

**Supplementary Information for Lee et al,
"Diverse cell-specific patterns of alternative polyadenylation in *Drosophila*"**

Full list of Fly Cell Atlas (FCA) Consortium members and affiliations p. 1

Supplementary Figures 1-8 p. 2-9

Fig S1. Illustrative patterns of FCA 10X-3' data clusters in the vicinity of 3' UTRs.

Fig S2. Neurons express downstream 3' isoforms relative to other celltypes.

Fig S3. Additional examples of 3' isoform dynamics in spermatocytes.

Fig S4. Genes that preferentially express distal 3' isoforms in late spermatocytes.

Fig S5. Additional examples of 3' isoform shifts in the intestinal stem cell lineage.

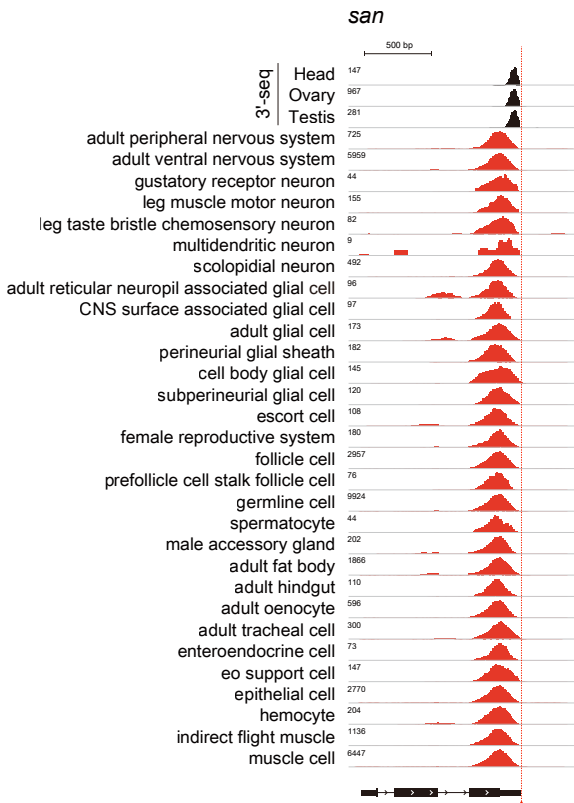
Fig S6. RBPs that are differentially expressed in ISCs and their progeny.

Fig S7. RBP and miRNA binding site enrichment analysis.

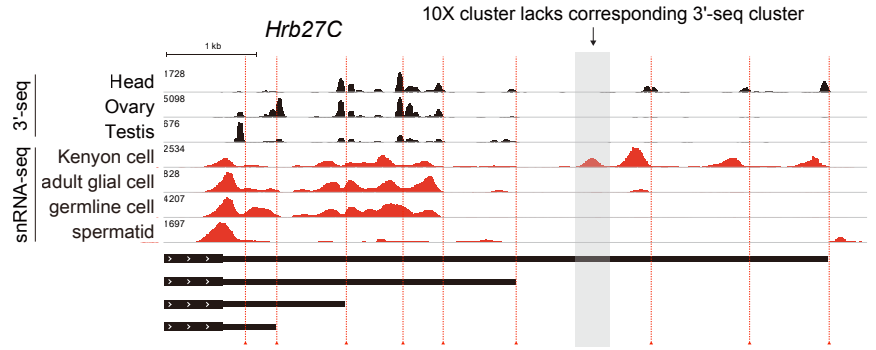
Fig S8. Most cleavage and polyadenylation (CPA) factors are coordinately induced in early spermatocytes.

Supplementary Table 1 p. 10

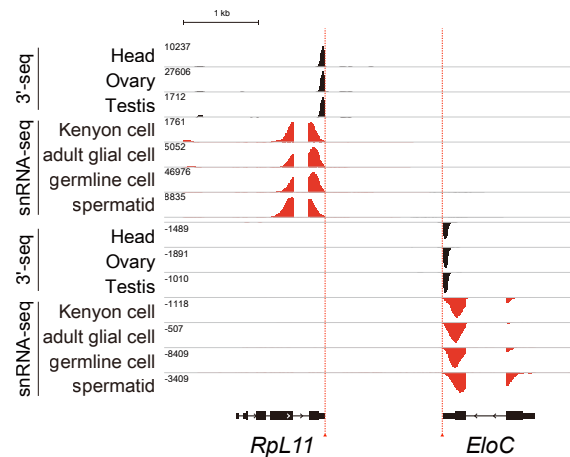
A Example of single-end transcript



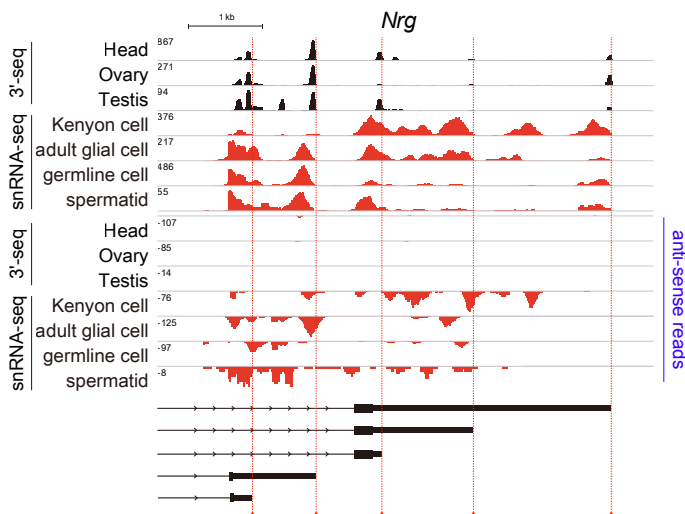
B Example of multiple tissue/cell-specific 3' UTR isoforms



