bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dazl regulates germ cell survival through a network of polyA-proximal mRNA interactions

Leah L. Zagore[1], Thomas J. Sweet[1], Molly M. Hannigan[1], Sebastien M. Weyn-Vanhentenryck[2], Chaolin Zhang[2], and Donny D. Licatalosi[1]*

[1] *Center for RNA Science and Therapeutics, Case Western Reserve University, Cleveland, Ohio, 44106, USA*
[2] *Center for Motor Neuron Biology and Disease, Columbia University, New York, New York, 10032, USA*

Keywords: germ cell survival; post-transcriptional regulation; 3'UTR, RNA networks; spermatogenesis; RNA binding protein; Dazl

Running title: mRNA regulation and germ cell survival

*To whom correspondence should be addressed
Donny D. Licatalosi
Center for RNA Science and Therapeutics
Case Western Reserve University
Cleveland, OH, 44106, USA
Tel: (216) 368-4757
Fax: (216) 368-2010
Email: ddl33@case.edu

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

The RNA binding protein Dazl is essential for gametogenesis, but its direct *in vivo* functions, RNA targets, and the molecular basis for germ cell loss in *DAZL* KO mice are unknown. Here, we generated transcriptome-wide maps of Dazl-RNA interactions in adult and juvenile mouse testes. In parallel, we used transgenic mice and fluorescence activated cell sorting to isolate *DAZL* knockout germ cells and identify mRNAs sensitive to Dazl deletion. Integrative analyses reveal that Dazl functions as a master regulator of germ cell survival by post-transcriptionally enhancing a vast network of genes necessary for cell cycle regulation and spermatogenesis. Strikingly, Dazl displays a strong positional bias for binding near polyA tails and multimerizes on a subset of its targets. These results reveal a mechanism for Dazl recruitment to its RNA targets and delineate a Dazl-dependent mRNA regulatory program essential for postnatal mammalian germ cell survival.

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

RNA binding proteins (RBPs) are potent post-transcriptional gene regulators. In the nucleus, RBPs can alter pre-mRNA processing to generate mRNAs with different coding and non-coding sequences. In the cytoplasm, RBPs influence mRNA localization, translation, and stability, typically through interactions with 3' untranslated regions (3'UTRs). Importantly, the combined regulatory impacts of RBPs on RNA stability and translation can greatly influence protein expression. Despite widespread recognition of the importance of RBPs in development, immunity, and disease (Scotti and Swanson, 2016), our understanding of how specific RBPs modulate transcriptional programs to control cell fate is extremely limited.

Regulation of mRNA processing and translation is especially relevant during spermatogenesis, the highly ordered process of postnatal male germ cell development that yields haploid spermatozoa (Licatalosi, 2016). During embryogenesis, primordial germ cells (PGCs) commit to a male or female developmental program (McLaren, 2001). Male germ cells proliferate briefly then arrest at G1/G0 as gonocytes that remain quiescent for the remainder of embryogenesis. The first few postnatal days are critical for the establishment of spermatogonial stem cells (Manku and Culty, 2015). During postnatal day one, mitotically-arrested spermatogonial precursors located in the center of testis cords re-enter the cell cycle and begin proliferating. A few days later, these cells extend projections towards the basement membrane and, after making contact, migrate to the basement membrane. After relocation, the germ cells differentiate, acquire "stem cell" properties, and rapidly expand in number. Continued sperm production throughout life depends on proper control of spermatogonial proliferation and differentiation (Manku and Culty, 2015). Many genes critical for this regulation have been identified, including genes encoding cell cycle checkpoint factors, regulators of the DNA damage response, transcription factors, and RBPs that affect mRNA export and stability (Costoya et al., 2004; Raverot et al., 2005; Hao et al., 2008; Takubo et al., 2008; Pan et al., 2009; Saga, 2010;

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Song and Wilkinson, 2014). However, the mechanisms that coordinately regulate expression of genes essential for germ cell maintenance are not well understood.

The functional significance of RBPs in germ cell development is well illustrated by the DAZ family of RBPs. The DAZ (deleted in azoospermia) proteins comprise a family of germ cell restricted RBPs (Daz, Dazl, and Boule) that are necessary for gametogenesis in worms, flies, mice, and humans (VanGompel and Xu, 2011). Their significance was first demonstrated in the 1990s, when *DAZ* was discovered in a region of the Y chromosome deleted in 10-15% of men with azoospermia (no sperm) (Reijo et al., 1995; Reijo et al., 1996). Deletion of *DAZL* in mice leads to a dramatic decrease in the number of surviving germ cells, and complete arrest as cells enter meiosis. Remarkably, transgenic expression of human *DAZL* or *DAZ* partially rescues the extensive germ cell loss in *DAZL KO* mice (Vogel et al., 2002), indicating functional conservation of DAZ RBPs across species. Despite the clear biological importance of DAZ proteins many critical questions remain, including the identities of their direct RNA targets, how these RNAs are regulated, and why loss of this regulation results in germ cell defects. In this study, we provide answers to these long-standing questions.

Dazl's cytoplasmic localization, co-sedimentation with polyribosomes, and association with polyA+ RNA in germ cells (Ruggiu et al., 1997; Tsui et al., 2000) suggest that it may regulate mRNA stability or translation. In addition, yeast two hybrid analysis of Dazl-interactors identified RBPs with cytoplasmic roles in mRNA regulation including Pum2, QK3, and the polyA-binding protein Pabpc1 (Moore et al., 2003; Fox et al., 2005; Collier et al., 2005). However, the scarcity and variable number of germ cells present in *DAZL* KO mice (Ruggiu et al., 1997; Schrans-Stassen et al., 2001; Saunders et al., 2003) present major challenges to investigating Dazl's direct *in vivo* function(s) in the male germline. Consequently, previous Dazl studies have mostly used reconstituted systems including transfection of somatic cell lines(Maegawa et al., 2002; Xu et al., 2013), artificial tethering to *in vitro* synthesized RNAs injected into oocytes or zebrafish (Collier et al., 2005; Takeda et al., 2009), and *in vitro*-derived germ cells (Haston et al.,

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2009; Chen et al., 2014) These efforts have suggested diverse functions for Dazl in different cell contexts, including roles in mRNA stabilization, stress granule assembly, mRNA localization, and translation (Fu et al., 2015). However, transfection assays have demonstrated that Dazl can have opposite effects on the same reporter RNA in different somatic cell lines(Xu et al., 2013). It is not clear whether these discrepancies are due to cell context and/or non-physiological levels or recruitment of Dazl to RNA targets.

While the direct *in vivo* functions are not defined, *in vitro* binding assays and X-ray crystallography of the Dazl RRM identified GUU as a high affinity binding site (Jenkins et al., 2011). However, the frequency of GUU across the transcriptome hampers bioinformatic predictions of functional Dazl-binding sites and RNA targets *in vivo*. Microarray analyses of RNAs that co-immunoprecipitate (IP) with Dazl (Reynolds et al., 2005; Reynolds et al., 2007; Chen et al., 2014) have suggested potential targets. Yet, few have been examined in *DAZL* KO mice, and these cannot account for the dramatic loss of germ cells seen in DAZL null mice. Furthermore, different investigators have arrived at alternate conclusions about Dazl's role as a translation repressor or activator based on immunofluorescence (IF) assays of wild type (WT) and *DAZL* KO germ cells (Reynolds et al., 2005; Chen et al., 2014). Moreover, neither group explored whether the observed differences in protein abundance are associated with changes in RNA levels.

Collectively, these observations underscore the need to identify Dazl's direct *in vivo* RNA targets in an unbiased and transcriptome-wide manner, and to devise a new strategy for the isolation of limiting *DAZL* KO germ cells for transcriptome-profiling.

In this study, we used an integrative approach to elucidate the direct RNA targets and *in vivo* functions of Dazl in male germ cells. High-resolution, transcriptome-wide mapping of Dazl-RNA interactions in testes reveal Dazl binding to a vast set of mRNAs, predominantly through polyA-proximal interactions. Extensive analyses of the Dazl-RNA maps strongly suggest that Dazl recruitment to its RNA targets is facilitated by local Pabpc1-polyA interactions. Using

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transgenic mice with fluorescently-labeled germ cells and FACS, we isolated germ cells from *DAZL* KO testes and wild type (WT) controls and used RNA-Seq to identify mRNAs that are sensitive to *DAZL*-deletion. Integration of the RNA-Seq and Dazl-RNA interaction datasets revealed that Dazl enhances expression of key cell cycle regulatory genes. Collectively, these data provide important new insights into the mechanism of Dazl recruitment to its RNA targets, the molecular basis for germ cell loss caused by *DAZL* deletion, and reveal an mRNA regulatory program that is essential for postnatal germ cell survival.

Results

Dazl binds GU-rich sequences in the testis transcriptome

To comprehensively map direct sites of Dazl-RNA interactions *in vivo*, HITS-CLIP libraries were generated from Dazl-RNA complexes purified from UV cross-linked adult mouse testes and sequenced using the Illumina platform (Figure 1A). The resulting CLIP reads from three biological replicates were filtered and mapped individually (Supplemental Table 1), and then intersected to reveal 11,297 genomic positions with overlapping CLIP reads in 3/3 libraries; hereafter designated as BR3 clusters (biologic reproducibility 3/3, Figure 1B). Remarkably, these interactions reveal Dazl directly binds over 3900 transcripts in adult testis.

We next identified CLIP peaks within each cluster (Supplemental Figure 1). Consistent with X-ray crystallographic studies (Jenkins et al., 2011), these sites were enriched for GUU motifs compared to shuffled control sequences (1.8-fold, χ^2=0, all peaks, Figure 1C). Using RNA-Seq data from age-matched testes, CLIP peaks were normalized to RNA levels and parsed into 10 bins based on CLIP:RNA-Seq ratios. A clear correlation was evident between CLIP:RNA-Seq read ratio, GUU-enrichment, and the proportion of peaks in each bin that contained GUU (Figure 1D). Separately, *de novo* motif analysis using the MEME suite (Bailey et al., 2015) identified GTT-containing motifs as the most enriched sequence elements in genomic regions corresponding to peaks with the highest CLIP:RNA-Seq ratios (Figure 1E).

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dazl binding sites were also examined using an independent read mapping and analysis pipeline that takes advantage of cross-link induced mutation sites (CIMS) that reflect sites of protein-RNA cross-linking (Zhang and Darnell, 2011). *De novo* motif analysis of the top 1000 CIMS deletion sites (±10nt) identified GUU-containing motifs as the most enriched sequence elements (Figure 1F). Examination of 6mers and 4mers associated with the top 1000 CIMS sites ±20nt relative to background sequences identified enrichment of GUU-containing sequences around deletions, particularly the GUUG motif, which was enriched 40-60 fold around CIMS deletion sites, with cross-linking occurring at U residues within GUU triplets (Figure G,H). Thus, two independent read mapping and bioinformatic workflows confirm that Dazl predominantly binds GUU-rich sequences *in vivo*.

Strikingly, 39% of all Dazl BR3 clusters contained multiple peaks, with a median spacing of ~60nt (Figure 1C, 1I; Supplemental Figure 1). Notably, motif enrichment values for adjacent peak sequences were nearly identical, regardless of relative CLIP tag density (data not shown) or 5' versus 3' position (Figure 1J). These observations are consistent with Dazl homodimerization, which was observed in yeast two hybrid and *in vitro* assays (Ruggiu and Cooke, 2000). We infer that Dazl multimerizes on a subset of its natural RNA targets.

Dazl predominantly binds GUU sites in close proximity to mRNA polyA tails

We next examined the distribution of Dazl-RNA contacts across the transcriptome. Strikingly, the majority of Dazl BR3 sites mapped to 3'UTRs of thousands of protein-coding genes (Figure. 2A; Supplemental Table 2). To improve the annotation of Dazl-3'UTR interactions we used PolyA-Seq to generate a quantitative global map of polyA site utilization in adult testes (Supplemental Figure 2A). Consistent with widespread transcription and alternative polyadenylation (APA) in the testis (Soumillon et al., 2013; Li et al., 2016), PolyA-Seq identified 28,032 polyA sites in RNAs from 16,431 genes (Figure 2B, Supplemental Figure 2B). Importantly, gene expression estimates from PolyA-Seq and RNA-Seq showed a strong

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

correlation (R=0.83, Supplemental Figure 2C), as did estimates of polyA site usage from PolyA-Seq and qRT-PCR (R=0.72, 10 candidates examined; Supplemental Figure 2D).

Having defined 3'UTRs expressed in the testis, metagene analyses were performed to examine the distribution of 3'UTR Dazl-RNA interactions relative to the upstream stop codon and downstream polyA site. Compared to the expected distribution if peaks were randomly distributed (dashed line in Figure 2C), no positional preference was observed when Dazl-3'UTR interactions were measured relative to the stop codon (Figure 2C, compare red line to dashed line). In stark contrast, a prominant positional bias for Dazl-RNA contacts was observed relative to the polyA site, with strong enrichment within ~150nt and the greatest number of interactions ~50nt upstream of the polyA site (Figure 2C, compare blue line to dashed line). This positional preference is also evident on 3'UTRs with multiple Dazl CLIP peaks, with broadening of Dazl binding in a 3'-to-5' direction as the number of peaks per 3'UTR increased (Figure 2D).

Further examination showed that 3'UTRs with Dazl BR3 CLIP sites had a higher proportion of uridine residues compared to a set of 3'UTRs lacking Dazl CLIP reads (Figure 2E, Supplemental Figure 3A). Moreover, the most enriched motifs in Dazl-bound versus control 3'UTRs were U-rich, with 2 of the top 7 most-enriched 4mers containing GUU (Figure 2F). Similarly, 3 of the top 4 6mers (and 7 of the top 20) contained GUU (Supplemental Figure 3B). Although both sets of 3'UTRs had enrichment of GUU upstream of polyA sites, GUU-enrichment was greater and extended over a broader region in Dazl-bound 3'UTRs (Figure 2G).

Collectively, transcriptome-wide mapping of direct, biologically-reproducible Dazl-RNA interactions in mouse testes indicates that Dazl predominantly binds U-rich 3'UTRs with preferential binding to GUU-containing sequences ~50nt upstream of polyA sites. In addition, the data suggests that Dazl multimerizes in a 3'-to-5' manner on a subset of its mRNA targets.

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

PolyA tracts specify positions of Dazl-RNA interactions

Enrichment of Dazl-RNA contacts near polyA tails, together with evidence of Dazl-Pabpc1 binding (Collier et al., 2005) suggest that Dazl recruitment to its RNA targets may be facilitated by local Pabpc1-polyA interactions. In addition to binding polyA tails, Pabpc1 interacts with genomic-encoded A-rich sequences in mRNA (Kini et al., 2016), including auto-regulatory interactions in the 5'UTR of its own mRNA (Hornstein et al., 1999). Interestingly, *PABPC1* was among a small set of 48 genes with Dazl BR3 CLIP sites in the 5'UTR (57 sites in 5'UTRs of 48 genes, compared to 6,397 sites in 3'UTRs of 3,249 genes), prompting further examination of these outliers (Figure 3A; Supplemental Table 2). For the majority (32/48 genes), CLIP read number was greatest in the 3'UTR and declined in a 3'-to-5' manner (for example *ACYP1*, *CHIC2*, and *CCNI*, Figure 3B-D, 5'low genes). *PABPC1* was one of only 16 genes where more than 50% of the total CLIP reads were located in the 5'UTR (Figure 3E-F, 5'high genes). Marked differences were observed in the sequence features associated with these two sets of 5'UTRs. Consistent with low CLIP coverage, 5'UTRs of 5'low genes were not enriched for GUU-containing sequences. In contrast, 5'UTRs of 5'high genes resembled 3'UTRs, whereby the most enriched motifs included polyU and polyU tracts interrupted by a single G (Figure 3G). Strikingly, the second most-enriched hexamer in 5'UTRs of 5'high genes was AAAAAA (Figure 3G, Supplemental Table 3). Thus, a common feature of Dazl-RNA interactions in 5'UTRs and 3'UTRs is a local polyA tract (either genomic-encoded or added to mRNA 3'ends, respectively). These observations implicate polyA tracts (more specifically, local polyA-Pabpc1 interactions) as a key determinant in specifying positions of Dazl-RNA interactions. They also shed light on the potential mechanism of Dazl recruitment to mRNA and the basis for the widespread polyA-proximal Dazl-RNA binding patterns observed across the germ cell transcriptome.

Dazl loss remodels the germ cell transcriptome

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Significant barriers to understanding Dazl's *in vivo* functions include the scarcity and variable number of germ cells in *DAZL* KO mice (Ruggiu et al., 1997; Lin and Page, 2005). To overcome these obstacles and collect *DAZL* KO germ cells for molecular analyses, we used Cre-lox to label germ cells with green fluorescent protein (GFP) followed by FACS. This approach (Zagore et al., 2015) utilizes the *Stra8-iCre* transgene that expresses Cre recombinase in postnatal germ cells (Sadate-Ngatchou et al., 2008), and the *IRG* transgene that expresses GFP following Cre-mediated recombination (De Gasperi et al., 2008). To generate animals with GFP+ Dazl-deficient postnatal germ cells, we used mixed background (CD1xC57BL/6J) breeders heterozygous for the null *DAZL*^{Tm1hgu} allele (Ruggiu et al., 1997) (see Methods). Testes of *Stra8-iCre*$^{+}$; *IRG*$^{+}$; *DAZL*$^{++}$ and *Stra8-iCre*$^{+}$; *IRG*$^{+}$; *DAZL*$^{Tm1hgu/Tm1hgu}$ mice were indistinguishable at postnatal day 0 (P0), however germ cells declined steadily thereafter in the latter (data not shown). Cre expression from *Stra8-iCre* commences ~P3, therefore we selected animals at P6 for tissue collection and FACS. At this age, spermatogonia are the only germ cells present. As expected, GFP+ spermatogonia were significantly reduced in *Stra8-iCre*$^{+}$; *IRG*$^{+}$; *DAZL*$^{Tm1hgu/Tm1hgu}$ animals compared to littermate controls (Figure 4B, C). Importantly, germ cell-restricted GFP expression in *Stra8-iCre*$^{+}$; *IRG*$^{+}$; *DAZL*$^{++}$ and *Stra8-iCre*$^{+}$; *IRG*$^{+}$; *DAZL*$^{Tm1hgu/Tm1hgu}$ mice was confirmed by IF microscropy (Figure 4A), and germ cell markers were significantly enriched in GFP+ cells (data not shown).

To determine how the absence of Dazl impacts global mRNA levels, RNA-Seq analysis was performed on GFP+ cells collected by FACS. Overall, the number of expressed genes and the distribution of their RPKM values were comparable between GFP+ WT and *DAZL* KO cells (11,739 and 12,269 expressed genes with RPKM>1, respectively). However, 1,462 transcripts had reduced RNA levels and 1,584 increased in *DAZL* KO cells (minimum 2-fold, adjusted $P<0.01$; Figure 4D, green and red dots, respectively; Figure 4E). Notably, *DDX4* and *SYCP3* were among the genes with reduced RNA in *DAZL* KO cells (86- and 26.5-fold, respectively). Therefore, the reduced levels of *Ddx4* and *Sycp3* proteins previously observed in *DAZL* KO

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

germ cells by IF and attributed to reduced translation (Reynolds et al., 2005; 2007) are associated with significant reductions in their corresponding mRNAs.

