Science Advances
AAAS

Supplementary Materials for

A developmentally programmed splicing failure contributes to DNA damage response attenuation during mammalian zygotic genome activation

Christopher D. R. Wyatt, Barbara Pernaute*, André Gohr, Marta Miret-Cuesta, Lucia Goyeneche, Quirze Rovira, Marion C. Salzer, Elvan Boke, Ozren Bogdanovic, Sophie Bonnal, Manuel Irimia*

*Corresponding author. Email: pernaute.barb@gmail.com (B.P.); mirimia@gmail.com (M.I.)

Published 13 April 2022, *Sci. Adv.* **8**, eabn4935 (2022)
DOI: 10.1126/sciadv.abn4935

The PDF file includes:

Figs. S1 to S23
Legends for tables S1 to S13
Legend for supplementary file S1

Other Supplementary Material for this manuscript includes the following:

Tables S1 to S13
Supplementary file S1

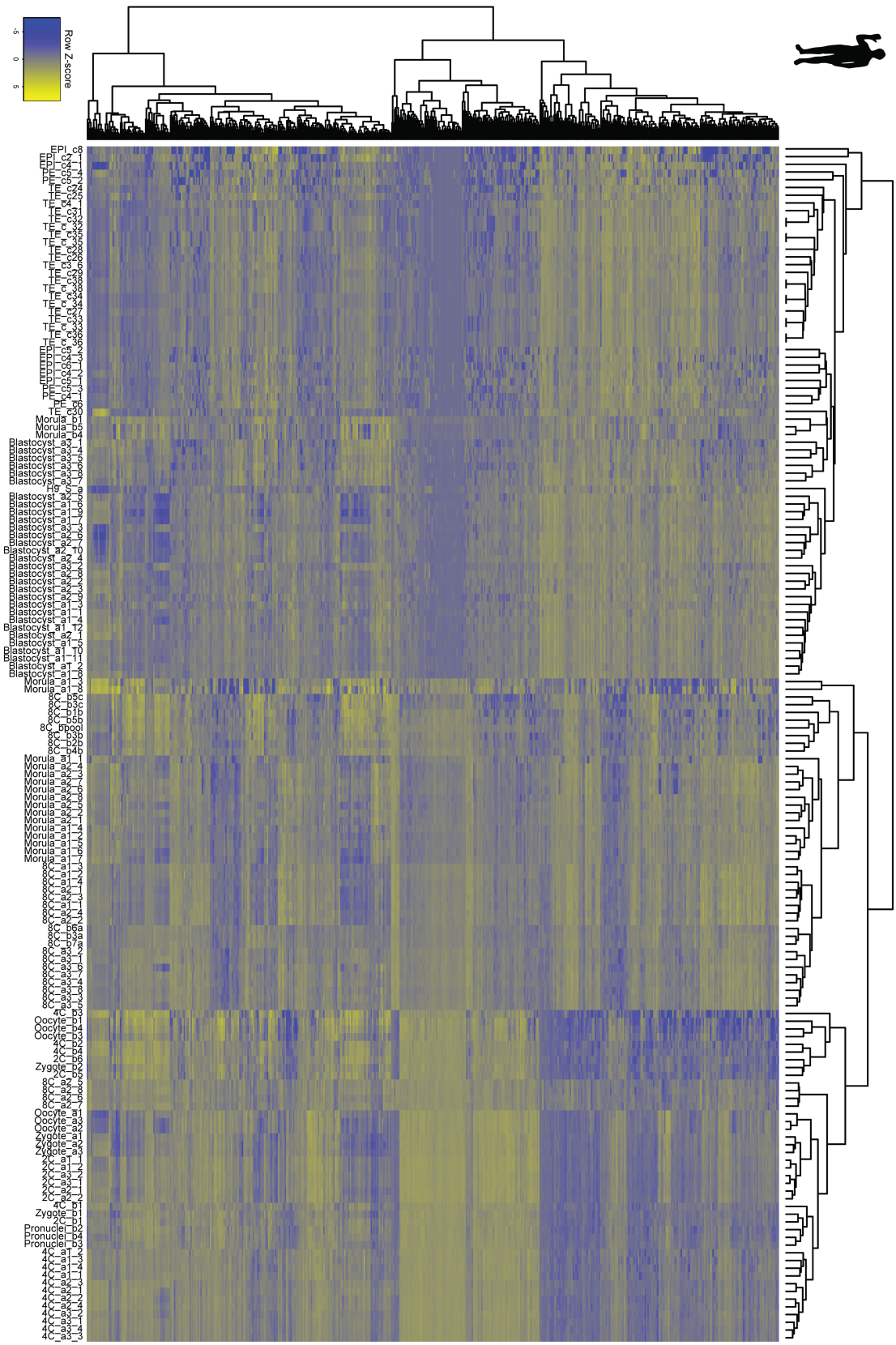
Row Z-score
-5
0
5

Fig. S1 - Hierarchical clustering of human early development samples. Heatmaps and unsupervised hierarchical clustering using default *heatmap.2* parameters of the top 500 most variably expressed genes as determined by *DESeq* for all studied human single cell or embryo pool RNA-seq samples. These clustering outputs were used to determine cell identity and locate anomalous samples (see Table S1 for details).

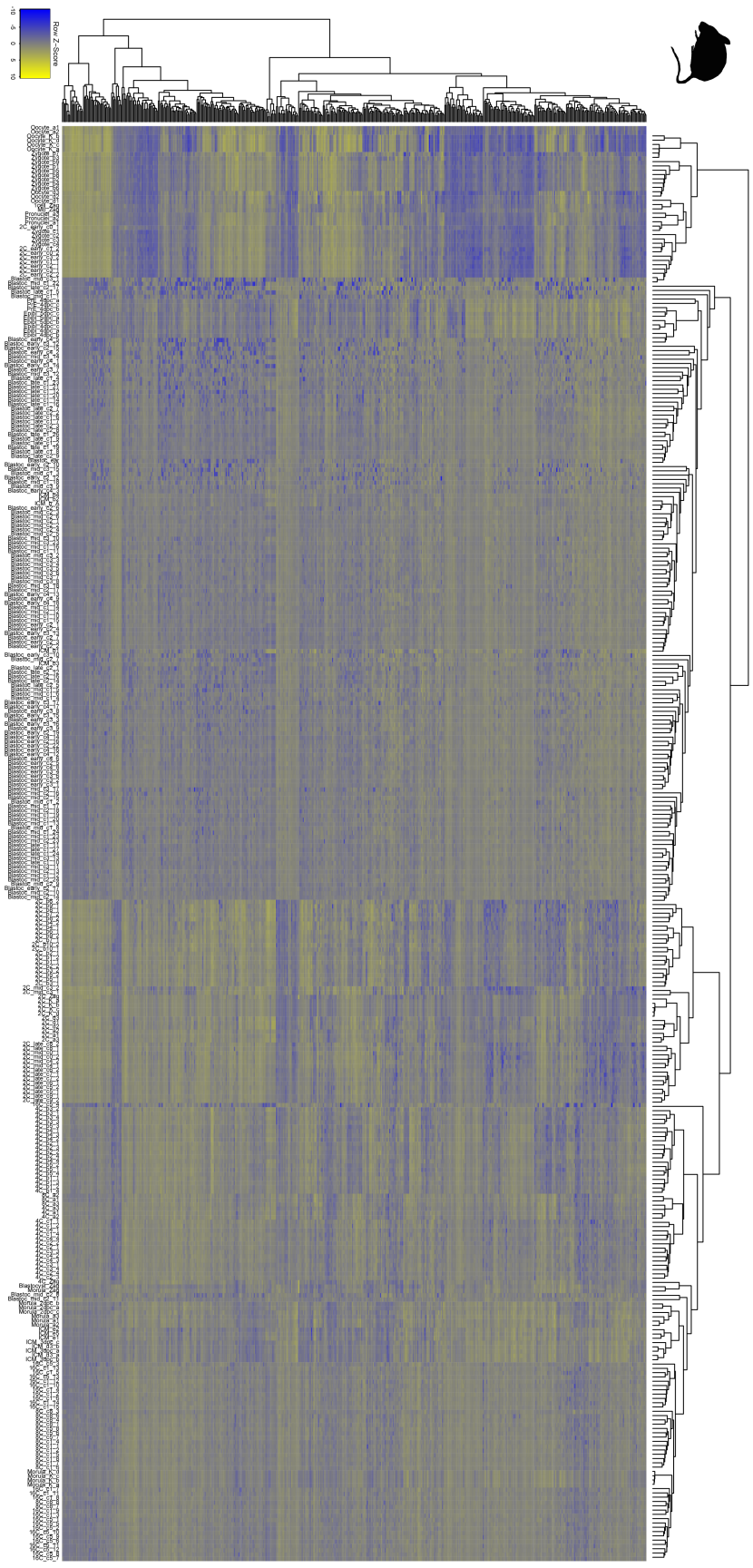
Row Z-Score

Fig. S2 - Hierarchical clustering of mouse early development samples. Heatmaps and unsupervised hierarchical clustering using default *heatmap.2* parameters of the top 500 most variably expressed genes as determined by *DESeq* for all studied mouse single cell or embryo pool RNA-seq samples. These clustering outputs were used to determine cell identity and locate anomalous samples (see Table S1 for details).

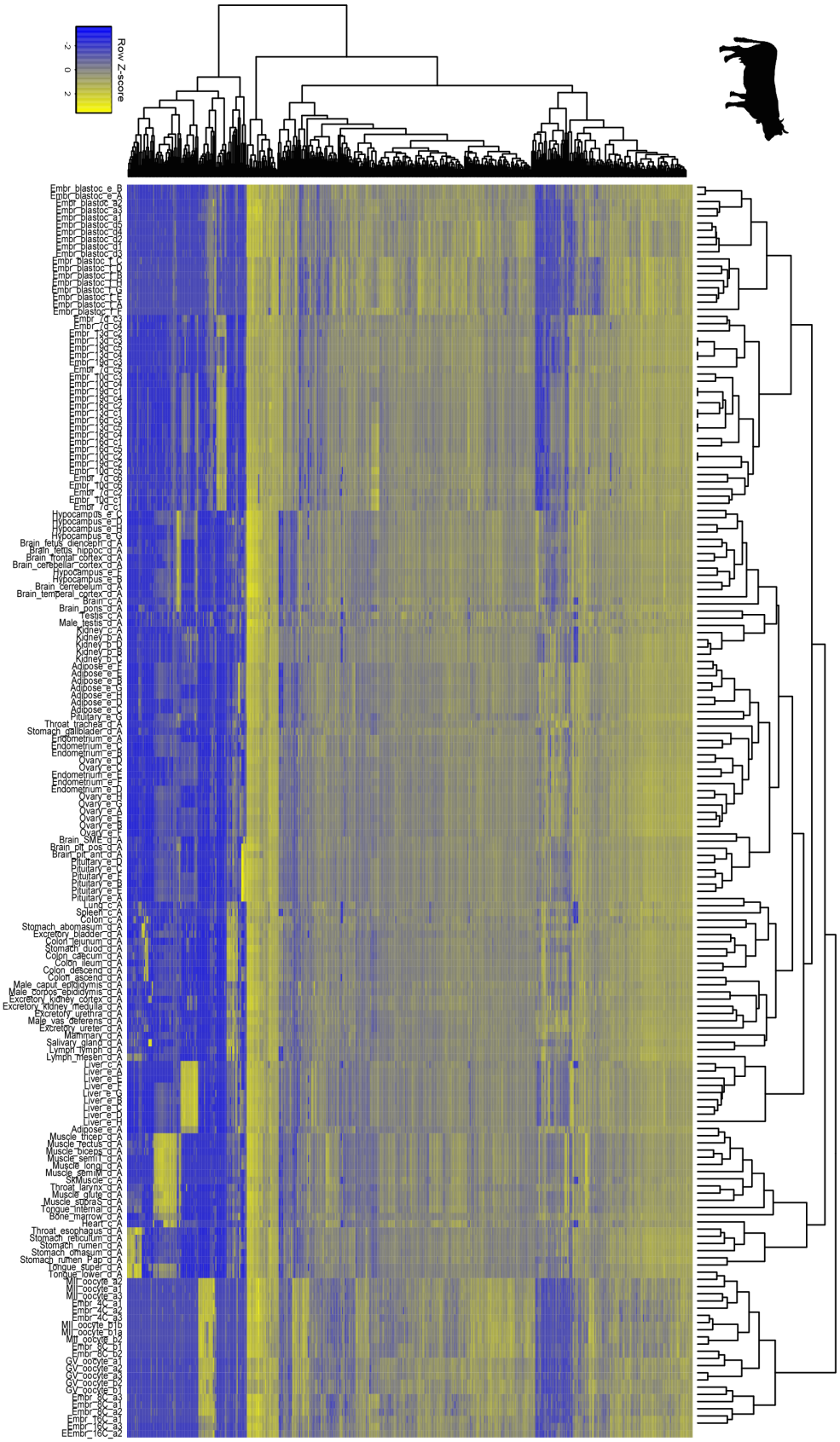
Row Z-score
-2
0
2

Fig. S3 - Hierarchical clustering of cow early development samples. Heatmaps and unsupervised hierarchical clustering using default *heatmap.2* parameters of the top 500 most variably expressed genes as determined by *DESeq* for all studied cow embryo and tissue RNA-seq samples. These clustering outputs were used to determine cell identity and locate anomalous samples (see Table S1 for details).