C Example of 10X read clusters that are separated by an intron

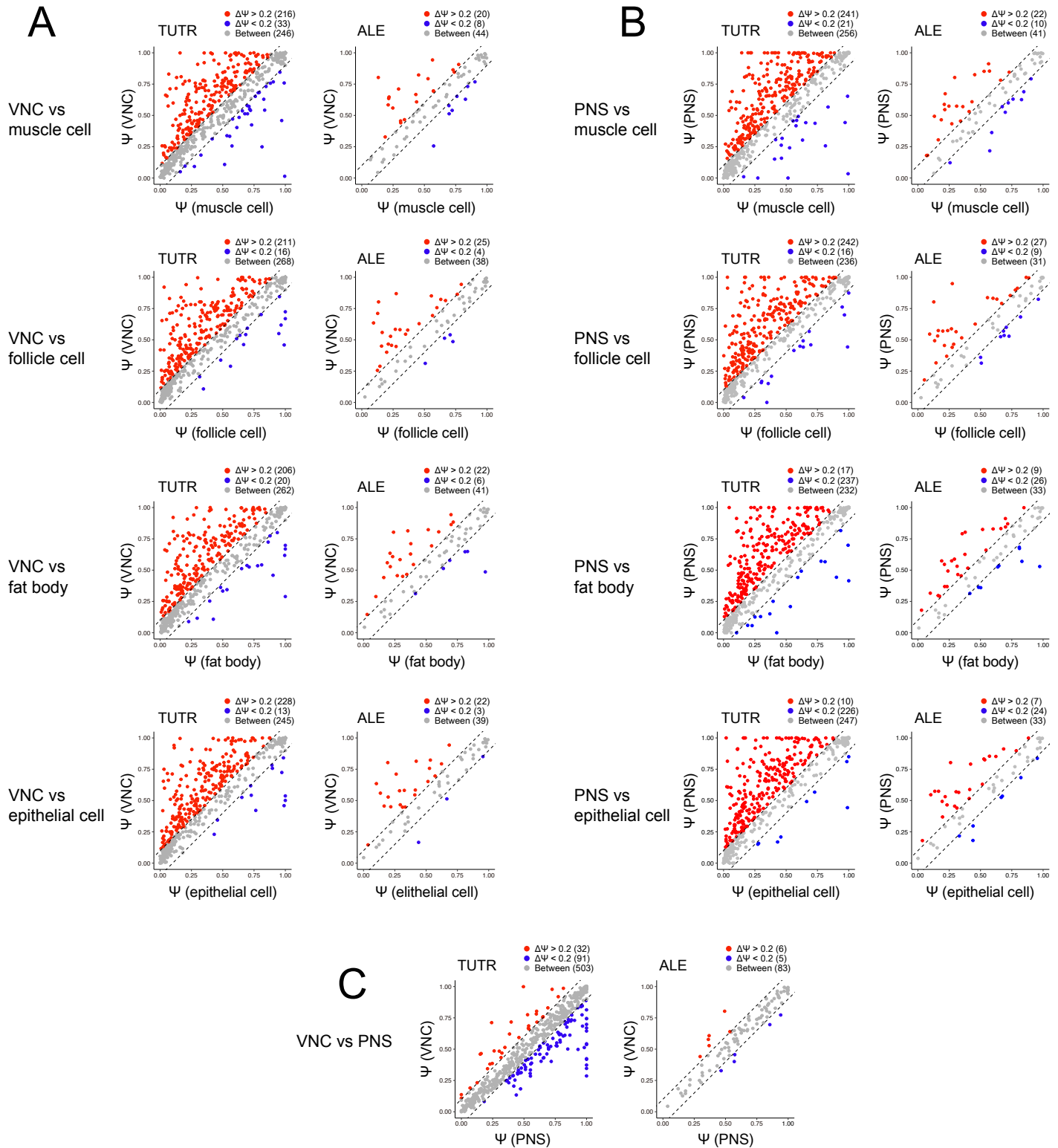


D Examples of antisense reads in 10X data



Supplementary Figure 1. Illustrative patterns of FCA 10X-3' data clusters in the vicinity of 3' UTRs.