To further explore why *DAZL* deletion results in rapid postnatal germ cell loss, we more closely examined the genes with steady state mRNA differences between WT and KO cells. Notably, 41% of *DAZL*-repressed genes (ie., up-regulated in KO) had little to no expression in WT cells (656 genes with RPKM<1 in WT), indicating that these genes are under negative regulation in WT cells. Gene ontology (GO) terms associated with this set were enriched for terms associated with the extracellular matrix and plasma membrane (Supplemental Figure 4A). These and other closely related GO terms were also enriched when GO analysis was performed on the complete set of 1,584 genes with increased RNA in *DAZL* KO cells (Supplemental Figure 4B). Thus, a disproportionate number of genes with increased expression in *DAZL* KO cells are normally expressed at very low levels in WT cells and encode extracellular and membrane-associated proteins.

In contrast, genes enhanced by *DAZL* (ie., down-regulated in KO) were among the most abundantly expressed, with ~50% of these among the top 20% of all genes expressed in WT cells (Figure 4E). In addition, a disproportionate number of *DAZL*-enhanced genes encode nuclear proteins with roles in cell cycle regulation (Supplemental Figure 4C). This includes clusters of genes and enriched GO terms associated with mitotic cell cycle control, DNA repair, regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle, meiotic cell cycle, and cell cycle checkpoints (Supplemental Figure 5). Also enriched were groups of genes with roles in RNA processing, RNA transport, transcription, DNA methylation, and a broad set of genes with the annotation 'spermatogenesis'. Contained in this last group are several genes essential for spermatogenesis including *DND1* (Yamaji et al., 2017), *PLZF/ZBTB16* (Costoya et al., 2004), *SOHLH2* (Hao et al., 2008) and *SOX3* (Raverot et al., 2005) (Supplemental Figure 5, group G).

Altogether, these observations indicate that genes with decreased or increased RNAs in

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DAZL KO cells differ considerably with respect to the distribution of RPKM values in WT cells and the functions of the encoded proteins. Importantly, the data shows that a large cohort of genes that are essential for cell cycle regulation and spermatogonial maintenance are normally expressed at very high levels in spermatogonia and have significantly reduced mRNA in the absence of Dazl.

Dazl directly enhances mRNA levels of a subset of in vivo targets

To investigate which RNA changes are primary or secondary consequences of *DAZL* loss, we mapped Dazl-RNA interactions in P6 testes using iCLIP. As with Dazl BR3 sites in adult testes, P6 BR3 regions were enriched for GUU sequences and the majority mapped to 3'UTRs (Figure 5A, B). Interestingly, when motif-enrichment analysis was performed on complete BR3 regions, AAAAAA was the second most enriched 6-mer (Figure 5B, left). However, AAAAAA enrichment dropped when motif analysis was performed on the 5'ends of BR3 sites (±5nt), where cross-link sites are enriched in iCLIP libraries (Konig et al., 2010) (Figure 5A, B). These observations provide further evidence that polyA tracts are enriched downstream of direct sites of Dazl-RNA interactions.

To accurately map positions of Dazl-RNA contacts in spermatogonia 3'UTRs, PolyA-Seq libraries were generated from GFP+ spermatogonia isolated by FACS from *Stra8-iCre*+;*IRG*+ testes. Of the 28,038 3'UTRs defined by PolyA-seq of whole adult testis (described above; Figure 2B), 16,502 were identified in the PolyA-seq data from GFP+ spermatogonia, corresponding to 10,370 genes. Intersecting the P6 iCLIP and spermatogonia PolyA-seq datasets showed that 84% of P6 Dazl BR3 sites (5,400/6,465) mapped to 3'UTRs of 2,290 genes, indicating Dazl directly binds to a vast network of RNAs in juvenile as well as adult germ cells. Furthermore, metagene analysis showed that Dazl-RNA interactions in spermatogonia-expressed 3'UTRs also had a polyA-proximal bias, with the greatest enrichment within 150nt of the polyA site (Figure 5C). Thus, PolyA-seq and transcriptome-wide mapping of Dazl-RNA

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

interactions confirm a polyA-proximal bias for Dazl-RNA binding in both adult and juvenile testes.

Interestingly, the majority of genes with Dazl-3'UTR interactions (74%) had higher RPKM values in WT compared to KO GFP+ cells (74% versus 26%, Figure 5D, left). Comparing genes with Dazl-3'UTR interactions to those displaying RNA differences between WT and KO cells identified a common set of 583 genes (2-fold or greater, $P<0.01$). Strikingly, a disproportionate number of these were *DAZL*-enhanced (501 genes with reduced RNA in KO) rather than *DAZL*-repressed (82 genes with increased RNA in KO) (Figure 5D, green and red dots, respectively, Figure 5E). Metagene analyses revealed striking differences in the distribution of Dazl binding sites in these 3'UTRs. Whereas BR3 sites in 3'UTRs of *DAZL*-repressed genes showed no positional preference (Figure 5D, red line), Dazl binding sites in 3'UTRs of *DAZL*-enhanced genes were enriched near the polyA tail, with the greatest number of interactions within 100nt of the polyA site (Figure 5D, green line). Moreover, GUU motifs and Dazl BR3 interaction sites were significantly more abundant in *DAZL*-enhanced rather than *DAZL*-repressed genes (Figure 5F). Altogether, these observations indicate that *DAZL*-enhanced genes are directly regulated by Dazl, while the *DAZL*-repressed genes represent indirect targets. Furthermore, they strongly suggest that Dazl functions directly *in vivo* as a positive post-transcriptional regulator of gene expression through polyA-proximal interactions.

We also examined 236 *DAZL*-insensitive genes with Dazl-3'UTR interactions but whose RNA levels did not differ by more than 20% between WT and *DAZL* KO cells (Figure 5D, blue dots). Remarkably, Dazl-RNA interactions in these 3'UTRs also displayed a strong polyA proximal bias (Figure 5D, blue line). Furthermore, we identified several examples where Dazl-3'UTR binding patterns were indistinguishable between *DAZL*-insensitive and *DAZL*-enhanced 3'UTRs (Supplemental Figure 6). These observations raise the possibility that a significant proportion of Dazl-3'UTR interactions may not have regulatory function and may represent opportunistic interactions reflecting the general mechanism of Dazl recruitment to its RNA

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

targets. They also indicate that polyA-proximal binding alone is not sufficient for Dazl to enhance mRNA levels and suggest that Dazl acts in a 3'UTR specific manner.

Dazl post-transcriptionally controls a network of cell cycle regulatory genes

To understand the primary causes of germ cell depletion in *DAZL* KO testes, we more closely examined the 501 *DAZL*-enhanced genes with Dazl-3'UTR interactions. Heirarchical clustering revealed 17 groups of significantly enriched GO terms, with the majority (12/17) associated with different aspects of cell cycle regulation, including chromatin modification, condensation, or segregation; synaptonemal complex assembly; mitotic spindle checkpoint; DNA replication, repair, or packaging; and mitotic or meiotic cell cycle regulation (Figure 6, groups with asterisks). The remaining groups contained genes and GO terms associated with RNA processing, ubiquitination, spermatogenesis, RNA polymerase II (pol II) transcription, and mRNA transport. Among the *DAZL*-enhanced genes are several required for spermatogenesis (Figure 6, groups A, F, and M), many with essential roles in spermatogonial maintenance or proliferation including *ATM* (Takubo et al., 2008), *DMRT1* (Zhang et al., 2016), *NXF2* (Pan et al., 2009), *SOHLH2* (Hao et al., 2008), *SOX3* (Raverot et al., 2005), and *PLZF/ZBTB16* (Costoya et al., 2004). These observations indicate that Dazl directly enhances postnatal germ cell survival by promoting high mRNA levels for a network of genes with key roles in cell cycle regulation and essential for spermatogenesis.

We identified several examples where genes encoding subunits of the same complex were among the 501 *DAZL*-enhanced genes (for example, gene cluster I in Figure 6). To further assess physical interactions between the protein products of Dazl's bound and regulated mRNA targets, we used the STRING database of protein-protein interactions (PPIs) (Szklarczyk et al., 2015). Focusing on high confidence PPIs supported by experimental evidence or curated databases (minimum interaction score 0.9), we observed a higher than expected number of the 501 genes encode protein that interact with one another (PPI enrichment p-value = 0). This

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

includes a broad network of interacting proteins including basal transcription factors, subunits of RNA pol II, RNA processing and export factors, chromatin binding proteins, epigenetic regulators, and several components of the proteasome and ubiquitination machineries (Figure 7). Together, GO and PPI enrichment analyses in combination with Dazl CLIP, PolyA-Seq, and RNA-Seq demonstrate that polyA-proximal Dazl-3'UTR interactions directly enhance expression of a network of genes whose protein products physically interact and have essential roles in cell cycle regulation and germ cell survival.

Discussion

The necessity of DAZ proteins for germ cell survival is well established in multiple species. However, the direct targets, regulatory roles, and biological functions of these RBPs remained unclear. Our integrative analyses combining transgenic mice, FACS, and a panel of unbiased, transcriptome-wide profiling tools provide important insights into the molecular and biological functions of this important family of RBPs.

Dazl controls a network of genes essential for germ cell survival

Our RNA-Seq analysis of WT and *DAZL* KO germ cells shows a critical requirement for *DAZL* in maintaining fidelity of the germ cell transcriptome. More specifically, we demonstrate that Dazl directly functions as a positive post-transcriptional regulator of a network of genes necessary for cell cycle regulation and germ cell survival. This includes genes with key roles in transcriptional control, chromatin modification, RNA processing, and several regulatory processes and checkpoints that control cell cycle progression. These observations provide important insights into the molecular basis for the dramatic decrease in postnatal *DAZL* KO germ cells during a period when WT gonocytes normally resume mitotic proliferation.

Global measurements of protein and mRNA stability have indicated that genes with short mRNA and protein half-lives are significantly enriched for the GO terms 'transcription', 'cell

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cycle', and 'chromatin modification' (Schwanhäusser et al., 2011), all of which are among the most-enriched GO terms associated with *DAZL*-enhanced genes with Dazl-3'UTR interactions. Notably, the *DAZL*-enhanced genes were among the most abundantly expressed genes in WT germ cells. Together, these observations indicate that Dazl function is required to maintain high steady state levels of mRNAs that are inherently unstable in germ cells, thus ensuring high concentrations of regulatory factors necessary to resume and/or sustain postnatal germ cell proliferation.

The enrichment of genes that encode core components of the transcription machinery and epigenetic regulators among those enhanced by *DAZL* indicate that some of the RNA differences between WT and *DAZL* KO cells may be secondary changes due to altered transcription. Consistent with this possibility, nearly 40% of genes with increased RNA in *DAZL* KO cells had little to no detectable expression in WT cells (RPKM<1). These observations indicate that in addition to directly regulating cytoplasmic mRNA levels, Dazl exerts indirect control over the germ cell transcriptome by regulating mRNAs for transcription factors and epigenetic regulators that, in turn, define which genes are transcribed. The presence of epigenetic regulators among the list of *DAZL*-enhanced genes with Dazl-3'UTR interactions (including *PIWIL2*, *TDRD9, DNMT1*, and *DNMT3B*) may also partially explain the defects in erasure and re-establishment of DNA methylation patterns in embryonic *DAZL* KO germ cells (Gill et al., 2011; Haston et al., 2009).

PolyA tracts facilitate Dazl-3'UTR interactions across the germ cell transcriptome

Transcriptome-wide mapping of direct Dazl-RNA interaction sites shows that Dazl binds mRNAs from a significant proportion of genes expressed in adult and juvenile (3,907 and 2,290 genes, respectively having Dazl-3'UTR interactions). Given the widespread transcription and high RNA levels in male germ cells (Soumillon et al., 2013), it is not unexpected or uncommon for a germ cell-expressed RBP to reproducibly bind transcripts from thousands of genes

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Hannigan et al., 2017). However, the Dazl CLIP maps are unusual in that the binding sites are highly concentrated near polyA sites. Our data suggest that widespread polyA-proximal binding reflects a general mechanism of Dazl recruitment and/or stabilization on mRNA. More specifically, multiple lines of evidence suggest a role for local Pabpc1-polyA interactions in specifying where Dazl-RNA interactions occur. This includes the prominent positional bias of Dazl-RNA interactions upstream of polyA sites; enrichment of polyA tracts in the rare set of 5'UTRs with high CLIP read levels; and enrichment of polyA tracts downstream of direct sites of Dazl-RNA cross-linking. Based on these observations and previous evidence of RNA-independent Dazl-Pabpc1 interactions (Collier et al., 2005), we propose that Dazl recruitment and/or stabilization on mRNA is mediated by local Pabpc1-polyA interactions. Considering the prevalence of GUU throughout the transcriptome, a requirement for local Pabpc1-polyA interactions in specifying sites of Dazl-RNA binding would increase the probability of Dazl loading on 3'UTRs, where the majority of cytoplasmic RBPs exert their regulatory functions.

Dazl enhances mRNA levels of specific targets through polyA-proximal interactions

Regulation of mRNA translation and stability are intricately coupled processes (Bicknell and Ricci, 2017), therefore it is not clear whether Dazl functions primarily to prevent mRNA decay, or as a translational enhancer that indirectly impacts mRNA levels. It is also possible that Dazl's primary function differs on distinct mRNAs. Additional studies are necessary to determine why some mRNAs with polyA-proximal interactions have reduced RNA levels in *DAZL* KO cells, while others are insensitive to *DAZL*-deletion. Notably, 3'UTRs of the former had the highest number of GUU sites per 3'UTR, despite modestly shorter lengths (Figure 5F). In addition, Dazl-RNA contacts on *DAZL*-insensitive genes were more concentrated in the -50nt region, with reduced upstream 3'UTR binding compared to contacts in *DAZL*-enhanced genes. This suggests that multiple Dazl-RNA contacts, potentially established via 3'-to-5' multimerization, are associated with enhanced RNA levels for a subset of genes. Binding to a broader segment of the 3'UTR may displace or neutralize mRNA-specific *cis*- or *trans*-acting

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

negative regulators that directly or indirectly promote mRNA decay. However, we also identified several examples where the Dazl-RNA binding on 3'UTRs of *DAZL*-enhanced and *DAZL*-insensitive genes were indistinguishable suggesting that Dazl functions in a 3'UTR-specific manner(Supplemental Figure 6). These findings also highlight the importance of combining RBP-RNA interaction maps with RNA-profiling to distinguish functional from potentially opportunistic RBP-RNA interactions.

In conclusion, our integrative analysis demonstrates that germ cell survival depends on an mRNA regulatory program controlled by Dazl. Importantly, the data also reveal the molecular basis of germ cell loss in *DAZL* KO mice. Given the functional conservation between mouse *DAZL*, human *DAZL*, and *DAZ* (Vogel et al., 2002), our study provides important insights into the molecular basis for azoospermia in 10-15% of infertile men with Y chromosome micro-deletions. Dazl's RNA targets extend far beyond germ cell-specific genes and include many that encode core components of macromolecular complexes present in all proliferating cells. Therefore, our findings may also be relevant to other human diseases as *DAZL* is a susceptibility gene for human testicular cancer (Ruark et al., 2013), and is amplified or mutated in nearly 30% of breast cancer patient xenografts examined in a single study (Eirew et al., 2015). We propose a general model (Supplemental Figure 7) whereby Dazl binds a vast set of mRNAs via polyA-proximal interactions facilitated by Pabpc1-polyA binding and post-transcriptionally enhances expression of a subset of mRNAs, namely a network of genes essential for cell cycle regulation and mammalian germ cell maintenance. These observations provide new insights into molecular mechanisms by which a single RBP is recruited to its RNA targets and coordinately controls a network of mRNAs to ensure germ cell survival.

Materials and Methods

Animals and tissue collection. C57BL/6J animals bearing the *DAZL*^{Tm1hgu} allele were rederived at the Case Transgenic and Targeting Facility, and bred with animals bearing the

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Stra8-iCre or *IRG* transgenes. Mixed background (CD1xC57BL/6J) *Stra8-iCre*$^{++}$; *DAZL*$^{Tm1hgu/+}$ males and *IRG*$^{+}$; *DAZL*$^{Tm1hgu/+}$ females were crossed to generate *DAZL* WT and KO offspring with GFP+ male germ cells. Day of birth was considered P0. For all procedures, animals were anesthetized by isoflourane inhalation and death confirmed by decapitation or cervical dislocation. HITS-CLIP, iCLIP, adult testis RNA-Seq, and adult testis Poly-Seq were performed using CD-1 animals purchased from Charles River labs. Spermatogonia PolyA-Seq libraries were generated from FACS-isolated cells from 8 week *Stra8-iCre*$^{+}$; *IRG*$^{+}$ C57BL/6J males as previously described (Zagore et al., 2015). All animal procedures were approved by the Institutional Animal Care and Use Committee at CWRU.

Cell isolation. Isolations of GFP+ cells was performed using dual fluorescence labeling as previously described (Zagore et al., 2015), except P6 cells were not stained with Hoechst and were collected using either BD Biosciences Aria or iCtye Reflection cytometry instrumentation.

Microscopy. P6 testes were decapsulated in cold HBSS, fixed overnight in 4% paraformaldehyde, and paraffin-embedded by the Histology Core Facility at CWRU. Slides containing 5uM sections were de-paraffinized with 3 5-minute washes in xylene followed by a 5-minute rinse in 100% ethanol. Tissue was rehydrated with 5-minute incubations in 100%, 95%, 70% and 50% ethanol followed by tap water. Antigen retrieval was performed with citrate buffer, pH 6.0 (10mM Sodium Citrate, .05% Tween 20), at 95C for 20 minutes. Slides were cooled with tap water for 10 minutes followed by a 5-minute wash in 1X PBS. Tissue was permeabilized for 10 minutes in 0.25% TritonX-100 in 1X PBS and rinsed in PBST (0.1% Tween 20 in 1XPBS). Slides were blocked in 1% BSA in PBST for 1 hour. Sections were incubated with 1:500 anti-GFP (Abcam) and 1:500 anti-Ddx4 (Abcam) for 1 hour followed by 3 5-minute washes in PBST. An additional one hour incubation was performed using 1:100 anti-mouse Alexa-488 (ThermoFisher) and 1:100 anti-rabbit Cy3 (Jackson ImmunoResearch). Slides were washed 3

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

times in PBST before staining with 0.5ug/mL DAPI for 5 minutes and rinsing in 1X PBS. Tissue was mounted on coverslips with Fluoromount G (Southern Biotech) and images captured using a Deltavision Deconvolution Microscope.

RNA-Seq. Adult testis RNA-Seq libraries (Tru-Seq, Illumina) were prepared from testes of two 8 week CD1 males. Illumina TruSeq/Clontech ultra low input library preparation with Nextera XT indexing was used to generate RNA-Seq libraries from GFP+ WT and *DAZL* KO cells purified by FACS. All RNA-Seq libraries were sequenced at CWRU Sequencing Core. Read mapping and gene expression quantification was performed using Olego and Quantas (Wu et al., 2013).

PolyA-Seq. *Sequencing library construction.* Adult testis PolyA-Seq libraries were generated from 8 week CD1 mice. Spermatogonia PolyA-Seq libraries were generated from 400ng of RNA from spermatogonia isolated by FACS from 8 week old *Stra8-iCre*$^{+}$; *IRG*$^{+}$ males as previously described (Zagore et al., 2015). PolyA+ RNA was selected by oligodT-hybridization (Dynal) and fragmented by alkaline hydrolysis. RNA fragments 50-100 nt long were gel purified from 10% PAGE/Urea gels, and used as template for reverse transcriptions with SuperScript III (Invitrogen). First strand cDNA was gel purified, then circularized with CircLigase (EpiCentre). Circularized DNA was used as the template for PCR. After cycle number optimization to obtain the minimal amount of PCR product detectable by Sybr Gold (Invitrogen), the PCR product was gel purified and used as template for a second PCR with primers containing Illumina adaptor sequences. Adult and spermatogonia PolyA-Seq libraries were sequenced at CWRU and UC Riverside, respectively.

