Tcfl5 mutants have fewer pachytene piRNAs

For the majority of pachytene piRNA genes, piRNA abundance in *Tcfl5*$^{em1/em1}$ testes (77 of 100 genes) and in *Tcfl5*$^{+/em1}$ primary spermatocytes (74 of 100 genes) was less than half that in C57BL/6 controls (Fig. 3C). The affected pachytene piRNA genes include loci with either or both A-MYB or TCFL5 bound to their promoters, as well as five bound by neither transcription factor. The loss of pachytene piRNAs in *Tcfl5* mutants likely reflects the combined effects of reduced transcriptional activation by TCFL5 and A-MYB and reduced expression of pachytene piRNA biogenesis proteins. Loss of initiator piRNAs may also contribute to reduced piRNA levels.

9

piRNA production begins when a PIWI protein, loaded with an initiator piRNA, cleaves a complementary piRNA precursor to generate a 5′ monophosphorylated pre-pre-piRNA. Binding of a second PIWI protein to the 5′ end of the pre-pre-piRNA allows it to be further processed into a PIWI-loaded responder piRNA that can initiate additional piRNA production at fully or partially complementary sites (Brennecke et al., 2007; Gainetdinov et al., 2018; Gunawardane et al., 2007; Han et al., 2015; Mohn et al., 2015; Ozata et al., 2019; Wu et al., 2020). Do piRNAs from TCFL5-activated pachytene piRNA genes initiate production of piRNAs from precursors whose transcription depends solely on A-MYB?

To test the idea that production of A-MYB-dependent piRNAs requires initiation by piRNAs from loci whose transcription is activated by TCFL5, we identified A-MYB-dependent piRNA precursor transcripts whose abundance declined less than twofold in in $Tcfl5^{em1/em1}$ but whose piRNA abundance declined twofold or more, suggesting a defect in piRNA biogenesis rather than transcriptional activation. Among these, we searched for piRNAs from TCFL5-dependent pachytene piRNA genes capable of directing cleavage of precursor transcripts from other A-MYB-dependent pachytene piRNA genes. Candidate piRNAs were required (1) to have complete seed complementarity (g2–g7) to the A-MYB-dependent precursor RNA; (2) to form at least ten additional base pairs within the region g8–g21 with that precursor; and (3) be able to generate the 5′ end of a piRNA by PIWI-protein-catalyzed precursor RNA cleavage. The 5′ ends of 29 piRNAs from five different TCFL5-dependent pachytene piRNA precursors are predicted to target seven A-MYB-dependent pachytene piRNA precursors. Four of the five TCFL5-dependent pachytene piRNA genes are confined to murine rodents and one was found only in mouse; none are conserved among eutheria generally. By contrast, five of the seven pachytene piRNA genes targeted by piRNAs from TCFL5-dependent pachytene piRNA genes are present at syntenic sites in other eutherian species; the remaining two are found in among murine rodents (Fig. 3D; Supplemental

10

Table S2; Fisher's exact test $p = 0.028$ between the two sets of pachytene piRNA genes). Conversely, 62 piRNAs from 12 A-MYB-dependent pachytene piRNA genes are predicted to target the precursors from eight TCFL5-dependent pachytene piRNA genes. Seven of the 12 A-MYB-dependent pachytene piRNA genes are present at syntenic sites in other eutheria, two are found among murine rodents, while three were found only in mouse. Among the eight TCFL5-dependent pachytene piRNA genes targeted by these 12 loci, none were present in non-murine species (Fig. 3E; Supplemental Table S2; Fisher's exact test $p = 0.037$ between the 12 A-MYB-dependent and the 8 TCFL5-dependent genes). These data suggest that pachytene piRNAs from TCFL5-dependent young loci initiate the processing of pachytene piRNA precursors from older loci that are generally controlled by A-MYB. Similarly, piRNAs from older pachytene piRNA genes initiate the production of piRNAs from the precursors of young loci controlled by TCFL5. Together, A-MYB and TCFL5 ensure the production of pachytene piRNAs during male meiosis I.

Pachytene piRNA-driven gene regulation is part of an incoherent feedforward loop established by the A-MYB/TCFL5 regulatory architecture

Loss of *pi6* piRNAs results in compromised sperm function—attributable to the increased abundance of six mRNAs directly targeted by *pi6* piRNAs (Wu et al., 2020). Intriguingly, although *pi6* pachytene piRNAs begin to accumulate in pachytene spermatocytes (Gainetdinov et al., 2018; Li et al., 2013; Wu et al., 2020; Özata et al., 2020), the increase in the abundance of the six mRNAs is delayed until secondary spermatocytes and subsequent stages (Wu et al., 2020). The initiation of the expression of *pi6*—a locus with conserved synteny among eutherians—by A-MYB (Fig. 3A; Supplemental Fig. S5A) and the delay in the increase in abundance of target mRNAs until later stages of spermatogenesis in *pi6* mutants (Wu et al., 2020) suggest that pachytene piRNAs do not completely repress their targets until their targets reach a

specific steady-state threshold, a signature feature of incoherent feedforward loops (Alon, 2007). Could then A-MYB/TCFL5 regulatory architecture amplify the production of pachytene piRNA targets in parallel to pachytene piRNA genes themselves? Our data identified TCFL5 peaks within 500 bp of the transcription start sites of five of the six genes whose mRNAs are directly targeted by *pi6* piRNAs, significantly more than detected for protein-coding genes generally (Chi-square test $p = < 1.9 \times 10^{-3}$; Supplemental Fig. S5B). Among those six genes, four were also bound by A-MYB (Chi-square test's $p = < 3.8 \times 10^{-6}$; Supplemental Fig. S5B).

Mice lacking the *pi18* pachytene piRNA gene are sterile, and *pi18* piRNAs directly target the *Golga2* mRNA, whose steady-state abundance increases in *pi18* mutant mice (Choi et al., 2021). Our data revealed that the promoter of the *pi18* pachytene piRNA gene was bound by A-MYB and that the promoter of *Golga2* featured both A-MYB and TCFL5 peaks near its transcription start site (Fig. 3A; Supplemental Fig. S5A, C). Together, these data suggest that A-MYB and TCFL5 drive the production of both pachytene piRNAs and their targets, generating an incoherent feedforward loop that generates a pulse of expression of piRNA target mRNAs that is subsequently extinguished by piRNA-directed, PIWI-protein catalyzed cleavage of the same mRNAs.

The regulatory architecture that drives pachytene piRNA production in rodents is conserved in rhesus macaque

In both rodents and primates, A-MYB activation of piRNA pathway proteins and pachytene piRNA genes creates a coherent feedforward loop that ensures the robust accumulation of pachytene piRNAs (Li et al., 2013; Özata et al., 2020). In mouse, TCFL5 similarly organizes a coherent feedforward loop that activates the production of pachytene piRNAs.

In addition to placental mammals, TCFL5 orthologs are found in the genomes of marsupials, monotremes, birds, reptiles, amphibians, and fishes (Fig. 4A). Thus, TCFL5

may participate in spermatogenesis across vertebrate species. Moreover, the A-MYB/TCFL5 regulatory architecture—reciprocal positive feedback circuits that reinforce *A-Myb* and *Tcfl5* transcription, sitting atop parallel A-MYB and TCFL5 coherent feedforward loops that drive pachytene piRNA production—is conserved in rhesus macaque (Fig. 4B, C; Supplemental Table S3). ChIP-seq of TCFL5 from adult rhesus testis identified 17,959 genomic regions with significant TCFL5 peaks (FDR < 0.05), comprising 4,048 annotated genes. The TCFL5-bound genes include six of the 15 rhesus orthologs of known piRNA biogenesis proteins and 30 of the 189 rhesus piRNA genes (Özata et al., 2020) (median distance from transcription start site to nearest TCFL5 peak = 0 bp) (Fig. 4C). Together, these data suggest that the A-MYB/TCFL5 transcriptional architecture was present in the last common ancestor of rodents and primates.

DISCUSSION

Our data, together with those from previous studies (Choi et al., 2021; Li et al., 2013; Wu et al., 2020; Özata et al., 2020), suggest a model for the evolutionarily conserved transcriptional architecture by which A-MYB and TCFL5 collaborate to regulate pachytene piRNA production during male meiosis I (Fig. 5). Together A-MYB and TCFL5 directly promote transcription of piRNA biogenesis machinery and the pachytene piRNA genes, whose transcripts are the sources piRNAs. Together and separately, A-MYB and TCFL5 participate in coherent feedforward loops that promote accumulation of pachytene piRNAs.

A-MYB and TCFL5 also promote transcription of chromatin-modifying enzymes and at least one enzyme required for the biosynthesis of acyl-CoA, the acyl carrier used by the transferases that modify histone lysines to facilitate active transcription of the pachytene piRNA genes. This transcriptional architecture is present in macaque, suggesting it is conserved among placental mammals. A-MYB also regulates male

13

meiosis in chickens (Li et al., 2013). The chicken genome encodes TCFL5, raising the possibility that the overall architecture of the A-MYB/TCFL5 regulatory circuit predates the divergence of mammals and birds.