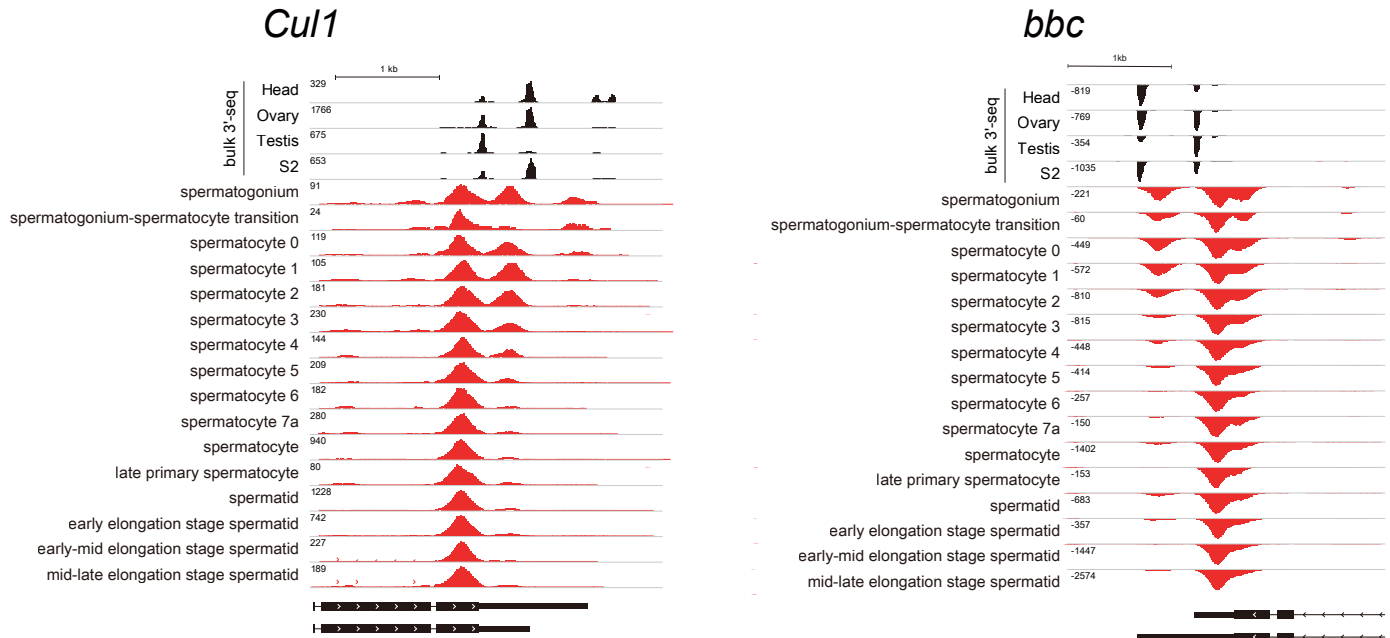
(A) Many genes stably express a single 3' UTR isoform, as determined by tissue-specific 3'-seq profiling, and this is mirrored in their highly reproducible, non-alternative, 10x read clusters across individual cell types. (B) Example of a gene with multiple tissue-specific tandem 3' UTR isoforms, which are seen in both bulk tissue-level 3'-seq and 10X data. Grey shaded region highlights 10X read clusters that lack corresponding 3'-seq signals. While these might be cell-specific 3' ends, we chose to be conservative and did not consider scRNA-peaks that were not in the vicinity of annotated 3'-seq termini. (C) Because 10x-3' clusters are located upstream of 3' ends, in cases where the terminal exon is very short, the single cell data can be separated from the 3' end by an intron. Thus, read clusters should be interpreted relative to the transcriptome and not the genome. (D) Examples of antisense reads in 10X data. The *Nrg* locus generates multiple types of ALE-APA and TUTR-APA 3' isoforms. The 3'-seq data indicates largely sense strand data in head/ovary/testis; however, 10X data of constituent cells in these tissues that express these isoforms also shows substantial antisense clusters, which may be an artifact of single cell sequencing.



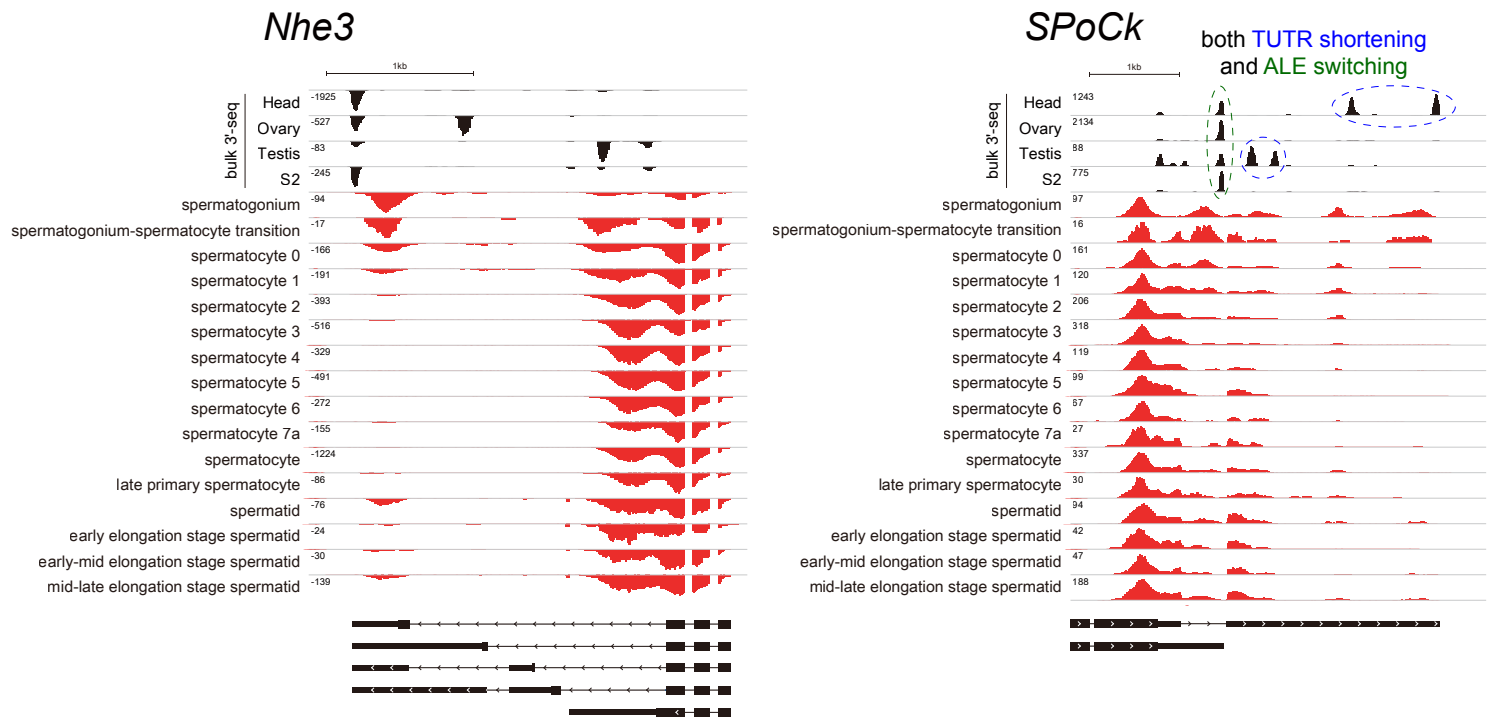
Supplementary Figure 2. Neurons express downstream 3' isoforms relative to other celltypes.