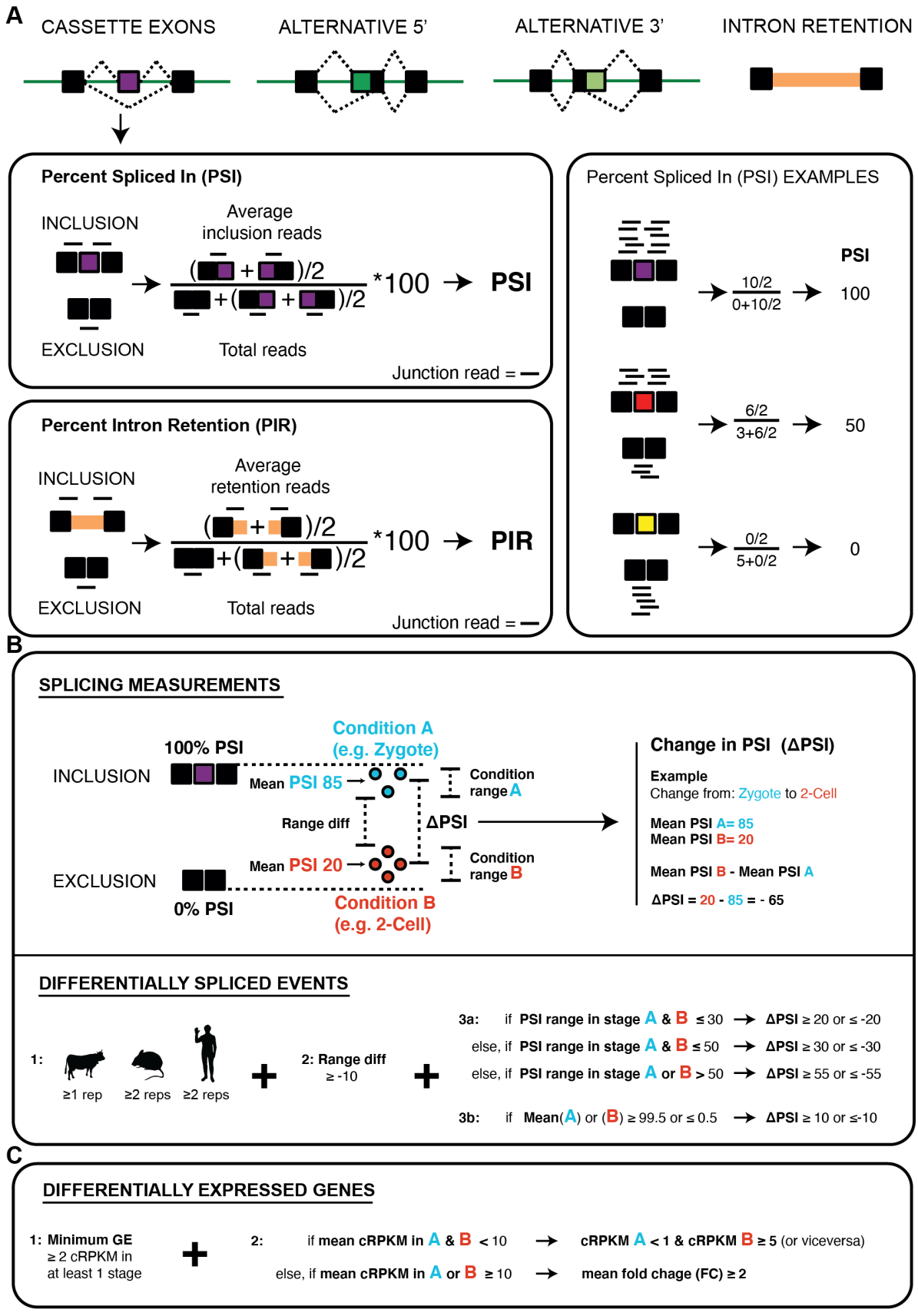
A
CASSETTE EXONS
ALTERNATIVE 5'
ALTERNATIVE 3'
INTRON RETENTION
Percent Spliced In (PSI)
INCLUSION
Average inclusion reads
*100
PSI
EXCLUSION
Total reads
Junction read =
Percent Spliced In (PSI) EXAMPLES
PSI
10/2
0+10/2
100
6/2
3+6/2
50
0/2
5+0/2
0
Percent Intron Retention (PIR)
INCLUSION
Average retention reads
*100
PIR
EXCLUSION
Total reads
Junction read =
B
SPLICING MEASUREMENTS
100% PSI
Condition A (e.g. Zygote)
INCLUSION
Mean PSI 85
Condition range A
Range diff
ΔPSI
Mean PSI 20
Condition range B
EXCLUSION
0% PSI
Condition B (e.g. 2-Cell)
Change in PSI (ΔPSI)
Example
Change from: Zygote to 2-Cell
Mean PSI A= 85
Mean PSI B= 20
Mean PSI B - Mean PSI A
ΔPSI = 20 - 85 = - 65
DIFFERENTIALLY SPLICED EVENTS
1:
≥1 rep
≥2 reps
≥2 reps
2: Range diff ≥ -10
3a: if PSI range in stage A & B ≤ 30 → ΔPSI ≥ 20 or ≤ -20
else, if PSI range in stage A & B ≤ 50 → ΔPSI ≥ 30 or ≤ -30
else, if PSI range in stage A or B > 50 → ΔPSI ≥ 55 or ≤ -55
3b: if Mean(A) or (B) ≥ 99.5 or ≤ 0.5 → ΔPSI ≥ 10 or ≤-10
C
DIFFERENTIALLY EXPRESSED GENES
1: Minimum GE ≥ 2 cRPKM in at least 1 stage
2: if mean cRPKM in A & B < 10 → cRPKM A < 1 & cRPKM B ≥ 5 (or viceversa)
else, if mean cRPKM in A or B ≥ 10 → mean fold chage (FC) ≥ 2

Fig. S4 - Schematic summary of AS quantification and differential AS and GE analyses. (**A**) Simplified schematic representation of how Percent Spliced In (PSI) and Percent Intron Retention (PIR) are calculated in *vast-tools* using solely RNA-Seq reads mapping to exon-exon (or exon-intron) junctions. See ref. 18 for further details. (**B**) Top: schematic representation of the main descriptive statistics considered for each stage (mean PSI and range of PSI of samples with sufficient read coverage) and the parameters used to call differential AS between two conditions (change in mean PSI between the two conditions [ΔPSI] and difference between the minimum and the maximum PSI in the two conditions [Range diff]). Bottom: definition of differentially spliced events between two stages. 1: minimum number of required samples with sufficient read coverage per stage, depending on the species (human: 2, mouse: 2, cow: 1). 2: Range diff ≥ -10. Either of two conditions, 3a: a minimum mean |ΔPSI| depending on the PSI range in each stage, or 3b: if mean PSI in either stage is very high (PSI ≥ 99.5) or very low (PSI ≤ 0.5), then a |ΔPSI| ≥ 10 is required. (**C**) Schematic representation of the definition of differentially expressed genes between two stages based on mean GE levels at each stage and mean fold change in GE.

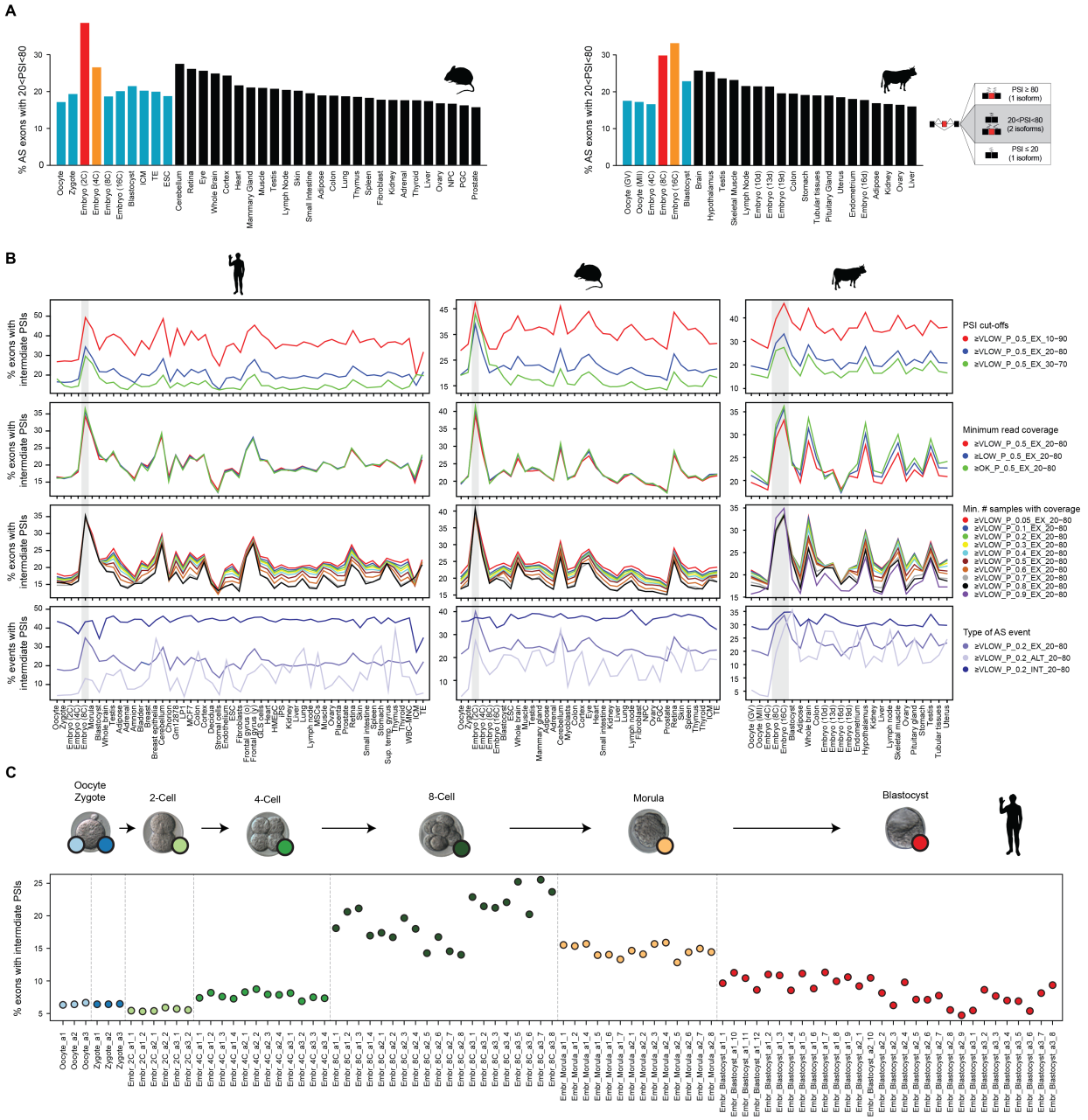

Fig. S5 - Relative exon skipping complexity is maximal at ZGA. (**A**) Percentage of exons with 20<PSI<80 (i.e. generating two substantial isoforms) in each sample was calculated for every stage and differentiated tissue, showing the highest relative exon skipping levels for the ZGA stage for mouse (left) and cow (right). (**B**) Percentage of splicing events with intermediate PSI levels for each species assessing the impact of different variables and cut-offs. From top to bottom: (i) different PSI cut-offs: 10-90, 20-80 (as in (a)) or 30-70. (ii) Minimum read coverage, as defined by *vast-tools* quality scores (≥VLOW, ≥LOW or ≥ OK). (iii) Minimum number of samples with coverage across the panel, from a minimum fraction of 0.05 to 0.9. (iv) Type of splicing event, including exons (EX), introns (INT) or alternative 3′ and 5′ splice sites (ALT). (**C**) Percentage of exons with 20<PSI<80 in individual blastomeres.

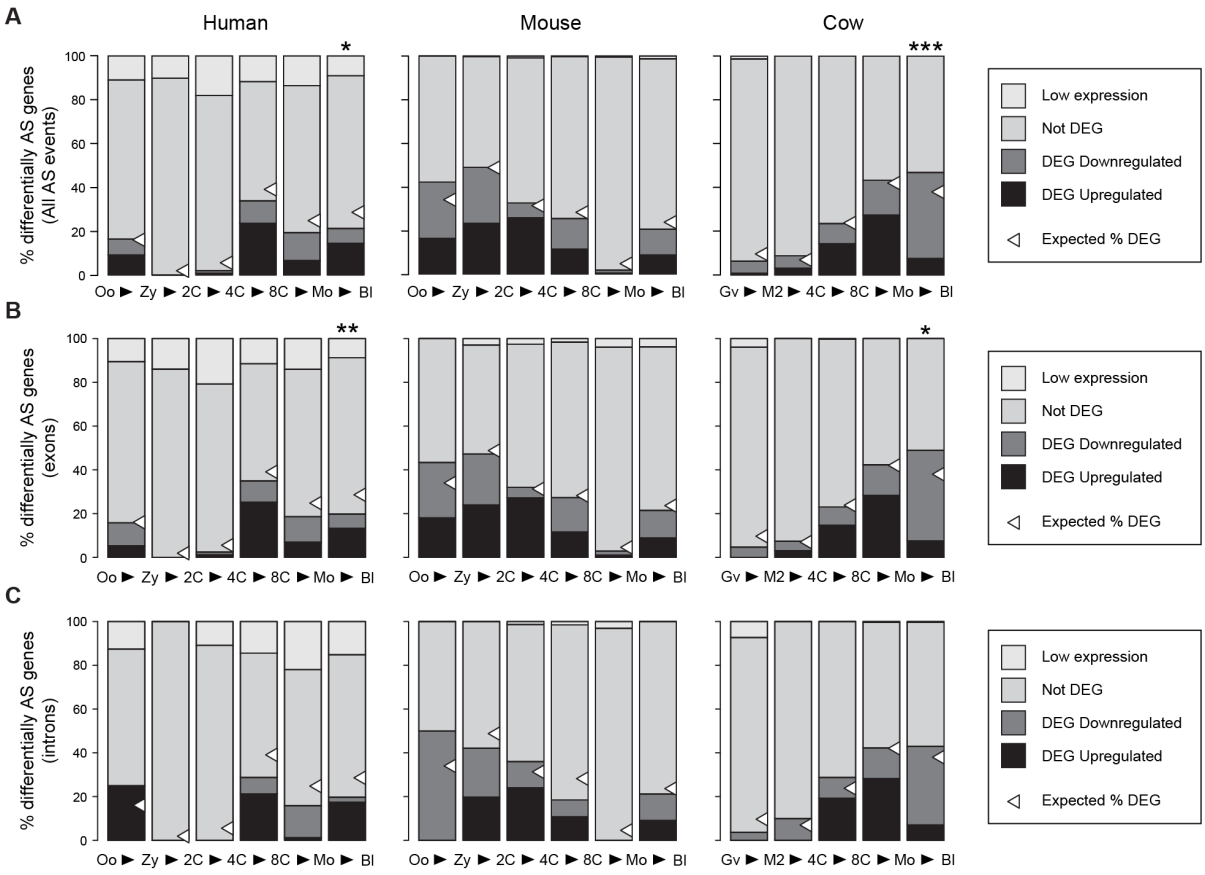

Fig. S6 - Overlap between genes regulated at the GE and AS levels. (**A-C**) Percentage of genes with differentially spliced AS events (A), exons (B) or introns (C) at each transition that also showed differential GE in the same transition. White triangles correspond to the percentage of differentially expressed genes (DEGs) expected by chance (i.e. the sum of the DEG Downregulated and DEG Upregulated categories); see Methods for details. * ($0.01 \leq P < 0.05$), ** ($0.001 \leq P < 0.01$) and *** ($P < 0.001$) indicate statistical significance (observed vs. expected) based on Proportion tests. "Low expression": genes with cRPKM < 2 in both stages of the transition and not tested for differential expression.