How do new pachytene piRNA-producing loci evolve? We suggest that three events promote the emergence of new pachytene piRNA loci: direct or indirect transcriptional activation of the locus by A-MYB or TCFL5; initiation of piRNA production from the newly transcribed locus by a piRNA from an older pachytene piRNA locus; and, finally, targeting of the same or another older pachytene piRNA gene by an initiator piRNA from the novel locus. Given that the majority of mouse pachytene piRNA genes are restricted to murine rodents or mice, cooption of non-coding transcripts into the piRNA pathway seems surprisingly easy. The resulting novel piRNAs are unlikely to provide a selectable advantage to the male mouse: surprisingly few piRNAs from even the most ancient pachytene piRNA genes appear to regulate genes required for the production of functional sperm (Choi et al., 2021; Wu et al., 2020). Supporting this view, human pachytene piRNA genes are among the most rapidly diverging sequences in human genome (Özata et al., 2020). In fact, the main function of pachytene piRNAs seems to be to promote their own biogenesis. Do pachytene piRNAs really exist mainly to ensure their own production? We speculate that a small number of piRNA species modestly enhance spermatogenesis or spermiogenesis, thereby ensuring the inheritance of the tens of thousands of selfish piRNAs with no utility in germ cell development. Do most pachytene piRNAs benefit the mouse in subtle ways that remain to be discovered, or are pachytene piRNAs simply tiny narcissists that have turned the ancestral transposon silencing function of piRNAs against their host, joining the ranks of the longer selfish genetic elements present in the mammalian genome? Distinguishing between these hypotheses promises to be both challenging and rewarding, like pachytene piRNAs themselves.

MATERIALS AND METHODS

Mice

Mice were maintained and used according to the guidelines of the Institutional Animal Care and Use Committee of the University of Massachusetts Medical School (A201900331). C57BL/6J mice (RRID: IMSR_JAX:000664) were used as wild-type controls. *A-Myb^repro9* (MGI3512907) were a gift from John Schimenti, and were maintained in a C57BL/6 background

To generate *Tcfl5^em1/em1* mutant mice, we transcribed two sgRNAs targeting sequences in *Tcfl5* intron 1 (5′-GCA GUC UGG GUA CUA GAUA G-3′) and intron 3 (5′-AUU CAC UCA AAC AAC AAG AG-3′) using T7 RNA Polymerase. After purification by electrophoresis through a denaturing 10% polyacrylamide/urea gel, both sgRNAs (20 ng/µl each) and Cas9 mRNA (50 ng/µl, TriLink Biotechnologies, L-7206) were injected into the pronucleus and cytoplasm of fertilized eggs (Transgenic Animal Modeling Core, University of Massachusetts Medical School), and 15–25 blastocysts were transferred into uterus of pseudo-pregnant ICR females at day E2.5. Pups were screened by PCR of genomic DNA extracted from tail tissue using primers (5′-TGC CTC ATC TGG GTG GAT ATC TGA-3′; 5′-TAC ACG AAA ACA AAA CTT AAG CGG T-3′) designed to detect deletion of *Tcfl5* exons 2 and 3.

Tcfl5^+/em2(flag)Pdz (*Tcfl5^+/FLAG*) was generated by inserting a 3XFLAG tag in-frame with TCFL5 at the endogenous *Tcfl5* locus (Cyagen, Santa Clara, California, USA). Briefly, two guide RNAs (5′-GCC GCG CCG CGC CGG CAT GTC GG-3′ and 5′-CCG CGC CGC GCC GGC ATG TCG GG-3′) targeting sequences at the 5′ end of the *Tcfl5* coding sequence, the donor vector containing the FLAG-HA-tag cassette, and Cas9 mRNA were injected into fertilized mouse eggs. F0 founders were bred to wild-type mice to identify animals with germ line transmission in the F1.

15

Isolation of mouse germ cells by FACS

Germ cell sorting was as described (Gainetdinov et al., 2018; Özata et al., 2020). Briefly, testis was de-capsulated, incubated with 1× Gey's Balanced Salt Solution (GBSS, Sigma, G9779) containing 0.4 mg/ml collagenase type 4 (Worthington; LS004188) at 33°C for 15 min, and the seminiferous tubules washed twice with 1× GBSS, and then incubated with 1× GBSS containing 0.5 mg/ml trypsin and 1 µg/ml DNase I at 33°C for 15 min. Next, the seminiferous tubules were maintained at 4°C on ice and gently pipetted up and down for 3 min through a Pasteur pipette to homogenize. Trypsin was inactivated with fetal bovine serum (FBS; f.c. 7.5% [v/v]), and the cell suspension was passed through a pre-wetted 70 µm cell strainer. Cells were recovered by centrifugation at 300 × g at 4°C for 10 min, resuspended in 1× GBSS containing 5% (v/v) FBS, 1 µg/ml DNase I, and 5 µg/ml Hoechst 33342 (Thermo Fisher, 62249), and incubated at 33°C for 45 min rotating at 150 rpm. Finally, propidium iodide (0.2 µg/ml, f.c.; Thermo Fisher, P3566) was added, and cells were filtered through a pre-wetted 40 µm cell strainer. Cell sorting (FACS Core, University of Massachusetts Medical School was as described (Bastos et al., 2005; Özata et al., 2020).

Testis histology and immunohistochemistry

Wild-type and *Tcfl5*$^{em1/em1}$ mutant testis tissue was fixed with Bouin's solution, and paraffin embedded tissue sectioned at 5 µm thickness, and then stained with hematoxylin and eosin (H&E) solutions (Morphology Core Facility, University of Massachusetts Medical School). Immunohistochemistry (IHC) was performed using a standard protocol. Briefly, testis sections were de-paraffinized with xylene, dehydrated with ethanol, and then boiled in 1 mM citrate buffer (pH 6.0) to retrieve antigens. To inactivate endogenous peroxidase, tissue sections were incubated with 3% (v/v) hydrogen peroxide at room temperature for 10 min, then blocked with 5% (v/v) horse serum (ImmPRESS HRP Anti-Rabbit IgG [Peroxidase] Polymer Detection Kit; Vector

16

labs; MP-7401). Sections were incubated with rabbit anti-TCFL5 (Sigma, HPA055223; 1:500 dilution), or rabbit anti-A-MYB (Sigma, HPA008791; 1:500 dilution) antibodies at 4°C overnight. Next, secondary HRP anti-rabbit antibody (Vector labs; MP-7401) was added and incubated at room temperature for 1 h, followed by incubation with chromogenic substrate (Fisher Scientific, TA-125-QHDX). Slides were counterstained with hematoxylin and covered with a #1 coverslip. IHC images were captured using a Leica DMi8microscope equipped with a 63✕, 1.4 NA oil immersion objective.

RNA fluorescent in situ hybridization

Paraffin embedded testes tissue was sectioned at 5 µm thickness, then de-paraffinized with xylene and dehydrated with ethanol. RNAscope Hydrogen Peroxide (H_2O_2) (ACD, 322335) was used to inactivate endogenous peroxidase and antigens retrieved using RNAscope Target Retrieval Reagents (ACD, 322000) followed by treatment with RNAscope Protease Plus (ACD, 322331) according to the manufacturer's directions. RNAscope 2.5 HD Detection Reagents-RED assay (ACD, 322360) were used to hybridize probes for *Tcfl5* (ACD, *Tcfl5* mRNA probe, NPR-0002511, targeting 202-1589 of NM_178254.3) and *A-MYB* (*A-MYB* mRNA probe, NPR-0002510, targeting 1165-2241 of NM_008651.3) mRNAs and to amplify and detect signal. Stained slides were sealed with EcoMount (Biocare Medical, EM897L) and imaged using a Leica DMi8 microscope equipped with a 63✕ 1.4 NA oil immersion objective. For simultaneous two-color detection of *Tcfl5* and *A-MYB* mRNAs, we used RNAscope two-color Fluorescent Detection v2 assay (ACD, 323110) and probes for *Tcfl5* (ACD, *Tcfl5* mRNA C3 probe, 815341-C3, targeting 202-1589 of NM_178254.3) and *A-MYB* (*A-MYB* mRNA C3 probe, 815331-C3, targeting 1165-2241 of NM_008651.3) mRNAs. Opal 520 reagent (Akoya Biosciences, FP1487001KT) was used to detect fluorescent signal for *Tcfl5* and Opal 690 reagent (Akoya Biosciences, FP1497001KT) for *A-Myb*. DAPI (1 mg/ml; Thermo Fisher Scientific, 62248) was used to counterstain nuclei. Stained slides were sealed

17

using gold antifade mountant solution (Fisher scientific, P36930), covered with a #1 coverslip, and Z-stack images captured using a Leica SP8 Lightning confocal microscope equipped with a 63✕, 1.4 NA oil immersion objective.