This figure depicts the relative usage of 3' isoforms (PSI) at individual genes in several annotated celltypes from the adult body FCA dataset. (A-C) The left panels in each group represent genes subject to tandem 3' UTR (TUTR) APA, the right panels in each group depict genes subject to alternative last exon (ALE) APA. Genes are colored red if they substantially express more extended/distal 3' isoforms for the Y-axis celltype compared to the X-axis celltype, blue if they have the reciprocal property, and grey if similar between celltypes. The numbers of genes in each category for each comparison are listed in each graph. (A) Comparison of the ventral nerve cord (VNC) to other major body celltypes. (B) Comparison of peripheral nervous system (PNS) to other major body celltypes. (C) Comparison of the VNC and PNS.

A TUTR shortening in the testis germline



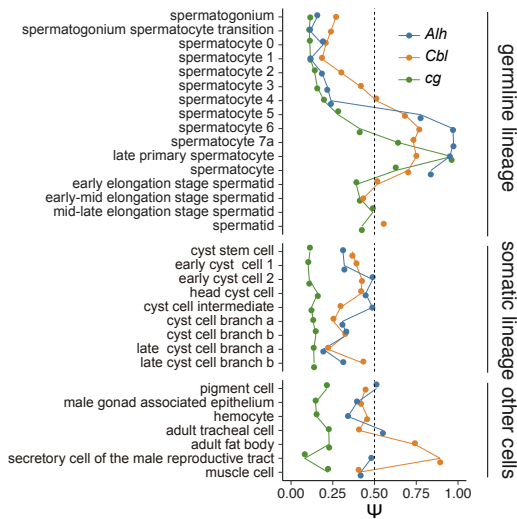
B Distal-to-proximal ALE switching in the testis germline



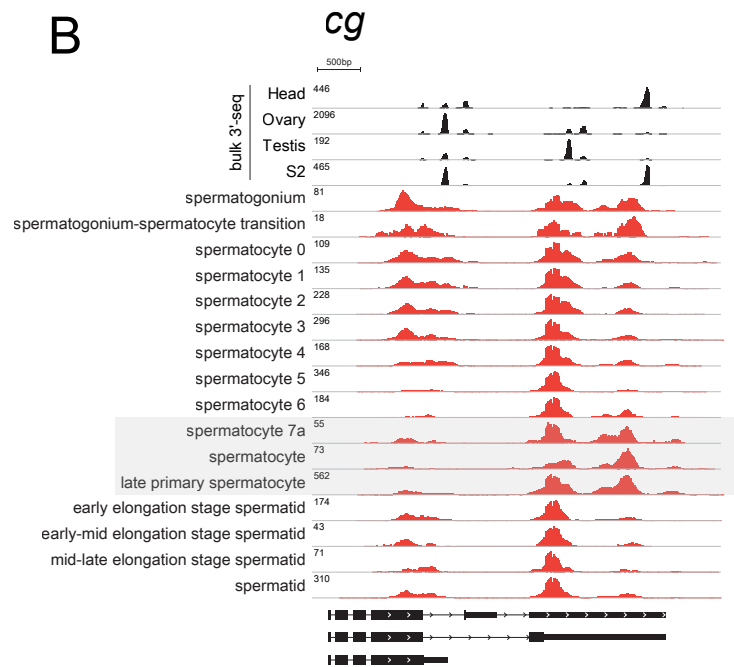
Supplementary Figure 3. Additional examples of 3' isoform dynamics in spermatocytes.

This figure shows IGV browser tracks for bulk 3'-seq data from head/ovary/testis/S2 cells (in black, top tracks) and 10X-3' scRNA-seq data from individual cell types in the testis; gene models are shown at the bottom. (A) Examples of switching from longer to shorter tandem 3' UTRs (TUTR) in early spermatocytes. (B) Examples of switching from distal to proximal alternative last exon (ALE) isoforms in early spermatocytes. Note that SPoCk is subject to both TUTR shortening and ALE switching in the germline lineage, and having lost expression of the distal ALE isoform in favor of the proximal ALE isoform in early spermatocytes, it actually regains partial expression of the distal ALE isoform in late spermatocytes and spermatids.