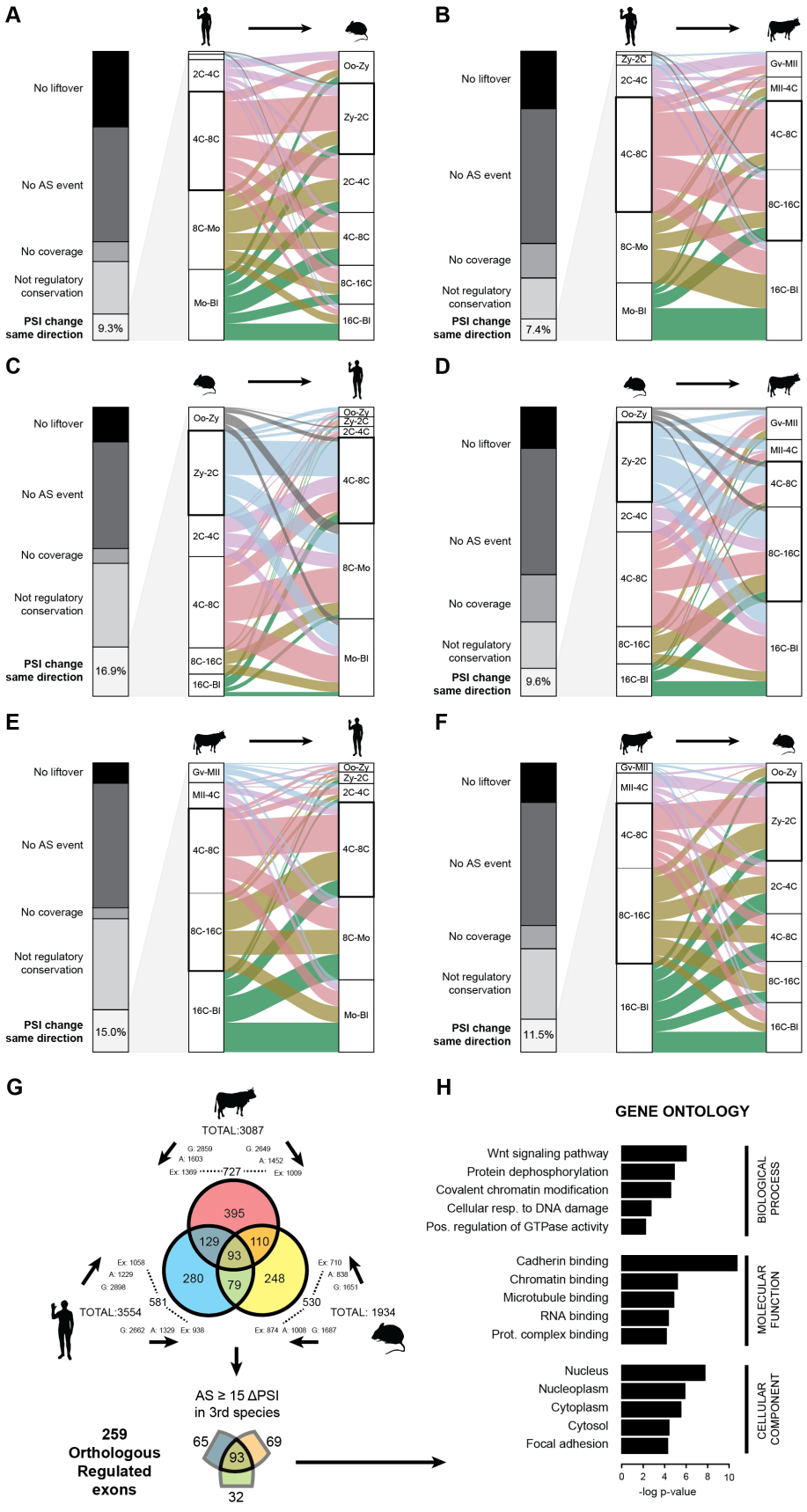
A
No liftover
No AS event
No coverage
Not regulatory conservation
PSI change same direction
9.3%
2C-4C
4C-8C
8C-Mo
Mo-Bl
Oo-Zy
Zy-2C
2C-4C
4C-8C
8C-16C
16C-Bl
B
No liftover
No AS event
No coverage
Not regulatory conservation
PSI change same direction
7.4%
Zy-2C
2C-4C
4C-8C
8C-Mo
Mo-Bl
Gv-MII
MII-4C
4C-8C
8C-16C
16C-Bl
C
No liftover
No AS event
No coverage
Not regulatory conservation
PSI change same direction
16.9%
Oo-Zy
Zy-2C
2C-4C
4C-8C
8C-16C
16C-Bl
Oo-Zy
Zy-2C
2C-4C
4C-8C
8C-Mo
Mo-Bl
D
No liftover
No AS event
No coverage
Not regulatory conservation
PSI change same direction
9.6%
Oo-Zy
Zy-2C
2C-4C
4C-8C
8C-16C
16C-Bl
Gv-MII
MII-4C
4C-8C
8C-16C
16C-Bl
E
No liftover
No AS event
No coverage
Not regulatory conservation
PSI change same direction
15.0%
Gv-MII
MII-4C
4C-8C
8C-16C
16C-Bl
Oo-Zy
Zy-2C
2C-4C
4C-8C
8C-Mo
Mo-Bl
F
No liftover
No AS event
No coverage
Not regulatory conservation
PSI change same direction
11.5%
Gv-MII
MII-4C
4C-8C
8C-16C
16C-Bl
Oo-Zy
Zy-2C
2C-4C
4C-8C
8C-16C
16C-Bl
G
TOTAL:3087
G: 2859
A: 1603
Ex: 1369
727
G: 2649
A: 1452
Ex: 1009
395
129
110
93
280
79
248
581
530
Ex: 1058
A: 1229
G: 2898
TOTAL:3554
G: 2662
A: 1329
Ex: 938
Ex: 710
A: 838
G: 1651
TOTAL: 1934
Ex: 874
A: 1008
G: 1687
AS ≥ 15 ΔPSI
in 3rd species
259
Orthologous
Regulated
exons
65
69
93
32
H
GENE ONTOLOGY
Wnt signaling pathway
Protein dephosphorylation
Covalent chromatin modification
Cellular resp. to DNA damage
Pos. regulation of GTPase activity
BIOLOGICAL PROCESS
Cadherin binding
Chromatin binding
Microtubule binding
RNA binding
Prot. complex binding
MOLECULAR FUNCTION
Nucleus
Nucleoplasm
Cytoplasm
Cytosol
Focal adhesion
CELLULAR COMPONENT
0
2
4
6
8
10
-log p-value

Fig. S7 - Evolutionary conservation analysis of alternative exon changes. (**A**-**F**) For each query and target species, stack plot of the level of conservation for each alternative exon changing at any transition in the query species. No liftover: absence of genomic conservation; No AS event: although the exon is found in the genome, no alternative exon is present in the *vast-tools* v1 annotation (i.e. likely a constitutive exon); No coverage: the exon does not have sufficient read coverage in at least the equivalent stages in the target species (see Methods); No regulatory conservation: the exon does not change in the same direction ($|\Delta PSI| \geq 15$) at any of the transitions in the target species. From the exons that change in the same direction in at least one transition of the target species, alluvial plot showing the transition of change in query species (left) and the observed highest change in the same direction in the target species (right). The transition overlapping with the major ZGA wave in each species is highlighted with a thicker block. (A) Human to mouse, (B) human to cow, (C) mouse to human, (D) mouse to cow, (E) cow to human, (F) cow to mouse. (**G**) For all exons differentially regulated in any pairwise stage comparison for each species (TOTAL), number of them with genome conservation [G], *vast-tools* AS identifier [A] and sufficient read coverage [Ex] in the target species (arrows). The intersect between Ex for both target species is shown at the edge of each circle of the Venn diagram (human: 581, mouse: 530, cow: 727). For these exons, the Venn diagram shows the overlap of exons that are differentially spliced in at least one pairwise stage comparison. For those exons regulated in two species, we then asked (bottom) whether the orthologous exon had a $|\Delta PSI| \geq 15$ in any pairwise stage comparison in the third species. This yielded a total of 259 orthologous exons with dynamic regulation in the three species. (**H**) GO term enrichment analysis for the genes harboring these 259 exons in human. Abbreviations: Oo, oocyte; Zy, zygote; Mo, morula; Bl, blastocyst; Gv, oocyte GV; MII, oocyte MII.

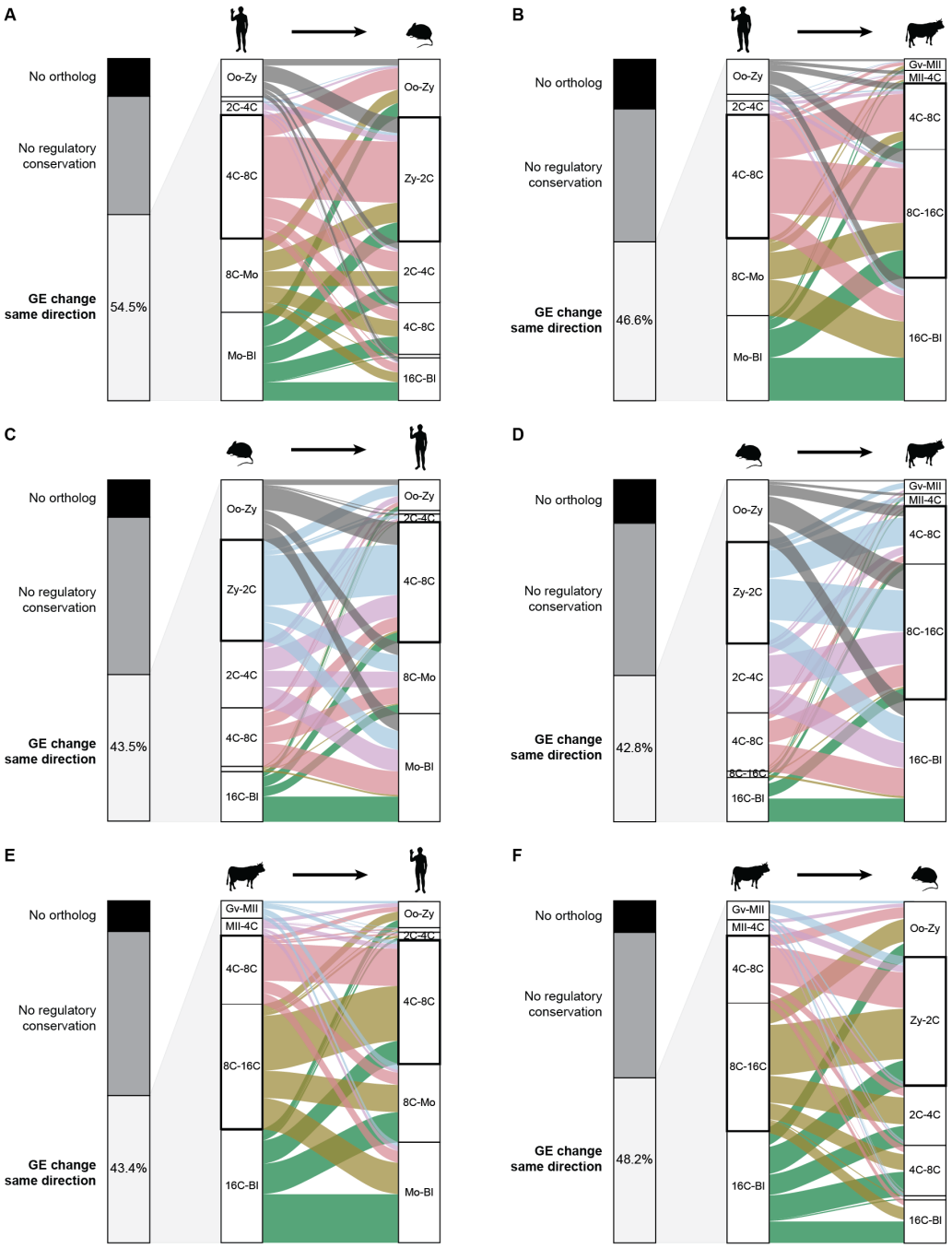

Fig. S8 - Pairwise evolutionary conservation of GE changes. (**A-F**) For each query and target species, stack plot of the level of conservation for each gene whose GE levels change at any transition in the query species. No ortholog: absence of one-to-one orthologs in the target species; No regulatory conservation: the orthologous gene does not significantly change in the same direction at any of the transitions in the target species. From the genes that change expression in the same direction in at least one transition of the target species, alluvial plot showing the transition of change in query species (left) and the observed highest change in the same direction in the target species (right). (A) Human to mouse, (B) human to cow, (C) mouse to human, (D) mouse to cow, (E) cow to human, (F) cow to mouse.

