Expanded View

Insulators recruit histone methyltransferase dMes4 to regulate chromatin of flanking genes

Priscillia Lhoumaud[+*], Magali Hennion[+*], Adrien Gamot[+*], Suresh Cuppadrah[^], Sophie Queille[+], Jun Liang[+], Gael Micas[+], Pauline Morillon[+], Serge Urbach[$], Olivier Bouchez[`], Dany Severac[~], Eldon Emberly[#], Keji Zhao[^] and Olivier Cuvier[+@]

Expanded Figure Legends

Figure E1. Genome-wide analyses highlight a role of Beaf32 in gene expression

A. Western blotting analysis of nuclear extracts using anti-Beaf32 antibodies in the presence (+) or absence (-) of peptide competitor ('competitor'), as indicated. Lower blot: Anti-histone H3 antibodies were used for loading control ('anti-H3').

B. Graph representing the results from 3 independent RT-qPCR analyses to measure the mRNA levels of Beaf32 (green) and actin (black) in Beaf32 knock-down ('Beaf32-KD') or wild-type ('Control') cells. The error bar corresponds to the standard deviation of the measure. The y axis shows the relative fold change in expression normalized to *actin* control.

C. Western blotting analysis of nuclear extracts prepared from Beaf32-KD or control cells using anti-Beaf32 or anti-actin control antibodies for loading control.

D. Venn diagram showing the intersection analysis between differentially expressed (DE) genes whose expression is affected in Beaf32-KD cells compared to WT cells (2,059 genes) as

measured in two independent DGE-Seq experiments (see Methods) and genes that are flanked (< 500bp from TSSs) by one of the 4,120 Beaf32 binding site (> 25 ChIPSeq reads; see Methods). The indicated p-value was obtained by fisher exact test. Also indicated are the percentages of DE genes depending on whether their promoters are bound by Beaf32 or not. Note that approximately 91% (3,749) of the 4,120 Beaf32 sites correspond to 'direct' Beaf32 binding sites (Liang et al., 2014) that harbor the CGATA motifs and/or that localize close to a TSS.

E. Plot of the correlation between differential expression between WT and Beaf32-KD cells and Beaf32 binding. All drosophila genes were scored according to their differential expression levels as measured by DGE-Seq (x-axis; see Methods) after ranking of genes according to the levels of Beaf32 binding to their promoters as measured by ChIP-Seq (y-axis; see Methods). The dotted lines represent the thresholds for Beaf32 binding (horizontal line) and for DE genes (vertical line).

F. Correlation between RNA-Seq and microarray analysis or RT-qPCR and for differentially expressed genes. Graph representing the percentage (%) of differentially regulated genes identified through DGE-Seq (2,059 genes) upon Beaf32-depletion ('Beaf32-KD') and most differentially expressed genes identified by microarray analyses (see Methods). The red bar marks the FDR<0.01 for the DGE-Seq including 85% of most differentially expressed genes identified by microarray analyses (p-value ~1e-28).

G. Graph showing the linear regression line for gene expression in WT- compared to Beaf32-KD-cells obtained by both DGE-Seq and by RT-qPCR (Pearson's correlation coefficient: 0.83; p-value~ 3.2e-5) for 18 genes tested using the absolute quantification component of pyQPCR (http://pyqpcr.sourceforge.net/)(see Methods). Each point corresponds to a gene whose expression was measured in triplicates as a ratio of expression between WT/Beaf32-KD cells,

either for DGE-Seq reads or for the number of copies as measured by RT-qPCR. On the x-axis is represented the ratio by DGE-seq and on the y-axis the ratio by RT-qPCR.

Figure E2. Nucleosome positioning defects upon insulator protein depletion

A. Heat map of nucleosome-positioning as measured by MNase-Seq (see Methods) for DE genes whose promoters were not bound by Beaf32 (left panel), bound by Beaf32 (middle panel) or for control genes whose promoters were bound by Beaf32 (right panel). Nucleosomes were aligned relative to TSS (x-axis; position 0) of genes. The dotted lines mark the lower limit for DE genes. NFR, Nucleosome Free Regions (see also Figure 1B).

B. Heat map representing the intersection analysis between the list of genes with similar levels of nucleosome positioning or changes in nucleosome-positioning defects ('ctrl' and 'var, respectively; as found upon Beaf32-KD as compared to control cells; see Methods) and genes whose promoters were bound by Beaf32 ('Beaf32') or not ('no Beaf32') and/or genes that were differentially expressed ('DE genes') or not (control) genes. Each intersection was tested using fisher exact test. See panel C for the percentages of genes corresponding to each intersection.

C. Histogram showing the relative enrichment of genes where most significant changes in nucleosome-positioning were detected between MNase-Seq read counts in Beaf32-KD compared to WT cells as in panel B. The error bars represent the standard deviation obtained by analyzing reads from 3 independent experiments in Beaf32-KD compared to WT.

D. Detection of aberrant RNASeq reads in Beaf32-KD compared to WT cells. Histogram showing the relative enrichment of genes showing 'aberrant' RNASeq reads (outside defined exons; see Methods) in Beaf32-KD compared to WT cells, as a function of: left, the binding in promoters of Beaf32 (red) or not (control, black); right: variations in nucleosome-positioning

('Beaf32-KD / WT', group of genes harboring most significant changes in nucleosome positioning; see Figure 1C) or not ('control'). The error bars represent the standard deviation obtained by analyzing reads from 4 independent RNASeq experiments in Beaf32-KD compared to WT.

Figure E3. Characterization of dMes-4 and insulator protein depletions

A. Beaf32 directly interacts with dMes-4. Data obtained by yeast 2 hybrid using the indicated various combinations of bait and target constructs expressing fusion protein Beaf32, dMes-4 and CP190 proteins compared to positive controls (Krev1-Ral) in cells replicated on nonselective (+Ura) or selective (-Ura) plates (see Methods).

B. Western blotting analysis of chromatin fractions purified through sucrose cushions ('chromatin'; see Methods) prepared from Beaf32-KD or control cells or of the corresponding nuclear extracts ('input'), using anti-Beaf32, anti-H3K36me2 antibodies or total anti-histone H3 for loading control.

C. Quantification of H3K36me2/3 levels in Beaf32-KD or control cells as compared to loading control from the western blotting analysis in Figure 2E and in panel B (see Methods).

D. Histogram showing the percentage of DE genes as determined by expression analysis in Beaf32-KD compared to WT cells as a function of the levels of histone acetylation (H4K16ac; High or Low, light red and red bar, respectively) within the corresponding promoter regions (-500 to 0 bp from TSS).