Western blotting

Frozen testis tissues were homogenized in a Dounce homogenizer using 20 strokes of pestle B in RIPA lysis buffer (25 mM Tris-HCl, pH 7.6, 150 mM NaCl, 1% (v/v) NP-40, 1% sodium deoxycholate, and 0.1% (w/v) SDS) containing 1✕ homemade protease inhibitor cocktail (1 mM 4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride (Sigma; A8456), 0.3 µm Aprotinin, 40 µm Betanin hydrochloride, 10 µm. E-64 (Sigma; E3132), 10 µm Leupeptin hemisulfate). Lysed tissue was then sonicated (Branson Digital Sonifier; 450 Cell Disruptor) to break nuclei. After sonication, the samples were centrifuged at 20,000 × g at 4°C for 30 min. The supernatant was transferred to a new tube, and protein concentration measured using the Pierce BCA Protein Assay Kit (ThermoFisher; 23225). Total protein (50 µg) from each sample was mixed with 1/4 volume of loading dye (106 mM Tris-HCl, pH 6.8, 141 mM Tris base, 2% SDS, 10% v/v glycerol, 0.51 mM EDTA, 0.22 mM SERVA Blue G and 0.175 mM Phenol Red) containing 0.2 M dithiothreitol and heated at 95°C for 6 min to denature proteins. Samples were resolved by electrophoresis through a 4–20% polyacrylamide gradient SDS gel (Thermo Fisher, XP04205BOX), and the gel transferred to PVDF membrane (Millipore, IPVH00010). The membrane was blocked with Blocking Buffer (Rockland Immunochemicals, MB-070) at room temperature for 1.5 h and incubated with rabbit polyclonal anti-A-MYB (Sigma, HPA008791; 1:1,000 dilution), rabbit polyclonal anti-TCFL5 (Abgent, AP17006b; 1:1000 dilution), or mouse monoclonal anti-FLAG (Sigma, F3165; 1:1,000) antibody at 4°C overnight. Next, the membrane was washed three times (30 min each) with 1✕ PBS-T (0.1% (v/v) Tween-20 in 1✕ PBS) at room temperature; incubated with secondary donkey anti-rabbit IRDye 680RD antibody (LI-

COR, 926-68073; 1:15,000 dilution) at room temperature for 30 min; washed three times (30 min each) with 1× PBS-T at room temperature, and signal detected using the Odyssey Infrared Imaging System (LI-COR). As a loading control, membrane was incubated with mouse anti-ACTIN antibody (Santa Cruz Biotechnology, sc-47778; 1:10,000 dilution) at room temperature for 2 h, followed by goat anti-mouse IRDye 800CW secondary antibody (LI-COR, 926-32210; 1:15,000 dilution) as described above.

Small-RNA library construction and analysis

Total RNA was extracted from frozen testis tissue using the mirVana miRNA isolation kit (Thermo Fisher, AM1560). Small RNA library construction was performed as described (Gainetdinov et al., 2018). Briefly, total RNA was resolved by electrophoresis through a denaturing 15% polyacrylamide gel to select 18–34 nt RNAs. Size-selected small RNA was oxidized with sodium periodate (NaIO$_4$) prior to 3′ adaptor ligation to suppress ligation to microRNAs and siRNAs. The 3′ adaptor (5′-rApp NNN TGG AAT TCT CGG GTG CCA AGG /ddC/-3′) was ligated in the presence of 50% (w/v) PEG8000 (Rigaku, 1008063). The 3′ adaptor ligated small RNAs were then purified by electrophoresis through a denaturing 15% polyacrylamide gel to eliminate the unligated 3′ adaptor; 42 to 60 nt small RNA containing the 3′ adaptor were recovered from the gel. Finally, 5′ adaptor was ligated to the RNA and cDNA generated. Libraries were sequenced (79 nt single-end reads) using a NextSeq500 (Illumina).

After removing 3′ adaptor sequence from the reads, we eliminated reads with PHRED score < 5. Remaining reads were then mapped to mouse genome (mm10) using piPipes(Han et al., 2015) employing Bowtie 2.2.5 (Langmead and Salzberg, 2012), allowing one mismatch. We quantified the length distribution and abundance of piRNA reads by normalizing to genome-mapping reads reported as reads per million (rpm). Spike-in oligonucleotides were identified allowing no mismatches. The absolute

19

quantity of piRNAs was calculated based on the read abundance of the spike-in RNAs. Small RNA sequencing statistics are provided in Supplemental Table S4.

Long RNA library construction and analysis

Total RNA from frozen testis samples and from sorted germ cells was extracted using mirVana miRNA isolation kit (Thermo Fisher, AM1560). Before RNA library construction from FACS-purified germ cells, we added 1 µl of 1:100 dilution of ERCC spike-in mix 1 (Thermo Fisher, 4456740, LOT00418382) to 0.5–1µg total RNA to enable subsequent molecular quantification of transcripts. Library construction was performed as described (Zhang et al., 2012) with a modified ribosomal RNA (rRNA) removal procedure. Briefly, 186 short antisense DNA oligos (ASOs; 50–80 nt) complementary to mouse rRNAs were designed and pooled (0.05 µM/each) (Adiconis et al., 2013; Morlan et al., 2012). Pooled rRNA ASO (1 µl) were added to 1 µg total RNA in the presence of 100 mM Tris-HCl (pH 7.4) and 200 mM NaCl, heated to 95°C and cooled to 22°C at −0.1°C/s then incubated at 22°C for 5 min. The DNA:RNA hybrids were digested with 10 U Thermostable RNase H (Epicentre, H39500) in 50 mM Tris-HCl (pH 7.4), 0.1 M NaCl, and 20 mM $MgCl_2$ at 45°C for 30 min. Next, samples were treated with 4 U Turbo DNase (Thermo Fisher, AM2238) at 37°C for 20 min. To enrich RNA >200 nt and remove tRNA, RNA samples were purified with an RNA Clean & Concentrator-5 (Zymo Research, R1015), and cDNA generated as described. Libraries were sequenced as 79 + 79 nt paired-end reads using a NextSeq500 (Illumina).

rRNA reads were eliminated using Bowtie 2.2.5 with default parameters (Langmead and Salzberg, 2012), and then the remaining reads were mapped to the mouse genome (mm10) using STAR 2.3 (ref. Dobin et al. (2013)). Results in SAM format were de-duplicated and transformed into BAM format using SAMtools 1.8 and Umitools (Fu et al., 2018; Li et al., 2009). Finally, mapped reads were assigned to protein-coding genes, long non-coding RNAs (lncRNAs) and piRNA genes using HTSeq

0.9.1 (ref. Anders et al. (2015)). Transcript abundance was reported as reads per million uniquely mapped reads per thousand nucleotides (RPKM) calculated using homemade Bash scripts.

Transcript abundance in molecules per cell was performed as described (Gainetdinov et al., 2018). Because each sample contained ~623,291,645 molecules of ERCC spike-in mix, the abundance of each gene = (number of mapped reads × 623291645)/(number of cells used to prepare the library × the number of reads mapping to the ERCC spike-in sequences). RNA sequencing statistics are provided in Supplemental Table S4. All analyses used Ensembl-v86 gene annotations.

Chromatin immunoprecipitation and sequencing

Chromatin Immunoprecipitation (ChIP) was performed as described (Li et al., 2013; Özata et al., 2020). Briefly, frozen testis tissue was minced on dry ice and incubated in 1.7 ml tubes in ice-cold PBS containing 2% (w/v) formaldehyde at room temperature for 30 min, slowly tumbling end-over-end. Fixed tissue was crushed in ChIP lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.1) with a Dounce homogenizer using 40 strokes pestle B (Kimble-Chase, Vineland, USA). To shear chromatin to 150–200 bp, samples were sonicated (Covaris, E220). Sonicated lysate was then diluted 1:10 with ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.1, 167 mM NaCl) and immunoprecipitated using 5 µg anti-TCFL5 (Sigma, HPA055223, lot # 000000907) or 5.5 µg anti-A-MYB (Sigma, HPA008791, lot # C105987) antibody. After immunoprecipitation, DNA was extracted with phenol:chloroform:isoamyl alcohol (25:24:1; pH 8), and ChIP-seq libraries generated as described(Zhang et al., 2012). Libraries were sequenced as 79 + 79 nt paired-end reads using a NextSeq500 (Illumina).

Raw reads were mapped to the mouse genome (mm10) using Bowtie 2.2.5 with parameters `—very-sensitive —no-unal —no-mixed —no-discordant -|10.`

Duplicate reads were removed using SAMtools 1.8 (ref. Li et al. (2009)). Normalized genome coverage files in bigWig format were generated using a homemade script. Data are reported as reads per million (rpm) using uniquely mapping reads normalized to read depth. Significant TCFL5 or A-MYB peaks (FDR < 0.05 for enrichment relative to input) were identified using MACS 1.1.2 (ref. Zhang et al. (2008)) with parameters `–keep-dup all -q 0.001`. ChIP sequencing statistics are provided in Supplemental Table S4.

The distance (bp) of transcription start site to the nearest TCFL5 or A-MYB peak summit for each gene was calculated. Genes that retain TCFL5 or A-MYB peak within the range of 500 bp around their transcription start site were considered as TCFL5- or A-MYB-regulated genes.