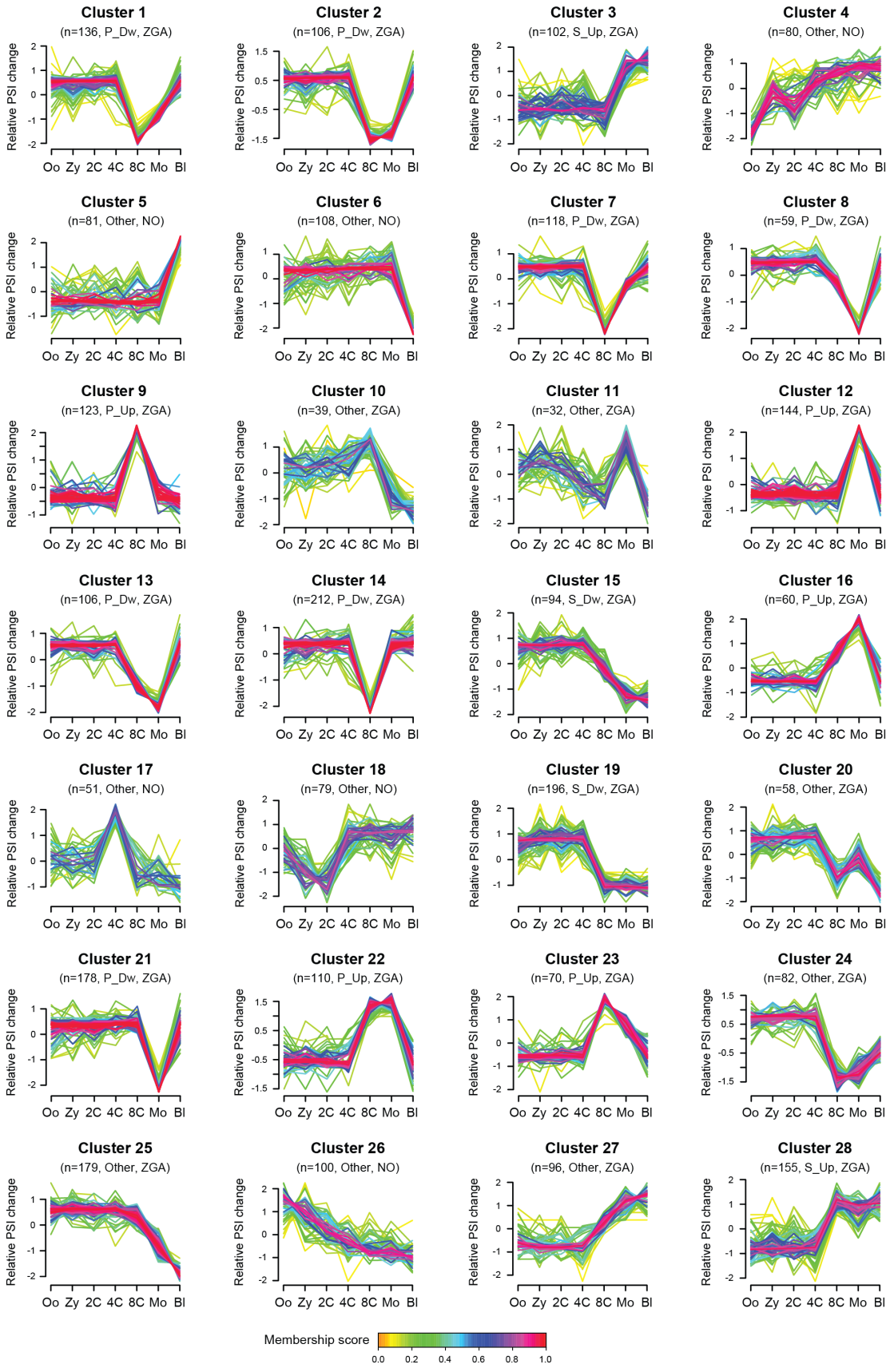
Cluster 1
(n=136, P_Dw, ZGA)
Cluster 2
(n=106, P_Dw, ZGA)
Cluster 3
(n=102, S_Up, ZGA)
Cluster 4
(n=80, Other, NO)
Cluster 5
(n=81, Other, NO)
Cluster 6
(n=108, Other, NO)
Cluster 7
(n=118, P_Dw, ZGA)
Cluster 8
(n=59, P_Dw, ZGA)
Cluster 9
(n=123, P_Up, ZGA)
Cluster 10
(n=39, Other, ZGA)
Cluster 11
(n=32, Other, ZGA)
Cluster 12
(n=144, P_Up, ZGA)
Cluster 13
(n=106, P_Dw, ZGA)
Cluster 14
(n=212, P_Dw, ZGA)
Cluster 15
(n=94, S_Dw, ZGA)
Cluster 16
(n=60, P_Up, ZGA)
Cluster 17
(n=51, Other, NO)
Cluster 18
(n=79, Other, NO)
Cluster 19
(n=196, S_Dw, ZGA)
Cluster 20
(n=58, Other, ZGA)
Cluster 21
(n=178, P_Dw, ZGA)
Cluster 22
(n=110, P_Up, ZGA)
Cluster 23
(n=70, P_Up, ZGA)
Cluster 24
(n=82, Other, ZGA)
Cluster 25
(n=179, Other, ZGA)
Cluster 26
(n=100, Other, NO)
Cluster 27
(n=96, Other, ZGA)
Cluster 28
(n=155, S_Up, ZGA)
Relative PSI change
Oo Zy 2C 4C 8C Mo Bl
Membership score
0.0 0.2 0.4 0.6 0.8 1.0

Fig. S9 - Mfuzz exon clusters in human. Profiles of scaled PSIs for each exon within each Mfuzz cluster in human. The color of each profile line corresponds to the "Mfuzz membership score" of the exon, i.e. how similar it is to the profile of the median values (from orange-yellow [low similarity] to purple-red [high similarity]). For each cluster, the number of exons (n), the most common pattern (Peak, Shift or Other; and its direction) and whether or not it is considered to change at ZGA. P_Dw, peak-down; P_Up, peak-up; S_Dw, shift-down; S_Up, shift-up.

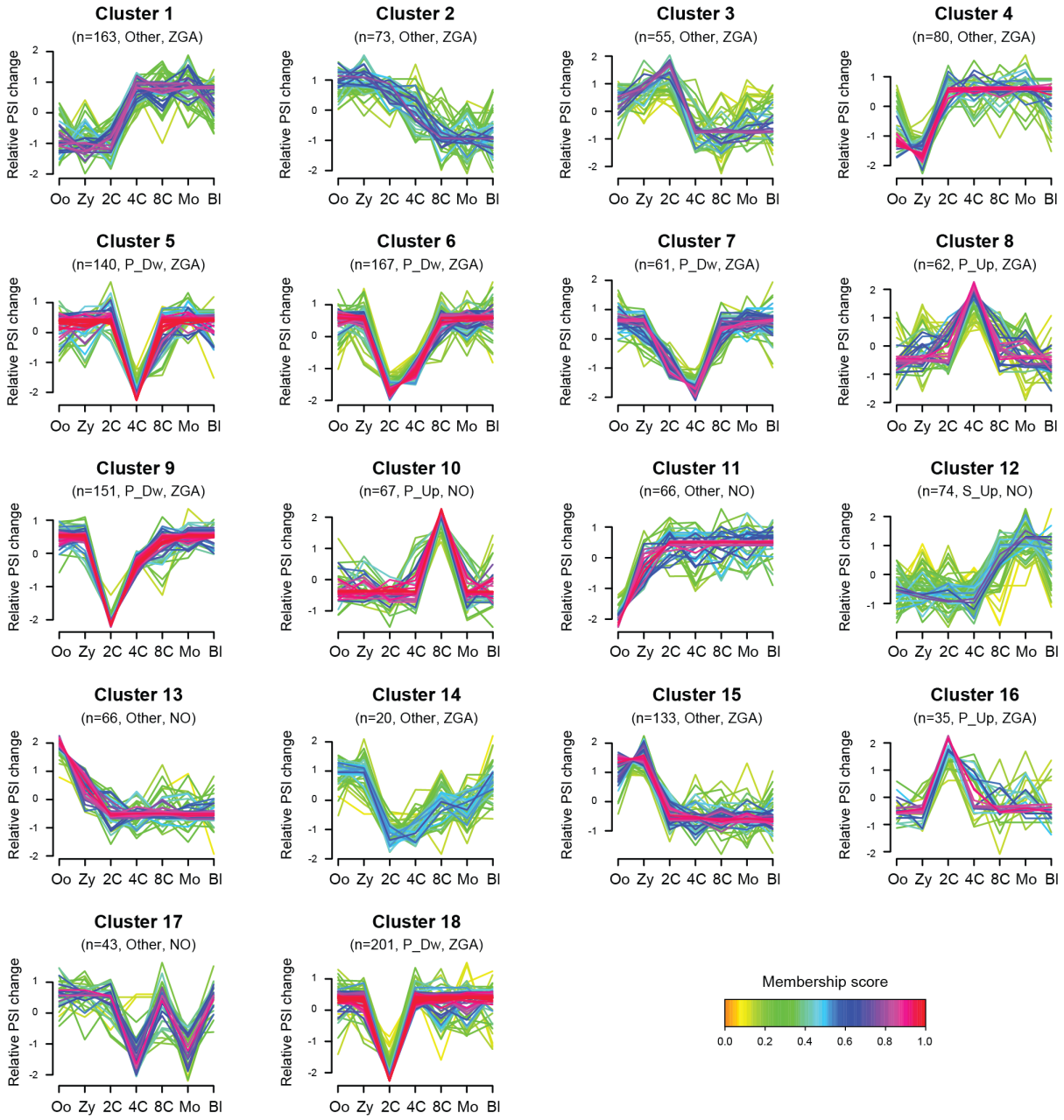

Fig. S10 - Mfuzz exon clusters in mouse. Profiles of scaled PSIs for each exon within each Mfuzz cluster in mouse. The color of each profile line corresponds to the "Mfuzz membership score" of the exon, i.e. how similar it is to the profile of the median values (from orange-yellow [low similarity] to purple-red [high similarity]). For each cluster, the number of exons (n), the most common pattern (Peak, Shift or Other; and its direction) and whether or not it is considered to change at ZGA. P_Dw, peak-down; P_Up, peak-up; S_Dw, shift-down; S_Up, shift-up.

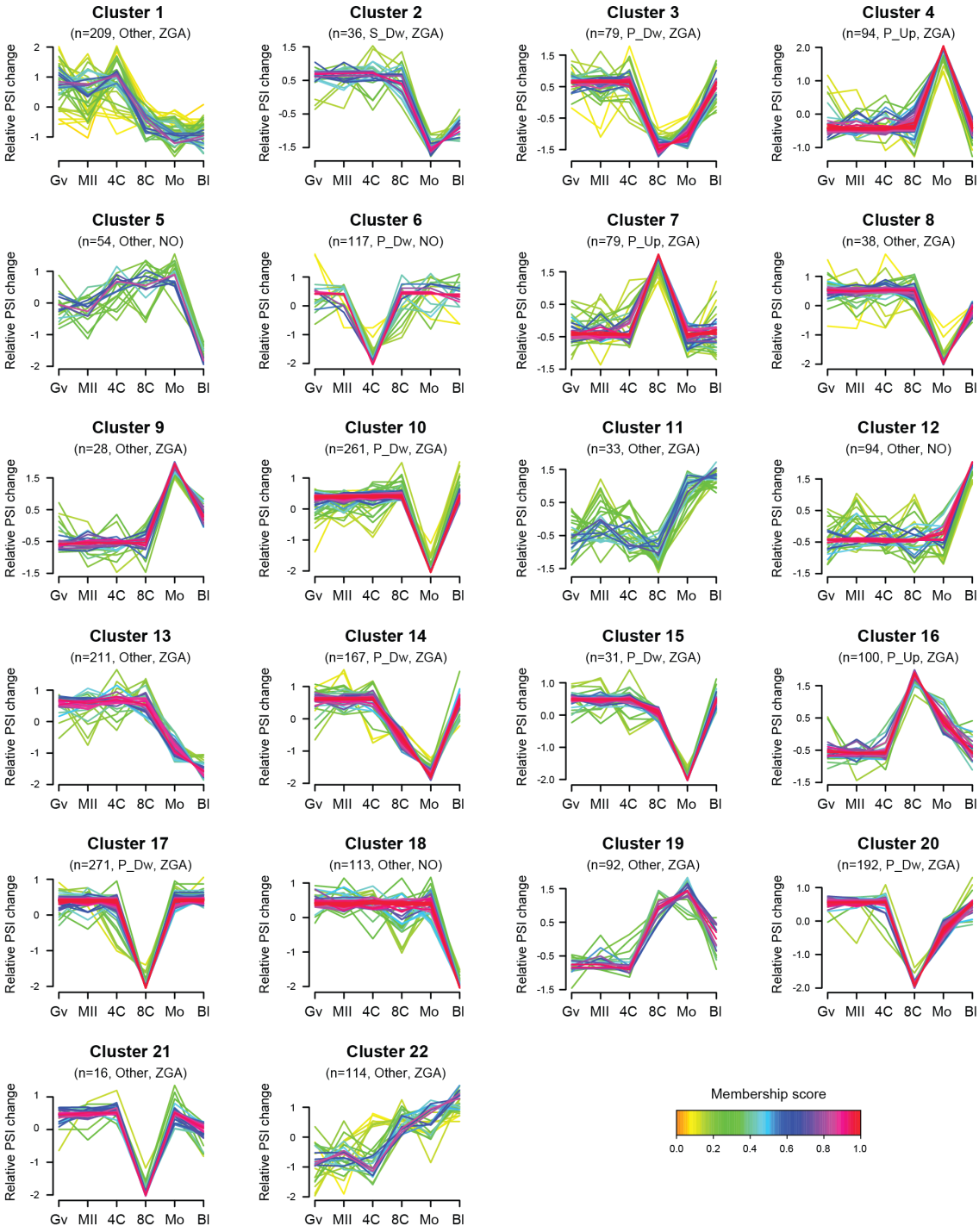

Fig. S11 - Mfuzz exon clusters in cow. Profiles of scaled PSIs for each exon within each Mfuzz cluster in cow. The color of each profile line corresponds to the "Mfuzz membership score" of the exon, i.e. how similar it is to the profile of the median values (from orange-yellow [low similarity] to purple-red [high similarity]). For each cluster, the number of exons (n), the most common pattern (Peak, Shift or Other; and its direction) and whether or not it is considered to change at ZGA. P_Dw, peak-down; P_Up, peak-up; S_Dw, shift-down; S_Up, shift-up.

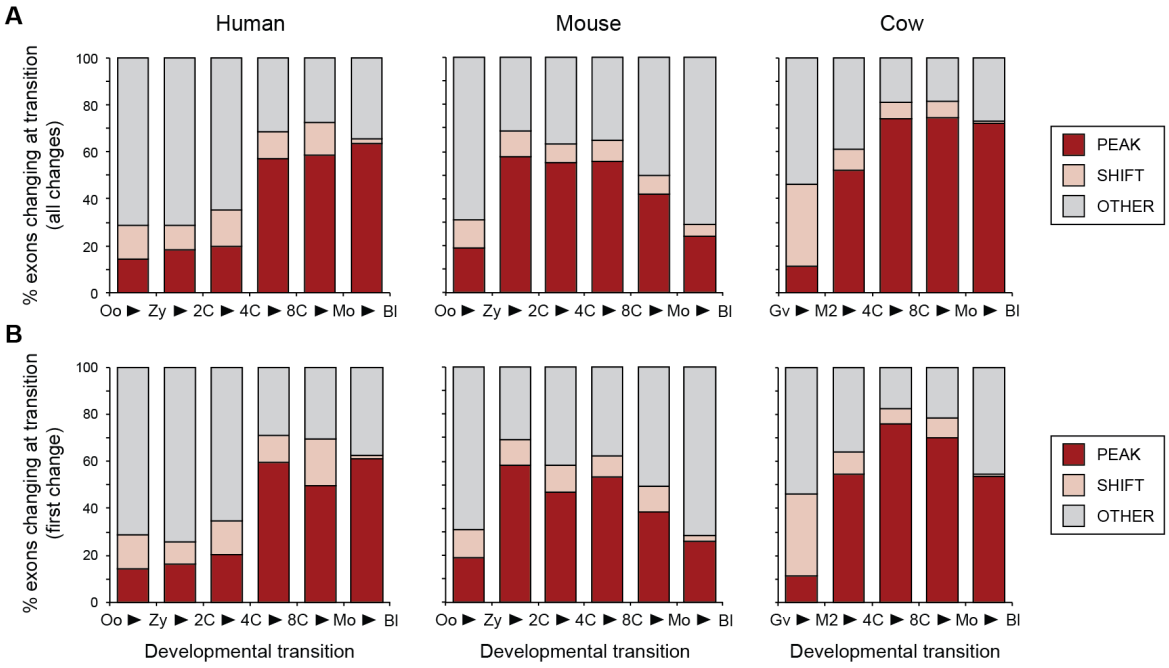

Fig. S12 - Proportion of exons by Mfuzz cluster type. (**A**,**B**) Proportions of each type of Mfuzz cluster (Peak, Shift or Other) for alternative exons that change at each transition, taken all transitions with a significant change into account (A) (as per Fig. 1E) or only considering the first transition at which a change is observed in the time course (B). Abbreviations: Oo, oocyte; Zy, zygote; Mo, morula; Bl, blastocyst; Gv, oocyte GV; MII, oocyte MII.

