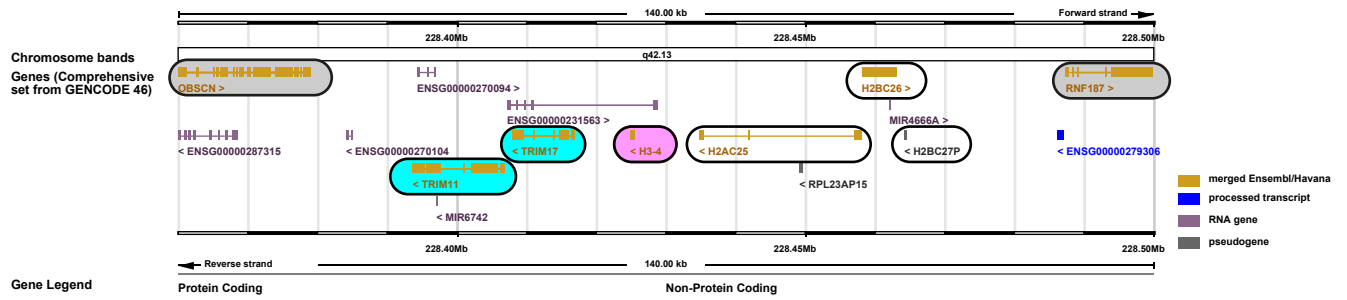
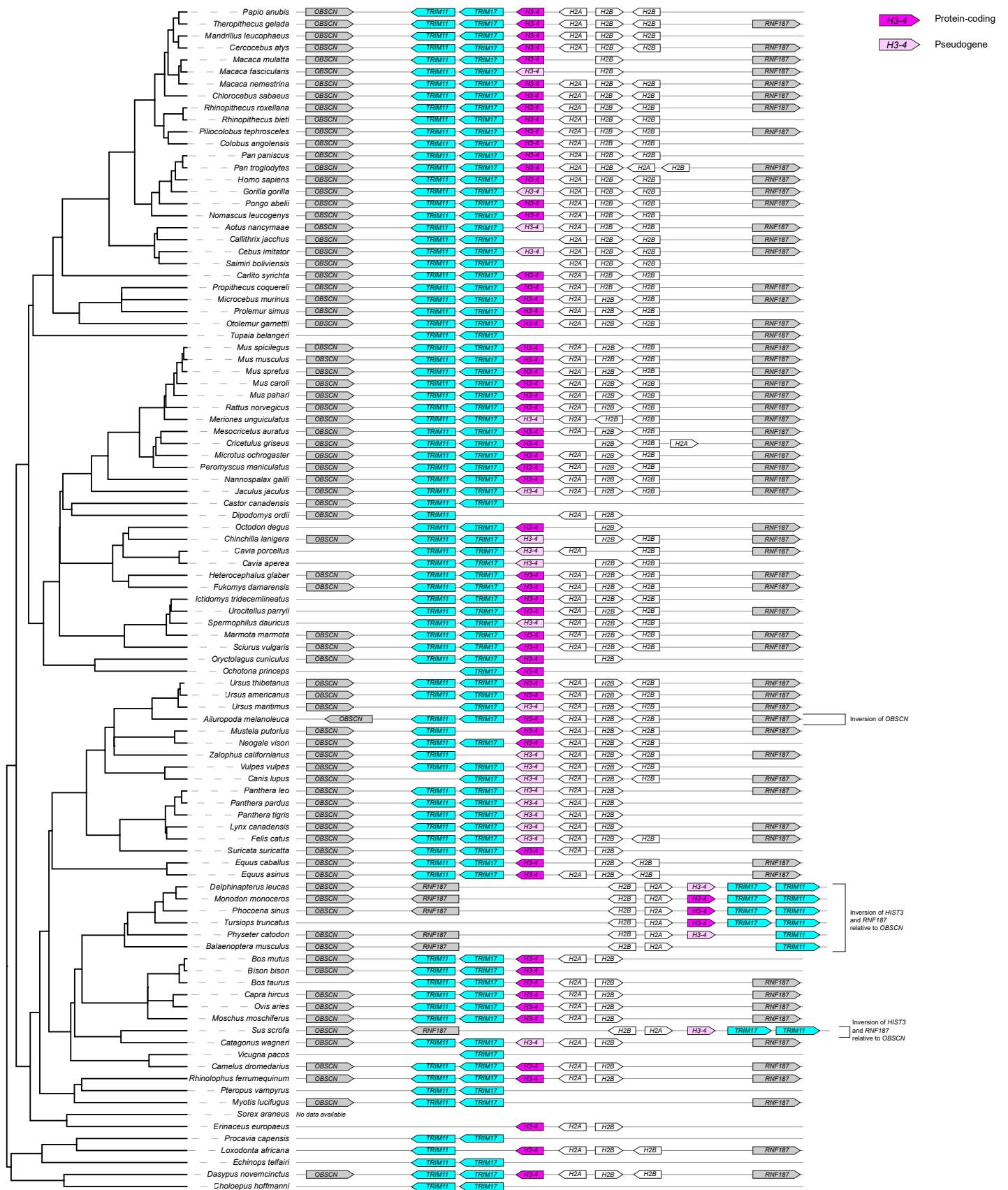