CUT&RUN Sequencing

Cleavage Under Targets and Release Using Nuclease (CUT&RUN) sequencing from FACS-purified germ cells was performed as described (Skene and Henikoff, 2017). Briefly, FACS-purified germ cells were lysed using nuclear extraction buffer (20 mM HEPES-KOH, pH 7.9, 10 mM KCl, 0.5 mM Spermidine, 0.1% Triton X-100, 20% glycerol) and nuclei recovered by centrifugation at 600 × g at 4°C for 3 min. Nuclei were immobilized on BioMag Plus Concanavalin A-coated paramagnetic beads (Bangs Laboratories, BP531) at room temperature for 20 min. After immobilizing on beads, nuclei were blocked in 20 mM HEPES-KOH, pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 0.1% BSA, 2 mM EDTA) at room temperature for 5 min. Then 2.5 µg anti-TCFL5 (Sigma, HPA055223 or Sigma, HPA055223, lot # 000000907) or 2.5 µg anti-A-MYB (Sigma, HPA008791, lot # C105987) antibodies were added to the beads and incubated at 4°C for 4 h, slowly tumbling end-over-end. CUT&RUN without antibody provided a negative control. After antibody incubation, beads were washed twice in wash buffer (20 mM HEPES-KOH, pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 0.1% BSA) and

incubated with 1.5 µl protein A-MNase fusion protein (pA-MN; 140 µg/ml) at 4°C for 2 h, slowly tumbling end-over-end. To remove unbound pA-MN, beads were washed twice in wash buffer. To liberate DNA bound by TCFL5 or A-MYB, 3 µl 100mM CaCl$_2$ was added to the beads and incubated at 0°C for 7 min. The reaction was stopped by adding an equal volume of 2✕ STOP buffer (200 mM NaCl, 20 mM EDTA, 4 mM EGTA, 50 µg/ml RNase A, 40 µg/ml glycogen) and then incubated at 37°C for 20 min. Nuclei were removed by centrifugation at 16,000 × g at 4°C. The supernatant containing CUT&RUN fragments was transferred to a fresh 1.7 ml tube. After extracting the DNA with phenol:chloroform:isoamyl alcohol (25:24:1; pH 8), CUT&RUN libraries were prepared as described(Zhang et al., 2012) with slight modifications. Briefly, end-repair using 0.5 U/µl T4 DNA polymerase (NEB, M0203) and adenylation using 0.5 U/µl *Taq* DNA polymerase (NEB, M0273) were carried out in a single reaction. Then, end-repaired and adenylated DNA was purified using AMPure beads. Finally, Illumina adapters (MP-Ada1: 5′-p GAT CGG AAG AGC ACA CGT CT-3′; MP-Ada2: 5′-p ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT-3′) were added using 600 U/µl T4 DNA ligase (Enzymatics Inc., L603-HC-L) at room temperature for 30 min.

CUT&RUN fragments were sequenced using 79 + 79 nt paired-end reads on a NextSeq500 (Illumina). Raw reads were mapped to the mouse genome (mm10) using Bowtie 2.2.5 (ref. Langmead and Salzberg (2012)) with parameters `—very-sensitive` `—no-unal —no-mixed —no-discordant -|10`. TCFL5 and A-MYB peaks were identified used S̲parse E̲nrichment A̲nalysis for C̲UT&R̲UN (SEARC (Meers et al., 2019)) in "norm relaxed" mode. CUT&RUN performed without antibody was used as the background control. Genes with a TCFL5 or A-MYB peak ±500 bp of their transcription start site were considered as TCFL5- or A-MYB-regulated. To be considered bound by TCFL5, a gene was required (1) to have a SEARC-identified peak within 500 bp of its transcription start site in both CUT&RUN replicates; or (2) to have a SEARC-identified peak within 500 bp of its transcription start site in one of the two CUT&RUN replicates

23

and a MACS2-identified peak within 500 bp of its transcription start site in two of three ChIP-seq replicates. To be considered bound by A-MYB, a gene was required (1) to have a SEARC-identified peak within 500 bp of its transcription start site in both CUT&RUN replicates; or (2) to have a SEARC-identified peak within 500 bp of its transcription start site in one of the two CUT&RUN replicates and a MACS2-identified peak within 500 bp of its transcription start site in ChIP-seq. CUT&RUN sequencing statistics are provided in Supplemental Table S4.

Statistics

All statistics were calculated using R console (https://www.rstudio.com); graphs were generated using R console or Prism 8.4.3 (GraphPad Software, LLC.). For box plots, boxes represent the first and third quartiles. Outliers are defined as data points with values > third quartile + 1.5 × IQR or lower than the first quartile − 1.5 × IQR, where IQR (interquartile range) is the maximum third quartile minus the first quartile minimum. Two-sided Mann-Whitney-Wilcoxon U and Fisher's exact tests was used to calculate p values.

DATA DEPOSITION

Sequencing data are available from the National Center for Biotechnology Information Sequence Read Archive using Gene Expression Omnibus accession number GSE166104.

CODE AVAILABILITY

Code used to identify putative piRNA-directed cleavage sites is available at https://github.com/weng-lab/GTBuster.

ACKNOWLEDGEMENTS

24

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Adiconis, X. et al. (2013) Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat Methods*, **10**, 623-629.

Alon, U. (2007) Network motifs: theory and experimental approaches. *Nat Rev Genet*, **8**, 450-461.

Anders, S., Pyl, P.T. & Huber, W. (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166-169.

Aravin, A. et al. (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*, **442**, 203-207.

Bastos, H. et al. (2005) Flow cytometric characterization of viable meiotic and postmeiotic cells by Hoechst 33342 in mouse spermatogenesis. *Cytometry A*, **65**, 40-49.

Bolcun-Filas, E. et al. (2011) A-MYB (MYBL1) transcription factor is a master regulator of male meiosis. *Development*, **138**, 3319-3330.

Brennecke, J. et al. (2007) Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. *Cell*, **128**, 1089-1103.

Choi, H., Wang, Z. & Dean, J. (2021) Sperm acrosome overgrowth and infertility in mice lacking chromosome 18 pachytene piRNA. *PLoS Genet*, **17**, e1009485.

Dobin, A. et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15-21.

Fu, Y. et al. (2018) Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. *BMC Genomics*, **19**, 531.

Gainetdinov, I. et al. (2018) A Single Mechanism of Biogenesis, Initiated and Directed by PIWI Proteins, Explains piRNA Production in Most Animals. *Mol Cell*, **71**, 775-790.e5.

Galán-Martínez, J. et al. (2022) TCFL5 deficiency impairs the pachytene to diplotene transition during spermatogenesis in the mouse. *Sci Rep*, **12**, 10956.

Girard, A. et al. (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, **442**, 199-202.

Goh, W.S. et al. (2015) piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. *Genes Dev*, **29**, 1032-1044.

Green, C.D. et al. (2018) A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Dev Cell*, **46**, 651-667.e10.

Grivna, S.T. et al. (2006) A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev*, **20**, 1709-1714.

Gunawardane, L.S. et al. (2007) A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5′ End Formation *in Drosophila*. *Science*, **315**, 1587-1590.

Han, B.W. et al. (2015) Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science*, **348**, 817-821.

Homolka, D. et al. (2015) PIWI Slicing and RNA Elements in Precursors Instruct Directional Primary piRNA Biogenesis. *Cell Rep*, **12**, 418-428.

Langmead, B. & Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**, 357-359.

Lau, N.C. et al. (2006) Characterization of the piRNA complex from rat testes. *Science*, **313**, 363-367.

Li, H. et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

Li, X.Z. et al. (2013) An Ancient Transcription Factor Initiates the Burst of piRNA Production during Early Meiosis in Mouse Testes. *Mol Cell*, **50**, 67-81.

Luong, A. et al. (2000) Molecular characterization of human acetyl-CoA synthetase, an enzyme regulated by sterol regulatory element-binding proteins. *J Biol Chem*, **275**, 26458-26466.

Maezawa, S. et al. (2018) Polycomb protein SCML2 facilitates H3K27me3 to establish bivalent domains in the male germline. *Proc Natl Acad Sci U S A*, **115**, 4957-4962.

Maruyama, O. et al. (1998) Cloning of TCFL5 encoding a novel human basic helix-loop-helix motif protein that is specifically expressed in primary spermatocytes at the pachytene stage. *Cytogenet Cell Genet*, **82**, 41-45.

Meers, M.P., Tenenbaum, D. & Henikoff, S. (2019) Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin*, **12**, 42.

Mohn, F., Handler, D. & Brennecke, J. (2015) Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science*, **348**, 812-817.

Morlan, J.D., Qu, K. & Sinicropi, D.V. (2012) Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PLoS One*, **7**, e42882.

Nebel, B.R., Amarose, A.P. & Hacket, E.M. (1961) Calendar of gametogenic development in the prepuberal male mouse. *Science*, **134**, 832-833.

Ozata, D.M. et al. (2019) PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet*, **20**, 89-108.

Özata, D.M. et al. (2020) Evolutionarily conserved pachytene piRNA loci are highly divergent among modern humans. *Nat Ecol Evol*, **4**, 156-168.