E. Box plot showing the results of ChIP performed in Beaf32-KD (red boxes) compared to WT-control (mock-depleted) cells (green boxes) in percent of input (y-axis) with anti- H3K36me2 antibodies or IgG control, for promoters harboring a Beaf32 site or not as indicated. ChIP data

were analyzed by qPCR analyses in triplicates and for three independent measures normalized to three control loci (see Methods).

F. Western blotting analysis of nuclear extracts using anti-dMes-4 antibodies. The arrow indicates the unique band detected corresponding to the molecular weight of dMes-4 (1423 AA).

G. Western blotting analysis of nuclear extracts prepared from dMes-4-KD or control cells using anti-dMes-4 or anti-actin control antibodies for loading control. The efficiency of dMes-4-KD was confirmed by RTqPCR analysis (see Figure 4A).

H. Quantification of dMes-4 depletion as compared to loading control from the western blotting analysis in panel F (see Methods).

Figure E4. dMes-4 and Beaf32 co-regulate specific Gene Ontologies

A. Examples of gene ontologies enriched among the differentially expressed genes as measured by DGE-Seq in Beaf32-KD or Mes4-KD compared to control cells and as identified through DGE-Seq (threshold for DE genes p-value < 1e-3).

B. The 4,120 Beaf32 peaks identified by ChIP-Seq (see Figure 1A) harbor two consensus motifs that are recognized by Beaf32 and DREF, in agreement with the involvement of both factors in co-regulating specific gene ontologies (GO) including the cell cycle (see panel A; see Expanded Table E1)(Emberly et al., 2008). Also indicated are the associated E-values as obtained by motif searches through MEME (see Methods).

Figure E5. Genome-wide correlations between genes regulated by dMes-4 and nucleosome positioning or H3K36me2/me3 levels

A. Heat map of nucleosome-positioning as measured by MNase-Seq in wild-type cells (see Methods) after ranking of genes whose promoters were bound by Beaf32 (right panel) or not (left panel) by their differential expression upon dMes-4KD as compared to control cells. Nucleosomes were aligned relative to the TSS (x-axis; position 0) of genes. The dotted lines mark the lower limit for DE genes. NFR, Nucleosome Free Regions (see Figure E2 for a similar analysis of DE genes upon Beaf32-KD).

B. Heat map representing the intersection analysis between the list of genes with most significant nucleosome-positioning defects upon Beaf32-KD as compared to control cells ('var'; see Figure E2B) or with similar levels of nucleosome positioning ('ctrl') and genes whose promoters were bound by Beaf32 ('Beaf32') or not ('no Beaf32'). Each intersection was tested using fisher exact test. See panel C-D for the levels of H3K36me2 and H3K36me3 corresponding to each intersection.

C. Box plot showing the levels of H3K36me2 (in ChIP-Seq reads; see Methods) for each intersection shown in panel B. Reads were counted within windows corresponding to gene bodies (from +500 to the end of each gene) in wild-type cells and for all genes or specifically for the genes differentially expressed upon dMes4-KD as compared to control genes ('DE genes') or for control genes with not change in expression in dMes4-KD.

D. Same as panel C except that H3K36me3 levels are indicated (in ChIP-Seq reads; see Methods). Reads were counted within windows corresponding to gene bodies (from +500 to the end of each gene) in wild-type cells and for all genes or specifically for the genes differentially expressed upon dMes4-KD as compared to control genes ('DE genes') or for control genes with not change in expression in dMes4-KD.

E. Impact of dMes-4 as a function of Beaf32 and dCTCF binding. Receiver operating characteristic (ROC) curve analysis (see Methods) of DE genes in dMes-4KD compared to WT cells allowing to test the influence of dMes-4 on differential expression for genes flanking a unique dCTCF binding site (< 500bp from TSS; blue curve) or a Beaf32 binding site (< 500bp from TSS; black curve). Y-axis, % of true positives (differentially expressed genes); X-axis, % of false positives (constant expression), depending on the calculated threshold for differential expression (see also Figure 3C for a combinatorial analysis).

Figure E5. Beaf32 binding correlates with H3K36me3 levels

A. Box plot showing the enrichment in H3K36 trimethylated levels according to the ranking of genes by quartiles of Beaf32 binding (~3650 genes/quartile; from 'q1'/highest to 'q4'/lowest-levels). y-axis, ChIP-Seq reads for H3K36me3.

B. Box plot showing the results of ChIP performed in Beaf32-KD (red boxes) compared to WT-control (mock-depleted) cells (green boxes) in percent of input (y-axis) with anti- H3K36me3 antibodies or IgG control for promoters harboring a Beaf32 site (see Methods for a list of genes; see Figure 5A for similar analysis of the same Beaf32 bound promoters as compared to control promoters). ChIP data were analyzed by qPCR analyses in triplicates and for three independent measures normalized to three control loci.

C. Averaged profiles of H3K36me3 levels (y-axis; in ChIP-Seq reads) according to quartiles of Beaf32 binding (see panel A). The arrow indicates the relatively low levels of H3K36me3 in promoter regions as compared to gene bodies.

D. Box plot showing the results of ChIP performed in Hypb-KD (red boxes) compared to WT-control (mock-depleted) cells (green boxes) in percent of input (y-axis) with anti- H3K36me2

antibodies or IgG control, for 16 promoters harboring a Beaf32 site or not (see Methods for a list of genes). ChIP data were analyzed by qPCR analyses in triplicates and for two independent measures normalized to three control loci. Note that the increase in H3K36me2 levels upon Hypb-KD supports its role in triggering the transition from H3K36me2 to H3K36me3 (Bell et al., 2007).

Figure E7. RNA splicing defects upon depletion of Beaf32 or dMes-4

A. Graph showing examples of RTqPCR measurements (in triplicate) using oligos that span exon-intron junctions to quantify the levels for immature RNAs in Beaf32-KD (red bars) compared to WT cells (black bars) after normalization to the levels of mature RNAs for the indicated genes that harbor a Beaf32 binding site or not (see also Figure E9B for a similar analysis to 16 Beaf32-bound or unbound genes in Beaf32-KD; see Methods for a list). y-axis, fold change normalized to mature RNA levels.

B. Histogram showing the percentage of genes with significant splicing defects as quantified by measuring 'retained introns' by counting RNASeq reads in introns normalized to RNASeq read counts in exons in three independent Beaf32-KD compared to WT cells (see Methods)(Hu et al., 2013), as a function of the H3K36-me2 or -me3 quartiles (see also Figure 5C).