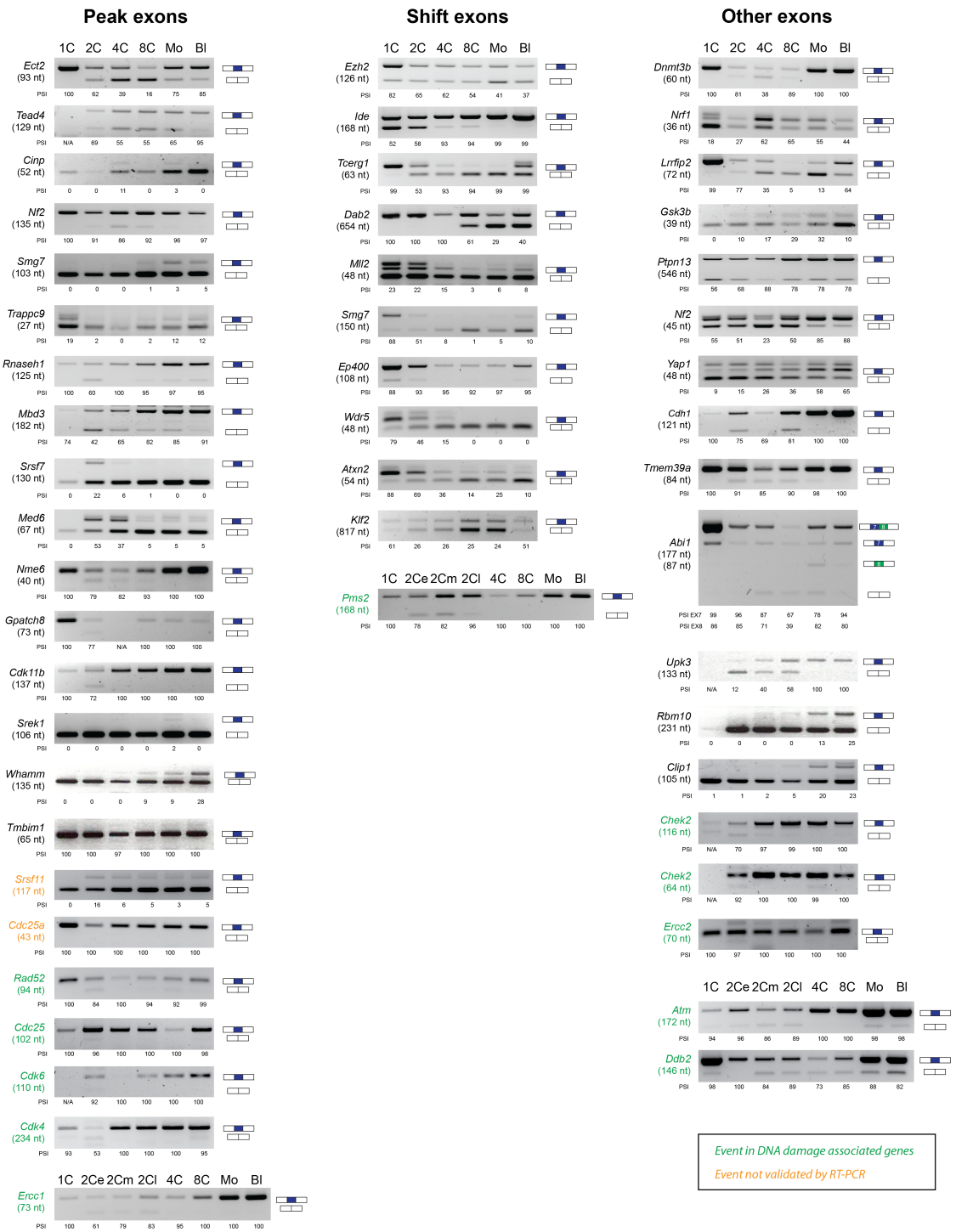

Fig. S13 - RT-PCR validation of exons during early development. RT-PCR assays were performed using independent RNA samples for each developmental stage to validate different types of exon profiles. Exon length in nts is indicated, as well as PSI quantifications based on relative band intensity (as measured with ImageJ). Exons in DDR genes are highlighted in green, and the two exons whose profiles were not validated, in orange. Note that variable amounts of cDNA template were used for different stages and PCR reactions; therefore changes in GE cannot be inferred from the variations in band intensity between stages.

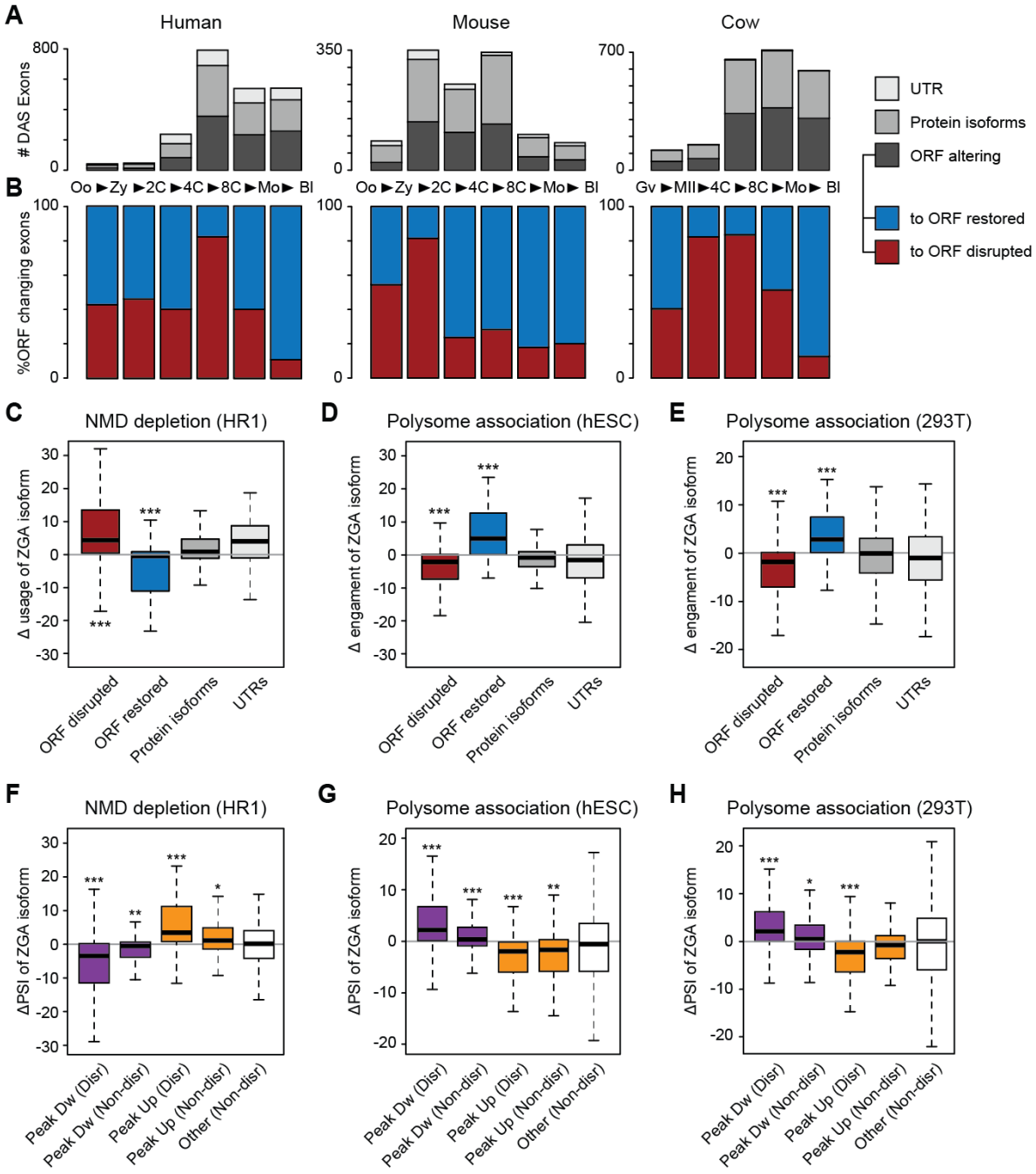

Fig. S14 - Differentially spliced exons at ZGA often generate non-functional proteins. (**A**) Number of differentially spliced exons at each transition by predicted impact on the ORF. "UTR", exon located in the 5′ or 3′ UTR; "Protein isoforms", both inclusion and exclusion isoforms are predicted to lead to functional protein isoforms; "ORF altering", either the sequence inclusion or exclusion lead to ORF disruption. Oo, oocyte; Zy, zygote; Mo, morula; Bl, blastocyst; Gv, oocyte GV; MII, oocyte MII. (**B**) For ORF altering exons, the percentage of those that disrupt (red) or restore (blue) the ORF specifically in that transition. (**C**) Change in usage of the ZGA isoform (whether inclusion or exclusion) upon NMD depletion in HR1 cells

based on the ORF impact predicted at the ZGA stage. (**D**-**E**) Change in ribosome engagement (PSI in high polysome fraction - PSI cytoplasmic fraction) of the ZGA isoform in embryonic stem cells (ESCs)(D) or HEK293 cells (E). (**F**) ΔPSI upon NMD depletion in HR1 cells of exons of different types of Mfuzz clusters changing at ZGA based on the impact on the ORF ("Disr", ORF disruptive; "Non-disr", non ORF disruptive). (**G**,**H**) Change in ribosome engagement in ESCs (G) or HEK293 cells (H) of different types of Mfuzz clusters changing at ZGA based on the impact on the ORF. * ($0.01 \leq P < 0.05$), ** ($0.001 \leq P < 0.01$) and *** ($0.001 < P$) indicate statistically significant differences respect to "Protein isoforms" (for c-e) or "Other (Non-disr)" (for F-G) based on Wilcoxon Sum-Rank tests. Data from E-MTAB-4008 [C,F], GSE100007 [D,G] and GSE69352 [E,H].

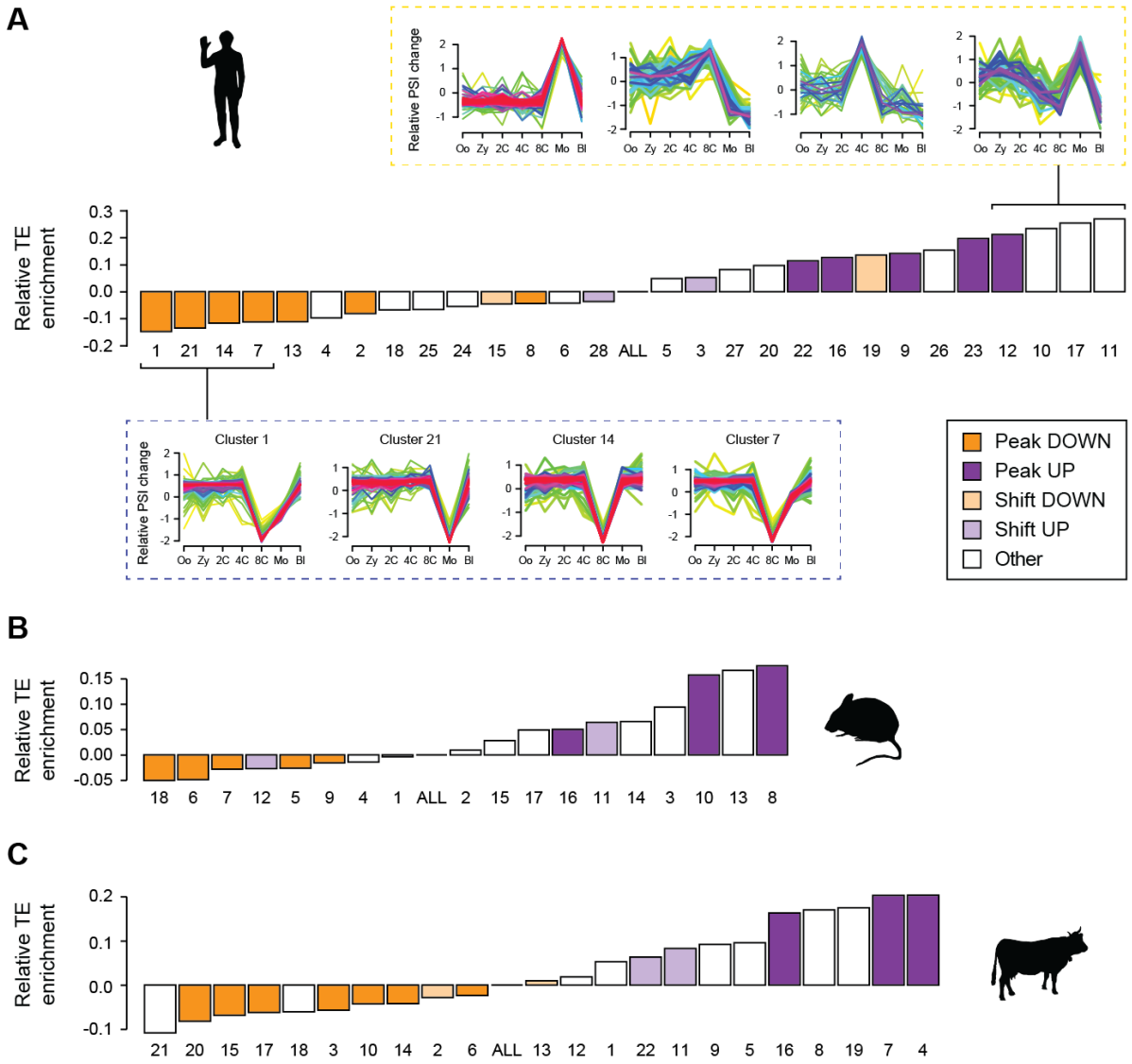

Fig. S15 - Overlap between exon types and transposable elements. (**A-C**) For each Mfuzz cluster for human (A), mouse (B) and cow (C), relative enrichment (respect to the average of all exons) of exons overlapping RepeatMasker repeats (excluding simple repeats) by at least one nucleotide. See Methods for further details. Peak, Shift or Other dynamics are indicated. Cluster IDs correspond to those in Supplementary figs. S9-S11.