Sabari, B.R. et al. (2015) Intracellular crotonyl-CoA stimulates transcription through p300-catalyzed histone crotonylation. *Mol Cell*, **58**, 203-215.

Sabari, B.R. et al. (2017) Metabolic regulation of gene expression through histone acylations. *Nat Rev Mol Cell Biol*, **18**, 90-101.

Shiromoto, Y. et al. (2013) GPAT2, a mitochondrial outer membrane protein, in piRNA biogenesis in germline stem cells. *RNA*, **19**, 803-810.

Skene, P.J. & Henikoff, S. (2017) An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife*, **6**, e21856.

Tang, H. & Goldberg, E. (2012) A-MYB (MYBL1) stimulates murine testis-specific Ldhc expression via the cAMP-responsive element (CRE) site. *Biol Reprod*, **86**, 30.

Vagin, V.V. et al. (2013) Minotaur is critical for primary piRNA biogenesis. *RNA*, **19**, 1064-1077.

Wu, P.H. et al. (2020) The evolutionarily conserved piRNA-producing locus pi6 is required for male mouse fertility. *Nat Genet*, **52**, 728-739.

Xu, W. et al. (2022) Transcription factor-like 5 is a potential DNA- and RNA-binding protein essential for maintaining male fertility in mice. *J Cell Sci*, **135**, jcs259036.

Yu, T. et al. (2021) Long first exons and epigenetic marks distinguish conserved pachytene piRNA clusters from other mammalian genes. *Nat Commun*, **12**, 73.

Zhang, Y. et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, **9**, R137.

Zhang, Z. et al. (2012) Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. *Silence*, **3**, 9.

Zhou, L. et al. (2017) BTBD18 Regulates a Subset of piRNA-Generating Loci through Transcription Elongation in Mice. *Dev Cell*, **40**, 453-466.e5.

MAIN FIGURE TITLES AND LEGENDS

Fig. 1. TCFL5 is expressed in pachytene spermatocytes during male meiosis I.

(A,B) *A-Myb* and *Tcfl5* mRNA (**A**) and protein abundance (**B**) at distinct developmental stages of mouse spermatogenesis. See Supplemental Fig. S1A for uncropped western blot images.

(C) Abundance of FLAG–TCFL5 protein in FACS-purified germ cell from *Tcfl5$^{+/FLAG}$* mouse. Each lane contained protein from 75,000 germ cells.

(D) mRNA Abundance of *Tcfl5* mRNA measured by single-cell sequencing by Green et al. (2018).

(E) Immunohistochemical detection of A-MYB in adult testis section. *A-Myb$^{-/-}$* testis section serve as antibody specificity controls.

(F) Immunohistochemical detection of TCFL5 in testis sections from staged mice. *Tcfl5$^{em1/em1}$* testis sections serve as antibody specificity controls. Spg: spermatogonia; SpI: primary spermatocytes; SpII: secondary spermatocytes; RS: round spermatids; ES: elongating spermatids; SC: Sertoli cells.

Fig. 2. *A-Myb* expression precedes that of *Tcfl5*.

(A,B) Temporal *A-Myb* and *Tcfl5* mRNA expression in a seminiferous tubule sections from adult mouse testis was detected. Two-color RNA fluorescent *in situ* hybridization (RNA-FISH) **(A)**. single-color RNA-FISH **(B)**.

(C) Strategy to generate *Tcfl5* knockout mice using a single guide RNA (sgRNA) and Cas9. Scissors indicate sites targeted by sgRNAs designed to delete *Tcfl5* exons 2 and 3. RNA-seq was used to measure the abundance of *Tcfl5* mRNA in adult testes.

(D) Abundance of TCFL5 protein in *Tcfl5$^{em1/em1}$* homozygous and *Tcfl5$^{+/em1}$* heterozygous mutant testes, relative to C57BL/6 wild-type testes. ACTIN serves as a loading control. Each lane contained 50 µg testis protein.

(E) Abundance of A-MYB and TCFL5 proteins in *A-Myb$^{-/-}$* and $^{Tcfl5em1/em1}$ homozygous mutant mouse testes. ACTIN serves as a loading control. Each lane contained 50 µg testis protein.

(F) A-MYB and TCFL5 ChIP-seq peaks at the promoters of the *A-Myb* and *Tcfl5* genes.

Fig. 3. Younger pachytene piRNA genes whose transcription is activated by TCFL5 produce piRNAs that initiate piRNA production in evolutionarily older pachytene piRNA genes.

(A) The distance from the nearest TCFL5 peak to the transcription start sites (TSS) of pachytene piRNA genes, genes encoding piRNA biogenesis proteins, and other protein-coding or non-coding genes obtained from CUT&RUN of FACS-purified primary spermatocytes. Horizontal lines: median; whiskers: maximum and minimum values, excluding outliers. Each dot represents distance (mean of two trials) from the nearest TCFL5 peak to the transcription start site of individual genes. Measurements with the same values are indicated by a single marker indicating the number of individual data points. *Gpat2*, which is required for piRNA biogenesis in flies (Vagin et al., 2013) and in

cultured mouse germline stem cells (Shiromoto et al., 2013), was not included in this analysis, because its participation in the piRNA pathway in vivo in mice remains to be established.

(B) TCFL5 CUT&RUN peak at the promoter of *Btbd18*. WT denotes C57BL/6.

(C) Steady-state abundance of pachytene piRNAs in primary spermatocytes (left) or whole testes (right) for $Tcfl5^{+/em1}$ heterozygotes compared to C57BL/6 controls. Each marker represents a pachytene piRNA gene.

(D, E) Pachytene piRNA-directed cleavage sites in pachytene piRNA precursor transcripts. In (**D**), each arrow points from the piRNAs derived from a TCFL5-dependent pachytene piRNA gene towards the corresponding cleavage site in a pachytene piRNA precursor transcripts whose transcription is activated by A-MYB. TCFL5-dependent genes were defined as those whose promoters were occupied by TCFL5 and whose precursor and piRNA abundance in $Tcfl5^{em1/em1}$ mutant mice were less than those in C57BL/6. In (**E**), each arrow points from the piRNAs derived from one of 12 A-MYB-dependent pachytene piRNA genes towards genes for which the abundance of pachytene piRNA precursor transcripts was reduced <2-fold in $Tcfl5^{em1/em1}$ mutant mice. A-MYB-dependent genes were defined as those whose promoters were occupied by A-MYB and whose expression was essentially unchanged in $Tcfl5^{em1/em1}$ mutant mice.

Fig. 4. The role of TCFL5 in piRNA production is conserved in rhesus macaque.

(A) Left: Percent identity of mouse TCFL5 to TCFL5 in other mammals. Right: Unrooted tree of TCFL5 protein sequences, aligned using Clustal Omega. The unrooted tree was constructed using Randomized Axelerated Maximum Likelihood with default parameters and visualized in Interactive Tree Of Life.

(B) A-MYB and TCFL5 ChIP-seq peaks at the promoters of the rhesus macaque *A-MYB* and *TCFL5* genes.

32

(C) The distance from the nearest TCFL5 ChIP-seq peak to the transcription start site (TSS) for rhesus genes separated by class. Vertical lines: median; whiskers: maximum and minimum values, excluding outliers. A single marker containing the number of data points marks multiple identical values.

Fig. 5. A model for the A-MYB/TCFL5 regulatory architecture of mouse male meiotic cells.

The model incorporates hypotheses from refs. (Choi et al., 2021; Li et al., 2013; Wu et al., 2020; Özata et al., 2020) and this study. The figure highlights the roles of A-MYB and TCFL5 in regulating pachytene piRNA genes and direct targets of pachytene piRNAs during mouse spermatogenesis.
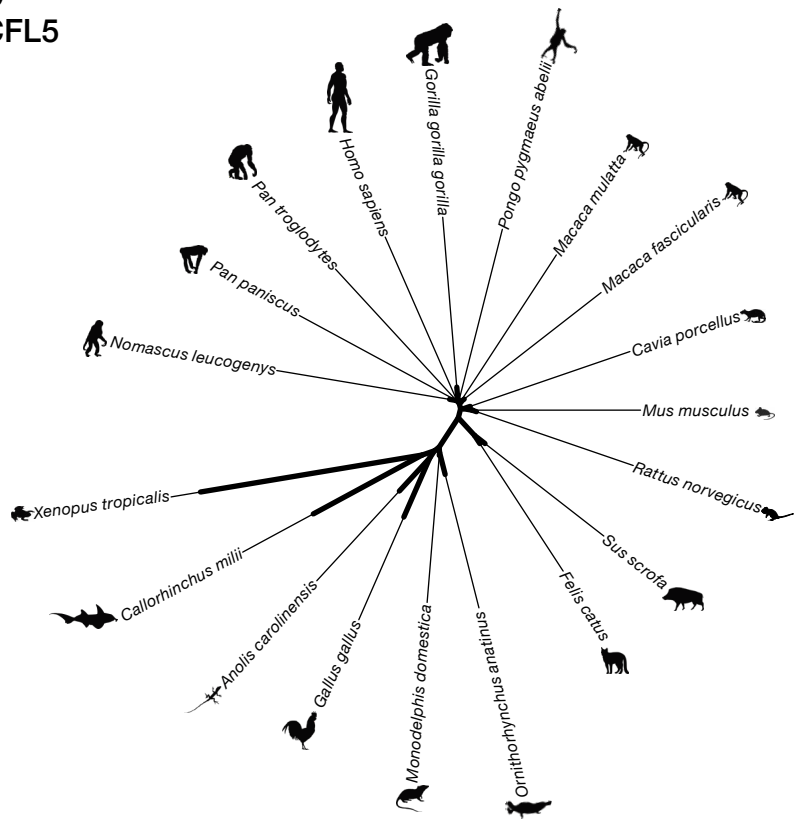
Yu et al., Figure 1