Figure E8. Characterization of DREF depletion

A. Graph representing the results from 3 independent RT-qPCR analyses to measure the mRNA levels of *actin*, *beaf32*, *dMes-4*, and *dref* upon knock-down of Beaf32 ('Beaf32-KD'; green bars), dMes-4 ('dMes-4-KD; red bars), DREF ('DREF-KD'; purple bars) or control ('control';

black bars). The error bar corresponds to the standard deviation of the measure. The y axis shows the relative fold change in expression normalized to *actin* control.

B. Western blotting analysis of nuclear extracts using anti-DREF antibodies. The arrow indicates the unique band detected corresponding to the molecular weight of DREF (713 AA).

C. Western blotting analysis of nuclear extracts prepared from DREF-KD or control cells using anti-DREF or anti-actin control antibodies for loading control. The efficiency of DREF-KD was confirmed by RTqPCR analysis (see panel A).

D. Quantification of DREF depletion as compared to loading control from the western blotting analysis in panel C (see Methods).

E. Box plot showing the results of ChIP performed in DREF-KD (red boxes) compared to WT-control (mock-depleted) cells (green boxes) in percent of input (y-axis) with anti- dMes-4 antibodies or IgG control, for 16 promoters harboring a Beaf32 site or not, as indicated (see Methods for a list of genes). ChIP data were analyzed by qPCR analyses in triplicates and for three independent measures normalized to three control loci (see Methods).

F. Co-immunoprecipitation experiment using anti-DREF antibodies performed in parallel with IgG control followed by western blotting analysis using anti- Beaf32, DREF or -histone H3 as loading control, compared to input nuclear extract (10% 'input').

Figure E9. Principal component analyses, RNA splicing defects upon depletion of Beaf32 or DREF.

A. Graph showing the genome-wide correlations as obtained by principal component analyses (PCA; see Methods)) among histone modifications and nucleosome positioning for genes that harbor or not a Beaf32/DREF binding site in their promoter (right and left panel, respectively).

The y- and x- axis represent the two main dimensions. Note that to generate dendograms (see Figure 7A) by Clustering Ascendant Hierarchical (CAH) all additional PCA dimensions of significance are taken into consideration (Josse et al., 2011). All ChIP-Seq (Beaf32, H3K36me2/me3, H4K16ac, H3K27me3) and MNase-Seq (nucleosome-positioning, NFR) data were obtained from exponentially growing Drosophila cells. Note that the higher correlation between Beaf32/DREF binding and H3K36 methylation levels in gene bodies ('H3K36me3') as compared to promoter regions ('H3K36me3p') in agreement with the specific impact of Beaf32-KD in the former regions. Note also that H3K36me2/3 are anti-correlated with H3K27me3 marks, in agreement with data obtained in *C.elegans* or vertebrates (Gaydos et al., 2012; Cai et al., 2013). See also Figure 7A.

B. RNA splicing defects upon depletion of Beaf32 or DREF. Graph showing the results from RTqPCR measurements in triplicate showing the fold change (y-axis) in mRNA levels (left panel) or levels of immature RNAs normalized to mRNA (right panel). Immature RNAs were quantified using oligos that span exon-intron junctions in DREF-KD (purple boxes), Beaf32-KD (green boxes) as compared to control cells (grey boxes) for 16 genes (see Methods for a list) that harbor a Beaf32 binding site or not (see Methods for a list).

Figure E10. Influence of IBP/dMes-4 on H3K27me3 deposition

A. Averaged H3K27me3 profiles for control promoters where no significant variations in H3K27me3 levels could be scored between Beaf32-KD and control ('WT') cells as measured by ChIP-Seq (see Methods). Note that such variations were systematically scored over +/- 2 Kbp regions surrounding all *Drosophila* TSSs. See Figure 7B for the complementary list of genes (990 genes) with significant changes in H3K27me3 levels upon Beaf32-KD.

B. Histogram representing the percentage of genes with no variation in H3K27me3 levels in Beaf32-KD/WT (y-axis) depending on the presence or absence of a Beaf32 binding site in their promoters. See Figure 7C for a similar analysis of the complementary list of genes.

C. Heat map representing the intersection analysis between the list of genes with variations in H3K27me3 levels ('var'; see Figure 7B) or no variation ('ctrl', see Figure E10A) and the list of genes whose promoters were bound by Beaf32 ('Beaf32') or not ('no Beaf32') and/or genes that were differentially expressed ('DE genes') or not (control) genes in Beaf32-KD or dMes4-KD (left and right heat maps, respectively). Each intersection was tested using fisher exact test.

D. Fold changes in H3K27me3 levels between Beaf32-KD and dMes4-KD. H3K27me3 levels were measured by qPCR from ChIP experiments performed in three independent triplicates in Beaf32-KD compared to control cells (x-axis) or in dMes4-KD compared to control cells (y-axis). The genes tested include those selected based on variations as scored from our ChIP-seq data (see Methods) as well as those genes tested in other ChIP assays (Figure 6; see Methods for a list).