A



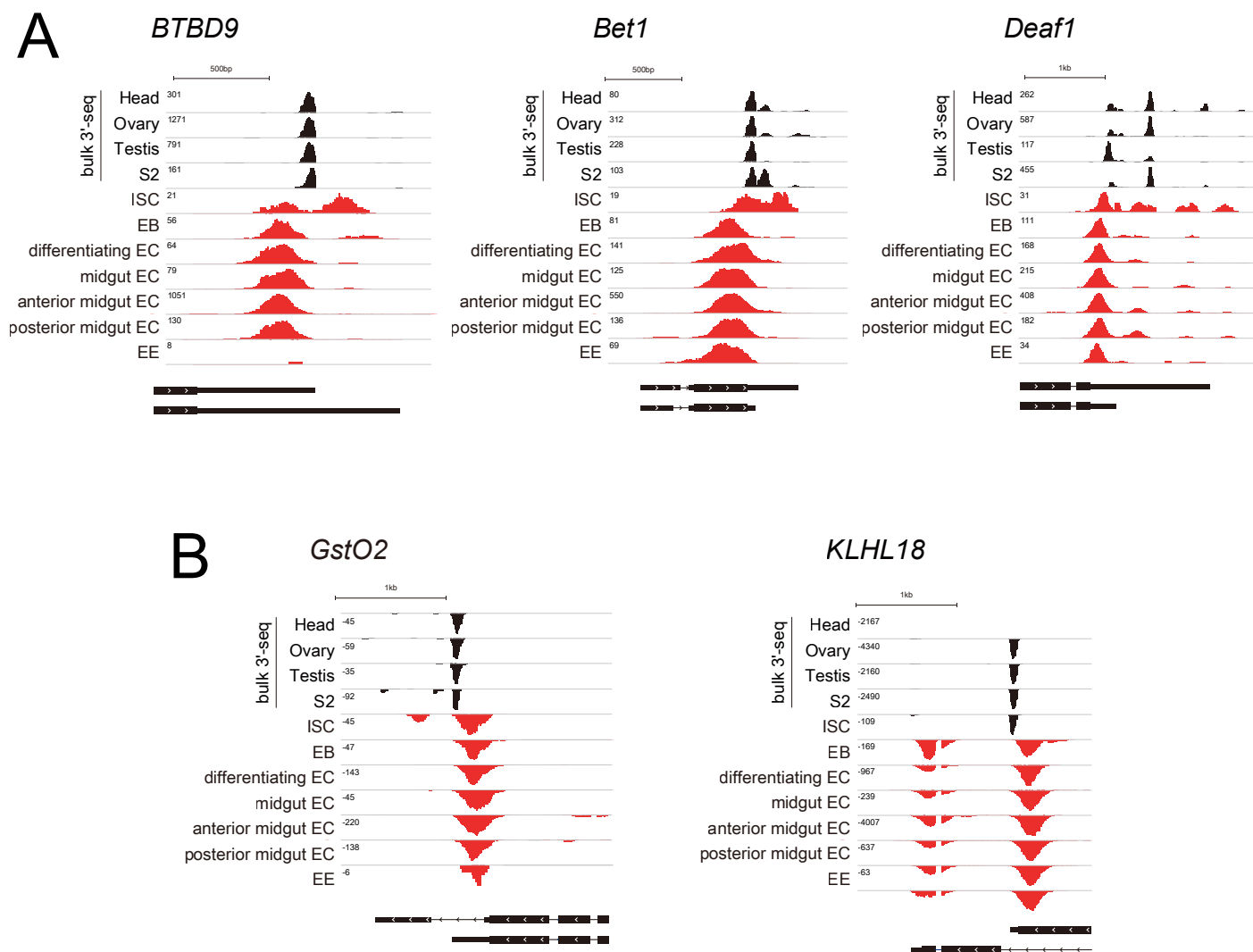
B



Supplementary Figure 4. Genes that preferentially express distal 3' isoforms in late spermatocytes.