Yu et al., Figure 2

A

A-Myb + *Tcfl5* | *A-Myb* | *Tcfl5* | DAPI DNA

Pach
Lep/Zyg
Stage X–XI
20 µm

B

A-Myb mRNA

mid-Pach
A-Myb mRNA
Stage VI–VII

mid-Pach
Lep
Stage IX

late-Pach
Zyg
Stage X–XI

Tcfl5 mRNA

Tcfl5 mRNA
Stage-VI

Stage IX

Stage X–XI
20 µm

C

C57BL/6, *Tcfl5^em1/em1*

Abundance (rpm)

Trial 1

Trial 2

20,753 bp

Tcfl5^em1/em1 Δ2,274 bp (11% of locus)

D

C57BL/6 | *Tcfl5^+/em1* | *Tcfl5^em1/em1*
Trial: 1 2 | 1 2 | 1 2

75 kDa—
50—
37—
—TCFL5

75—
50—
37—
—ACTIN

TCFL5/ACTIN   1.0   0.75   0.11

E

C57BL/6 | *A-Myb^-/-* | *Tcfl5^em1/em1*
Trial: 1 2 | 1 2 | 1 2

A-MYB—
—100 kDa
—75

ACTIN—
—50
—37

A-MYB/ACTIN   1   0.08   0.6

C57BL/6 | *A-Myb^-/-* | *Tcfl5^em1/em1*
Trial: 1 2 | 1 2 | 1 2

TCFL5—
—75 kDa
—50

ACTIN—
—50
—37

TCFL5/ACTIN   1   0.2   0.09

F

Occupancy (rpm)

A-MYB ChIP
3.2
$p = 5.5 \times 10^{-49}$
$p = 6.4 \times 10^{-242}$
0

TCFL5 ChIP
1.5
$p = 4.2 \times 10^{-81}$
$p = 5.6 \times 10^{-38}$
0

32,795 bp
20,753 bp

A-Myb   *Tcfl5*

Yu et al., Figure 3

Yu et al., Figure 4

A

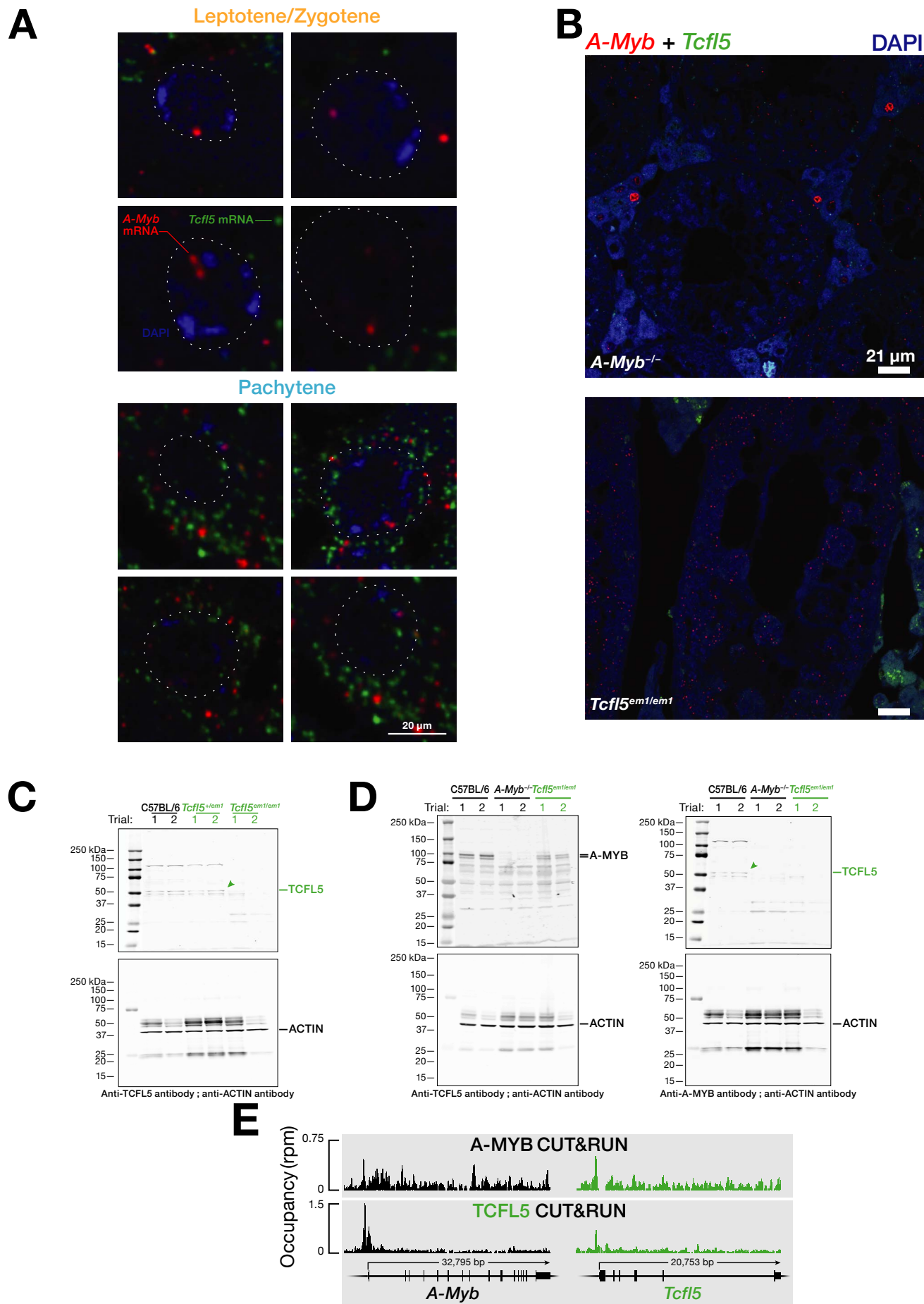|		% Identity to mouse TCFL5
	Callorhinchus milii	46
	Xenopus tropicalis	40
	Anolis carolinensis	63
	Gallus gallus	61
	Monodelphis domestica	76
	Ornithorhynchus anatinus	72
	Felis catus	81
	Sus scrofa	78
	Cavia porcellus	93
	Rattus norvegicus	97
	Macaca mulatta	90
	Macaca fascicularis	89
	Nomascus leucogenys	89
	Pongo pygmaeus abelii	89
	Gorilla gorilla gorilla	83
	Pan paniscus	89
	Pan troglodytes	89
	Homo sapiens	89