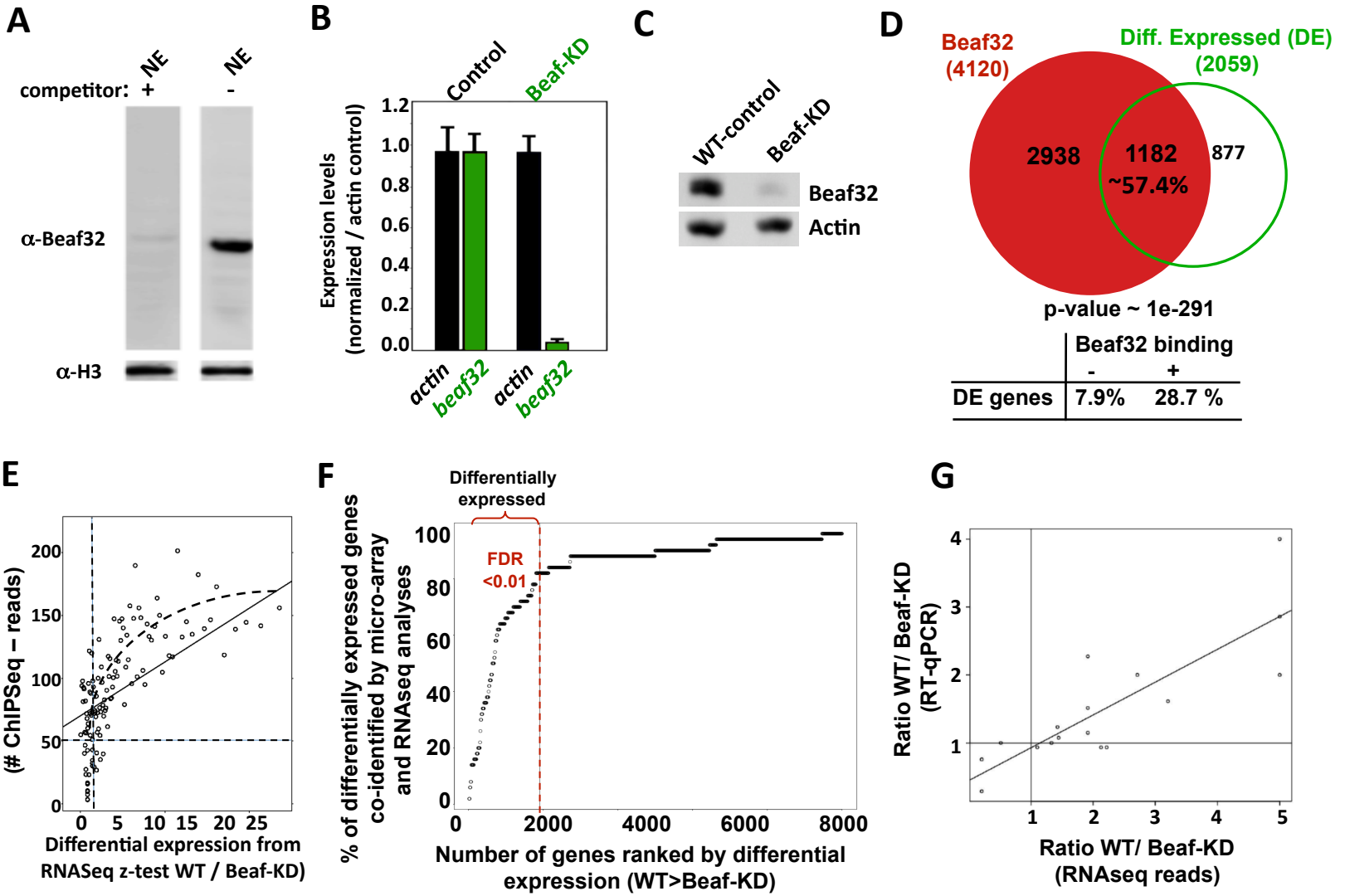
E. Control of primers by quantitative PCR standard curves. The graph shows an example of the standard curve used to quantify gene expression or ChIP analysis by RT-qPCR and q-PCR analyses, respectively, as done for every oligo pair used in this study. The quantification was automatically done using the pyQPCR software to calculate linear regressions and PCR efficiency (calculation details are available at http://pyqpcr.sourceforge.net/; see Methods) with triplicate measurements for each of 4 concentrations used to generate a standard curve for either cDNAs (for gene expression measurements) or input DNAs (for ChIP measurements and determination of copy number for gene expression). x-axis: Quantity of DNA (in Log scale). y-axis: Ct (threshold cycle) of qPCR.

Expanded Tables

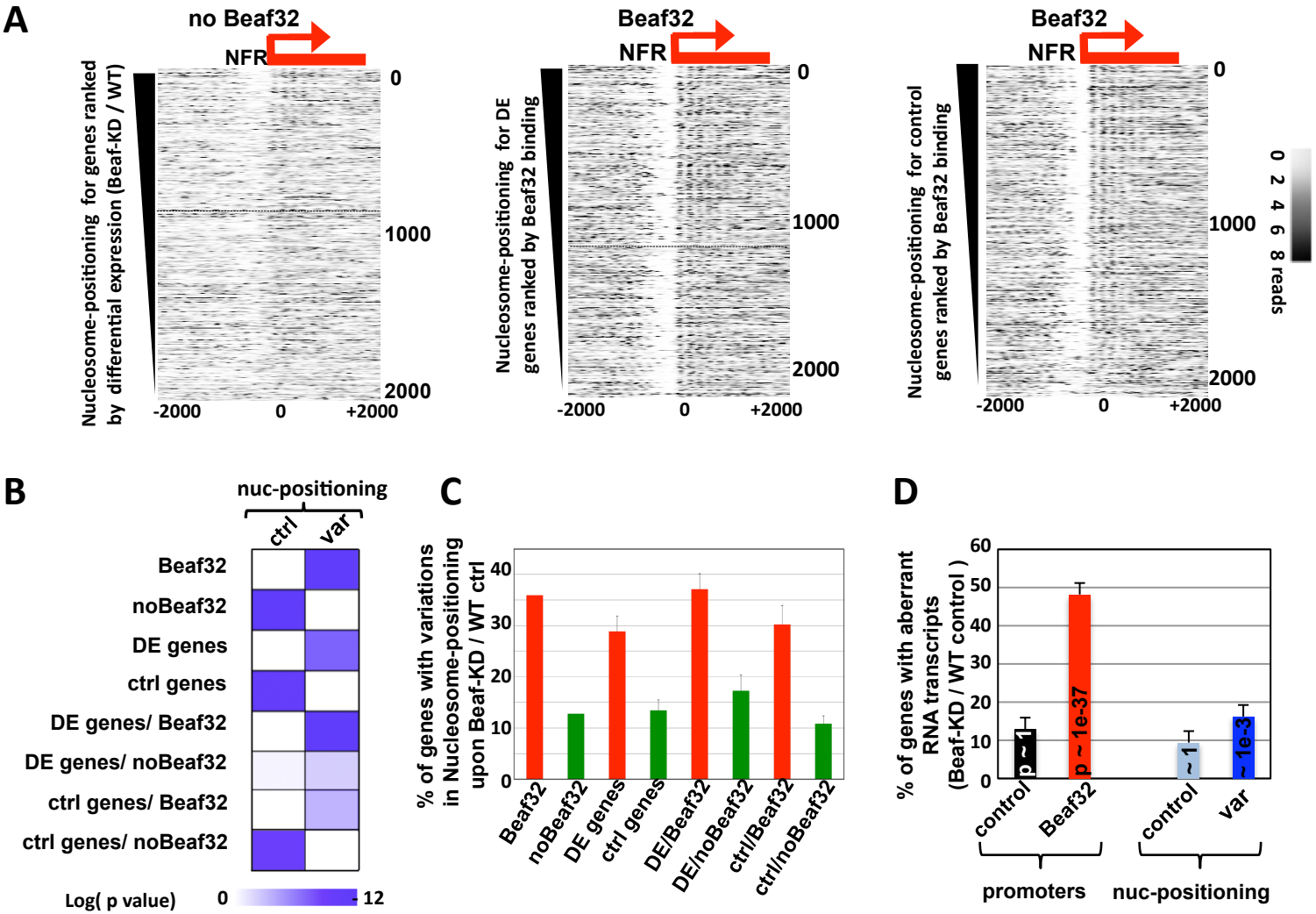
Table E1: Table listing the most significant Gene Ontologies (GO) represented among the most significantly differentially expressed genes in Beaf32-KD compared to WT cells identified by RNA-Seq from three combined experiments. 'Number (TOT)' is the number of genes for each GO, 'FREQ' the total frequency of genes