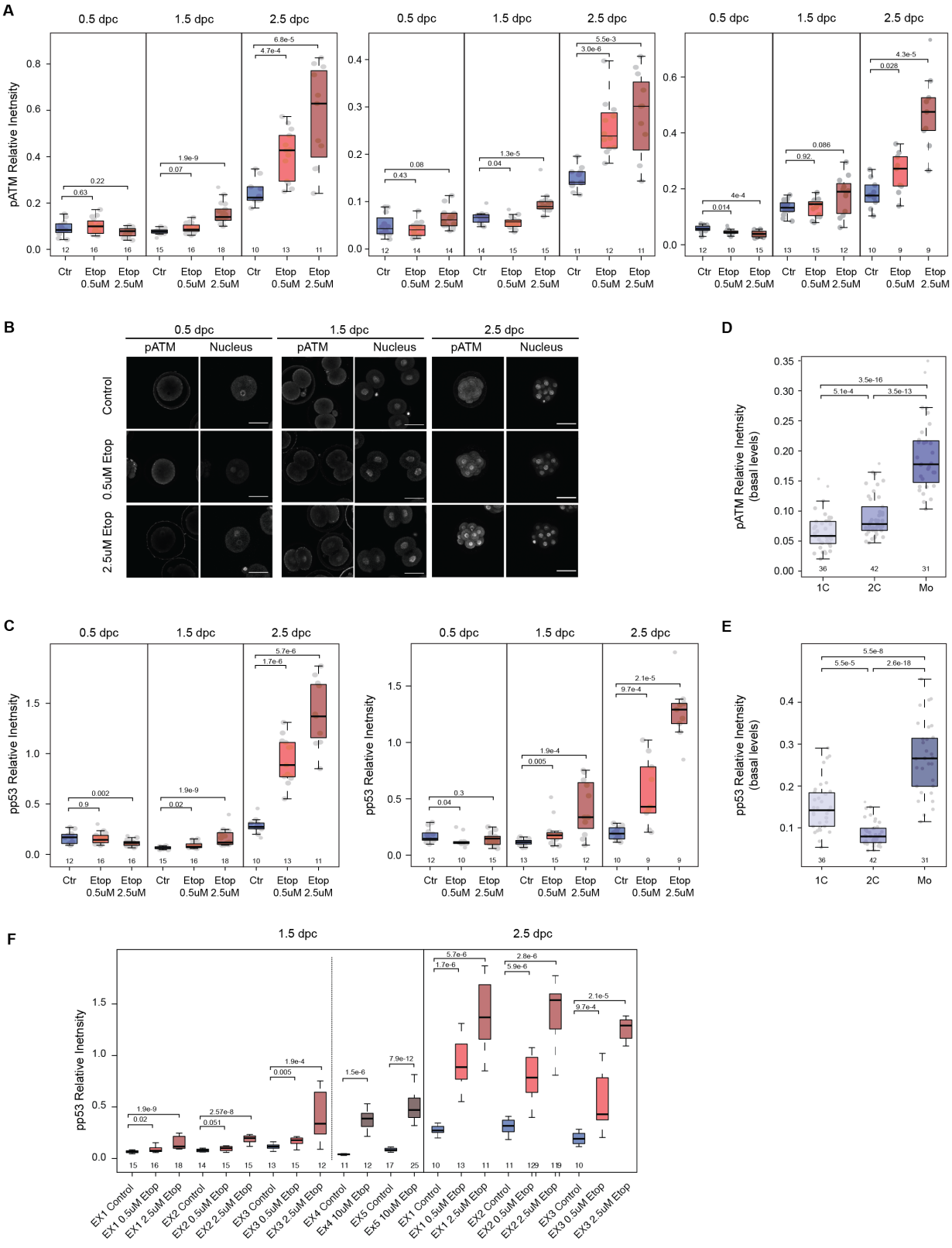
A
0.5 dpc
1.5 dpc
2.5 dpc
pATM Relative Inetnsity
Ctr
Etop 0.5uM
Etop 2.5uM
B
pATM
Nucleus
Control
0.5uM Etop
2.5uM Etop
C
pp53 Relative Inetnsity
D
pATM Relative Inetnsity (basal levels)
1C
2C
Mo
E
pp53 Relative Inetnsity (basal levels)
F
1.5 dpc
2.5 dpc
pp53 Relative Inetnsity
EX1 Control
EX1 0.5uM Etop
EX1 2.5uM Etop
EX2 Control
EX2 0.5uM Etop
EX2 2.5uM Etop
EX3 Control
EX3 0.5uM Etop
EX3 2.5uM Etop
EX4 Control
EX4 10uM Etop
EX5 Control
EX5 10uM Etop

Fig. S16 - Early pre-implantation mouse embryos show lower ATM pathway activation in response to etoposide (**A**) Three independent experiments showing quantification of nuclear phospho-ATM (Ser1981) levels from embryos treated for 1h with etoposide or left untreated for control is shown for each stage. (**B**) Phospho-ATM immunostaining in embryos treated at different developmental stages for 1h with etoposide or left untreated. (**C**) Second and third replicate of the experiment shown in Fig. 3B. Quantification of nuclear phospho-p53 (Ser15) levels from embryos treated for 1h with etoposide or left untreated for control is shown for each stage. (**D**,**E**) Data from three independent experiments showing quantification of nuclear phospho-ATM (D) or phospho-p53 (E) basal levels from embryos at different developmental stages. (**F**) Boxplot showing phopho-p53 quantification in 2C embryos treated for 1h with etoposide at a high concentration (10uM). All experiments performed at 1.5dpc and 2.5dpc are shown for comparison of the induction levels. For all boxplots, each dot represents the quantification of the average relative intensity detected by immunofluorescence of all cells of a single embryo. P-values from Wilcoxon Rank Sum tests are shown for each comparison. The number of embryos per condition is shown. Scale bar in (B) represents 50um.

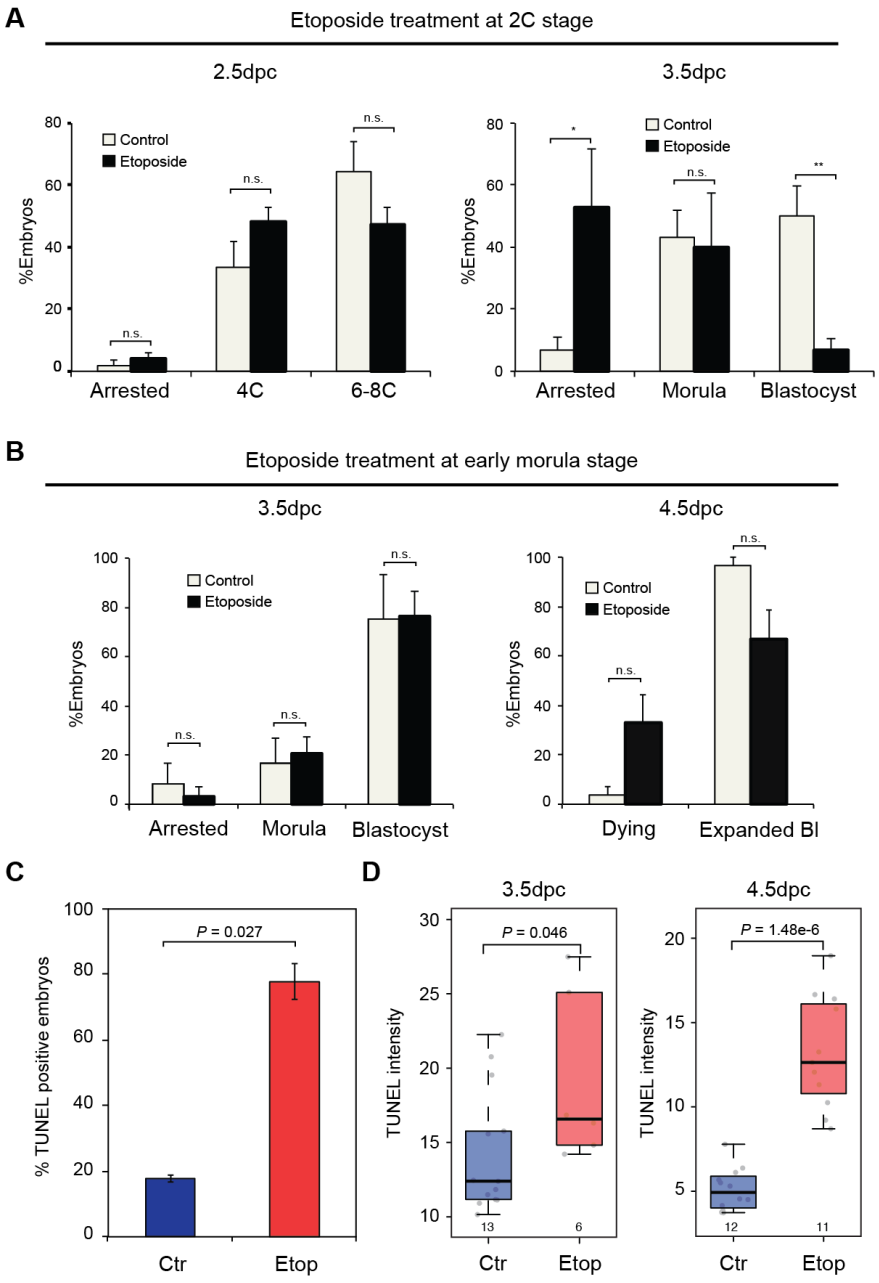

Fig. S17 - Etoposide treatment at ZGA results in a delayed but higher rate of embryo arrest than treatment at morula stage. (**A**) Bar plots corresponding to stack plots in Figs. 4B,C. Proportion of embryos found arrested or developed to the stated stage 24h or 48h post treatment with 0.5uM etoposide for 1h at 2C stage. The average of 6 independent experiments with a total number of embryos analysed of n=105 (control) and n=113 (treated) is shown. (**B**) Bar plots corresponding to stack plots in Figs. 4H,I. Proportion of embryos found arrested or developed to the stated stage 24h or 48h post 0.5uM etoposide treatment for 1h at early morula stage. Average of 2 independent experiments is shown, with a total number of embryos analysed of n=45 (control) and n=44 (treated). (**C**) Percentage of embryos showing TUNEL positive staining at 3.5dpc from the experiment in (B) and Fig. 4H. Average of two independent

experiments is shown with a total number of embryos of: n=22 (control), n=30 (etoposide). P-value corresponds to a proportion test with both experiments combined. Error bars show the range of the two experiments. (**D**) TUNEL staining intensity in morphologically normal blastocysts from Figs. 4H,I. Each dot represents TUNEL intensity in one embryo. Boxplots show staining levels of embryos from one experiment with a total number of embryos of: n=13 (control, 3.5dpc), n=6 (etoposide, 3.5dpc), n=12 (control, 4.5dpc) and n=11 (etoposide, 4.5dpc). P-values from Wilcoxon Rank Sum tests are shown. In (A) and (B) all embryos not reaching the expected developmental stage at a given time point either because of cell death or cell cycle arrest were classified as "Arrested". In (A), and (B) * $P<0.05$; ** $P<0.01$ two-sided Student's T-test. SEM is shown.

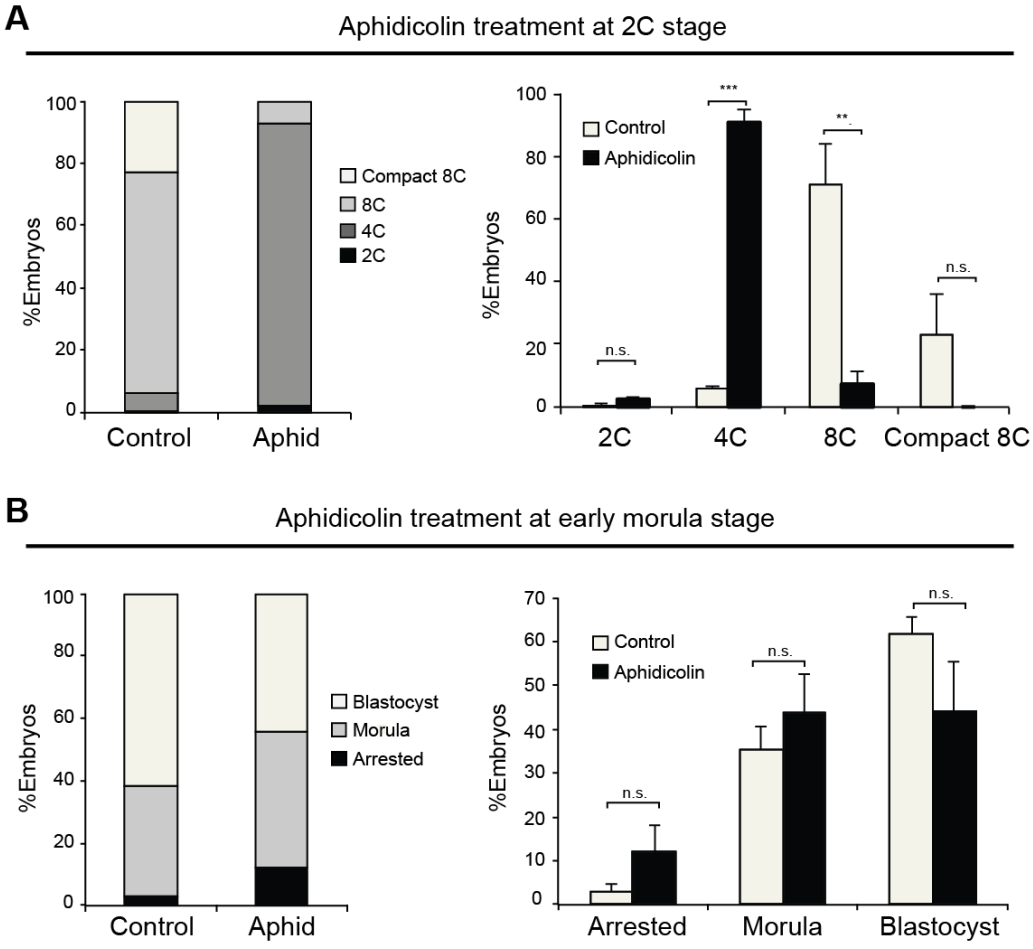

Fig. S18 - Mouse embryos have high sensitivity to aphidicolin at the time of ZGA. (**A**) Stack plot and its corresponding bar plot representing the proportion of 2C embryos reaching the stated developmental stage after 16h treatment with 0.25ug/ml aphidicolin followed by 8h recovery in KSOM media without drug. Average of 4 independent experiments is shown, with a total number of embryos analysed of n=133 (control) and n=133 (treated). $^{**}P<0.01$; $^{***}P<0.001$ two-sided Student's T-test. SEM is shown. (**B**) Stack plot and its corresponding bar plot representing the proportion of early morulas reaching the stated developmental stage after 16h treatment with 0.25ug/ml aphidicolin followed by 8h recovery in KSOM media without drug. Average of 4 independent experiments is shown, with a total number of embryos analysed of n=109 (control) and n=111 (treated). Two-sided Student's T-test. SEM is shown.