Read processing, filtering, and mapping. Reads consisting of two or more consecutive N's or all poly(A) were filtered out, then remaining reads were processed to remove adapter and poly(A) sequences. Remaining reads were mapped to mouse genome mm9 with GSNAP allowing 2 mismatches. Reads with identical genomic footprint and 4N's were then collapsed into single

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reads as likely PCR duplicates. Mapped reads that were likely internal priming events rather than true poly(A) tails were discarded if they had 6 or more consecutive A's or 7 or more A's in the 10 nt window downstream of the read 3' end (cleavage and polyadenylation site). We accepted only 3' ends with 1 of 14 hexamers within 50 nt upstream of the 3' end (as described in (Martin et al., 2012).

Clustering of reads into poly(A) sites. To focus on high confidence poly(A) sites, we accepted read 3' ends that were ≥ 1 tag per million reads (TPM). Due to heterogeneity in cleavage site selection, reads that fell within 10 nt of each other were clustered into poly(A) sites, and a TPM was calculated for the entire cluster (polyA site region).

Identification of 3'UTRs. PolyA sites regions from adult PolyA-Seq were intersected with RefSeq genes. Intergenic sites within 10kb of an upstream stop codon were assigned to the upstream gene. For each polyA site region mapped to a protein coding gene, the closest upstream stop codon was identified and used to define 28,032 3'UTRs in adult testes (only polyA sites with at least 5% of the total PolyA-Seq reads in a gene were considered). To identify 3'UTRs expressed in spermatogonia, mapped reads from spermatogonia PolyA-Seq libraries were intersected with 100nt regions corresponding to the 28,032 adult testis polyA sites, identifying 16,502 sites with read counts greater than zero in both spermatogonia PolyA-Seq replicate libraries.

qRT-PCR validation. qRT-PCR was performed as previously described (Zagore et al. 2015). Primer sequences are available in Supplemental Table 2.

Dazl HITS-CLIP from adult testes. *Library construction.* Dazl HITS-CLIP libraries were generated from testes from three biologic replicate 8-week old mice. Testes were detunicated in cold HBSS and seminiferous tubules UV-irradiated on ice. All steps were performed as previously described (Licatalosi et al., 2008; Licatalosi et al., 2012). Libraries were sequenced at the CWRU Sequencing Core and resulting reads mapped using Bowtie and Tophat (Trapnell et al., 2012) and filtered as previously described (Hannigan et al., 2017).

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Identification of CLIP peaks. To identify peak regions in BR3 HITS-CLIP regions, the sum of all CLIP read lengths in a cluster was divided by the length of the cluster footprint to obtain an 'expected' density. Peaks are defined as regions of at least 10 nt long where the observed CLIP read density exceeded the expected density. 11,224 out of 11,297 had identifiable peaks (99.4% of BR3 clusters, 18364 peaks total).

Peak normalization and binning. For each BR3 peak, the sum of all observed-expected values per nt was calculated and divided by the average RNA-Seq read density per nt of the BR3 cluster. Only peaks within clusters with an average RNA-Seq read density per nt greater than 10 were considered for binning (13,106 peaks).

Genomic distribution of BR3 positions. To annotate BR3 sites, RefSeq coding regions, 5'UTRs, 3'UTRs, and introns were downloaded from the UCSC genome browser and intersected individually with BR3 coordinates.

Examination of genes with 5'UTR BR3 sites. For this analysis, only BR3 sites that mapped unambiguously to a single type of RefSeq gene fragment were examined (see Supplemental Figure 2).

Dazl iCLIP from P6 testes. *Library construction and analysis.* Individual Dazl iCLIP libraries were generated from three P6 testes and processed similar to HITS-CLIP libraries with the following changes. No 5'linker ligation step was performed, and the RT primer contained iSp18 spacers and phosphorylated 5' end permitting circularization of first strand cDNA (after gel purification) to generate a PCR template without linearization (as described by (Ingolia, 2010). Libraries were sequenced at the CWRU Sequencing Core. Identical reads within each library were collapsed, and then mapped using Olego (Wu et al., 2013). Reads overlapping repetitive elements were discarded. After removing exon-spanning reads (0.06% of total), MultiIntersectBED (Quinlan and Hall, 2010) was used to identify genomic regions with overlapping CLIP reads from 3/3 libraries, with 6465 regions identified of 10nt or more.

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genomic distribution of BR3 positions. BR3 regions were intersected with RefSeq genes and 3'UTRs of RefSeq protein coding genes. Only 2/6465 BR3 regions mapped to coordinates overlapping coding and non-coding RNAs annotated in RefSeq.

Examination of sequence features. *Sequence analyses*. Enriched motifs within CLIP regions and UTRs were performed using the EMBOSS tool Compseq. To generate z-scores, shuffled control sets were generated for each dataset analyzed using the EMBOSS tools Shuffleseq (10 shuffled versions of each sequence in each dataset). CIMS analysis was performed as previously described (Zhang and Darnell, 2011).

Distribution of Dazl-RNA contacts in 3'UTRs. Metagene analysis of Dazl-3'UTR interactions was performed on a subset of 3'UTRs defined by PolyA-Seq. 3'UTRs that overlapped with any intron sequence annotated in RefSeq were omitted. Only BR3 sites in genes with a single 3'UTR were analyzed. The resulting 5,284 BR3 HITS-CLIP peaks (adult testis) were examined in 2022 3'UTRs, and 3,321 BR3 iCLIP regions (P6 testis) were examined in 1,604 3'UTRs. To calculate distances, the midpoint of each BR3 site was measured relative to the upstream stop codon and the downstream polyA site. Positions of binding were then examined in 20nt windows relative to the stop codon or polyA site. For each 3'UTR, the expected distribution was determining by counting the number of BR3 sites in the 3'UTR and the number of 20nt bins in the 3'UTR, to determine the likelihood of any 20nt bin in a given 3'UTR containing a BR3 site. The control set of 3771 3'UTRs was identified by intersecting the 28032 3'UTRs from adult testis (identified by PolyA-Seq) with all adult testis CLIP reads, and selecting those with zero CLIP read coverage and no overlap with RefSeq introns.

***GO analyses*.** GO term analyses were performed with the Cytoscape application BiNGO (Saito et al., 2012; Cline et al., 2007) using a hypergeometric statistical test and Benjamini & Hochberg FDR correction (significance level of 0.05) to identify enriched terms after multiple testing correction.. GO Slim settings were used to process the 656 set of genes (RPKM<1 in WT),

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

while GO Full was used for the 1462 and 1584 sets of genes (Dazl-enhanced and repressed, respectively). A set of 13,659 genes with RPKM>1 in either WT or *DAZL* KO samples was used as the background gene set for enrichment. GO analysis of 501 Dazl-bound and enhanced genes was performed using GO Miner (Zeeberg et al., 2003).

Protein-protein-interactions. The STRING database (Szklarczyk et al., 2015) was used to identify PPIs between protein products of the 501 Dazl-enhanced genes with Dazl-3'UTR interactions. Selected parameters were experiments and databases (for interaction sources), and highest confidence (0.900) (for minimum required interaction score).

All deep sequencing datasets associated with this study are available at NCBI Gene Expression Omnibus accession number GSE108997.

Acknowledgements. We are grateful to Marco Conti for providing *DAZL*$^{Tm1hgu/+}$ mice, Jo Ann Wise for comments on the manuscript, and the following CWRU core facilities: Transgenic and Targeting Facility (mouse rederivation); Genomics (deep sequencing); Tissue Resources (embedding); Virology, Next Generation Sequencing, and Imaging Core (IF microscopy); Cytometry and Microscopy (FACS); Applied Functional Genomics (low input RNA-Seq libraries). This work was supported by funds from the National Institutes of Health to LLZ (T32 GM08056) and DDL (R01 GM107331).

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME Suite. Nucleic Acids Res *43*, W39-49.

Bicknell, A. A., and Ricci, E. P. (2017). When mRNA translation meets decay. Biochem Soc Trans *45*, 339-351.

Chen, H. H., Welling, M., Bloch, D. B., Muñoz, J., Mientjes, E., Chen, X., Tramp, C., Wu, J., Yabuuchi, A., Chou, Y. F., Buecker, C., Krainer, A., Willemsen, R., Heck, A. J., and Geijsen, N. (2014). DAZL limits pluripotency, differentiation, and apoptosis in developing primordial germ cells. Stem Cell Reports *3*, 892-904.

Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R., Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., Pico, A. R., Vailaya, A., Wang, P. L., Adler, A., Conklin, B. R., Hood, L., Kuiper, M., Sander, C., Schmulevich, I., Schwikowski, B., Warner, G. J., Ideker, T., and Bader, G. D. (2007). Integration of biological networks and gene expression data using Cytoscape. Nat Protoc *2*, 2366-2382.

Collier, B., Gorgoni, B., Loveridge, C., Cooke, H. J., and Gray, N. K. (2005). The DAZL family proteins are PABP-binding proteins that regulate translation in germ cells. EMBO J *24*, 2656-2666.

Costoya, J. A., Hobbs, R. M., Barna, M., Cattoretti, G., Manova, K., Sukhwani, M., Orwig, K. E., Wolgemuth, D. J., and Pandolfi, P. P. (2004). Essential role of Plzf in maintenance of spermatogonial stem cells. Nat Genet *36*, 653-659.

De Gasperi, R., Rocher, A. B., Sosa, M. A., Wearne, S. L., Perez, G. M., Friedrich, V. L. J., Hof, P. R., and Elder, G. A. (2008). The IRG mouse: a two-color fluorescent reporter for assessing Cre-mediated recombination and imaging complex cellular relationships in situ. Genesis *46*, 308-317.

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Eirew, P., Steif, A., Khattra, J., Ha, G., Yap, D., Farahani, H., Gelmon, K., Chia, S., Mar, C., Wan, A., Laks, E., Biele, J., Shumansky, K., Rosner, J., McPherson, A., Nielsen, C., Roth, A. J., Lefebvre, C., Bashashati, A., de Souza, C., Siu, C., Aniba, R., Brimhall, J., Oloumi, A., Osako, T., Bruna, A., Sandoval, J. L., Algara, T., Greenwood, W., Leung, K., Cheng, H., Xue, H., Wang, Y., Lin, D., Mungall, A. J., Moore, R., Zhao, Y., Lorette, J., Nguyen, L., Huntsman, D., Eaves, C. J., Hansen, C., Marra, M. A., Caldas, C., Shah, S. P., and Aparicio, S. (2015). Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. Nature *518*, 422-426.

Fox, M., Urano, J., and Reijo Pera, R. A. (2005). Identification and characterization of RNA sequences to which human PUMILIO-2 (PUM2) and deleted in Azoospermia-like (DAZL) bind. Genomics *85*, 92-105.

Fu, X. F., Cheng, S. F., Wang, L. Q., Yin, S., De Felici, M., and Shen, W. (2015). DAZ Family Proteins, Key Players for Germ Cell Development. Int J Biol Sci *11*, 1226-1235.

Gill, M. E., Hu, Y. C., Lin, Y., and Page, D. C. (2011). Licensing of gametogenesis, dependent on RNA binding protein DAZL, as a gateway to sexual differentiation of fetal germ cells. Proc Natl Acad Sci U S A *108*, 7443-7448.

Hannigan, M. M., Zagore, L. L., and Licatalosi, D. D. (2017). Ptbp2 Controls an Alternative Splicing Network Required for Cell Communication during Spermatogenesis. Cell Rep *19*, 2598-2612.

Hao, J., Yamamoto, M., Richardson, T. E., Chapman, K. M., Denard, B. S., Hammer, R. E., Zhao, G. Q., and Hamra, F. K. (2008). Sohlh2 knockout mice are male-sterile because of degeneration of differentiating type A spermatogonia. Stem Cells *26*, 1587-1597.

Haston, K. M., Tung, J. Y., and Reijo Pera, R. A. (2009). Dazl functions in maintenance of pluripotency and genetic and epigenetic programs of differentiation in mouse primordial germ cells in vivo and in vitro. PLoS One *4*, e5654.

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hornstein, E., Harel, H., Levy, G., and Meyuhas, O. (1999). Overexpression of poly(A)-binding protein down-regulates the translation or the abundance of its own mRNA. FEBS Lett *457*, 209-213.

Ingolia, N. T. (2010). Genome-wide translational profiling by ribosome footprinting. Methods Enzymol *470*, 119-142.

Jenkins, H. T., Malkova, B., and Edwards, T. A. (2011). Kinked beta-strands mediate high-affinity recognition of mRNA targets by the germ-cell regulator DAZL. Proc Natl Acad Sci U S A *108*, 18266-18271.

Kini, H. K., Silverman, I. M., Ji, X., Gregory, B. D., and Liebhaber, S. A. (2016). Cytoplasmic poly(A) binding protein-1 binds to genomically encoded sequences within mammalian mRNAs. RNA *22*, 61-74.

Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D. J., Luscombe, N. M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat Struct Mol Biol *17*, 909-915.

Li, W., Park, J. Y., Zheng, D., Hoque, M., Yehia, G., and Tian, B. (2016). Alternative cleavage and polyadenylation in spermatogenesis connects chromatin regulation with post-transcriptional control. BMC Biol *14*, 6.

Licatalosi, D. D. (2016). Roles of RNA-binding Proteins and Post-transcriptional Regulation in Driving Male Germ Cell Development in the Mouse. Adv Exp Med Biol *907*, 123-151.

Licatalosi, D. D., Mele, A., Fak, J. J., Ule, J., Kayikci, M., Chi, S. W., Clark, T. A., Schweitzer, A. C., Blume, J. E., Wang, X., Darnell, J. C., and Darnell, R. B. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature *456*, 464-469.

Licatalosi, D. D., Yano, M., Fak, J. J., Mele, A., Grabinski, S. E., Zhang, C., and Darnell, R. B. (2012). Ptbp2 represses adult-specific splicing to regulate the generation of neuronal precursors in the embryonic brain. Genes Dev *26*, 1626-1642.

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Lin, Y., and Page, D. C. (2005). Dazl deficiency leads to embryonic arrest of germ cell development in XY C57BL/6 mice. Dev Biol *288*, 309-316.

Maegawa, S., Yamashita, M., Yasuda, K., and Inoue, K. (2002). Zebrafish DAZ-like protein controls translation via the sequence 'GUUC'. Genes Cells *7*, 971-984.

Manku, G., and Culty, M. (2015). Mammalian gonocyte and spermatogonia differentiation: recent advances and remaining challenges. Reproduction *149*, R139-57.

Martin, G., Gruber, A. R., Keller, W., and Zavolan, M. (2012). Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. Cell Rep *1*, 753-763.

McLaren, A. (2001). Mammalian germ cells: birth, sex, and immortality. Cell Struct Funct *26*, 119-122.

Moore, F. L., Jaruzelska, J., Fox, M. S., Urano, J., Firpo, M. T., Turek, P. J., Dorfman, D. M., and Pera, R. A. (2003). Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (Deleted in AZoospermia) and DAZ-like proteins. Proc Natl Acad Sci U S A *100*, 538-543.

Pan, J., Eckardt, S., Leu, N. A., Buffone, M. G., Zhou, J., Gerton, G. L., McLaughlin, K. J., and Wang, P. J. (2009). Inactivation of Nxf2 causes defects in male meiosis and age-dependent depletion of spermatogonia. Dev Biol *330*, 167-174.

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Raverot, G., Weiss, J., Park, S. Y., Hurley, L., and Jameson, J. L. (2005). Sox3 expression in undifferentiated spermatogonia is required for the progression of spermatogenesis. Dev Biol *283*, 215-225.

Reijo, R., Alagappan, R. K., Patrizio, P., and Page, D. C. (1996). Severe oligozoospermia resulting from deletions of azoospermia factor gene on Y chromosome. Lancet *347*, 1290-1293.

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Reijo, R., Lee, T. Y., Salo, P., Alagappan, R., Brown, L. G., Rosenberg, M., Rozen, S., Jaffe, T., Straus, D., Hovatta, O., and et, A. (1995). Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. Nat Genet *10*, 383-393.

Reynolds, N., Collier, B., Bingham, V., Gray, N. K., and Cooke, H. J. (2007). Translation of the synaptonemal complex component Sycp3 is enhanced in vivo by the germ cell specific regulator Dazl. RNA *13*, 974-981.

Reynolds, N., Collier, B., Maratou, K., Bingham, V., Speed, R. M., Taggart, M., Semple, C. A., Gray, N. K., and Cooke, H. J. (2005). Dazl binds in vivo to specific transcripts and can regulate the pre-meiotic translation of Mvh in germ cells. Hum Mol Genet *14*, 3899-3909.

Ruark, E., Seal, S., McDonald, H., Zhang, F., Elliot, A., Lau, K., Perdeaux, E., Rapley, E., Eeles, R., Peto, J., Kote-Jarai, Z., Muir, K., Nsengimana, J., Shipley, J., Bishop, D. T., Stratton, M. R., Easton, D. F., Huddart, R. A., Rahman, N., and Turnbull, C. (2013). Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14. Nat Genet *45*, 686-689.

Ruggiu, M., and Cooke, H. J. (2000). In vivo and in vitro analysis of homodimerisation activity of the mouse Dazl1 protein. Gene *252*, 119-126.

Ruggiu, M., Speed, R., Taggart, M., McKay, S. J., Kilanowski, F., Saunders, P., Dorin, J., and Cooke, H. J. (1997). The mouse Dazla gene encodes a cytoplasmic protein essential for gametogenesis. Nature *389*, 73-77.

Sadate-Ngatchou, P. I., Payne, C. J., Dearth, A. T., and Braun, R. E. (2008). Cre recombinase activity specific to postnatal, premeiotic male germ cells in transgenic mice. Genesis *46*, 738-742.

Saga, Y. (2010). Function of Nanos2 in the male germ cell lineage in mice. Cell Mol Life Sci *67*, 3815-3822.

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Saito, R., Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., Lotia, S., Pico, A. R., Bader, G. D., and Ideker, T. (2012). A travel guide to Cytoscape plugins. Nat Methods *9*, 1069-1076.

Saunders, P. T., Turner, J. M., Ruggiu, M., Taggart, M., Burgoyne, P. S., Elliott, D., and Cooke, H. J. (2003). Absence of mDazl produces a final block on germ cell development at meiosis. Reproduction *126*, 589-597.

Schrans-Stassen, B. H., Saunders, P. T., Cooke, H. J., and de Rooij, D. G. (2001). Nature of the spermatogenic arrest in Dazl -/- mice. Biol Reprod *65*, 771-776.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337-342.

Scotti, M. M., and Swanson, M. S. (2016). RNA mis-splicing in disease. Nat Rev Genet *17*, 19-32.

Song, H. W., and Wilkinson, M. F. (2014). Transcriptional control of spermatogonial maintenance and differentiation. Semin Cell Dev Biol

Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthes, P., Kokkinaki, M., Nef, S., Gnirke, A., Dym, M., de Massy, B., Mikkelsen, T. S., and Kaessmann, H. (2013). Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. Cell Rep *3*, 2179-2190.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., and von Mering, C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res *43*, D447-52.

Takeda, Y., Mishima, Y., Fujiwara, T., Sakamoto, H., and Inoue, K. (2009). DAZL relieves miRNA-mediated repression of germline mRNAs by controlling poly(A) tail length in zebrafish. PLoS One *4*, e7513.