	TOTAL		INSULATOR		
GO_TERM	**NUMBER**	**FREQ**	**NUMBER**	**FREQ**	**P_VALUE**
cell cycle	402	0.0397	103	0.0829	4.00E-14
system development	1671	0.1651	288	0.2317	2.13E-11
cell differentiation	1070	0.1057	199	0.1601	8.06E-11
homeostatic process	202	0.02	58	0.0467	1.44E-10
ribonucleotide binding	893	0.0818	176	0.1271	1.61E-10
purine nucleotide binding	960	0.0879	186	0.1343	1.99E-10
biopolymer metabolic process	3492	0.3451	512	0.4119	2.79E-08
transferase activity	511	0.0468	107	0.0773	2.81E-08
protein localization	365	0.0361	80	0.0644	5.50E-08
nucleoside, nucleotide	1640	0.1621	261	0.21	3.83E-07
regulation of cell size	56	0.0055	20	0.0161	3.99E-06
appendage development	256	0.0253	56	0.0451	4.87E-06
organelle fission	177	0.0175	41	0.033	2.03E-05
programmed cell death	207	0.0205	46	0.037	2.04E-05
cytoskeletal protein binding	215	0.0197	48	0.0347	2.78E-05
hydrolase activity, acting on acid	605	0.0554	107	0.0773	6.06E-05
cation binding	1481	0.1356	231	0.1668	6.11E-05
cell-cell signaling	236	0.0233	49	0.0394	6.57E-05
fusome organization	14	0.0014	8	0.0064	6.98E-05
cellular response to stress	177	0.0175	39	0.0314	9.31E-05
synapse organization	64	0.0063	18	0.0145	0.0003372
cellular macromolecule loc	280	0.0277	53	0.0426	0.000342
nucleus organization	39	0.0039	13	0.0105	0.0003796
enzyme binding	63	0.0058	18	0.013	0.0004036
asymmetric cell division	57	0.0056	15	0.0121	0.0019148
stem cell division	49	0.0048	13	0.0105	0. 0033381