The eutherian-specific histone H3.4 promotes germ cell development and reproductive fitness

Pavel A. Komarov, Philipp Bammer, Ching-Yeu Liang, Hans-Rudolf Hotz, Grigorios Fanourgakis, Sunwoo Chun, Hubertus Kohler, Tim-Oliver Buchholz, Jean-Francois Spetz and Antoine H.F.M Peters

Supplementary Figures

aHuman *HIST3***b**

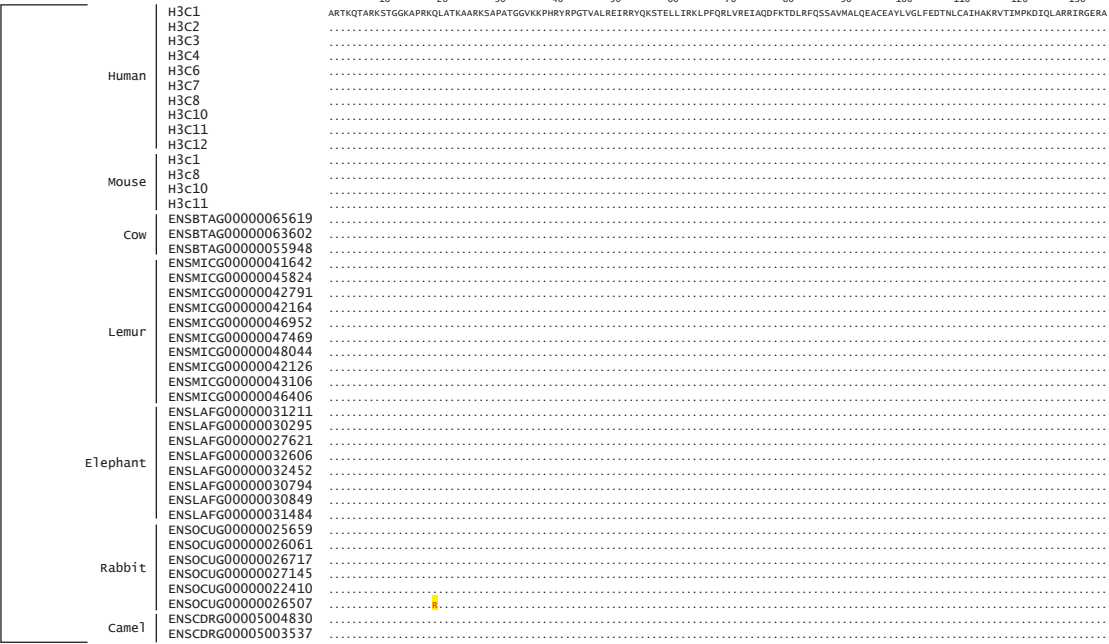
Supplementary Figure 1. Conservation of *H3.4*-containing gene clusters in selected mammalian species

a, Ensembl genome browser snapshots of the *H3-4* locus and neighboring genes in the human genome (Release 112). The link to human genome browser view is: https://may2024.archive.ensembl.org/Homo_sapiens/Location/View?db=core;r=1:228360000-228500000