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Takubo, K., Ohmura, M., Azuma, M., Nagamatsu, G., Yamada, W., Arai, F., Hirao, A., and Suda, T. (2008). Stem cell defects in ATM-deficient undifferentiated spermatogonia through DNA damage-induced cell-cycle arrest. Cell Stem Cell *2*, 170-182.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562-578.

Tsui, S., Dai, T., Warren, S. T., Salido, E. C., and Yen, P. H. (2000). Association of the mouse infertility factor DAZL1 with actively translating polyribosomes. Biol Reprod *62*, 1655-1660.

VanGompel, M. J. W., and Xu, E. Y. (2011). The roles of the DAZ family in spermatogenesis: More than just translation? Spermatogenesis *1*, 36-46.

Vogel, T., Speed, R. M., Ross, A., and Cooke, H. J. (2002). Partial rescue of the Dazl knockout mouse by the human DAZL gene. Mol Hum Reprod *8*, 797-804.

Wu, J., Anczukow, O., Krainer, A. R., Zhang, M. Q., and Zhang, C. (2013). OLego: fast and sensitive mapping of spliced mRNA-Seq reads using small seeds. Nucleic Acids Res *41*, 5149-5163.

Xu, X., Tan, X., Lin, Q., Schmidt, B., Engel, W., and Pantakani, D. V. (2013). Mouse Dazl and its novel splice variant functions in translational repression of target mRNAs in embryonic stem cells. Biochim Biophys Acta *1829*, 425-435.

Yamaji, M., Jishage, M., Meyer, C., Suryawanshi, H., Der, E., Yamaji, M., Garzia, A., Morozov, P., Manickavel, S., McFarland, H. L., Roeder, R. G., Hafner, M., and Tuschl, T. (2017). DND1 maintains germline stem cells via recruitment of the CCR4-NOT complex to target mRNAs. Nature *543*, 568-572.

Zagore, L. L., Grabinski, S. E., Sweet, T. J., Hannigan, M. M., Sramkoski, R. M., Li, Q., and Licatalosi, D. D. (2015). RNA Binding Protein Ptbp2 Is Essential for Male Germ Cell Development. Mol Cell Biol *35*, 4030-4042.

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Zeeberg, B. R., Feng, W., Wang, G., Wang, M. D., Fojo, A. T., Sunshine, M., Narasimhan, S., Kane, D. W., Reinhold, W. C., Lababidi, S., Bussey, K. J., Riss, J., Barrett, J. C., and Weinstein, J. N. (2003). GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol *4*, R28.

Zhang, C., and Darnell, R. B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat Biotechnol *29*, 607-614.

Zhang, T., Oatley, J., Bardwell, V. J., and Zarkower, D. (2016). DMRT1 Is Required for Mouse Spermatogonial Stem Cell Maintenance and Replenishment. PLoS Genet *12*, e1006293.

Figure Legends

Figure 1. HITS-CLIP identification of Dazl-RNA contacts in the adult testis. **A.** Autoradiograph of nitrocellulose membrane containing radio-labeled cross-linked Dazl-RNA complexes purified from adult testes. Arrowhead indicates Dazl cross-linked to minimal RNA fragments. Bracket indicates region of nitrocellulose excised for cDNA preparation, which corresponds to Dazl cross-linked to RNA fragments of 35-50 nt **B.** Venn diagram showing the overlap between CLIP reads from each biologic replicate, and the identification of 11,297 genomic coordinates containing overlapping clusters of CLIP reads from 3/3 biologic replicates. **C.** Number of BR3 clusters that contain different numbers of Dazl CLIP peaks. **D.** Bins of CLIP peaks normalized to RNA-Seq, with GUU enrichment (green) and proportion of peaks in each bin that contain GUU (grey), with Chi square distribution values for GUU enrichment (χ^2) at right. **E.** Top motifs present in Bin 1, identified by MEME (top, middle) and DREME (bottom). **F.** Top motifs associated with CIMS sites identified by MEME. **G.** Top ten 6-mers and 4-mers associated with CIMS sites. **H.** Positioning of UGUUGU, GUUG, and UGUU motifs relative to CIMS deletion sites. **I.** Distances between the midpoints of adjacent peaks in multi-peak clusters. **J.** Motif enrichment comparison (z-scores) of 5'- or 3'-positioned peaks in BR3 clusters with two peaks (n=2770).

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2. Dazl predominantly binds 3'UTRs at polyA-proximal sites. **A.** Distribution of BR3 CLIP sites in different genic regions, with the majority of interactions in mRNA 3'UTRs. BR3 clusters that mapped to regions with overlapping coding and non-coding RNAs were annotated as nc/mRNA. **B.** Number of genes with one or more polyA site, as determined by PolyA-seq analysis. **C.** Metagene analysis of the number of Dazl CLIP peaks in 20nt bins relative to the stop codon (left panel, red line) or the polyA site (right panel, blue line). Dashed line represents the expected random distribution. **D.** Top panel corresponds to a higher magnification view of the right panel in C. Bottom three panels show the distribution of CLIP peaks relative to the polyA site in 3'UTRs with one, two, and more than two peaks in the entire 3'UTR. **E.** Nucleotide frequency in the 500nt upstream of the polyA sites of control 3'UTRs (no Dazl CLIP reads, top panel), and 3'UTRs with Dazl-RNA interactions (bottom panel). **F.** Motif enrichment analysis of 3'UTRs with and without Dazl CLIP peaks (y and x axes, respectively), where values correspond to z-scores. **G.** GUU enrichment (observed/expected) in 3'UTRs of genes with and without Dazl CLIP peaks (green and grey lines, respectively).

Figure 3. Local polyA-tracts contribute to Dazl-RNA interactions outside of 3'UTRs. **A.** Comparison of BR3 CLIP tag density in the 5' and 3'UTRs of genes with 5'UTR BR3 CLIP sites. Color-coding represents the percentage of BR3 tags in the gene that map to either the 5' or 3'UTR. Dashed line denotes cutoff between genes with less than (above) or greater than (below) 50% of the CLIP reads in the 5'UTR. **B-F.** Representative examples of genes listed in A, with RNA-Seq read density and CLIP read density in BR3 regions shown. **G.** Motif enrichment analysis of 5'UTRs with more or less than 50% of total BR3 CLIP reads in the 5'UTR.

Figure 4. Genetic-labeling, isolation, and transcriptome-profiling of GFP+ WT and *DAZL* KO germ cells. **A.** Immunofluorescence analysis of Ddx4 and GFP expression in *Stra8-iCre*$^{+}$; *IRG*$^{+}$; *DAZL*$^{++}$ (*DAZL* WT) and *Stra8-iCre*$^{+}$; *IRG*$^{+}$; *DAZL*$^{Tm1hgu/Tm1hgu}$ (*DAZL* KO) testes at P6 (left and right panels, respectively). Sections from biological replicates are shown in the top and bottom

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

rows. Scale bars represent 25µM. **B**. Flow cytometry plots of GFP- and GFP+ cells from WT and KO mice. **C**. Percentage of GFP+ cells isolated by FACS from P6 WT and *DAZL* KO mice. **D.** Scatter plot of RPKM values for genes in WT and KO cells. Green and red dots denote genes with decreased and increased RNA levels in Dazl KO cells compared to WT controls. **E.** Box plot showing the distribution of RPKM values for green, red, and grey genes indicated in panel D.

Figure 5. Dazl directly enhances mRNA levels through polyA-proximal binding **A.** Distribution of 6,465 BR3 Dazl-RNA sites from iCLIP analysis on P6 testes. **B.** Top 10 motifs (based on z-score ranking) for sequences of BR3 regions (left) or 10 nucleotide windows centered at the 5'end of each BR3 CLIP region. **C.** Metagene analysis of the distribution of BR3 Dazl-RNA sites in 3'UTRs of spermatogonial-expressed 3'UTRs (as in Figure 2C). **D.** Left: RPKM levels of 2290 genes with Dazl-3'UTR interactions in GFP+ WT and GFP+ KO cells. Green, and red circles correspond to genes with RNA level differences between GFP+ WT and KO cells of 2-fold or more ($P<0.01$). Blue dots correspond to genes with RPKM values that do not differ more than 20% between WT and KO cells. Right: metagene analysis showing the distribution of BR3 Dazl-RNA sites in 3'UTRs of Dazl-enhanced, Dazl-repressed, and Dazl-insensitive genes (green, red, and blue, respectively). **E.** Venn diagram comparing genes with RNA level changes between WT and *DAZL* KO cells (2-fold or more, $P<0.01$) and the set of 2290 with Dazl-3'UTR interactions. **F.** Examination of 3'UTR lengths (left) and frequency of GUU per 3'UTR (right) for genes with increased RNA, unchanged RNA, or decreased RNA leves (green, blue, and red, respectively) in GFP+ Dazl KO cells compared to WT controls. P values (Wilcoxon rank sum test) for pairwise comparisons indicated above.

Figure 6. Dazl binds and regulates a broad network of essential genes. **A.** Heirarchical clustering of enriched GO terms and genes associated with the set of 501 genes that are enhanced by Dazl and have Dazl-3'UTR interactions. Subgroups of overlapping GO categories

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

are indicated at left, while subgroups of genes are indicated with letters at bottom. **B.** Lists of all genes in each letter group indicated at bottom of panel A.

Figure 7. A disproportionate number of Dazl target genes encode networks of interacting proteins. **A.** Protein-protein interaction network for genes with Dazl-3'UTR interactions and reduced RNA levels in Dazl KO cells compared to WT controls.

Supplemental Figures

Supplemental Figure 1. BR3 CLIP clusters with different numbers of peaks. BR3 clusters containing one (A), two (B), or three (C, D) peaks are shown. The footprint of the BR3 region is shown at the top. Reads are color coded to reflect each of the three CLIP libraries sequenced. Density indicates the number of overlapping reads per nucleotide within the BR3 region. Peak baseline corresponds to the total number of CLIP bases in the cluster divided by the length of the cluster. Peaks correspond to regions where the CLIP density exceeds the peak baseline.

Supplemental Figure 2. PolyA-seq identification of polyA sites of testis mRNAs. **A.** Overview of the PolyA-Seq library construction strategy. **B.** Characterization of polyA site usage in different genes. **C.** Comparison of RPKM and TPM values derived from RNA-Seq and PolyA-Seq, respectively, from adult testes. **D.** Comparison of proximal polyA site estimates from qRT-PCR and PolyA-Seq.

Supplemental Figure 3. A. Nucleotide frequency in the last 500 nucleotides of 3'UTRs with Dazl BR3 sites (n=2022) and control 3'UTRs that have no CLIP tags (n=3771). B. Enrichment of 6-mers present in the 3'UTRs of genes with Dazl-3'UTRs compared to control genes. Each dot represents an individual 6 mer. X-axis corresponds to difference in observed-to-expected ratio for bound versus control 3'UTRs, while y-axis corresponds to the enrichment z-score for the 6-mer in the Dazl-bound 3'UTRs. Shown at right are the top 20 6-mers ranked by difference in the observed-to-expected ratios between bound and control 3'UTRs.

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplemental Figure 4. A. Enriched GO terms associated with the set of 656 genes with little to no expression in WT cells (RPKM<1) and upregulated in *DAZL* KO cells (*P*<0.01). Results from a GO Slim analysis performed in BinGO are shown, where circle size denotes number of genes in enriched category and circle color reflects P value. **B.** Top 50 enriched GO terms associated with all genes with decreased RNA levels in KO cells (ranked by P value). **C.** Top 50 enriched GO terms associated with all genes with increased RNA levels in KO cells (ranked by P value).

Supplemental Figure 5. Heirarchical clustering of enriched GO terms (*P*<0.01) and genes associated with the set of 1462 genes that have decreased RNA levels in *DAZL* KO cells.

Supplemental Figure 6. Genome browser images showing P6 BR3 CLIP sites in 3'UTRs of Dazl-enhanced and Dazl-insensitive genes (**A** and **B**, respectively). Adult PolyA-seq and RNA-Seq tracks are also shown. CLIP scale denotes the maximum number of overlapping CLIP tags in the region shown. **C**. Average RPKM values for the genes shown in panels A and B.

Supplemental Figure 7. Model for *DAZL*-dependent germ cell maintenance. **A.** Seminiferous tubules from WT and *DAZL* KO animals with different numbers of germ cells present. **B.** Dazl binds a broad set of mRNAs via polyA-proximal interactions facilitated by Pabpc1-polyA interactions. In the absence of Dazl, some genes (*ORF1* and *ORF2*) have reduced mRNA levels, while other genes with Dazl-3'UTR interactions (*ORF3*) are insensitive. *ORF4* represents a *DAZL*-insensitive gene without Dazl-3'UTR binding. *ORF1* mRNAs have multiple Dazl-GUU contacts, whereas *ORF2* mRNAs have a single polyA-proximal Dazl-RNA interaction. Figure legend displayed at bottom.

Supplemental Tables

Supplemental Table 1. Numbers of raw and filtered HITS-CLIP reads at each step of the annotation workflow.

bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplemental Table 2. Annotation of 11,297 BR3 clusters in RefSeq fragments. Clusters that overlap two different fragments in the same transcript (such as coding and UTR sequence) are denoted with "-". Clusters that map to positions with alternative annotations (such as intron in one isoform and exon in another) are denoted with "/".

Supplemental Table 3. Top 10 most-enriched 4-mers (top) and 5-mers (bottom) in 5'UTRs of 5'low and 5'high genes. At left, are 10 top motifs ranked by z-score in 5'low genes. At right, are 10 top motifs ranked by z-score in 5'high genes.

Supplemental Table 4. Primers used in this study

Supplemental Files

Supplemental File 1. Adult testis HITS-CLIP BR3 clusters. Sheet 1 ('11297 BR3 clusters') contains genomic coordinates of 11,297 BR3 HITS-CLIP clusters (mm9) with CLIP reads per library, cluster length, and genomic annotation. **Sheet 2** ('13106 binned BR3 peaks') contains coordinates and values for normalized and binned peaks, as in Figure 1D, E. **Sheet 3** ('48 genes 5'UTR BR3 clusters') contains list of 48 genes with 5'UTR BR3 clusters, sorted by the percentage of total CLIP reads that map to the 5'UTR.

Supplemental File 2. Adult testis PolyA-Seq and 3'UTR metagene data. Sheet 1 ('28032 Adult Poly-Seq 3'UTRs') contains genomic coordinates of 3'UTRs, corresponding gene, frequency of polyA site usage, and number of 3'UTRs per gene. **Sheet 2** ('3907 genes Dazl-3'UTR') contains list of genes with Dazl-RNA interactions in 3'UTRs defined by PolyA-Seq of adult testis. **Sheet 3** ('Metagene') contains 5284 CLIP peaks in 2022 3'UTRs used for metagene analysis, with coordinates of all features and distance of each peak midpoint to 3'UTR beginning and end.

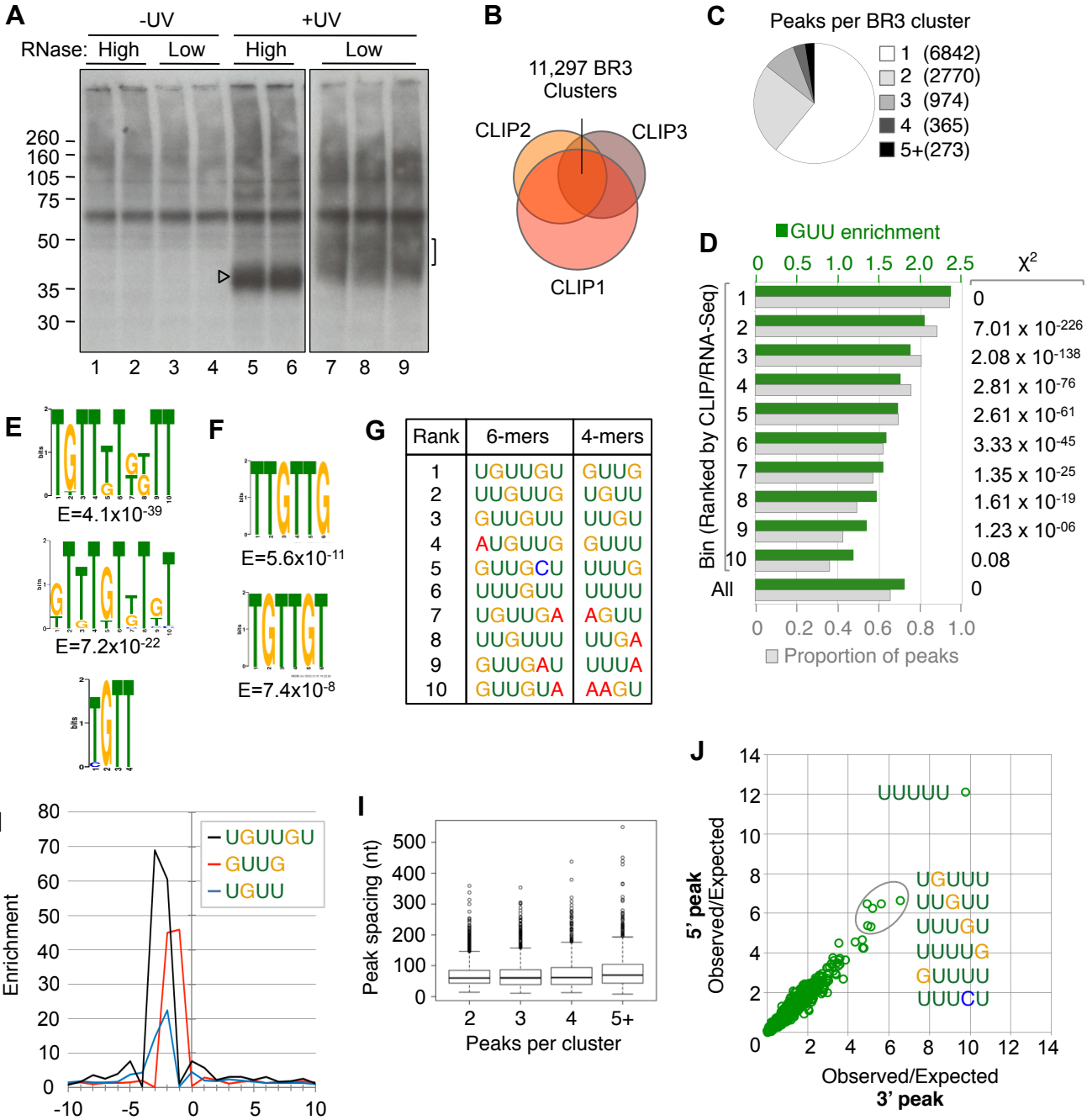
bioRxiv preprint doi: https://doi.org/10.1101/273292; this version posted February 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplemental File 3. Spermatogonia PolyA-Seq and 3'UTR metagene data. Sheet 1 ('16502 SG Poly-Seq 3'UTRs') contains genomic coordinates of 3'UTRs expressed in spermatogonia. **Sheet 2** ('2290 genes Dazl-3'UTR') contains list of genes with Dazl-RNA interactions in 3'UTRs defined by PolyA-Seq of spermatogonia. **Sheet 3** ('Metagene') contains 3321 P6 BR3 sites in 1604 3'UTRs used for metagene analysis, with coordinates of all features and distance of each BR3 midpoint to 3'UTR beginning and end.