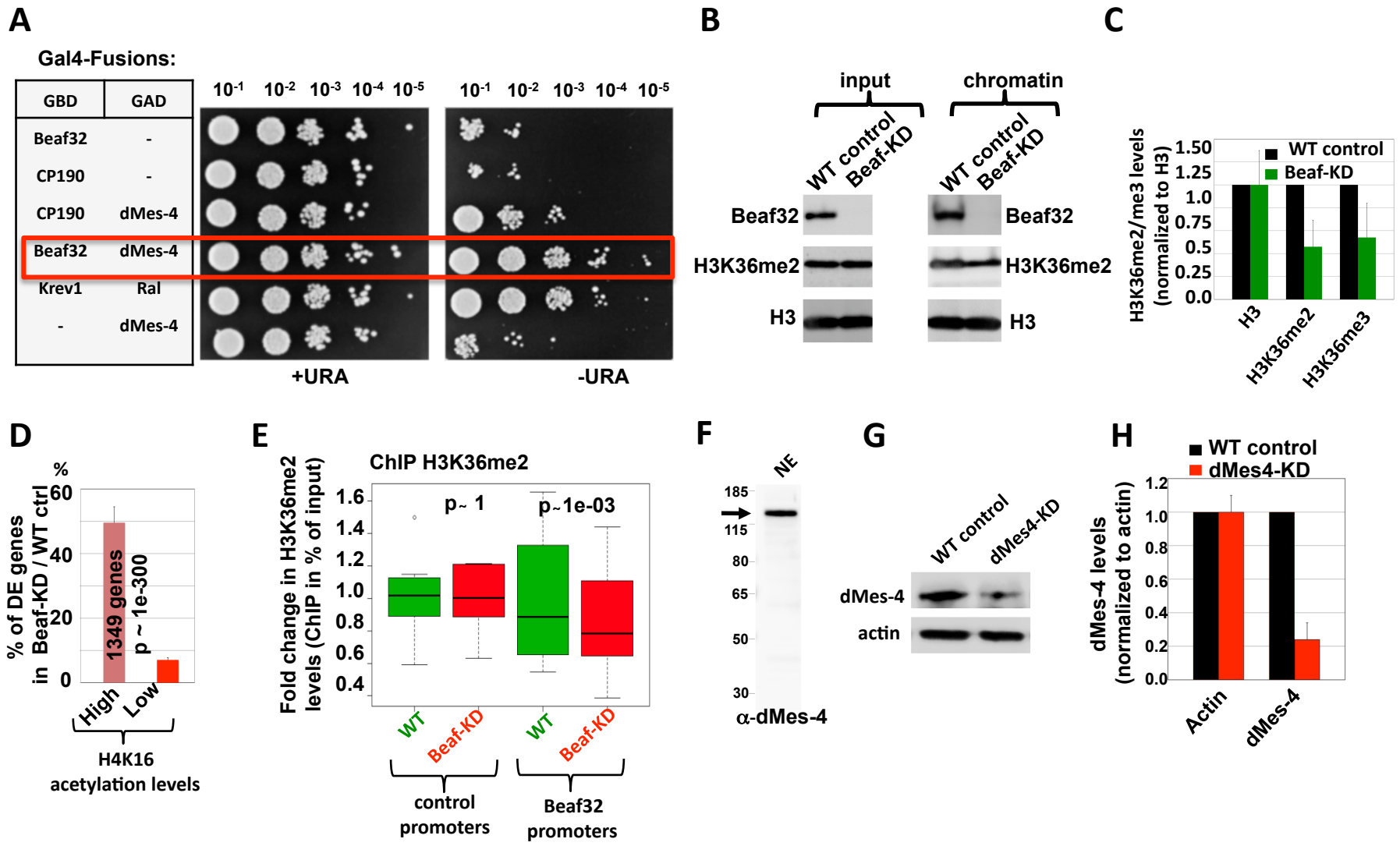
Lhoumaud_ Figure E1



Lhoumaud_ Figure E2



Lhoumaud_ Figure E3



Lhoumaud_ Figure E4

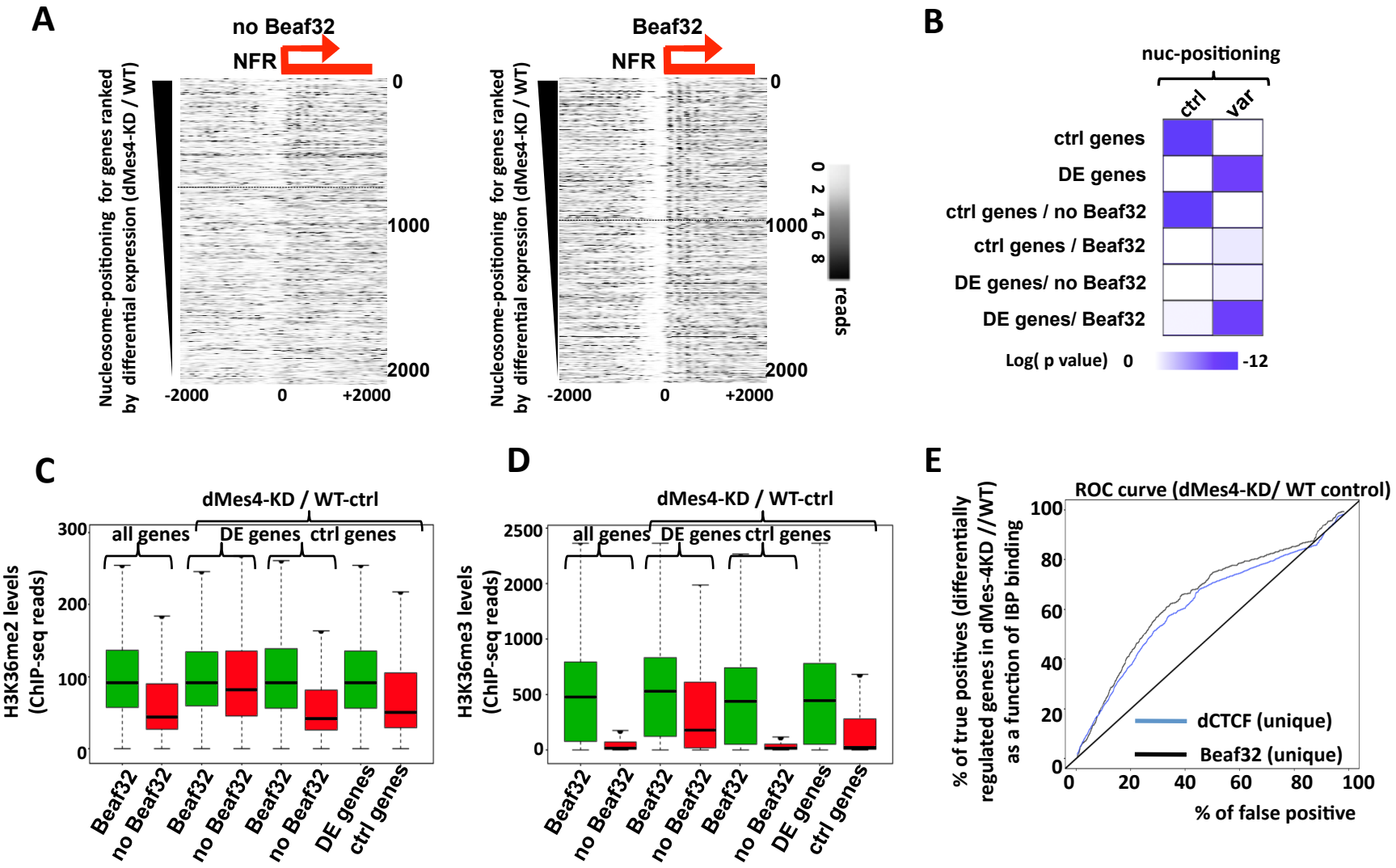
A

Gene Ontology (GO)	Depletion	p-value	Number of genes
Cell proliferation	dMes-4	2.24E-08	88
	Beaf32	9,18E-08	78
Cell-death	dMes-4	6.24E-05	57
	Beaf32	1.90E-08	54
Nucleoside triphosphate metabolic process	dMes-4	~1	
	Beaf32	~1	