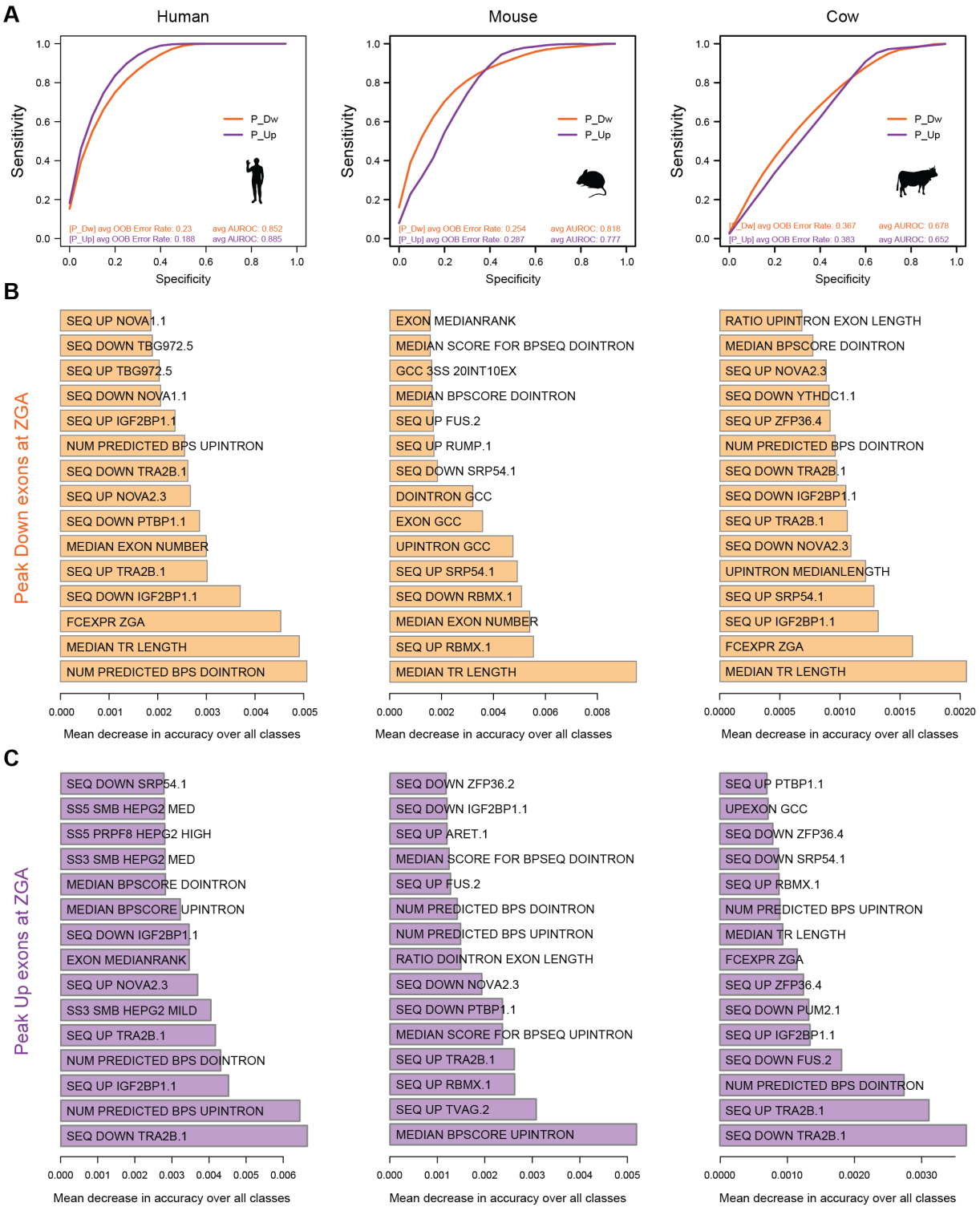

Fig. S19 - Random Forest classification of exons with peak dynamics at ZGA. (**A**) Receiver operating characteristic curve (ROC) for the Random Forest classification of exons with peak-down (orange, P_Dw) or peak-up (purple, P_Up) versus a matched set of background exons with similar pre-ZGA PSI distributions for each species. Average OOB error rate and area under the

ROC are indicated for each set for 1000 random Random Forest models. (**B**,**C**) Most discriminative features for peak-down (B) or peak-up (C) exons for each species quantified as mean decrease in accuracy. Precise definitions for each feature can be obtained on https://gitlab.com/aghr/matt.

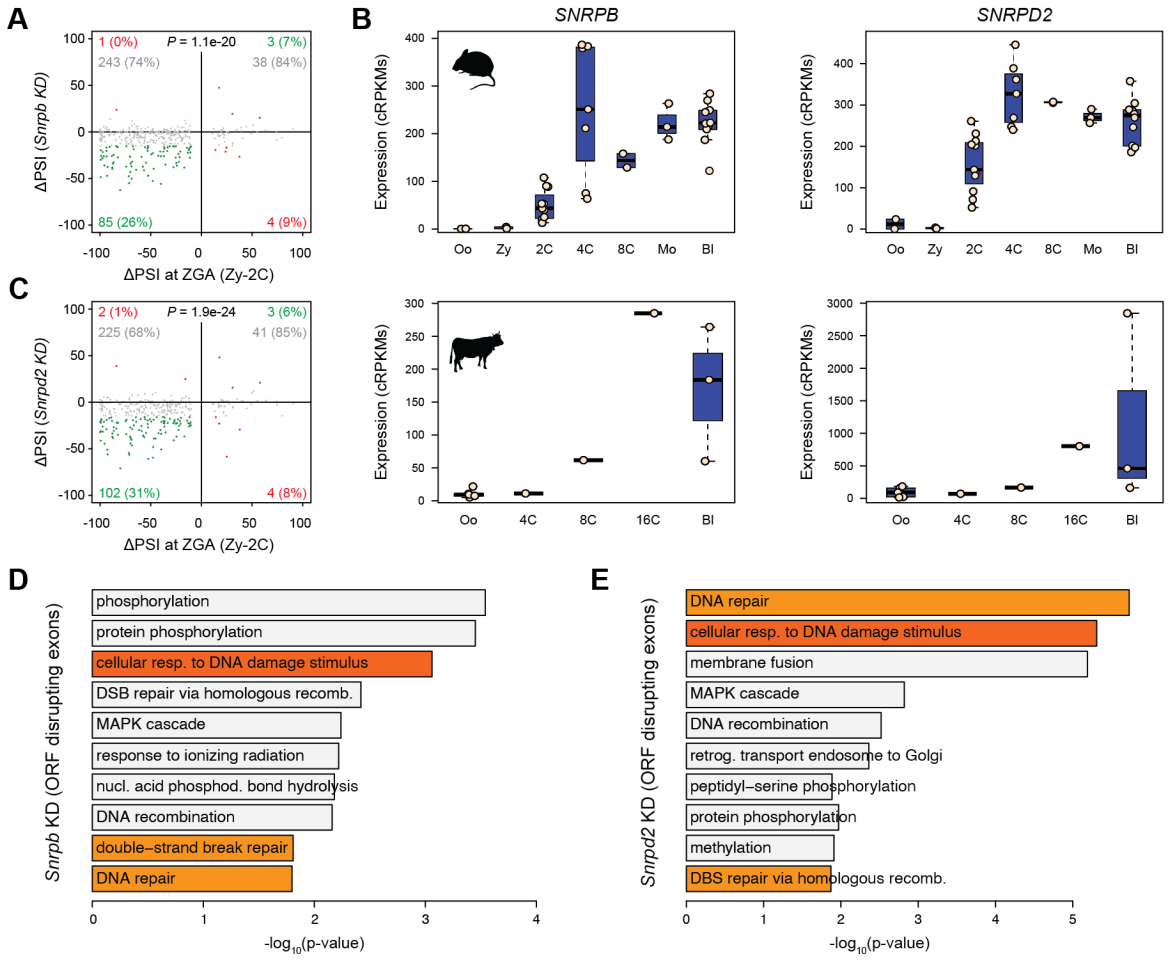

Fig. S20 - *Snrpb* and *Snrpd2* are dynamically regulated at ZGA and their depletion leads to ZGA-like splicing patterns. (**A**,**C**) Scatter plots showing ΔPSI of exons from Mfuzz clusters that peak at ZGA upon *Snrpb* (A) or *Snrpd2* (C) knockdown in mouse embryonic stem cells (Y-axis) versus ΔPSI between 1C and 2C stages (ZGA; X-axis). Exons with |ΔPSI| ≥ 15 upon knockdown are highlighted in green/red and those with |ΔPSI| < 15 are depicted in grey. Total numbers as well as the percentage among exons with increased or decreased PSI at ZGA are indicated for each exon category. P-value corresponds to a two-sided Binomial test between Q1+Q3 versus Q2+Q4. RNA-seq data from GSE168728. (**B**) cRPKM metric is used to quantify GE for *Snrpb* and *Snrpd2* in mouse and cow. (**D**,**E**) Enriched GO terms for genes containing exons predicted to cause ORF disruption upon *Snrpb* (D) or *Snrpd2* (E) knockdown in mouse embryonic stem cells. Orange bars indicate DNA damage/repair related categories.

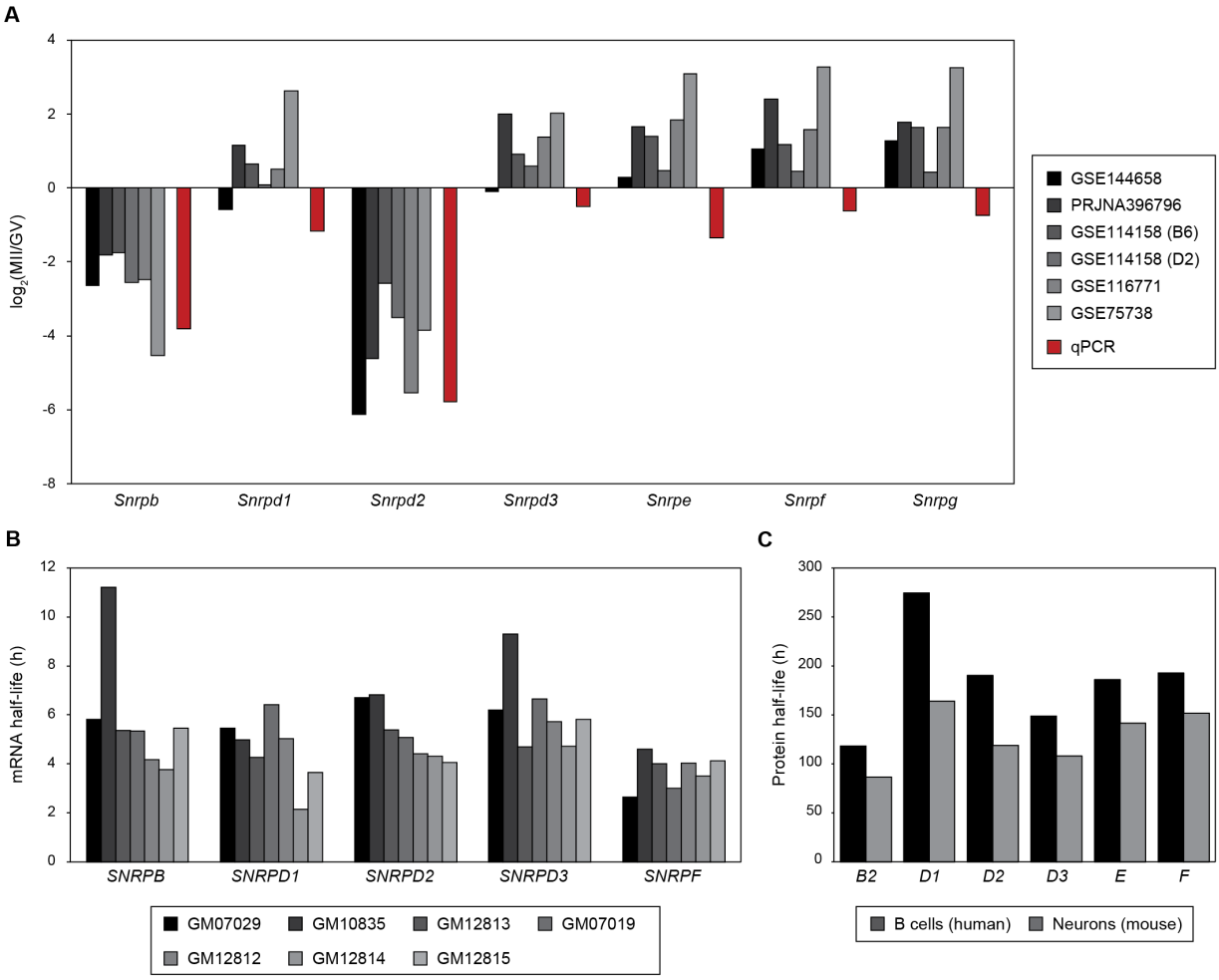

Fig. S21 - *Snrpb* and *Snrpd2* are specifically downregulated during oogenesis. (**A**) Change in mRNA steady state levels between GV and MII stage mouse oocytes (log2 fold change of cRPKM values) for the different Sm ring components as quantified from five different RNA-seq experiments (grey scale bars) and qPCR assays from independent RNA samples (red bars). RNA-seq experiment identifiers are indicated on the legend; GSE114158 includes data for B6 and D2 strains. Further information about the RNA-seq data is provided in Table S1. qPCR data represents relative expression values normalized against total RNA content (**B**) RNA half-lives (in hours) for different Sm ring components in various human lymphoblastoid cell lines (Duan et al. Sci Rep 3, 1318 (2013)). (**C**) Protein half-lives (in hours) for different Sm ring components in human B cells and mouse neurons (Mathieson et al., Nat Commun 9, 689 (2018)).