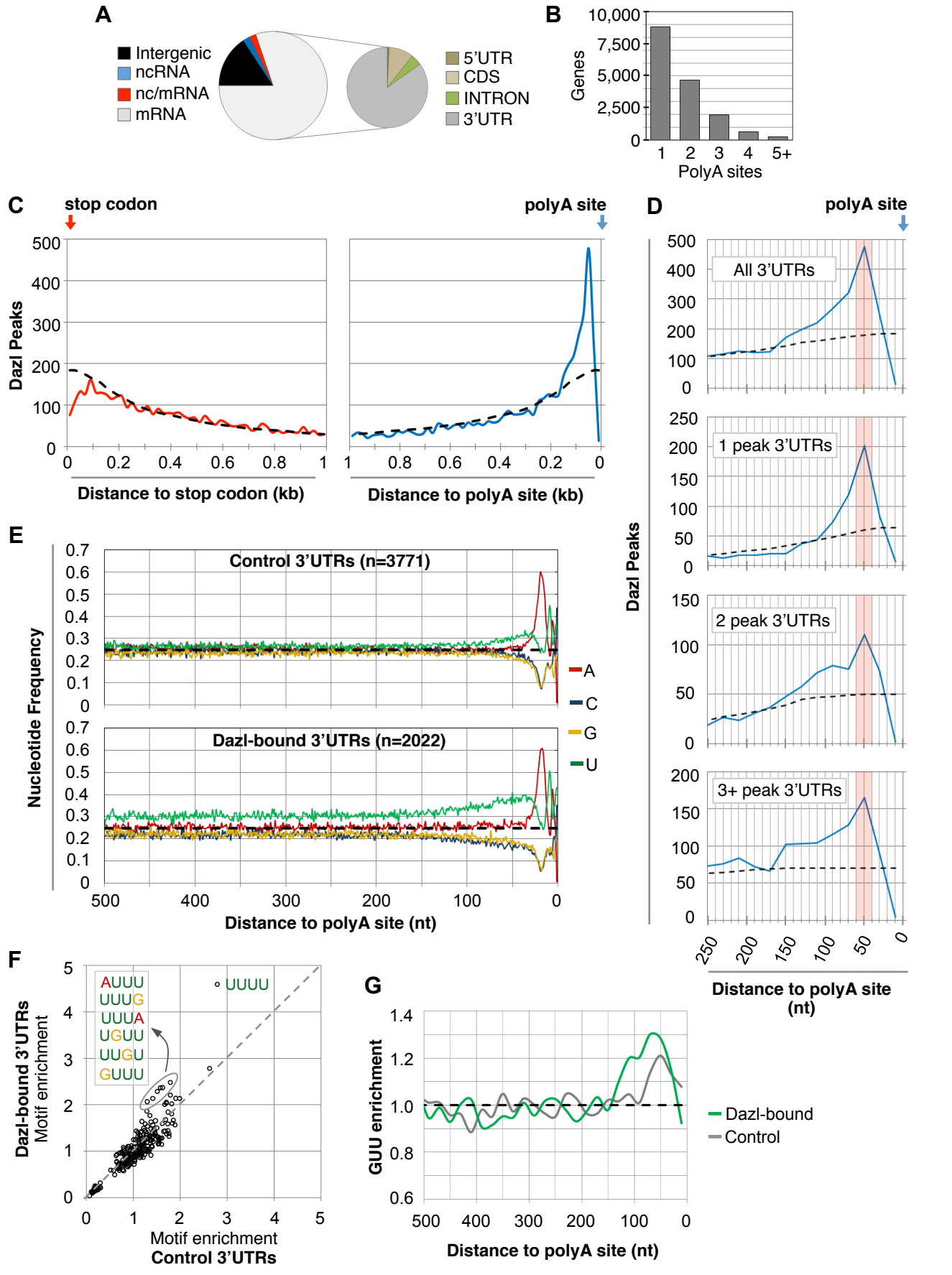
Supplemental File 4. RNA-seq data for GFP+ cells. Sheet 1 ('Gene expression') contains RPKM values for all genes. **Sheet 2** ('1462 genes decreased in KO'). **Sheet 3** ('1584 genes increased in KO'). **Sheet 4** ('656genes') contains genes with average RPKM<1 in WT GFP+ cells. **Sheets 5** ('307enhanced metagene'), **6** ('236insensitive metagene'), and **7** ('54repressed metagene') contain BR3, 3'UTR, and distances for metagene analyses performed using P6 CLIP and spermatogonia PolyA-Seq data on DAZL-enhanced, insensitive, and repressed genes, respectively.

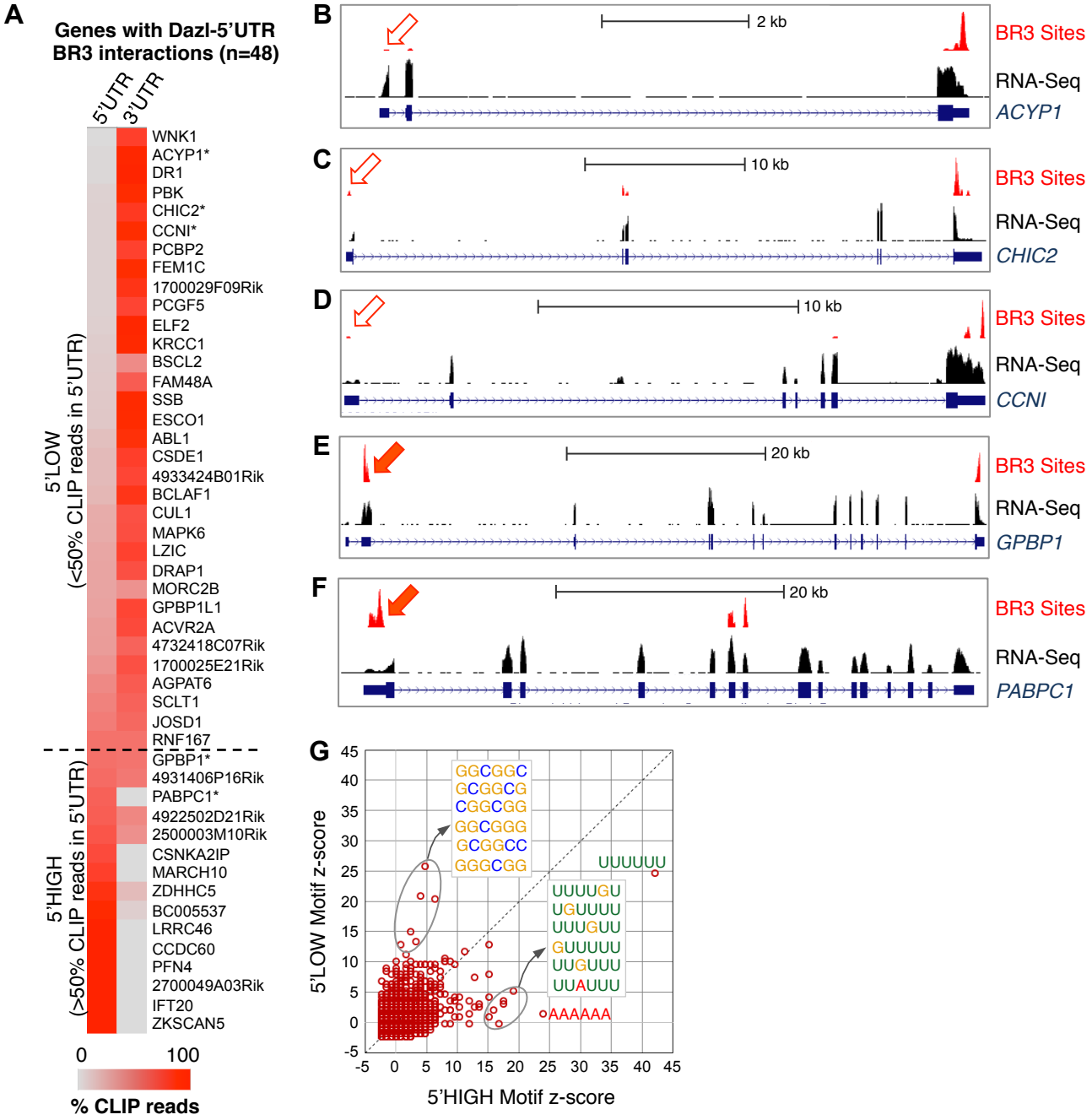
Supplemental File 5. P6 iCLIP data. Sheet 1 (P6 BR3) contains coordinates of P6 BR3 clusters, with number of CLIP reads per library and genomic annotation.

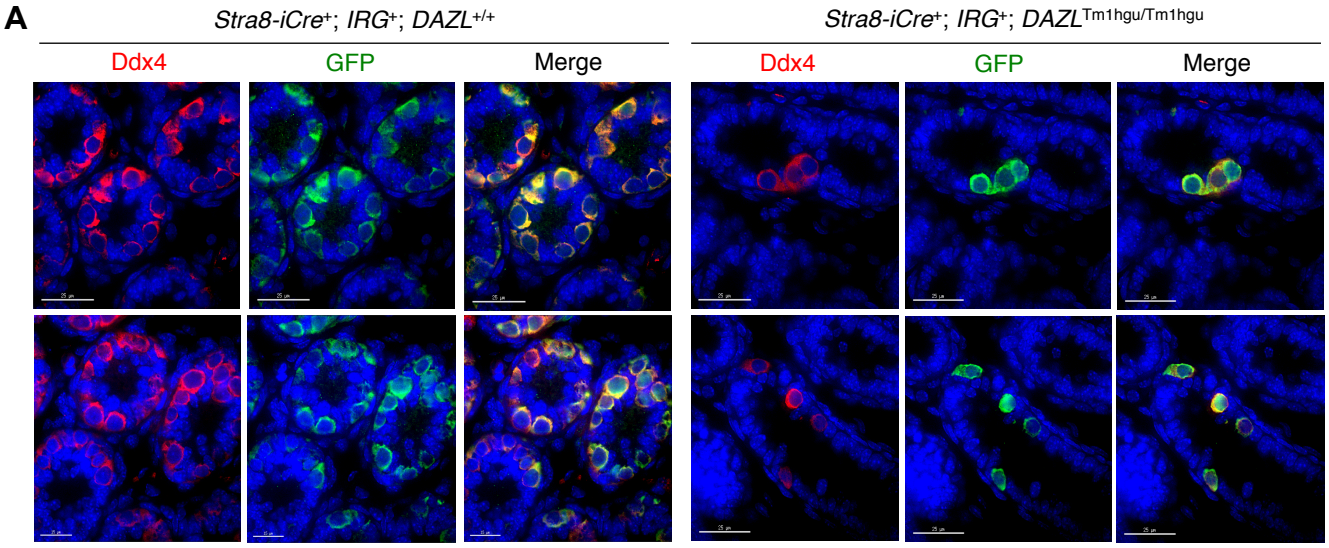
Supplemental File 6. GO enrichment results.

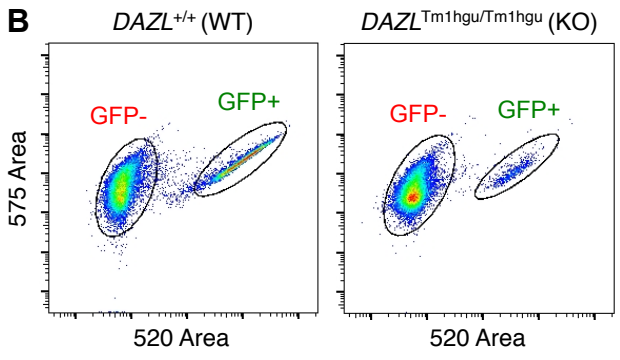

A
-UV
+UV
RNase: High Low High Low
260
160
105
75
50
35
30
1 2 3 4 5 6 7 8 9
B
11,297 BR3 Clusters
CLIP2
CLIP3
CLIP1
C
Peaks per BR3 cluster
1 (6842)
2 (2770)
3 (974)
4 (365)
5+(273)
D
GUU enrichment
0 0.5 1.0 1.5 2.0 2.5
X2
Bin (Ranked by CLIP/RNA-Seq)
1 0
2 7.01 x 10-226
3 2.08 x 10-138
4 2.81 x 10-76
5 2.61 x 10-61
6 3.33 x 10-45
7 1.35 x 10-25
8 1.61 x 10-19
9 1.23 x 10-06
10 0.08
All 0
0 0.2 0.4 0.6 0.8 1.0
Proportion of peaks
E
E=4.1x10-39
E=7.2x10-22
F
E=5.6x10-11
E=7.4x10-8
G
Rank | 6-mers | 4-mers
1 UGUUGU GUUG
2 UUGUUG UGUU
3 GUUGUU UUGU
4 AUGUUG GUUU
5 GUUGCU UUUG
6 UUUGUU UUUU
7 UGUUGA AGUU
8 UUGUUU UUGA
9 GUUGAU UUUA
10 GUUGUA AAGU
H
Enrichment
UGUUGU
GUUG
UGUU
-10 -5 0 5 10
Distance to CIMS deletion
I
Peak spacing (nt)
2 3 4 5+
Peaks per cluster
J
5' peak Observed/Expected
UUUUU
UGUUU
UUGUU
UUUGU
UUUUG
GUUUU
UUUCU
Observed/Expected 3' peak

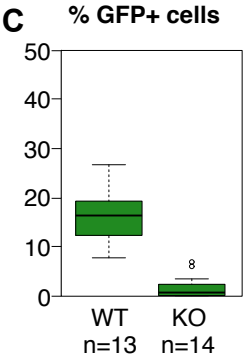
Zagore_Fig2

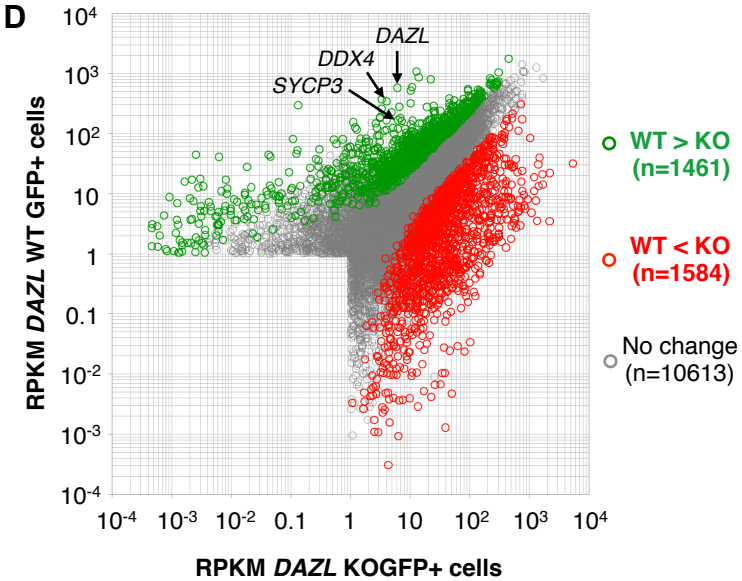

A
Intergenic
ncRNA
nc/mRNA
mRNA
5'UTR
CDS
INTRON
3'UTR
B
Genes
10,000
7,500
5,000
2,500
0
1
2
3
4
5+
PolyA sites
C
stop codon
polyA site
Dazl Peaks
Distance to stop codon (kb)
Distance to polyA site (kb)
D
polyA site
All 3'UTRs
1 peak 3'UTRs
2 peak 3'UTRs
3+ peak 3'UTRs
Dazl Peaks
Distance to polyA site (nt)
E
Control 3'UTRs (n=3771)
Dazl-bound 3'UTRs (n=2022)
Nucleotide Frequency
A
C
G
U
Distance to polyA site (nt)
F
Dazl-bound 3'UTRs
Motif enrichment
AUUU
UUUG
UUUA
UGUU
UUGU
GUUU
UUUU
Motif enrichment
Control 3'UTRs
G
GUU enrichment
Dazl-bound
Control
Distance to polyA site (nt)

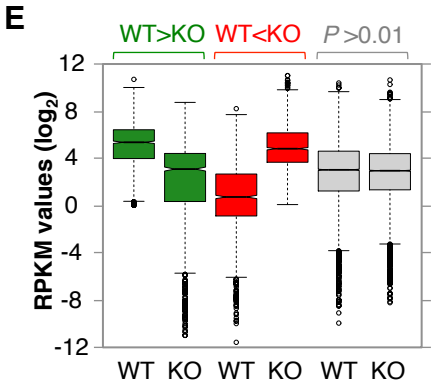

A
Genes with Dazl-5'UTR BR3 interactions (n=48)
5'UTR
3'UTR
5'LOW (<50% CLIP reads in 5'UTR)
5'HIGH (>50% CLIP reads in 5'UTR)
WNK1
ACYP1*
DR1
PBK
CHIC2*
CCNI*
PCBP2
FEM1C
1700029F09Rik
PCGF5
ELF2
KRCC1
BSCL2
FAM48A
SSB
ESCO1
ABL1
CSDE1
4933424B01Rik
BCLAF1
CUL1
MAPK6
LZIC
DRAP1
MORC2B
GPBP1L1
ACVR2A
4732418C07Rik
1700025E21Rik
AGPAT6
SCLT1
JOSD1
RNF167
GPBP1*
4931406P16Rik
PABPC1*
4922502D21Rik
2500003M10Rik
CSNKA2IP
MARCH10
ZDHHC5
BC005537
LRRC46
CCDC60
PFN4
2700049A03Rik
IFT20
ZKSCAN5
0
100
% CLIP reads
B
2 kb
BR3 Sites
RNA-Seq
ACYP1
C
10 kb
BR3 Sites
RNA-Seq
CHIC2
D
10 kb
BR3 Sites
RNA-Seq
CCNI
E
20 kb
BR3 Sites
RNA-Seq
GPBP1
F
20 kb
BR3 Sites
RNA-Seq
PABPC1
G
5'LOW Motif z-score
5'HIGH Motif z-score
GGCGGC
GCGGCG
CGGCGG
GGCGGG
GCGGCC
GGGCGG
UUUUUU
UUUUGU
UGUUUU
UUUGUU
GUUUUU
UUGUUU
UUAUUU
AAAAAA