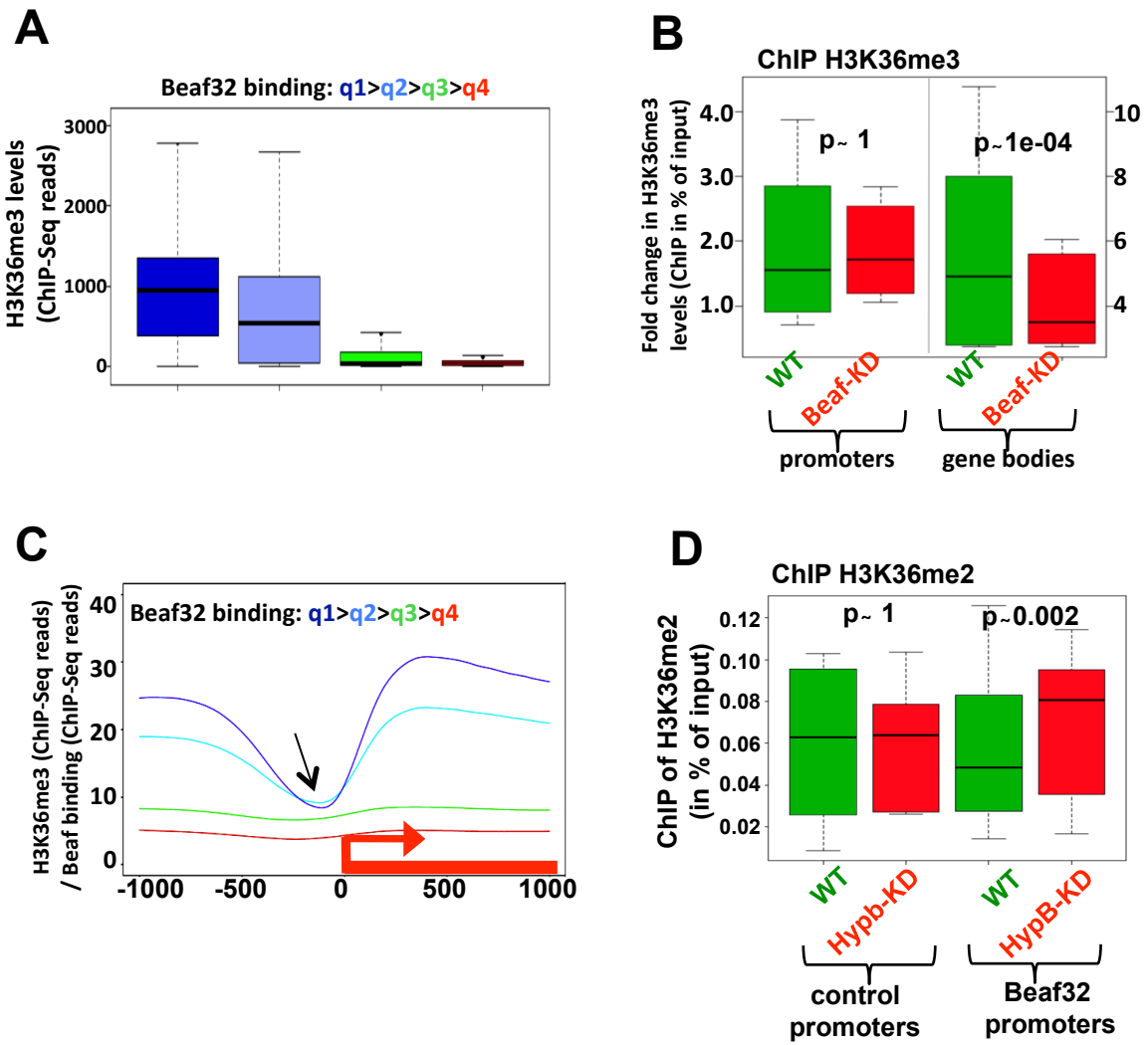
B

Factor	Consensus	Enrichment (E-value)
Beaf32	CGATA	1^{e}-60
DREF	TATCGATA	1^{e}-59