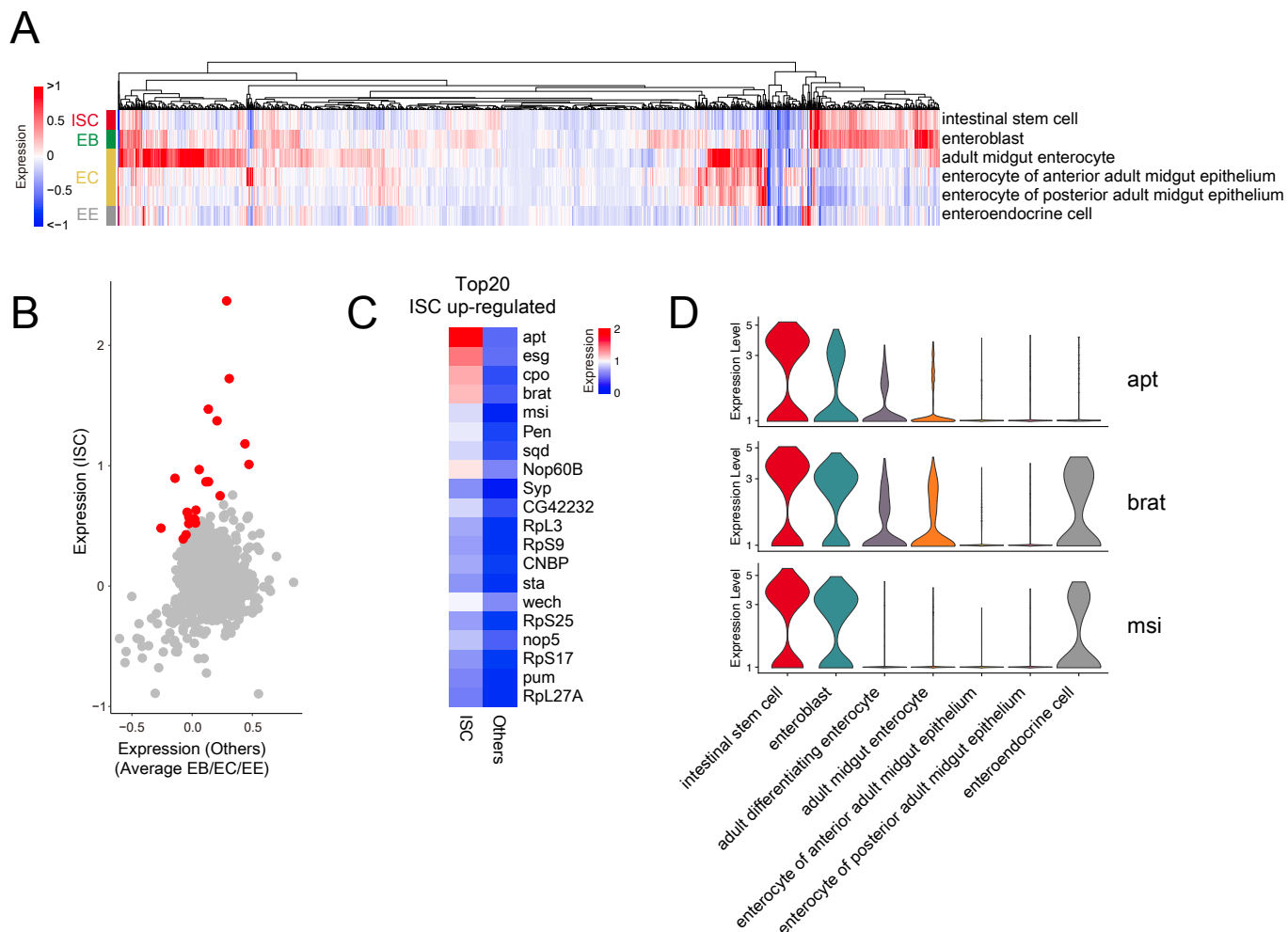
(A) Although we find global 3' shortening (both TUTR-APA and ALE-APA) across the testis germline lineage, a subcluster of three genes exhibits opposite proximal-to-distal ALE switching in late spermatocytes. (B) IGV browser tracks illustrating the distal 3' isoform shifts at *cg* in late spermatocytes. This opposite behavior indicates diverse mechanisms for 3' processing in the testis germline.

Supplementary Figure 4
Lee et al



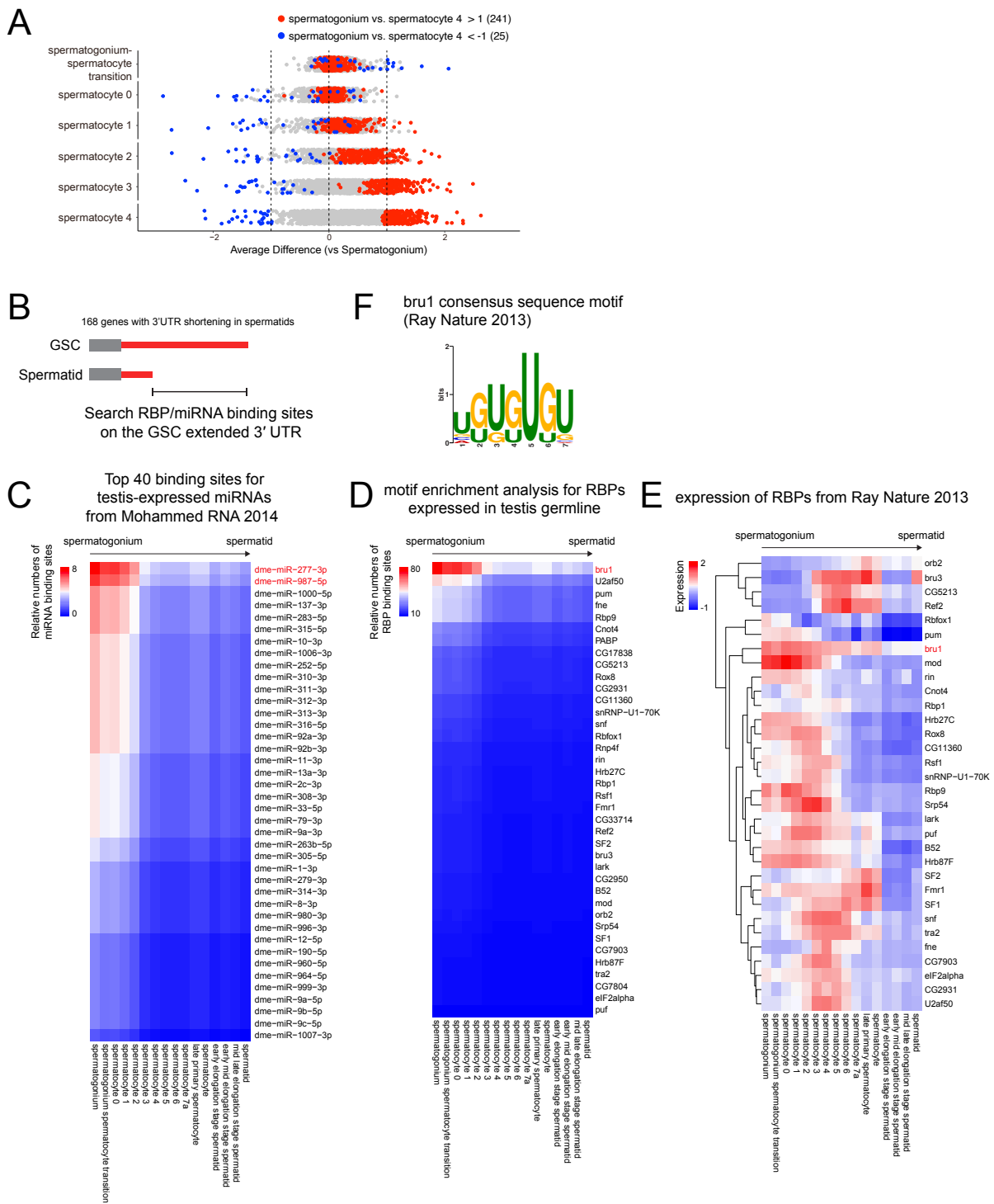
Supplementary Figure 5. Additional examples of 3' isoform shifts in the intestinal stem cell lineage.