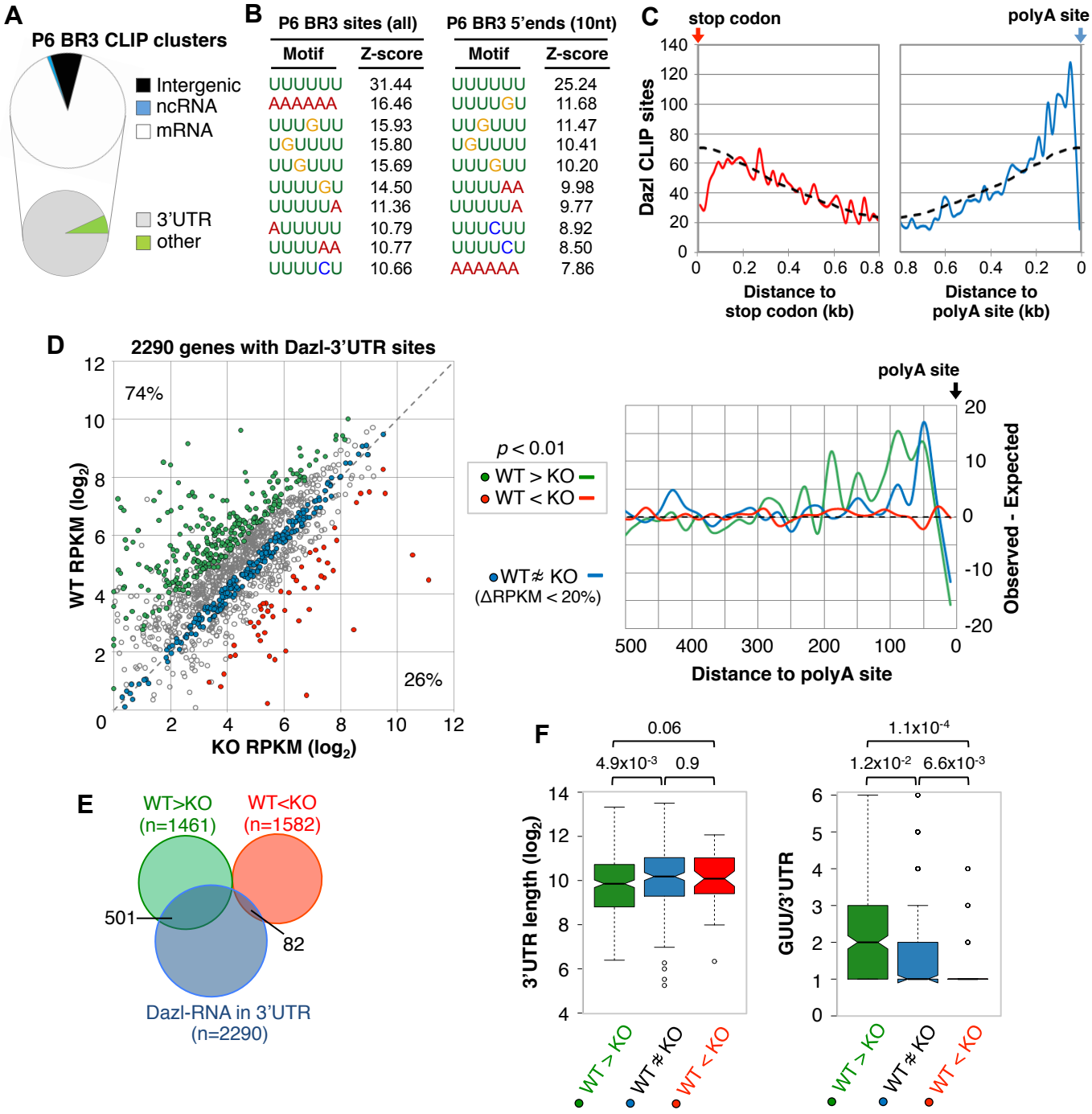

A
Stra8-iCre+; IRG+; DAZL+/+
Stra8-iCre+; IRG+; DAZLTm1hgu/Tm1hgu
Ddx4
GFP
Merge
Ddx4
GFP
Merge

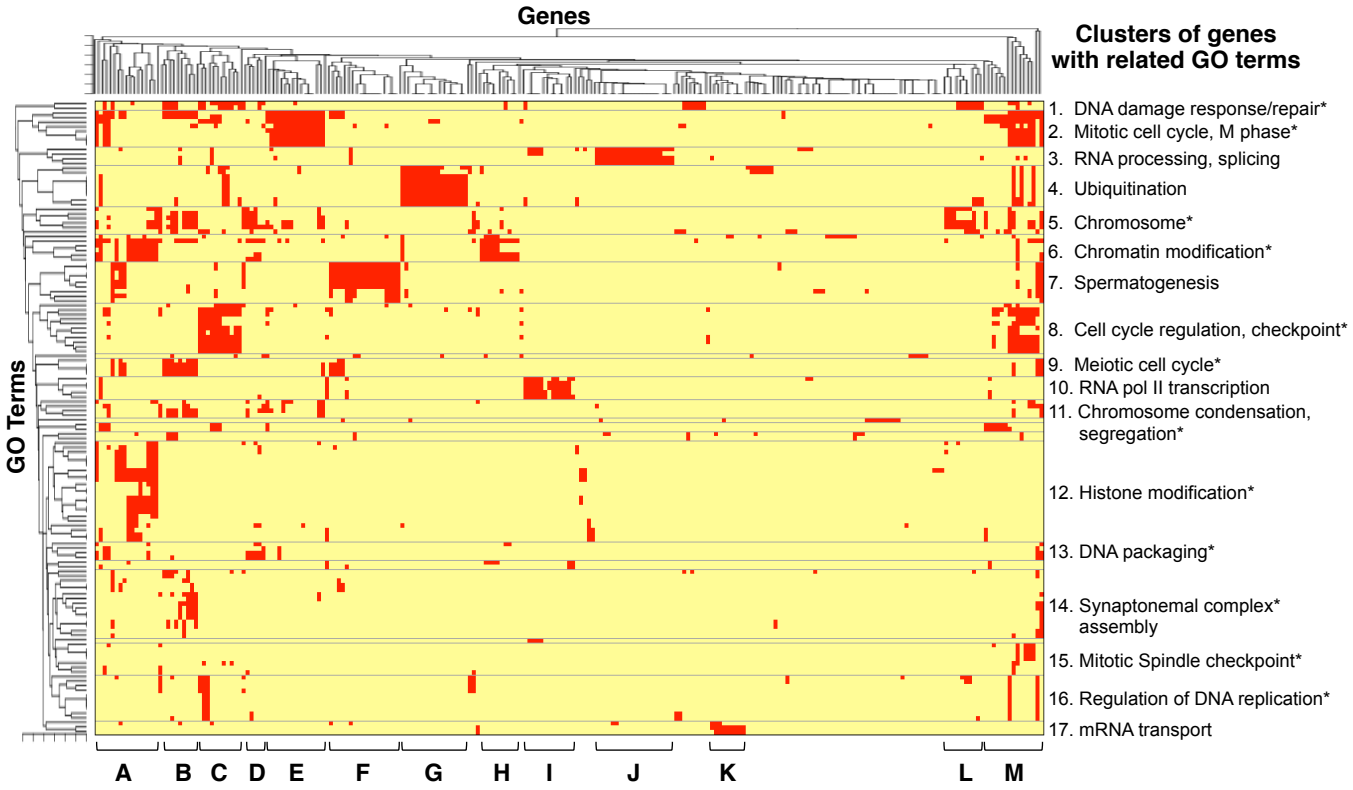

B
DAZL+/+ (WT)
DAZLTm1hgu/Tm1hgu (KO)
GFP-
GFP+
GFP-
GFP+
575 Area
520 Area
520 Area


C
% GFP+ cells
50
40
30
20
10
0
WT
n=13
KO
n=14

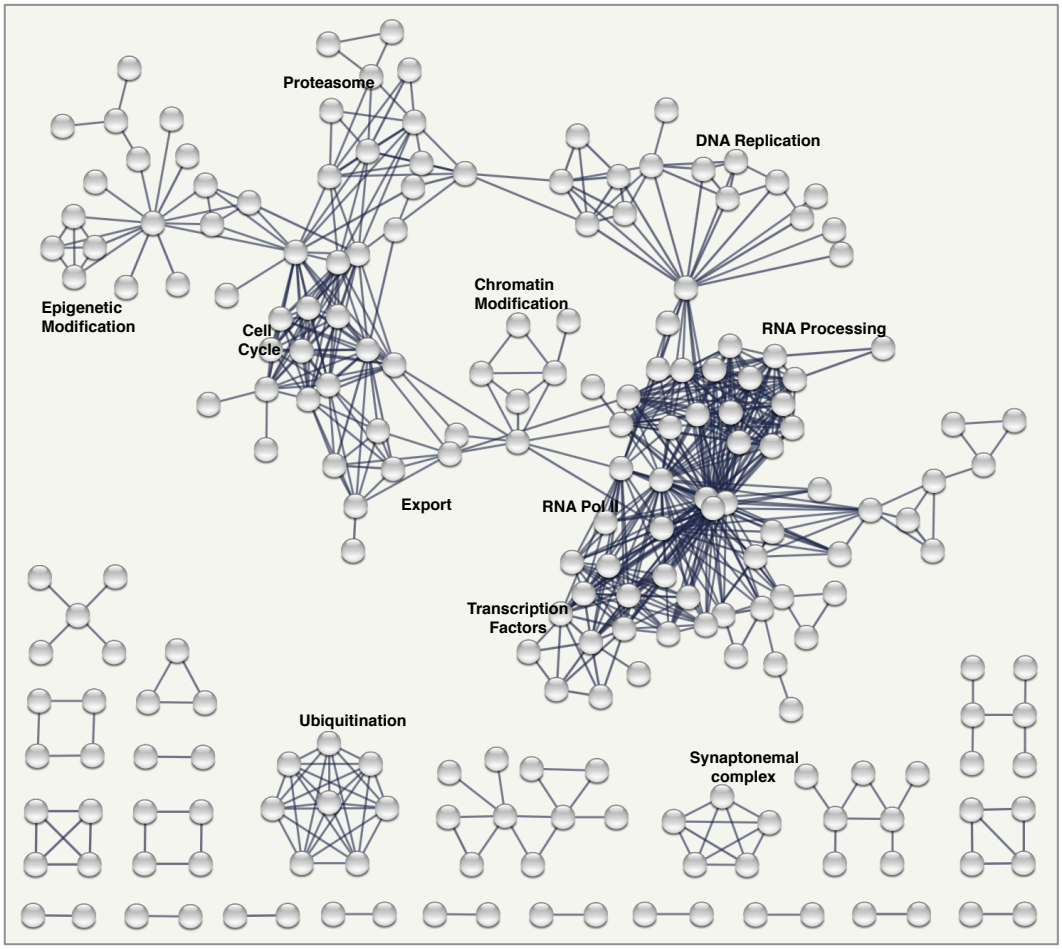

D
DAZL
DDX4
SYCP3
WT > KO
(n=1461)
WT < KO
(n=1584)
No change
(n=10613)
RPKM DAZL WT GFP+ cells
RPKM DAZL KOGFP+ cells

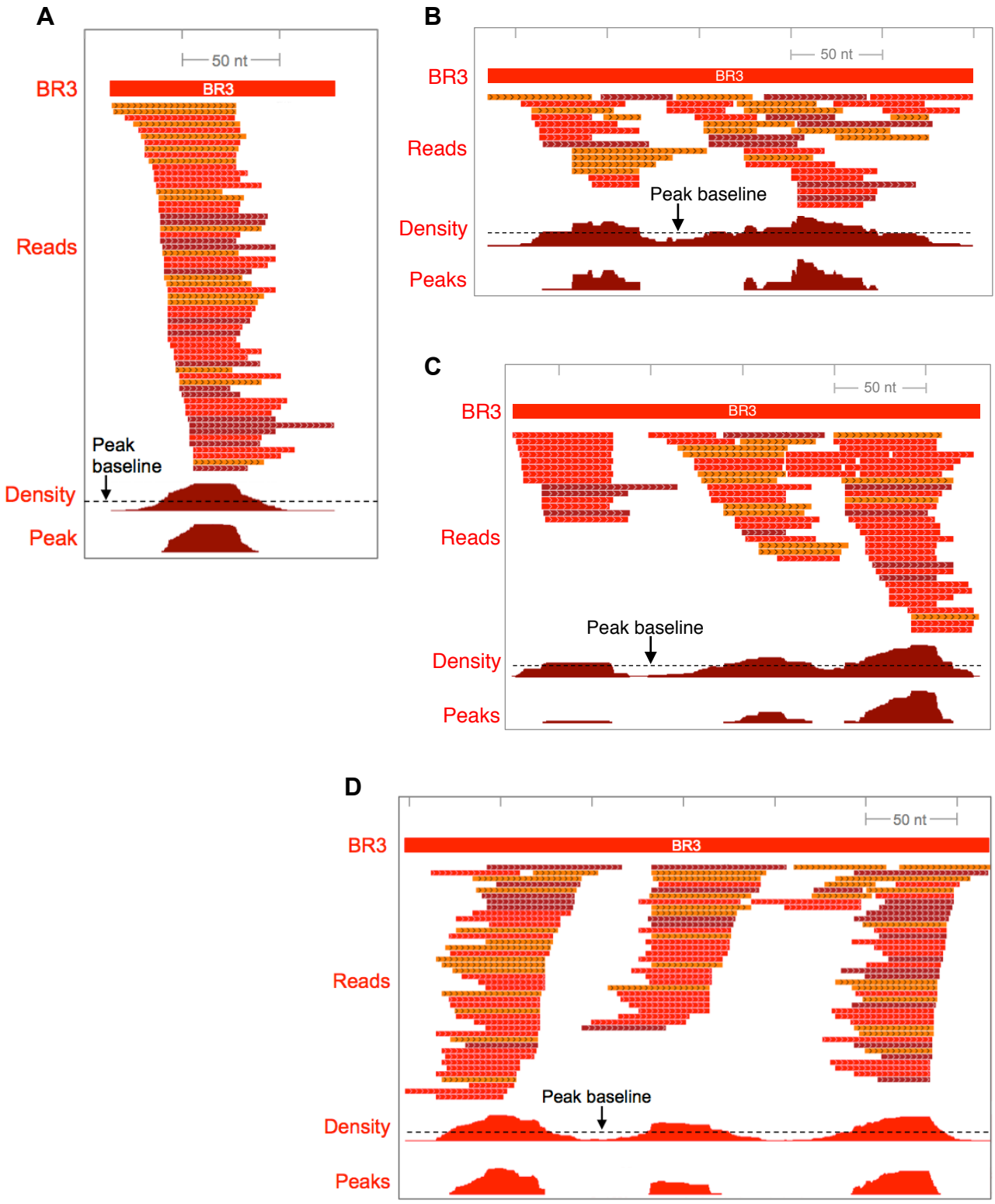

E
WT>KO
WT<KO
P >0.01
RPKM values (log2)
12
8
4
0
-4
-8
-12
WT KO WT KO WT KO

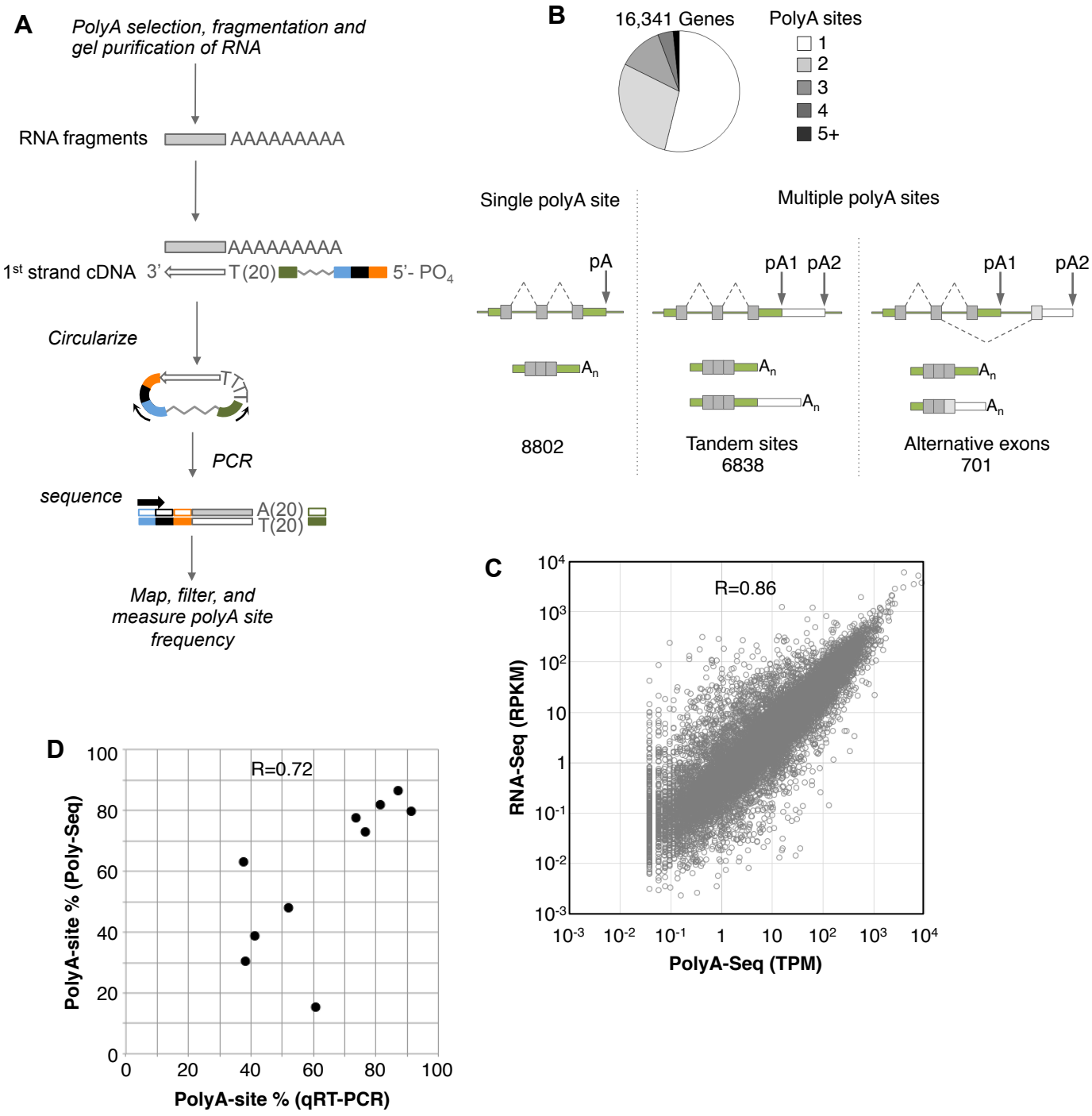

A
P6 BR3 CLIP clusters
Intergenic
ncRNA
mRNA
3'UTR
other
B
P6 BR3 sites (all)
Motif Z-score
UUUUUU 31.44
AAAAAA 16.46
UUUGUU 15.93
UGUUUU 15.80
UUGUUU 15.69
UUUUGU 14.50
UUUUUA 11.36
AUUUUU 10.79
UUUUAA 10.77
UUUUCU 10.66
P6 BR3 5'ends (10nt)
Motif Z-score
UUUUUU 25.24
UUUUGU 11.68
UUGUUU 11.47
UGUUUU 10.41
UUUGUU 10.20
UUUUAA 9.98
UUUUUA 9.77
UUUCUU 8.92
UUUUCU 8.50
AAAAAA 7.86
C
stop codon
polyA site
Dazl CLIP sites
Distance to stop codon (kb)
Distance to polyA site (kb)
D
2290 genes with Dazl-3'UTR sites
74%
26%
WT RPKM (log2)
KO RPKM (log2)
p < 0.01
WT > KO
WT < KO
WT ≈ KO
(ΔRPKM < 20%)
polyA site
Observed - Expected
Distance to polyA site
E
WT>KO
(n=1461)
WT<KO
(n=1582)
501
82
Dazl-RNA in 3'UTR
(n=2290)
F
0.06
4.9x10-3
0.9
1.1x10-4
1.2x10-2
6.6x10-3
3'UTR length (log2)
GUU/3'UTR
WT > KO
WT ≈ KO
WT < KO