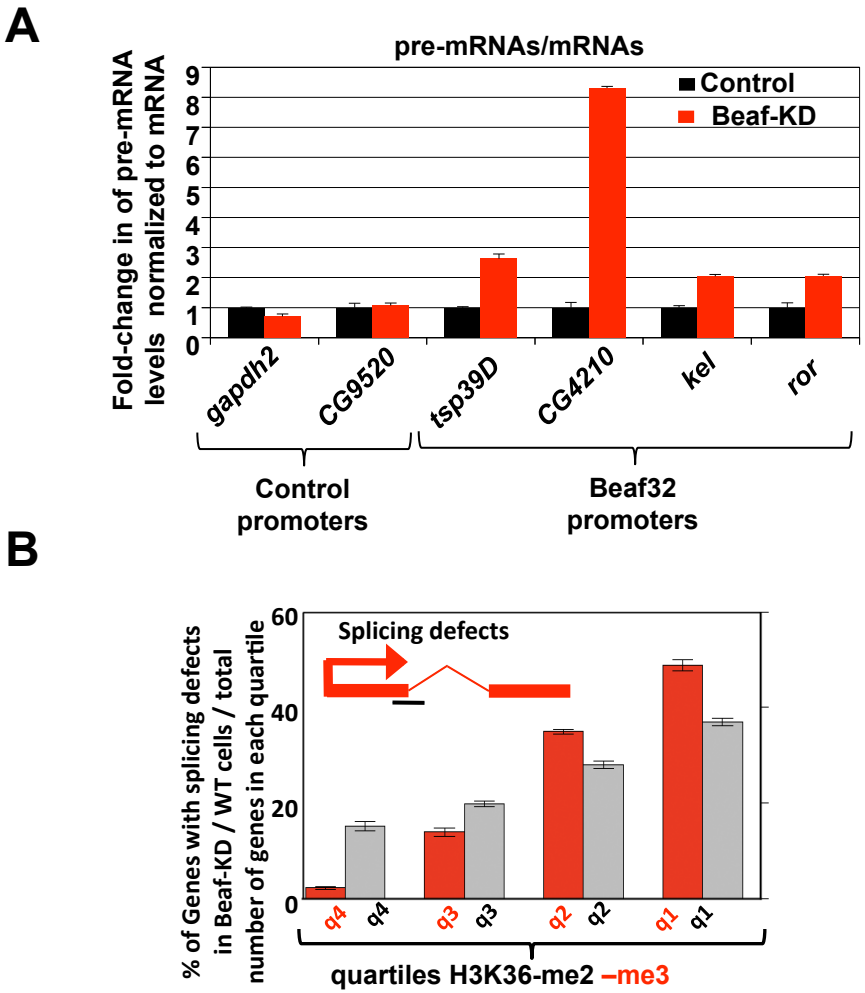
Lhoumaud_ Figure E5



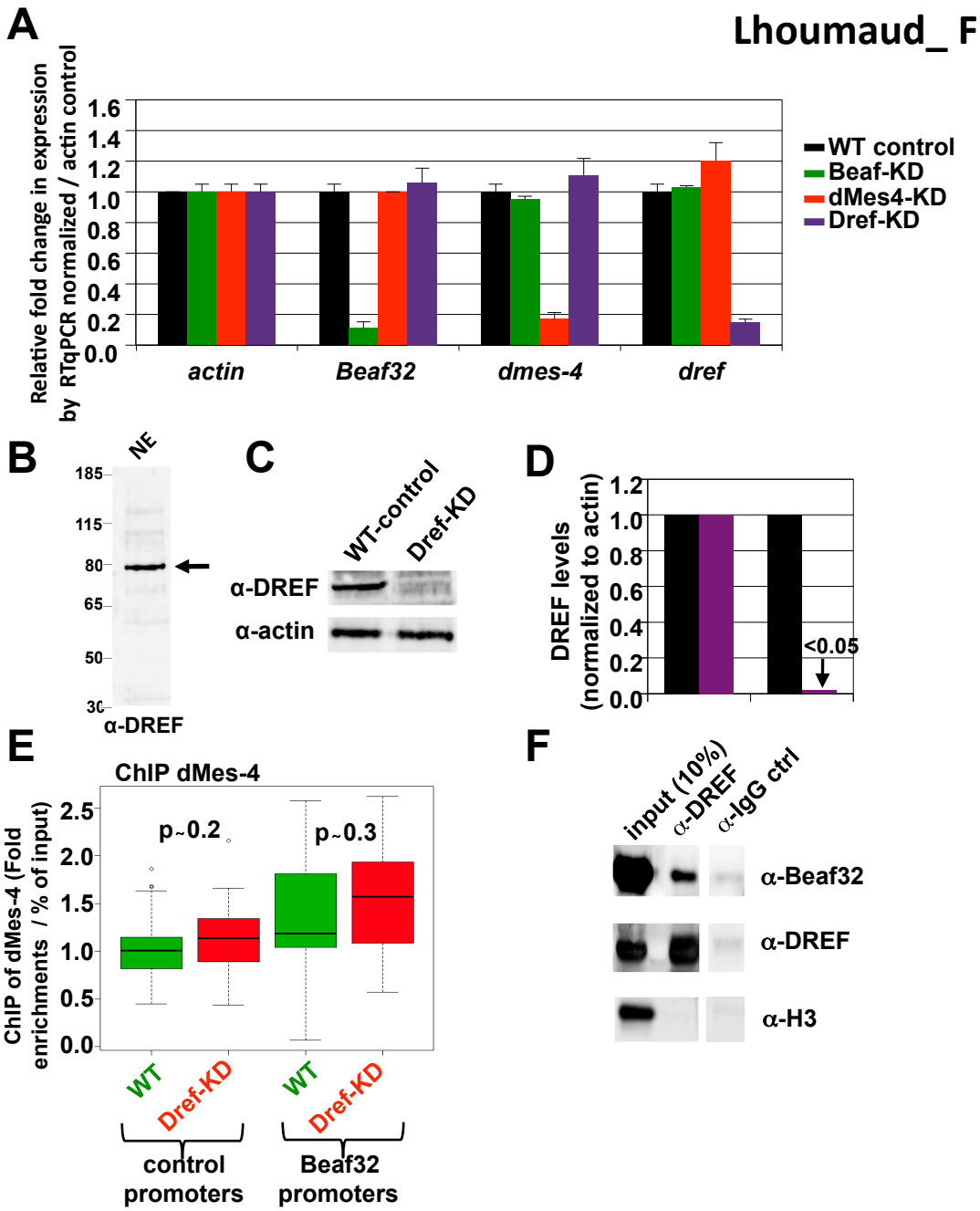
Lhoumaud_ Figure E6



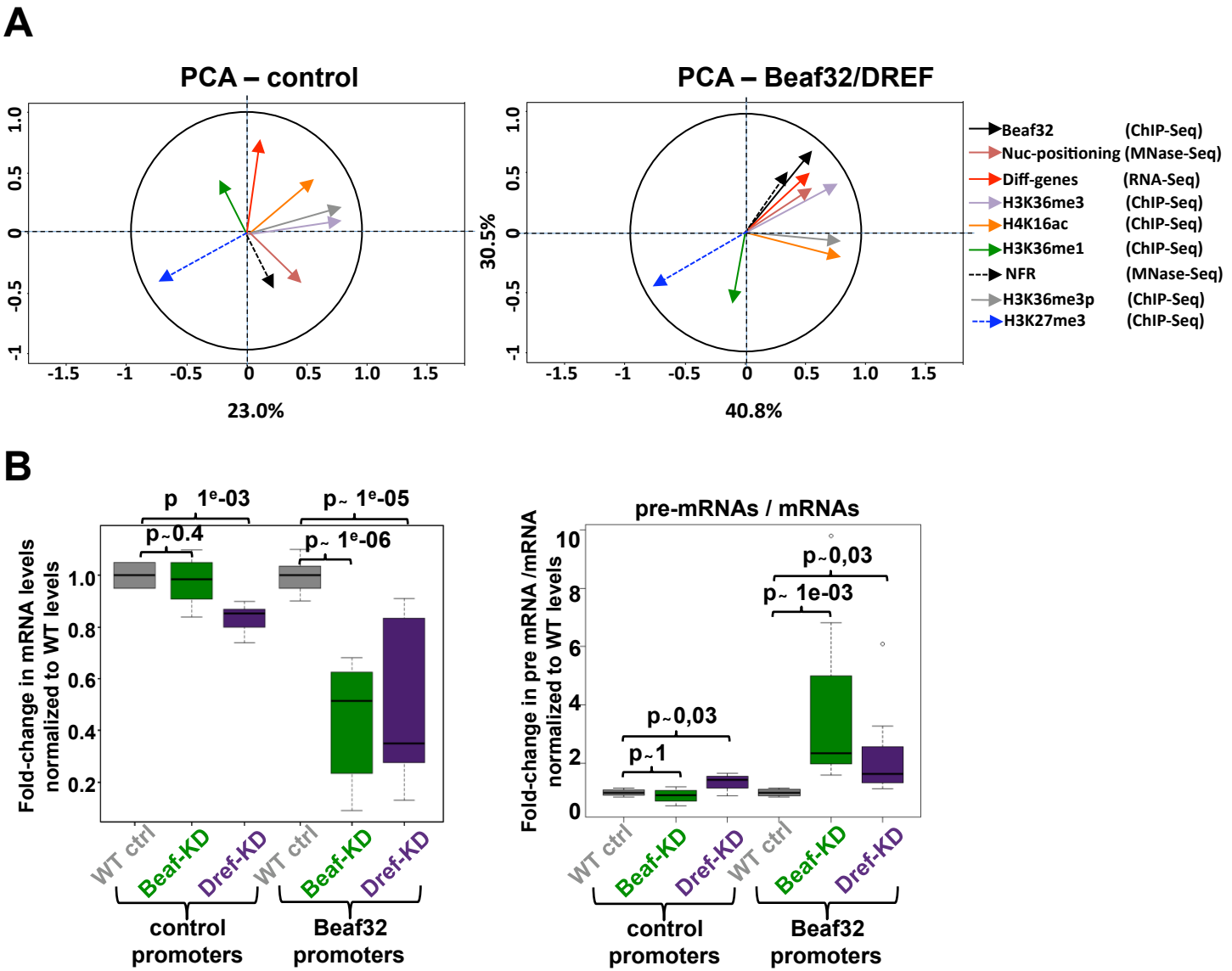
Lhoumaud_ Figure E7



Lhoumaud_ Figure E8



Lhoumaud_ Figure E9



Lhoumaud_ Figure E10