B

C

Yu et al., Figure 5

Yu et al., Supplemental Figure S1

A

Anti-A-MYB antibody ; anti-ACTIN antibody

Anti-TCFL5 antibody ; anti-ACTIN antibody

B

Tcfl5 allele

gRNA

gRNA

Homology

Targeting vector

3XFlag

C

Anti-FLAG antibody ; anti-ACTIN antibody

D

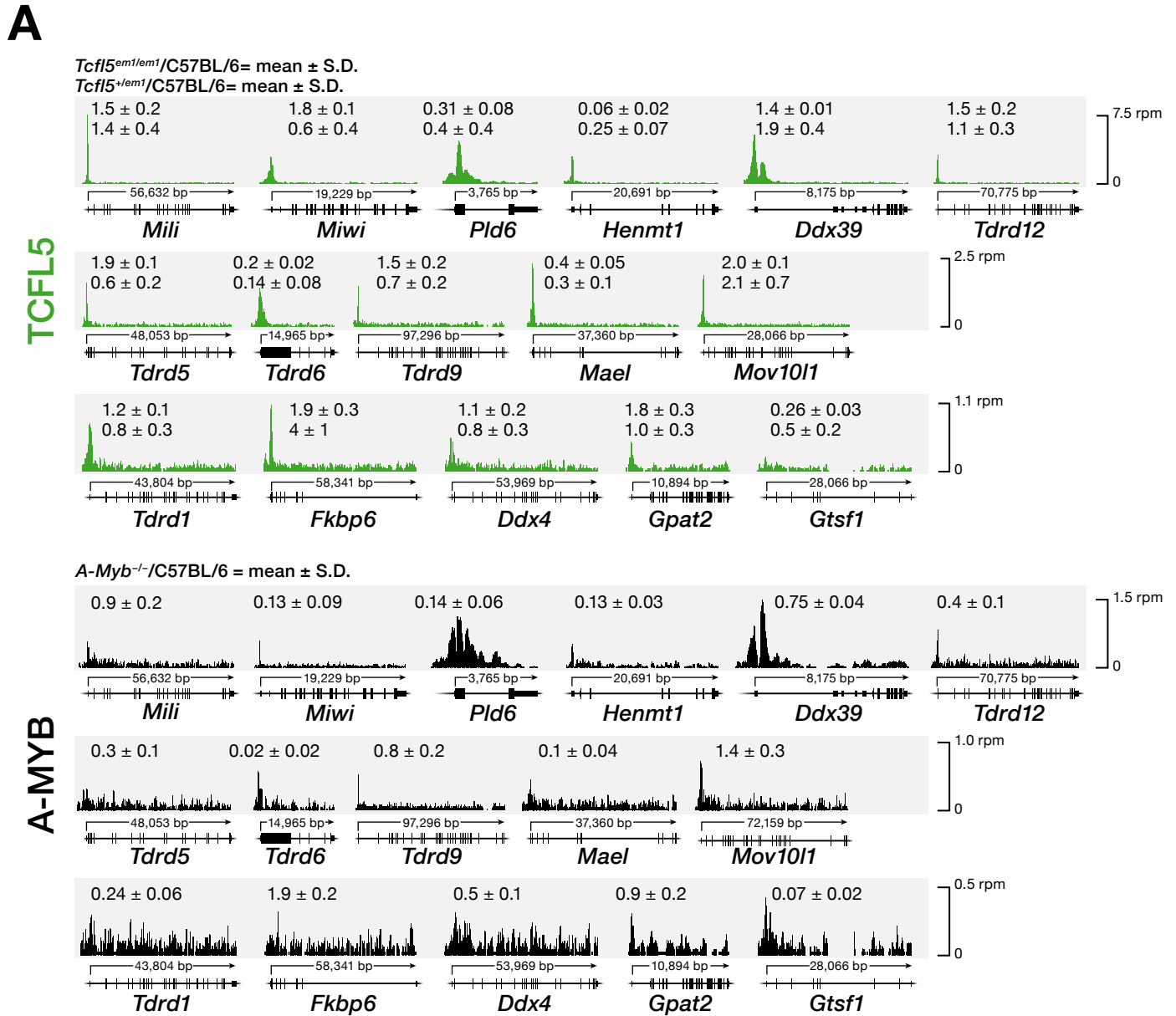SpI cells (Maezawa et al., *PNAS* 2018)

H3K4me3 Signal

ATAC Signal

20,753 bp
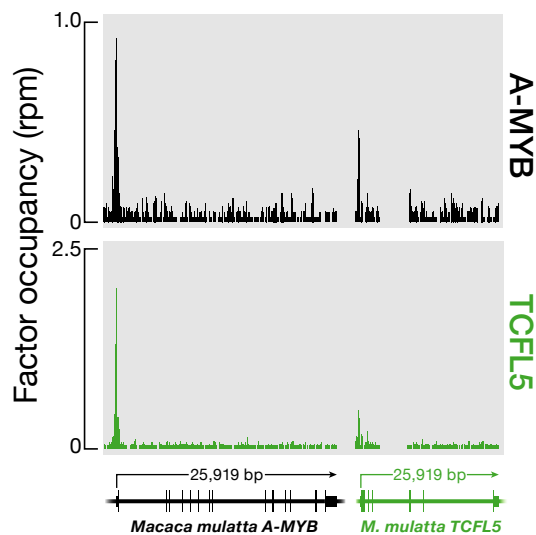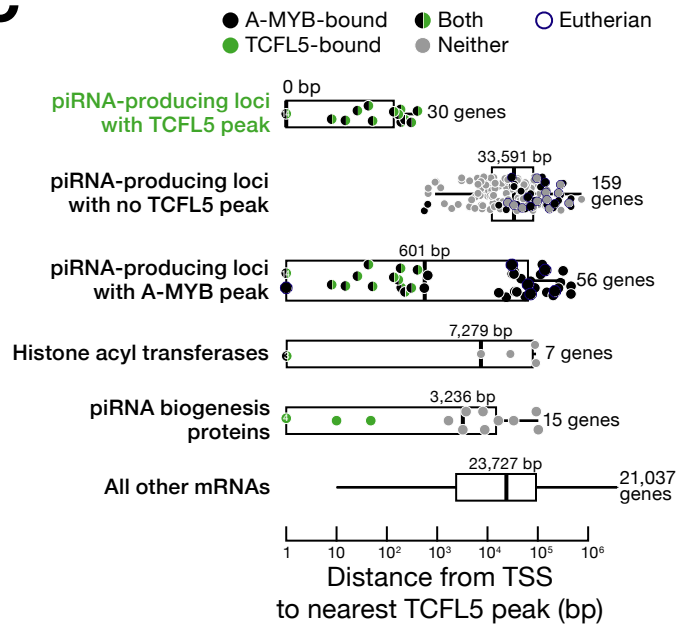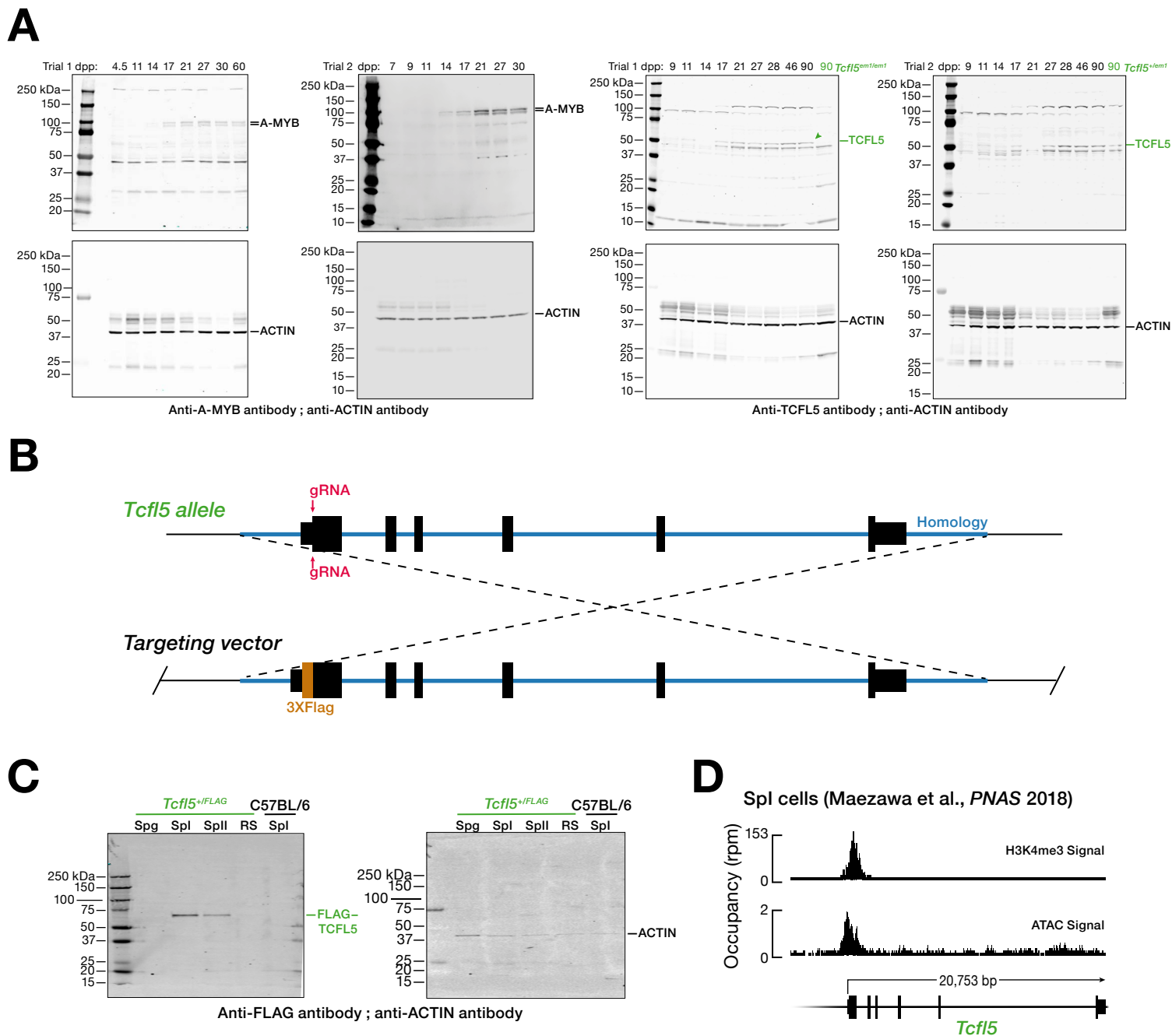
Tcfl5