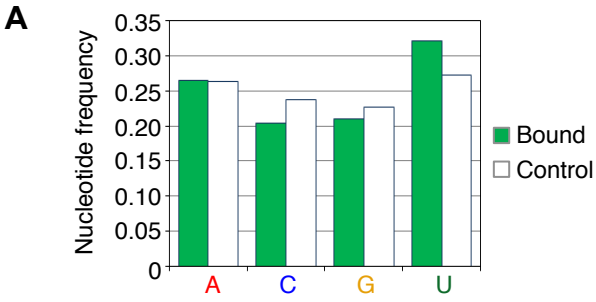


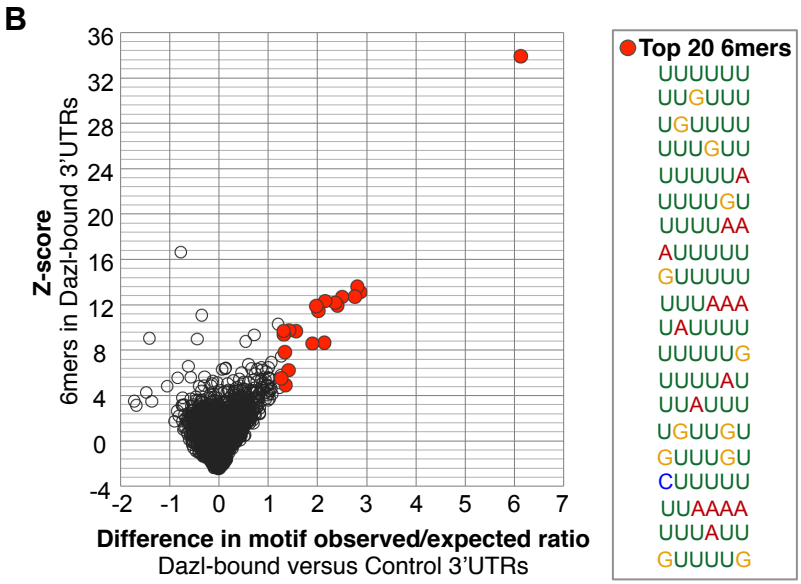
A	B	C	D	E	F	G	H	I	J	K	L	M
HELLS	PMS2	RAD17	RAD18	CENPF	TAF6	HDAC6	HUWE1	GTF2H3	SRPK1	IGF2BP1	SMCHD1	E2F6
TAF1	TOPBP1	CLSPN	CENPA	PES1	ASZ1	ARRB2	YEATS4	POLR2G	PPIL1	FYTTD1	CECR2	TAF2
CDCA5	RAD51	NAE1	MCM2	PAPD5	DDX4	UBE4B	MYST2	POLR2E	PHF5A	NXF2	PRDX1	CRLF3
CHMP1A	RPA1	RBM38	SMARCC1	NCAPG2	FKBP6	UBA6	PHF16	POLR2A	PPP1R8	RANBP2	UHRF1	DBF4
TRIP13	CPEB1	CCND1	SMARCA5	DSN1	TAF7L	SIAH1B	PHF17	POLR2B	ESRP1	EIF5A	PARG	CCNY
KDM1B	STAG3	RFWD3	TOP2A	CDCA8	CELF1	CDC34	CBX2	MED17	ZRSR2	EIF5A2	ESCO1	CDC25A
PIWIL2	TEX11	DTL		RCC2	ZBTB16	RNF216	CHD1L	ELL	SRSF9	NUPL1	ESCO2	TIPIN
TDRD9	SYCE2	USP28		NEK1	CADM1	PSME3	NASP	GTF2A1	SNRNP40	GLE1	RPA2	PLK1
WDR5	SYCP1	OBFC2A		KLHL21	TLE3	PSMD2	2410022L05RIK	GTF2E2	SF3B3	NUP210	FEN1	ATM
ASH2L		MAPK14		KIF11	SPATA5	PSMD5	MBTD1	GTF2E1	SF3A3		SMG1	CHFR
CXXC1		XPC		WEE1	SOHLH2	USP7		TAF4B	SF3A1			PCID2
JARID2				CCNA2	RNF114	USP19		TAF5	PUF60			APC
WHSC1				CEP55	GSR	USP34		TAF9B	NAA38			MAD2L1
BAZ2A				INCENP	NR6A1	CUL5			LSM4			STRA8
DNMT1				NEK2	KIT	USP14			LSM3			SYCP3
DNMT3B					SOX3	UCHL1			AQR			
					HOOK1	USP26			HNRNPU			
					RNF17				ROD1			
									CPSF4			
									PDCD11			

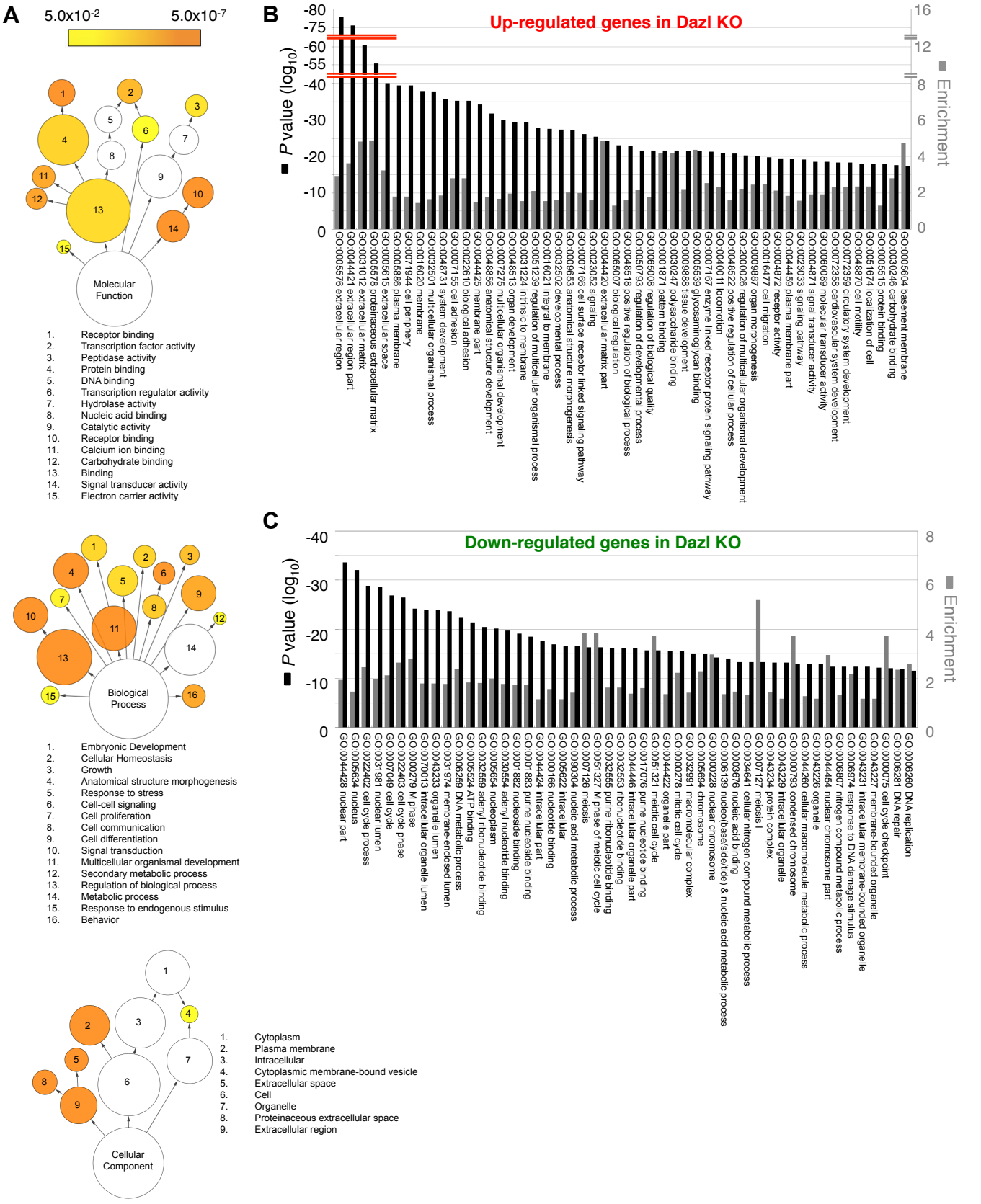

Proteasome
DNA Replication
Epigenetic Modification
Chromatin Modification
Cell Cycle
RNA Processing
Export
RNA Pol II
Transcription Factors
Ubiquitination
Synaptonemal complex


A
50 nt
BR3
BR3
Reads
Peak
baseline
Density
Peak
B
50 nt
BR3
BR3
Reads
Peak baseline
Density
Peaks
C
50 nt
BR3
BR3
Reads
Peak baseline
Density
Peaks
D
50 nt
BR3
BR3
Reads
Peak baseline
Density
Peaks

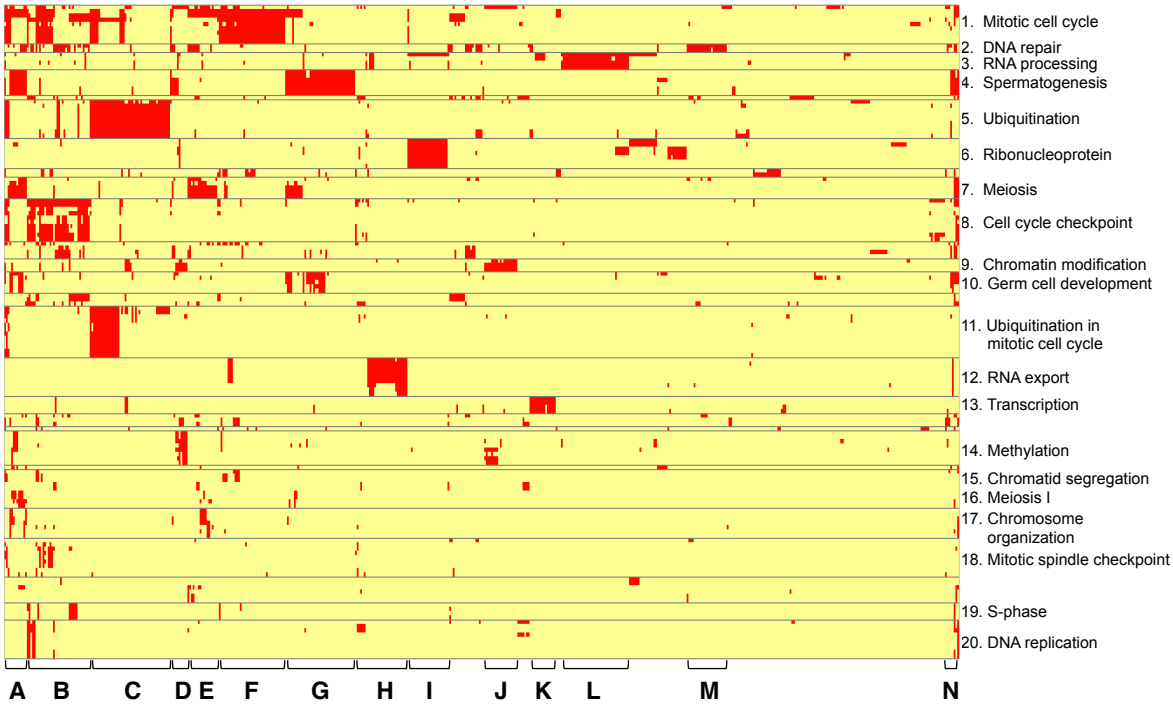

A
PolyA selection, fragmentation and gel purification of RNA
RNA fragments
AAAAAAAAA
AAAAAAAAA
1st strand cDNA
3'
T(20)
5'- PO4
Circularize
PCR
sequence
A(20)
T(20)
Map, filter, and measure polyA site frequency
B
16,341 Genes
PolyA sites
1
2
3
4
5+
Single polyA site
Multiple polyA sites
pA
pA1
pA2
An
8802
Tandem sites
6838
Alternative exons
701
C
R=0.86
RNA-Seq (RPKM)
PolyA-Seq (TPM)
D
R=0.72
PolyA-site % (Poly-Seq)
PolyA-site % (qRT-PCR)


A
Nucleotide frequency
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0
A
C
G
U
Bound
Control

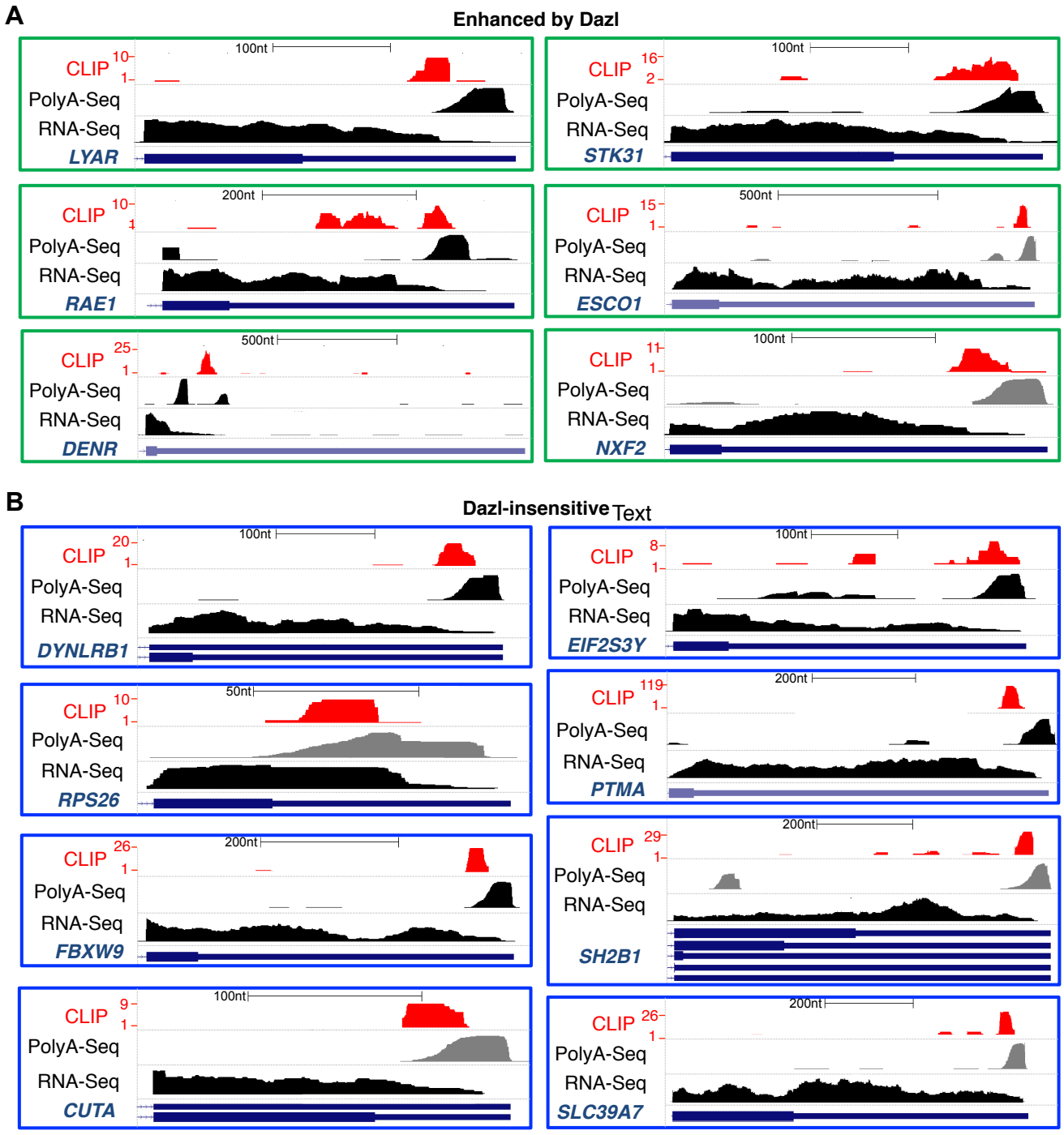

B
Z-score
6mers in Dazl-bound 3'UTRs
36
32
28
24
20
16
12
8
4
0
-4
-2
-1
0
1
2
3
4
5
6
7
Difference in motif observed/expected ratio
Dazl-bound versus Control 3'UTRs
Top 20 6mers
UUUUUU
UUGUUU
UGUUUU
UUUGUU
UUUUUA
UUUUGU
UUUUAA
AUUUUU
GUUUUU
UUUAAA
UAUUUU
UUUUUG
UUUUAU
UUAUUU
UGUUGU
GUUUGU
CUUUUU
UUAAAA
UUUAUU
GUUUUG