b, Organization of *HIST3* - *TRIM* clusters in eutherian species based on genome data in ENSEMBL (release 112) and listed in Supplementary Data 1. The order of species was determined by TimeTree (<https://timetree.org>). *TRIM17* and *TRIM11* are displayed in cyan. *H3-4* protein-coding orthologs and pseudogenes are presented in purple and light-purple, respectively. Neighboring *H2A* and *H2B* genes are in white and flanking *OBSCN* and *RNF187* genes are shown in gray. Additional non-histone genes, if present within or outside the cluster, are not displayed. If a gene was not annotated with a name but only e.g. with an ENSEMBL ID, we performed a BLASTP search to determine its identity and to define its name. The coordinates of *H3-4* orthologs and details on sequences in different species are available in Supplementary Data 1. For species where we were unable to find *H3-4* orthologs, we included the configuration of *TRIM*, other histone and flanking genes. We used the same color annotation for the different genes as shown in Fig. 1b.

a

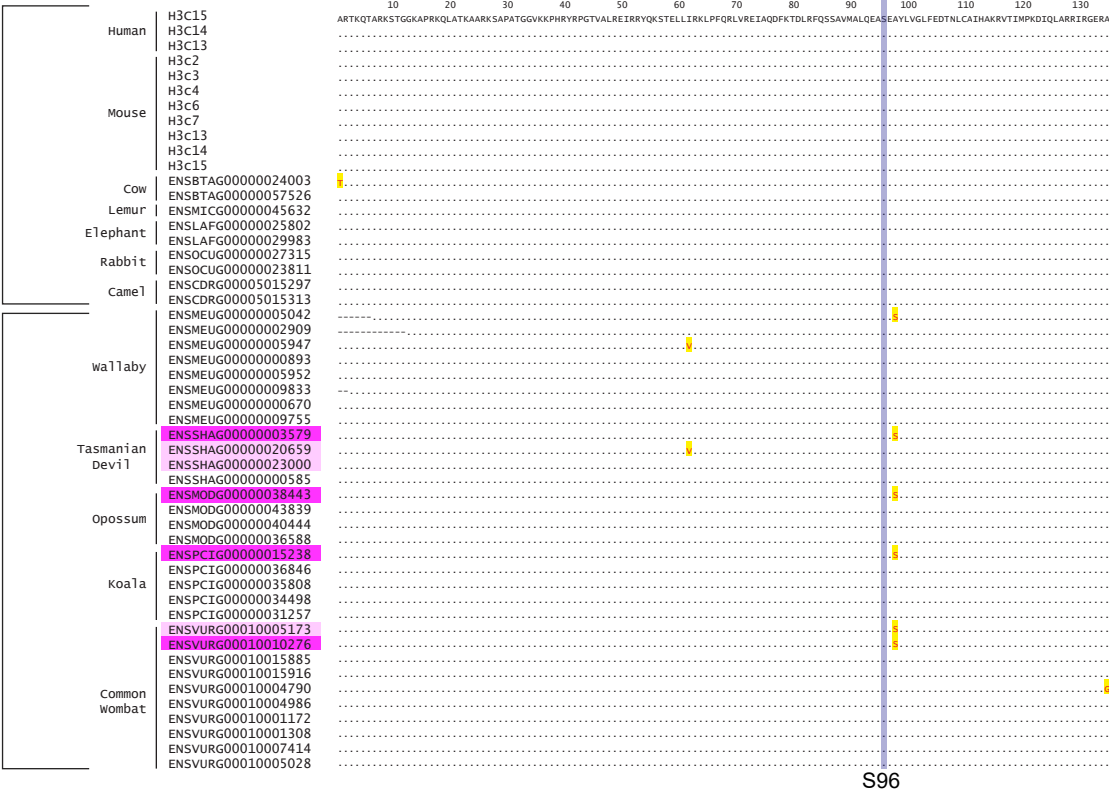
Eutherians



b

Eutherians

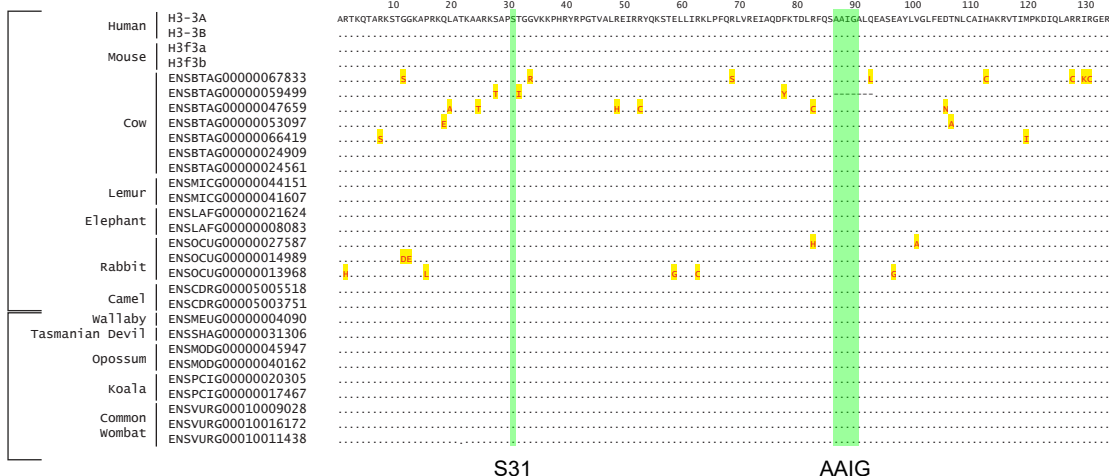
Marsupials



c

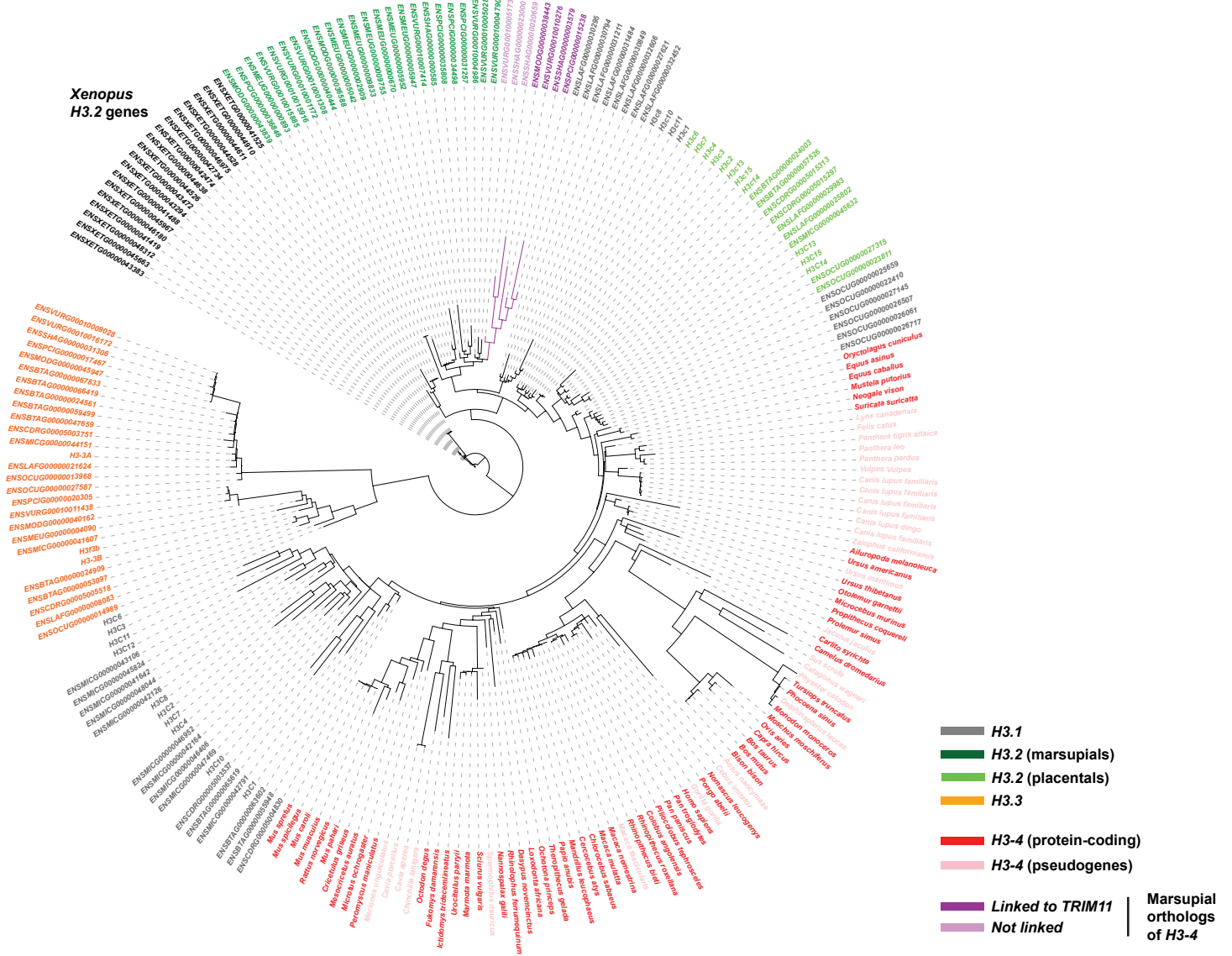
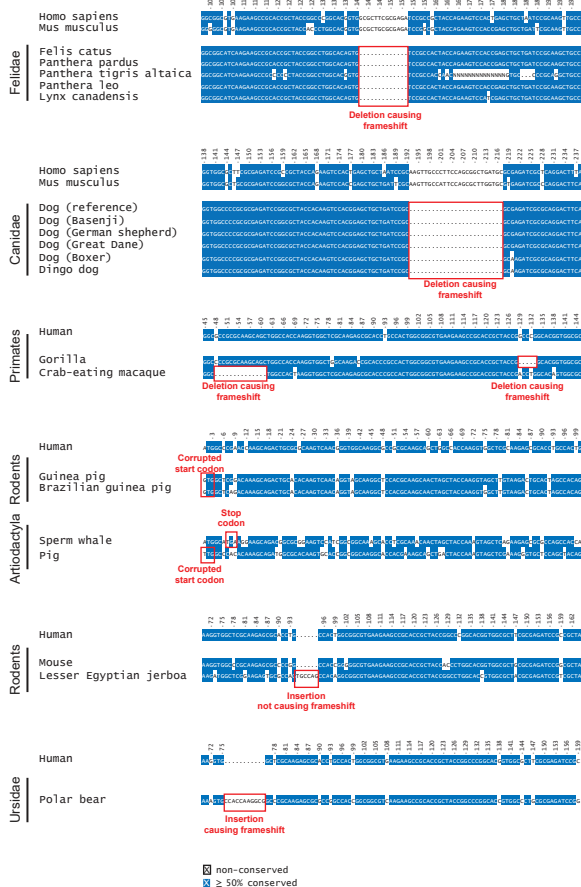
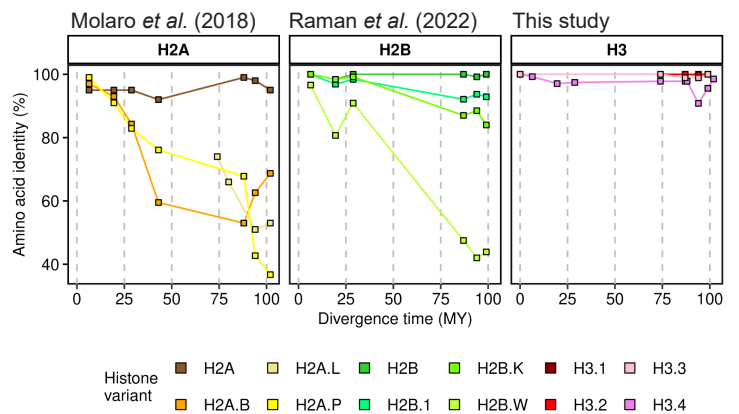
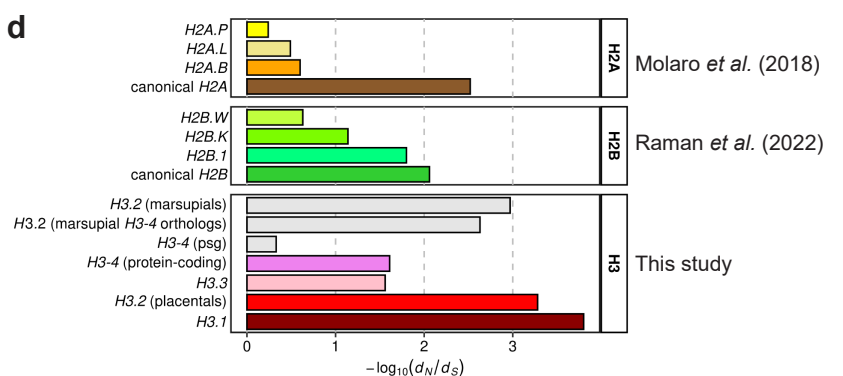
Eutherians