This figure shows IGV browser tracks for bulk 3'-seq data from head/ovary/testis/S2 cells (in black, top tracks) and 10X-3' scRNA-seq data from individual cell types in the gut; gene models are shown at the bottom. Intestinal stem cell (ISC), entero-blast (EB), enterocyte (EC), enteroendocrine cell (EE). (A) Examples of switching from longer to shorter tandem 3' UTRs (TUTR) in the differentiated progeny of ISCs. (B) Examples of switching from distal to proximal alternative last exon (ALE) isoforms in the differentiated progeny of ISCs. All of these gut celltypes express some amount of both ALE isoforms, but ISCs express greater relative levels of the distal ALE isoforms of both genes, compared to all other gut celltypes.



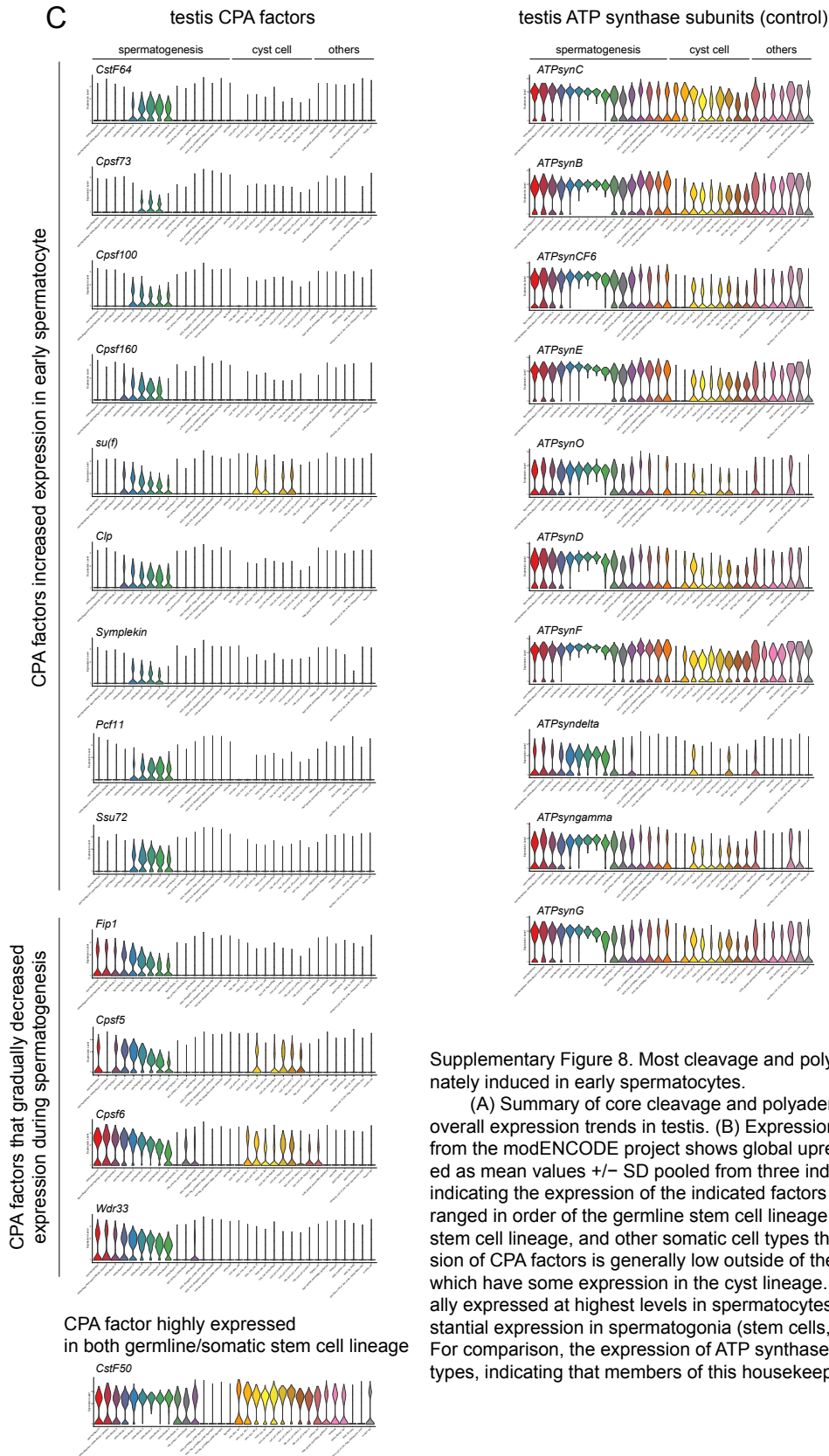
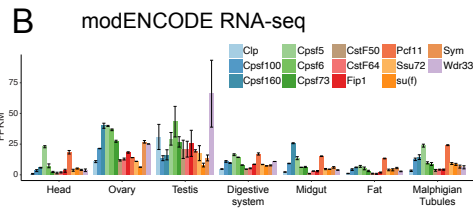
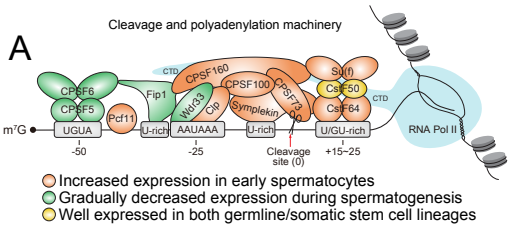
Supplementary Figure 6. RBPs that are differentially expressed in ISCs and their progeny.