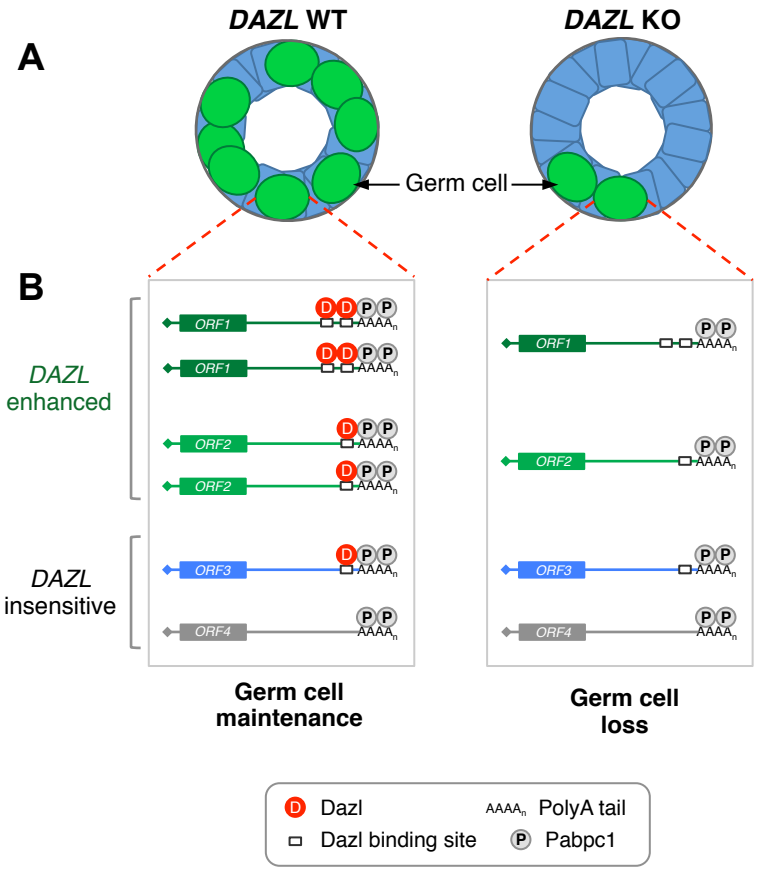

A
5.0x10-2
5.0x10-7
Molecular Function
1. Receptor binding
2. Transcription factor activity
3. Peptidase activity
4. Protein binding
5. DNA binding
6. Transcription regulator activity
7. Hydrolase activity
8. Nucleic acid binding
9. Catalytic activity
10. Receptor binding
11. Calcium ion binding
12. Carbohydrate binding
13. Binding
14. Signal transducer activity
15. Electron carrier activity
B
Up-regulated genes in Dazl KO
P value (log10)
Enrichment
GO:0005576 extracellular region
GO:0044421 extracellular region part
GO:0031012 extracellular matrix
GO:0005578 proteinaceous extracellular matrix
GO:0005615 extracellular space
GO:0005886 plasma membrane
GO:0071944 cell periphery
GO:0016020 membrane
GO:0032501 multicellular organismal process
GO:0048731 system development
GO:0007155 cell adhesion
GO:0022610 biological adhesion
GO:0044425 membrane part
GO:0048856 anatomical structure development
GO:0007275 multicellular organismal development
GO:0048513 organ development
GO:0031224 intrinsic to membrane
GO:0051239 regulation of multicellular organismal process
GO:0016021 integral to membrane
GO:0032502 developmental process
GO:0009653 anatomical structure morphogenesis
GO:0007166 cell surface receptor linked signaling pathway
GO:0023052 signaling
GO:0044420 extracellular matrix part
GO:0065007 biological regulation
GO:0048518 positive regulation of biological process
GO:0050793 regulation of developmental process
GO:0065008 regulation of biological quality
GO:0001871 pattern binding
GO:0030247 polysaccharide binding
GO:0009888 tissue development
GO:0005539 glycosaminoglycan binding
GO:0007167 enzyme linked receptor protein signaling pathway
GO:0040011 locomotion
GO:0048522 positive regulation of cellular process
GO:2000026 regulation of multicellular organismal development
GO:0009887 organ morphogenesis
GO:0016477 cell migration
GO:0004872 receptor activity
GO:0044459 plasma membrane part
GO:0023033 signaling pathway
GO:0004871 signal transducer activity
GO:0060089 molecular transducer activity
GO:0072358 cardiovascular system development
GO:0072359 circulatory system development
GO:0044870 cell motility
GO:0051674 localization of cell
GO:0005515 protein binding
GO:0030246 carbohydrate binding
GO:0005604 basement membrane
C
Down-regulated genes in Dazl KO
P value (log10)
Enrichment
GO:0044428 nuclear part
GO:0005634 nucleus
GO:0022402 cell cycle process
GO:0031981 nuclear lumen
GO:0007049 cell cycle
GO:0022403 cell cycle phase
GO:0000279 M phase
GO:0070013 intracellular organelle lumen
GO:0043233 organelle lumen
GO:0031974 membrane-enclosed lumen
GO:0006259 DNA metabolic process
GO:0005524 ATP binding
GO:0032559 adenyl ribonucleotide binding
GO:0005654 nucleoplasm
GO:0030554 adenyl nucleotide binding
GO:0001882 nucleoside binding
GO:0001883 purine nucleoside binding
GO:0044424 intracellular part
GO:0000166 nucleotide binding
GO:0005622 intracellular
GO:0090304 nucleic acid metabolic process
GO:0007126 meiosis
GO:0051327 M phase of meiotic cell cycle
GO:0032555 purine ribonucleotide binding
GO:0032553 ribonucleotide binding
GO:0044446 intracellular organelle part
GO:0017076 purine nucleotide binding
GO:0051321 meiotic cell cycle
GO:0044422 organelle part
GO:0000278 mitotic cell cycle
GO:0032991 macromolecular complex
GO:0005694 chromosome
GO:0000228 nuclear chromosome
GO:0006139 nucleo(base/side/tide) & nucleic acid metabolic process
GO:0003676 nucleic acid binding
GO:0034641 cellular nitrogen compound metabolic process
GO:0007127 meiosis I
GO:0043234 protein complex
GO:0043229 intracellular organelle
GO:0000793 condensed chromosome
GO:0044260 cellular macromolecule metabolic process
GO:0043226 organelle
GO:0044454 nuclear chromosome part
GO:0006807 nitrogen compound metabolic process
GO:0006974 response to DNA damage stimulus
GO:0043231 intracellular membrane-bounded organelle
GO:0043227 membrane-bounded organelle
GO:0000075 cell cycle checkpoint
GO:0006281 DNA repair
GO:0006260 DNA replication
Biological Process
1. Embryonic Development
2. Cellular Homeostasis
3. Growth
4. Anatomical structure morphogenesis
5. Response to stress
6. Cell-cell signaling
7. Cell proliferation
8. Cell communication
9. Cell differentiation
10. Signal transduction
11. Multicellular organismal development
12. Secondary metabolic process
13. Regulation of biological process
14. Metabolic process
15. Response to endogenous stimulus
16. Behavior
Cellular Component
1. Cytoplasm
2. Plasma membrane
3. Intracellular
4. Cytoplasmic membrane-bound vesicle
5. Extracellular space
6. Cell
7. Organelle
8. Proteinaceous extracellular space
9. Extracellular region

Zagore_Fig_S5



A	B	C	D	E	F	G	H	I	J	K	L	M	N
CCNB1	TIPIN	UBE2C	HERC2	MSH6	CENPF	DAZL	XPO1	PES1	EED	CCNH	MOV10	UNG	BAX
MAD2L1	CDT1	BUB1B	FANCD2	TOPBP1	RCC1	CCNB1IP1	THOC6	PDCD11	JARID2	GTF2H3	SRPK1	LIG1	DDX25
PLK1	RAD17	CDC20	PRKDC	EXO1	HELLS	MEI1	DDX39	FTSJ3	ASH2L	TAF6	TFIP11	RFC4	BRCA2
SYCP3	CLSPN	ANAPC1	KDM1B	PMS2	KIF18A	PIWIL1	THOC1	BOP1	CXXC1	POLR2G	SUGP1	RFC5	MSH2
EHMT2	NAE1	ANAPC4	PRMT7	RPA1	ESPL1	CCNA1	XPO7	WDR55	SUV39H2	POLR2E	SRSF9	BARD1	STRA8
MAEL	CDCA5	PSMD13	SUV39H1	RAD50	NEK2	DDX4	EIF5A2	UTP20	WHSC1	POLR2B	SNRNP40	PARG	
PIWIL2	CHMP1A	PSME3	SMARCA4	RAD51	AHCTF1	FKBP6	RANBP2	UTP18	EZH1	POLR2A	SNRNP25	ESCO1	
TDRD9	ATM	PSMD5	BAZ2A	TEX11	NUP37	CLGN	RANBP17	TBL3	WDR5	GTF2F1	SF3B3	ESCO2	
SPO11	ZW10	PSMD2	DNMT1	CPEB1	NUP43	ASZ1	NUPL1	RRP9	EYA2	GTF2F2	SF3A3	PALB2	
DMC1	APC	PSMD11	DNMT3B	SYCE2	NCAPG2	BOLL	NUP93	RRP1B	HAT1	ELL	SF3A1	TDP2	
TRIP13	BUB1	PSMD1		SYCP1	NCAPG	CELF1	NUP88	RBFA	HUWE1	TAF5	PUF60	SETX	
RAD51C	KNTC1	PSMC5		SGOL2	NCAPD2	RAD18	NUP50	POP5	TAF6L	TAF4B	PRPF3	TDP1	
TEX15	PCID2	PSMC1		STAG3	NCAPD3	DND1	NUP210	NOP2	NOC2L	GTF2E2	PPP1R8	SMC5	
	CENPE	PSMB7		FANCA	TIMELESS	TDRD7	NUP188	NOP14	MYST2	GTF2A1	PPIL1	ZSWIM7	
	CEP192	PSMB3		SMC1B	NEK1	KIT	LRPPRC	NOL9	KAT2A	GTF2E1	PLRG1	CIB1	
	MRE11A	PSMB1		KLHDC3	PAPD5	ZBTB16	NUP155	NHP2	YEATS4		PHF5A	UHRF1	
	ERCC3	PSMB2		PSMC3IP	KIF2C	TAF7L	XPOT	IMP4	PHF17		NAA38	TRP53BP1	
	DTL	CHFR			NDC80	MAST2	EIF5A	EXOSC8	ACTL6A		LSM4	FANCI	
	USP28	CCNF			KIF20B	SOX3	NXF2	EMG1	PHF16		LSM3	BRIP1	
	INTS3	USP9X			KIF11	HOOK1	NUP160	DIS3			HNRNPU	EEF1E1	
	UIMC1	TAF1			HAUS4	STRBP	NUP133	DDX56			ESRP2	PARP1	
	XPC	TAF9			HAUS8	ADAD1	FYTTD1	DCAF13			ESRP1	CHD1L	
	MAPK14	HDAC6			NUDC	RNF17	GLE1	DDX51			EFTUD2	MUM1	
	OBFC2A	USP3			KLHL21	HTT					DDX41		
	APBB1	ARRB2			DSN1	HMGA1					CSTF2		
	CRLF3	PAN2			INCENP	ZFP541					CSTF1		
	RB1	GCLC			CDCA8	TXNRD3					CPSF3		
	APEX1	UBE4B			CDC6	TLE3					AQR		
	DBF4	USP14			AURKB	SPATA5					CPSF1		
	FBXO31	UCHL1			ZC3HC1	SOHLH2					CPSF4		
	CCND1	USP34			WEE1	SBF1					ROD1		
	HRAS1	USP24			RCC2	RNF114					WDR77		
	RFWD3	2700078E11RIK			NEK3	NR6A1					LSM2		
	RBM38	CUL5			MASTL	NANOS3					ZRSR2		
	DGKZ	USP7			FSD1	MORC1					SNRPE		
	HINFP	USP26			CEP55	MEA1					SIP1		
		SIAH1B			CDCA2	JAG2					GEMIN5		
		UBA6			AURKA	GSR					DDX20		
		USP19			CCNA2	CADM1					GEMIN4		
		WWP2				D1PAS1							
		USP5											
		RNF216											
		RNF180											
		NUB1											
		CDC34											
		FBXO2											


A
Enhanced by Dazl
CLIP
PolyA-Seq
RNA-Seq
LYAR
STK31
RAE1
ESCO1
DENR
NXF2
B
Dazl-insensitive Text
DYNLRB1
EIF2S3Y
RPS26
PTMA
FBXW9
SH2B1
CUTA
SLC39A7
C
DAZL-enhanced
DAZL-insensitive
RPKM
WT
KO
LYAR
RAE1
DENR
STK31
ESCO1
NXF2
EIF2S3Y
RPS26
FBXW9
PTMA
SLC39A7
DYNLRB1
CUTA
SH2B1


A
DAZL WT
DAZL KO
Germ cell
B
DAZL
enhanced
DAZL
insensitive
ORF1
ORF2
ORF3
ORF4
AAAAn
Germ cell
maintenance
Germ cell
loss
Dazl
Dazl binding site
PolyA tail
Pabpc1

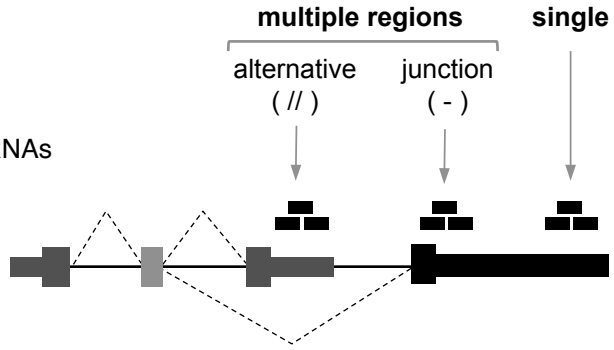
	Monomer 1	Monomer 2	Monomer 3	Total
Reads	7,009,499	28,586,273	26,925,643	62,521,415
Mapped with Bowtie	3,150,466	14,891,083	14,317,694	
Unique coordinate & barcode combo	482,940	271,995	237,336	
Unique within library	387,476	207,270	171,757	766,503
Unique across all 6 libraries	322,741	117,782	99,499	**540,022**
Mapped with Tophat	262,253	222,216	218,315	702,784
Unique starts across all 6 libraries	105,884	48,449	49,009	203,342
Exon spanning tags	*1437*	*440*	*209*	*2086*
Removal of exon spanning tags	104,447	48009	48800	**201,256**
Filtered Bowtie and Tophat combined	427,188	165,791	148,299	741,278
Removal of multi-hit	369,403	146,479	131,550	**647,432**
Exon spanning tags	*1437*	*440*	*209*	*2086*
Including exon spanning tags	370,840	146,919	131,759	**649,518**

9524 / 11297 map to annotated genes in RefSeq

237 map to non-coding RNAs
("NR_" RefSeq IDs)

225 map to overlapping coding and non-coding RNAs
("NR_" and "NM_" RefSeq IDs)

9062 map to protein coding genes
("NM_" RefSeq IDs only)



Genic location	BR3 Clusters	% BR3	5'UTR	3'UTR	CDS	Intron
3UTR	6397	70.59		6397		
3UTR // 5UTR	3	0.03	3	3		
3UTR // CDS-3UTR	10	0.11		10	10	
3UTR // CDS-INTRON	1	0.01		1	1	1
3UTR // INTRON	157	1.73		157		157
3UTR // INTRON-3UTR	5	0.06		5		5
3UTR // INTRON-CDS-3UTR	3	0.03		3	3	3
3UTR-INTRON	1	0.01		1		1
5UTR	57	0.63	57			
5UTR // CDS	1	0.01	1		1	
5UTR // INTRON	7	0.08	7			7
5UTR-CDS	37	0.41	37		37	
5UTR-CDS // INTRON	3	0.03	3		3	3
5UTR-CDS // INTRON-5UTR-CDS	3	0.03	3		3	3
5UTR-CDS-INTRON	3	0.03	3		3	3
5UTR-INTRON	4	0.04	4			4
CDS	741	8.18			741	
CDS // 3UTR	1	0.01		1	1	
CDS // INTRON	16	0.18			16	16
CDS // INTRON-CDS	1	0.01			1	1
CDS-3UTR	775	8.55		775	775	
CDS-3UTR // CDS-3UTR-INTRON	2	0.02		2	2	2
CDS-3UTR // CDS-INTRON	4	0.04		4	4	4
CDS-3UTR // INTRON	15	0.17		15	15	15
CDS-3UTR-INTRON	5	0.06		5	5	5
CDS-INTRON	226	2.49			226	226
INTRON	340	3.75				340
INTRON-3UTR	10	0.11		10		10
INTRON-3UTR // INTRON-CDS-3UTR	1	0.01		1	1	1
INTRON-5UTR	2	0.02	2			2
INTRON-5UTR // 5UTR	1	0.01	1			1
INTRON-5UTR-CDS	6	0.07	6		6	6
INTRON-5UTR-CDS-INTRON	3	0.03	3		3	3
INTRON-5UTR-INTRON	1	0.01	1			1
INTRON-CDS	126	1.39			126	126
INTRON-CDS // 3UTR	2	0.02		2	2	2
INTRON-CDS // INTRON-CDS-3UTR	1	0.01		1	1	1
INTRON-CDS-3UTR	62	0.68		62	62	62
INTRON-CDS-3UTR-INTRON	1	0.01		1	1	1
INTRON-CDS-INTRON	28	0.31			28	28

Top 10 motifs in 5'LOW genes

	Z-score	
Motif	**5'low**	**5'high**
GCGG	6.42	0.07
UUUU	6.32	20.00
CGGC	5.48	0.81
GGCG	5.48	-1.32
CCCC	4.91	3.68
CCUC	4.40	3.50
CUCC	4.35	5.72
GGAG	4.07	0.90
CCGC	3.93	1.74
CGGG	3.60	-0.30

Top 10 motifs in 5'HIGH genes

	Z-score	
Motif	**5'low**	**5'high**
UUUU	6.32	20.00
AAAA	1.15	6.74
UUUC	2.52	6.74
GUUU	1.39	6.19
CUUU	2.89	5.91
UUUG	2.05	5.91
CUCC	4.35	5.72
AUUU	0.17	5.45
UUCU	2.33	5.26
UGUU	0.78	5.07

Top 10 motifs in 5'LOW genes

	Z-score	
Motif	**5'low**	**5'high**
GGCGG	14.53	2.91
GCGGC	13.25	3.54
UUUUU	11.32	33.23
CGGCG	8.60	1.33
CCCCC	7.64	2.91
CCGCC	7.48	5.75
CGGCC	7.48	3.54
CCUCC	6.52	7.64
CCCUC	6.52	4.49
UUUUC	6.36	10.49

Top 10 motifs in 5'HIGH genes

	Z-score	
Motif	**5'low**	**5'high**
UUUUU	11.32	33.23
GUUUU	2.83	13.33
UGUUU	1.55	13.33
UUUUG	2.83	13.33
AAAAA	1.55	13.02
UUUCU	4.75	13.02
AUUUU	2.67	12.07
UUUGU	2.19	11.44
UUUUA	1.07	11.12
UUUUC	6.36	10.49

qRT-PCR primers	Sequence
Cpeb2 forward	CGTTATCAGCAGCCAAAACA
Cpeb2 reverse	GGGTTTTGGTGAAAAACAAGA
eIF4b forward	CGACGACTGCACTGAGTAGG
eIF4b reverse	CGGGGTGGAGACTAGGAGA
Arf4 forward	TGAAGCTGGATATCTAACCAAGG
Arf4 reverse	TTTGTTGTAACAAGCCTAGACCAA
Ndfip2 forward	GACTGCTGAGAGCAGAAGGAA
Ndfip2 reverse	GCGTAAACCACCCTGTCATT
Pacrgl forward	TCACGGCATCATTCATTCAT
Pacrgl reverse	TGATGATTCTTCCGAACTTGG
Terf1 forward	TCAACAACCGAACAAGTGTCA
Terf1 reverse	AGCTGAATCAAACCCATCCA
Fndc3a forward	AGTGTGCTGCAGTCATCCTG
Fndc3a reverse	GGCAATCAGAATGGAGAAAAA
Celf1 forward	CTTGTGAACACACAGCCAGTC
Celf1 reverse	AAGCACACAGAGGAACAAAGC
Hook1 forward	CACCCGAAGAAACCTCTCTG
Hook1 reverse	GAGGCACTTCCGTTGTGTTT
Dazap1 forward	CGCATCTGGACTGAGGTTTT
Dazap1 reverse	TGTGCTGATGGGTCAGAGAA

HITS-CLIP primers	Sequence
5' RNA linker	AGGGAGGACGAUGCGG
3' RNA lnker	5'-P GUGUCAGUCACUUCCAGCGG 3'-puromycin
5' DNA primer	AGGGAGGACGATGCGG
3' DNA primer	CCGCTGGAAGTGACTGACAC
DSFP5	AATGATACGGCGACCACCGACTATGGATACTTAGTCAGGGAGGACGATGCGG
DSFP3	CAAGCAGAAGACGGCATACGACCGCTGGAAGTGACTGACAC
SSP1 - sequencing primer	CTATGGATACTTAGTCAGGGAGGACGATGCGG

p6 iCLIP primers	Sequence
3' RNA lnker	5'-P GUGUCAGUCACUUCCAGCGG 3'-puromycin
CLIP RT barcode A	GNNNNCGTGATCCGCATCGTCCTCCCTCCCTATAGTGAGTCGTATTA-iSp18-CACTCA-iSp18-CCGCTGGAAGTGACTGACAC
CLIP RT barcode B	GNNNNACATCGCCGCATCGTCCTCCCTCCCTATAGTGAGTCGTATTA-iSp18-CACTCA-iSp18-CCGCTGGAAGTGACTGACAC
CLIP RT barcode C	GNNNNGCCCTACCGCATCGTCCTCCCTCCCTATAGTGAGTCGTATTA-iSp18-CACTCA-iSp18-CCGCTGGAAGTGACTGACAC
5' DNA primer	AGGGAGGACGATGCGG
3' DNA primer	CCGCTGGAAGTGACTGACAC
DSFP5	AATGATACGGCGACCACCGACTATGGATACTTAGTCAGGGAGGACGATGCGG
DSFP3	CAAGCAGAAGACGGCATACGACCGCTGGAAGTGACTGACAC
SSP1 - sequencing primer	CTATGGATACTTAGTCAGGGAGGACGATGCGG

PolyA-Seq primers	Sequence
RT primer	GNNNNCAACTCCGCATCGTCCTCCCTCCCTATAGTGAGTCGTATTA-iSp18-CACTCA-iSp18-CCGCTGGAAGTGACTGACACTTTTTTTTTTTTTTTTTTTTTVN
5' DNA primer	AGGGAGGACGATGCGG
3' DNA primer	CCGCTGGAAGTGACTGACAC
DSFP5	AATGATACGGCGACCACCGACTATGGATACTTAGTCAGGGAGGACGATGCGG
DSFP3	CAAGCAGAAGACGGCATACGACCGCTGGAAGTGACTGACAC
SSP1 - sequencing primer	CTATGGATACTTAGTCAGGGAGGACGATGCGG