Yu et al., Supplemental Figure S2

Yu et al., Supplemental Figure S3

A

B
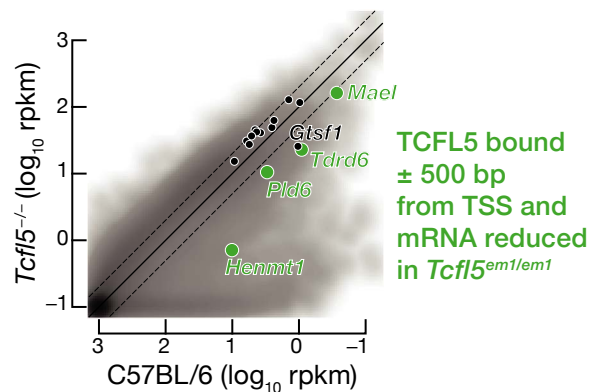
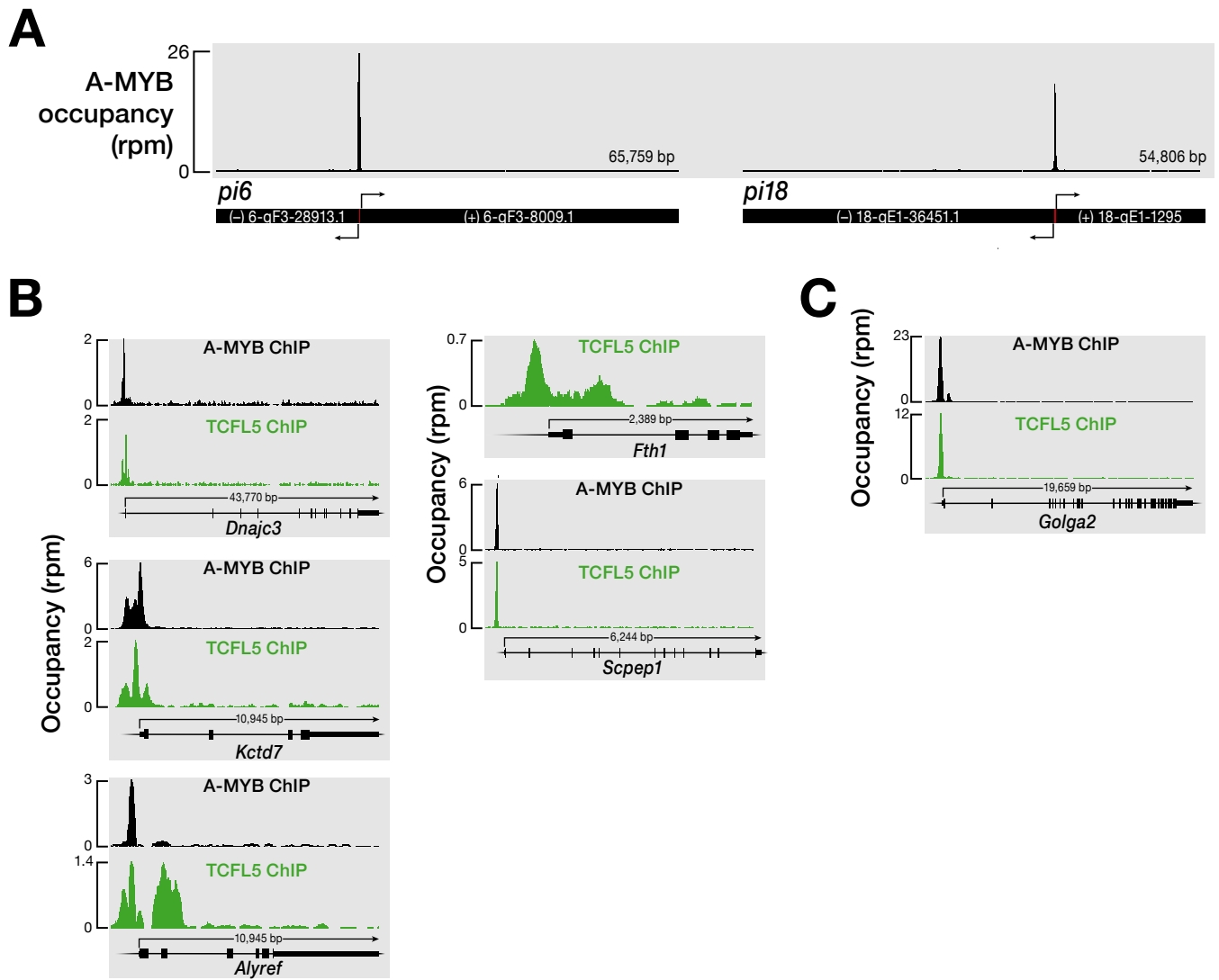Yu et al., Supplemental Figure S4
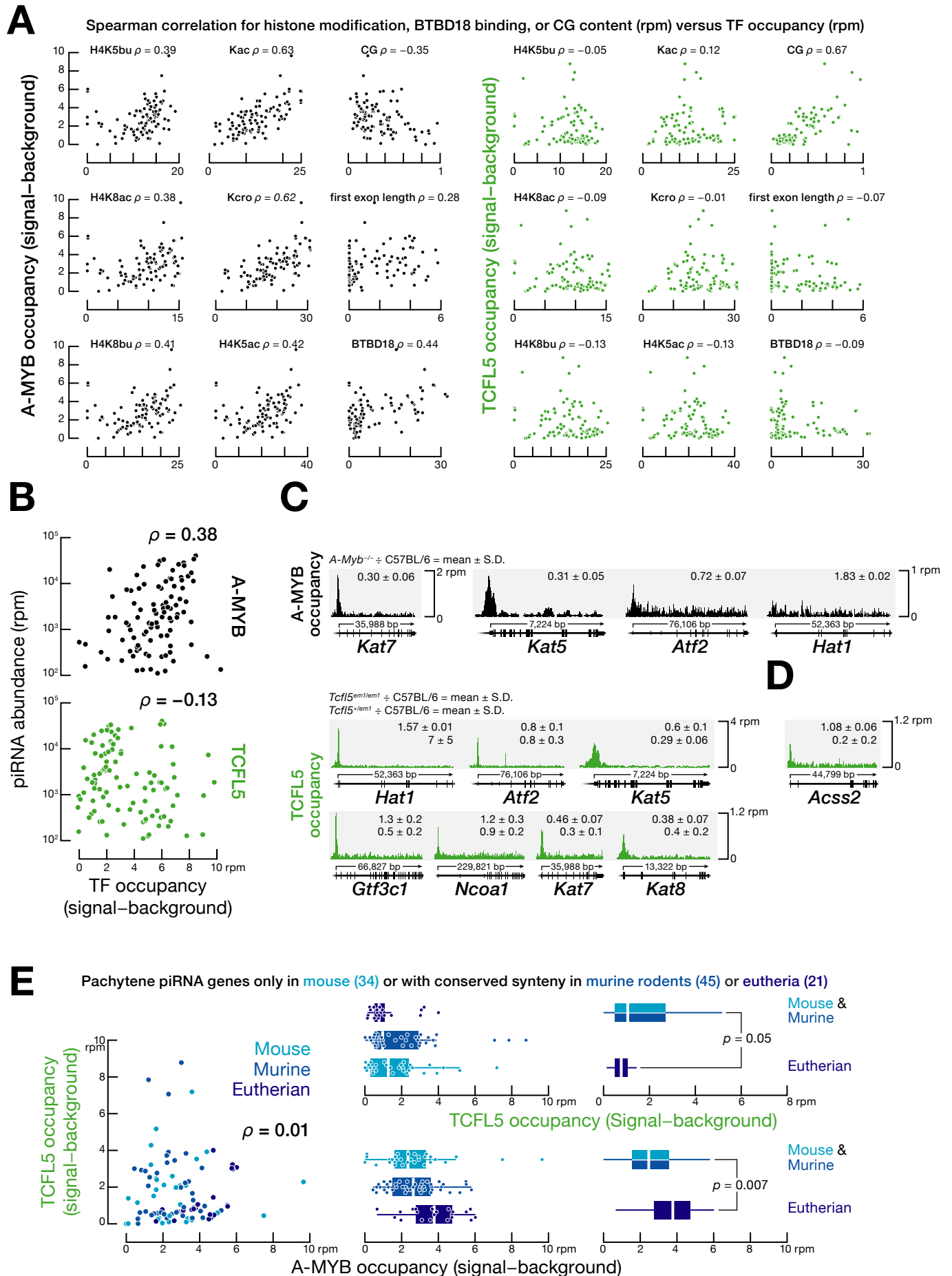
Yu et al., Supplemental Figure S5

A-MYB/TCFL5 regulatory architecture ensures the production of pachytene piRNAs in placental mammals

Tianxiong Yu, Adriano Biasini, Katharine Cecchini, et al.

Supplemental Material	http://rnajournal.cshlp.org/content/suppl/2022/10/14/rna.079472.122.DC1
P<P	Published online October 14, 2022 in advance of the print journal.
Accepted Manuscript	Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.
Open Access	Freely available online through the *RNA* Open Access option.
Creative Commons License	This article, published in *RNA*, is available under a Creative Commons License (Attribution 4.0 International), as described at http://creativecommons.org/licenses/by/4.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *RNA* go to:
http://rnajournal.cshlp.org/subscriptions