(A) Expression of RBPs in the intestinal stem cell (ISC) lineage, which gives rise to enteroblasts (EBs) and enterocytes (ECs), as well as enteroendocrine (EE) cells. (B) RBPs that are upregulated in ISCs compared to other lineage cell types. (C) Heat map of the top 20 RBPs that are upregulated in ISCs. (D) Violin plots showing the expression of RBP candidates that are highest in ISCs and/or their immediate progeny (EBs).



Supplementary Figure 7. RBP and miRNA binding site enrichment analysis.

(A) Comparison of RBP expression in various early spermatocyte stages with the spermatogonium (germline stem cell) stage emphasizes a dominant trend of RBP induction. (B) Strategy to evaluate regulatory impacts of directional 3'UTR dynamics in the testis germline. (C) Conserved miRNA sites for testis-expressed miRNAs enriched within GSC-extended 3' UTRs, included miR-277 and mir-987 as top hits. (D) RBP sites within GSC-extended 3' UTRs shows bru1 as the predominant hit. (E) Expression of RBPs from (D) in testis germline cell types shows that bru1 is correspondingly enriched in GSCs and is lost during differentiation. (F) Reported Bru1 binding site.



Supplementary Figure 8. Most cleavage and polyadenylation (CPA) factors are coordinately induced in early spermatocytes.

(A) Summary of core cleavage and polyadenylation (CPA) machinery and their overall expression trends in testis. (B) Expression of CPA factors in bulk tissue datasets from the modENCODE project shows global upregulation in gonads. Data are presented as mean values \pm SD pooled from three independent experiments. (C) Violin plots indicating the expression of the indicated factors in FCA testis data. Celltypes are arranged in order of the germline stem cell lineage for spermatogenesis, a somatic cyst stem cell lineage, and other somatic cell types that comprise the testis. (Left) Expression of CPA factors is generally low outside of the germline, except for *Cpsf6* and *su(f)* which have some expression in the cyst lineage. Otherwise, all CPA factors are generally expressed at highest levels in spermatocytes, although some of them have substantial expression in spermatogonia (stem cells, i.e. *Cpsf6*, *Wdr33* and *Fip1*). (Right) For comparison, the expression of ATP synthase subunits is shown across testis cell types, indicating that members of this housekeeping complex can be reliably detected

Supplementary Table 1. RT-PCR primer sequences used in this study

Name	Sequence (5' -> 3')
mei-P26-Primer1-F	CGTGTGCCCTCCTTTAATGT
mei-P26-Primer1-R	CGCATTCTACATCCGTTTT
mei-P26-Primer2-F	ACTTAGGCCCCAAAGCAAAT
mei-P26-Primer2-R	TGCGAAAATGGAAAGGAAAG
mei-P26-Primer3-F	TCCGAGGGCTATGTGGTTAC
mei-P26-Primer3-R	CGACTGATGGACGATTGTTG
ssg-Primer1-F	CCTATGACAACCACCCAGT
ssg-Primer1-R	GGTTTTGCTTTTGCTTTTGC
ssg-Primer2-F	GAATGGAGAAAGCCAATCCA
ssg-Primer2-R	TTTTGAAATTGCGGTTGATG
ssg-Primer3-F	CGAGCATCGTTCCTCTATCC
ssg-Primer3-R	TGCTTCTTCTACGGCTTCCT
Oaf-short-F	TTCGCATCACCAGTCATCAT
Oaf-short-R	AGCGGAAAAAGTGTGAGCAT
Oaf-long-F	CATTTGGGAATGGTGAATG
Oaf-long-R	AAACAGCCATTTTCGAGTGG
Lig-short-F	GCACACCCGCCAATACATTG
Lig-short-R	ACATGCGGTTCAACAAATCA
Lig-long-F	CGCGAACGACTTTTATAGCC
Lig-long-R	CGCTGCTCTTGCTGTAGATG
Bbc-short-F	ATTCACCGTCGATGTTTTGG
Bbc-short-R	CTGTCTCCCAGCAGCCTTAG
Bbc-long-F	AGCGCACTACCTACCTTGGA
Bbc-long-R	CCGTGCGAGCTACAAACATA