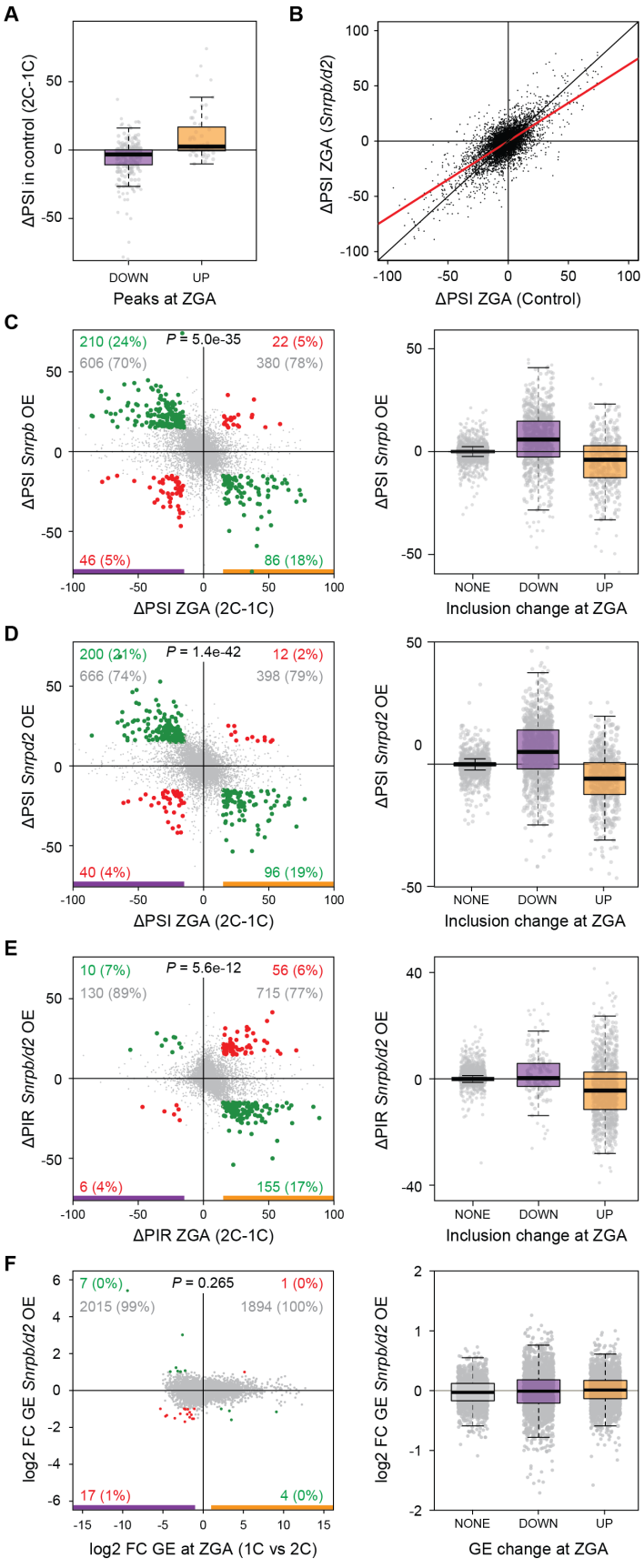
A
ΔPSI in control (2C-1C)
50
0
-50
DOWN
UP
Peaks at ZGA
B
ΔPSI ZGA (Snrpb/d2)
100
50
0
-50
-100
-100
-50
0
50
100
ΔPSI ZGA (Control)
C
210 (24%)
P = 5.0e-35
22 (5%)
606 (70%)
380 (78%)
46 (5%)
86 (18%)
ΔPSI Snrpb OE
ΔPSI ZGA (2C-1C)
NONE
DOWN
UP
Inclusion change at ZGA
D
200 (21%)
P = 1.4e-42
12 (2%)
666 (74%)
398 (79%)
40 (4%)
96 (19%)
ΔPSI Snrpd2 OE
ΔPSI ZGA (2C-1C)
Inclusion change at ZGA
E
10 (7%)
P = 5.6e-12
56 (6%)
130 (89%)
715 (77%)
6 (4%)
155 (17%)
ΔPIR Snrpb/d2 OE
ΔPIR ZGA (2C-1C)
Inclusion change at ZGA
F
7 (0%)
P = 0.265
1 (0%)
2015 (99%)
1894 (100%)
17 (1%)
4 (0%)
log2 FC GE Snrpb/d2 OE
log2 FC GE at ZGA (1C vs 2C)
GE change at ZGA

Fig. S22 - Partial reversion of ZGA transcriptomic changes by earlier *Snrpb/d2* expression. (**A**) ΔPSI between 1C and 2C stages (ZGA) in control samples for peak-up and peak-down exons. (**B**) ΔPSI between 1C and 2C stages in control mCherry (X axis) and *Snrpb*/*d2* (Y axis) injected embryos for all exons with sufficient read coverage in the four samples. The diagonal (x=y) is shown as a black line and a linear regression for all data points is shown in red. (**C**,**D**) Left: Changes in exon inclusion levels at 2C stage upon early-induced expression of *Snrpb* (C) or *Snrpd2* (D) (Y-axis) respect to the observed change at ZGA (2C-1C; X-axis). Exons with |ΔPSI| ≥ 15 upon *Snrpb* or *Snrpd2* expression and at ZGA are highlighted in green/red, while exons with no change are depicted in grey. Total numbers as well as the percentage among exons with increased or decreased PSI at ZGA are indicated for each exon category. Right: Distribution of ΔPSI upon *Snrpb* or *Snrpd2* induced expression for all exons with increased (UP), decreased (DOWN) or no change (NONE) in PSI at ZGA. (**E**) Left: Changes in intron retention levels at 2C stage upon early-induced expression of *Snrpb* and *Snrpd2* (Y-axis) respect to the observed change at ZGA (2C-1C; X-axis). Introns with |ΔPSI| ≥ 15 upon *Snrpb*/*d2* expression and at ZGA are highlighted in green/red, while introns with no change are depicted in grey. Total numbers as well as the percentage among introns with increased or decreased PSI at ZGA are indicated for each intron category. Right: Distribution of ΔPSI upon *Snrpb* or *Snrpd2* induced expression for all introns with increased (UP), decreased (DOWN) or no change (NONE) in PSI at ZGA. (**F**) Left: Changes in GE levels at 2C stage upon early induced expression of *Snrpb/d2* (Y-axis) respect to the observed change at ZGA (2C-1C; X-axis). Genes with |log2(FC)| ≥ 1 upon *Snrpb/d2* expression and at ZGA are highlighted in green/red, while genes with no change are depicted in grey. Total numbers as well as the percentage among genes with increased or decreased expression at ZGA are indicated for each category. Right: Distribution of log2(FC) upon *Snrpb/d2* induced expression for all exons with increased (UP), decreased (DOWN) or no change (NONE) in GE at ZGA. *P*-values in C-F correspond to a two-sided Binomial test between Q2+Q4 versus Q1+Q3.

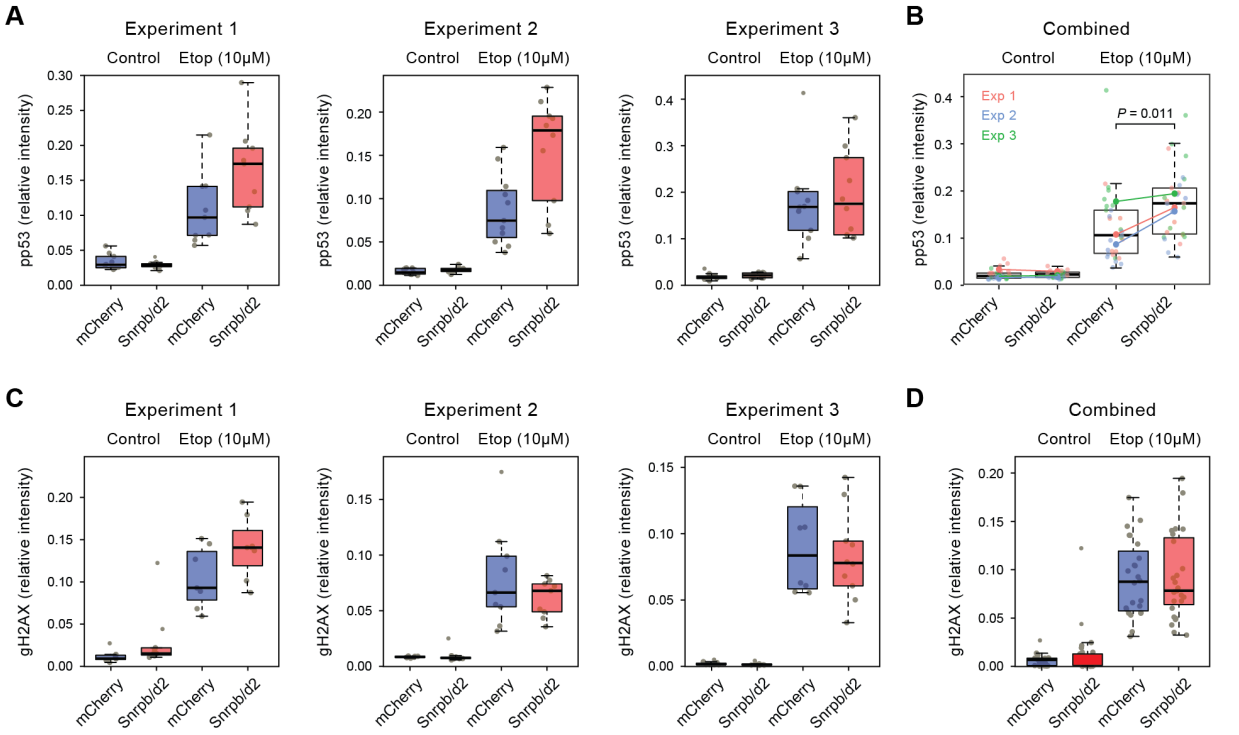

Fig. S23 - DNA damage response is modulated by earlier *Snrpb/d2* expression. (**A-D**) Quantification of phospho-p53 (Ser15, pp53) (A,B) or gamma-H2AX (C,D) immunostaining levels from 2C embryos injected at pronuclear stage with *Snrpb/d2* or mCherry mRNA and treated for 1.5h with 10uM etoposide or left untreated. Each plot represents an independent experiment and each dot represents the average relative intensity of both cells of a single embryo. Combined data from all three individual experiments for phopho-p53 (B) shows the average for each of the experiments separately. Total number of embryos for each condition and experiment in (A): Experiment 1: n=10 (mCherry untreated), n=9 (Snrpb/d2, untreated), n=9 (mCherry, etoposide) and n=9 (Snrpb/d2, etoposide); Experiment 2: n=9 (mCherry untreated), n=8 (Snrpb/d2, untreated), n=11 (mCherry, etoposide) and n=10 (Snrpb/d2, etoposide); Experiment 3: n=10 (mCherry untreated), n=9 (Snrpb/d2, untreated), n=10 (mCherry, etoposide) and n=10 (Snrpb/d2, etoposide). Total number of embryos for each condition and experiment in (C): Experiment 1: n=7 (mCherry untreated), n=9 (Snrpb/d2, untreated), n=7 (mCherry, etoposide) and n=8 (Snrpb/d2, etoposide); Experiment 2: n=8 (mCherry untreated), n=11 (Snrpb/d2, untreated), n=9 (mCherry, etoposide) and n=9 (Snrpb/d2, etoposide); Experiment 3: n=11 (mCherry untreated), n=10 (Snrpb/d2, untreated), n=8 (mCherry, etoposide) and n=10 (Snrpb/d2, etoposide); Combined (d): n=26 (mCherry untreated), n=30 (Snrpb/d2, untreated), n=24 (mCherry, etoposide) and n=27 (Snrpb/d2, etoposide).

Supplementary Table Legends

Table S1 - RNA-Seq samples used in this study. Several spreadsheets containing the publicly available RNA-Seq datasets used in this study for human, mouse and cow. For each species, we provide a sheet containing the early development data and another the cell and tissue types. Each table includes the SRA identifier, number of reads, length, source, etc. Furthermore, the groups of samples that were pooled together to increase read depth are indicated. For discarded samples (red font), the reason for removal is indicated. An additional sheet ("RBP_experiments_Human") contains the information about the experiments of splicing factor depletion, another one the information about the mouse GV and MII oocytes ("Oogenesis"), and another one ("Pool_read_stats") contains the summary statistics of the sample pooling.

Table S2 - Alternative exons not included in vast-tools. For each species, the list of alternative exons that were not present in vast-tools and that were added as an additional module. For each exon, the basic information is provided, including coordinates (hg19, mm9 or bosTau6), whether or not they are annotated in Ensembl ("Non_annot"), whether they are part of any Mfuzz cluster and the type of cluster ("ClusterID" and "ClusterType"), and the average PSI in each developmental stage (NA, insufficient read coverage).

Table S3 - Alternative exons dynamically regulated during early embryogenesis of the three species. Exons that showed dynamic inclusion levels in the three studied species (Fig. S7g). Each row corresponds to an orthologous exon and the information is provided for the three species, including coordinates (hg19, mm9 and bosTau6), vast-tools ID, PSI in each developmental stage and species and predicted impact on the ORF.

Table S4 - Differentially regulated exons in early development. For each species, the list of exons that showed differential regulation in any transition and/or that were included in the Mfuzz clusters. In addition to the basic exon information, the table includes: summary statistics of the inclusion levels across differentiated tissues and early development samples, conservation status in the other two species and predicted protein impact.

Table S5 - Significantly enriched motifs for RNA binding proteins. For each genomic region (upstream intron, exon or downstream intron), the motifs from cisBP-RNA that are significantly enriched over the background (as estimated by Matt) for peak-down (P_Dw) and peak-up (P_Up) exons at ZGA are shown for each species. Color code: dark red, the motif is depleted in P_Dw exons; light red, the motif is enriched in P_Dw exons; dark blue, the motif is depleted in P_Up exons; light blue, the motif is enriched in P_Up exons.

Table S6 - Correlations between GE of splicing factors and PSIs at the single-cell level. Summary statistics (correlation coefficient and p-value) for the correlation analysis between the expression of 196 splicing factors and the inclusion levels of peak-down and peak-up exons at ZGA in individual ZGA blastomeres (8C in human and 2C in mouse, respectively).

Table S7 - Association between splicing factor depletion and ZGA. Summary statistics for the overlap between peak-down and peak-up exons at ZGA and exons changing ($|\Delta PSI| > 15$) upon depletion of specific splicing factors. For each experiment (X*), the number of exons in

quadrants Q1-Q4 (based on the direction of the overlap) is provided, as well as the percentage of those in Q1 and Q3 ("Perc_Q1_Q3") and the p-value of a two-sided Binomial associated this percentage. Moreover, for the total number of peak exons at ZGA with read coverage ("Cov_P_Up" and "Cov_P_Dw"), it provides the percentage of exons in each quadrant.

Table S8 - Gene ontology results for DNA damage specific GO terms (related to Fig. 2). Tab1: For the three species we show the proportion test results for the two main enriched DRR related terms: "DNA repair" (GO:0006281) and "cellular response to DNA damage stimulus" (GO:0006974). Showing the observed proportion (colD), observed hits (colE), Expected proportion (colF), Expected hit (colG), Fold change between observed and expected hits (colH), Type of GO term (colH; BP-biological process), Genes in the observed set (colI; also in list form after row number 15), Event name (colK) and the proportion test p-values (colL). Tabs 2-4 show the entire list of genes associated with both DNA repair (GO:0006281) and cellular response to DNA damage stimulus (GO:0006974) compiled from Biomart, with orthologous annotations from human, mouse and cow merged.

Table S9 - Exons used for genomic feature analyses. For each species, the list of genes used as input for *Matt cmpr_exons* and associated information (including PSI in the pre-ZGA and post-ZGA stages and the group to which they belong).

Table S10 - Primers used for RT-PCR validations and qPCR analysis.

Table S11 - Statistics for in-house RNA-Seq samples. Statistics associated with the RNA-Seq data for embryos injected with *Snrpb* and/or *Snrpd2* or a control.

Table S12- PSI quantifications for the RNA-Seq data for embryos injected with *Snrpb* and/or *Snrpd2* or a control using *vast-tools* (v1.3.0).

Table S13 - Nomalized cRPKM values for the RNA-Seq data for embryos injected with *Snrpb* and/or *Snrpd2* or a control using *vast-tools* (v1.3.0).

Supplementary file 1 - Output files from *Matt cmpr_exons* analyses for human, mouse and cow exon sets. Full PDF reports obtained from *Matt* for each species. The reports are concatenated, and the species investigated is stated in the fifth page of each report (pages 5, 112 and 205; highlighted in yellow).