Marsupials



Supplementary Figure 2. Protein sequences of H3.1, H3.2 and H3.3 in selected placental and marsupial mammals

Amino acid sequence alignments of H3.1 (a), H3.2 (b) and H3.3 (c) proteins in selected placental and marsupial species. Marsupial *H3.2* genes that are putative orthologs of placental *H3-4* and are linked to *TRIM11* genes are highlighted in purple, while those not located adjacent to *TRIM-11* genes are shown in light-purple. For H3.2 proteins, S96 and S98 residues are highlighted in blue and yellow respectively. For H3.3 proteins, S31 and AAIG residues are highlighted in green. Moreover, variant amino acids are highlighted in yellow.

a**b****c****d**

Supplementary Figure 3. Analysis of sequence divergence and evolutionary dynamics of *H3* genes

a, Cladogram showing the results of maximum-likelihood analysis of *H3* genes. For *H3-4* protein-coding genes (red) and pseudogenes (light-red), sequences from all species were used, if available (Supplementary Data 1). The *H3.1*, *H3.2* and *H3.3*-encoding genes (gray, green and orange, respectively) were retrieved from the selected species (Supplementary Data 2). Marsupial *H3.2* genes that are putative orthologs of placental *H3-4* and are linked to *TRIM11* genes are highlighted in purple, while those not located adjacent to *TRIM11* genes are shown in light-purple.

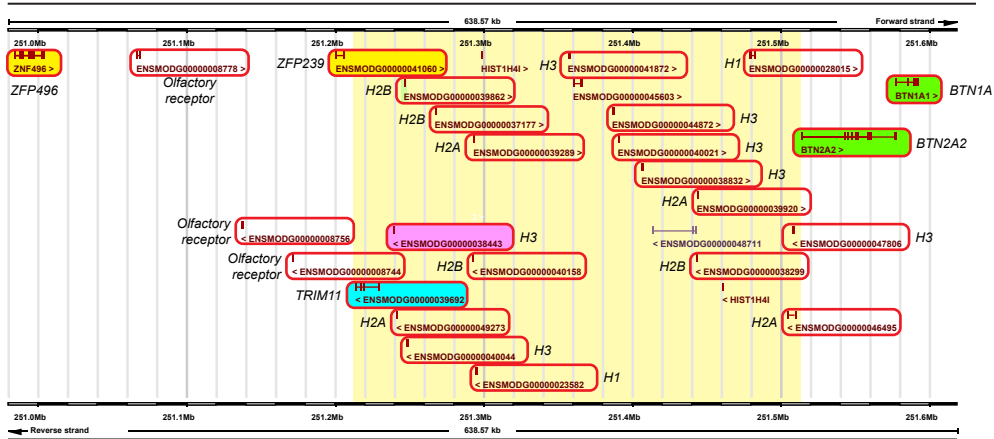
b, Examples of truncations, insertions and nonsense mutations of DNA sequences of *H3-4* pseudogenes in different species. Human *H3-4* and mouse *H3f4* sequences are given as references on the top.

c, Analysis of protein sequence variation of *H3* variants relative to human. As comparative references, variation rates of several *H2A* and *H2B* variants were added from published studies^{7,8}. Details of this analysis are available in Supplementary Data 3.

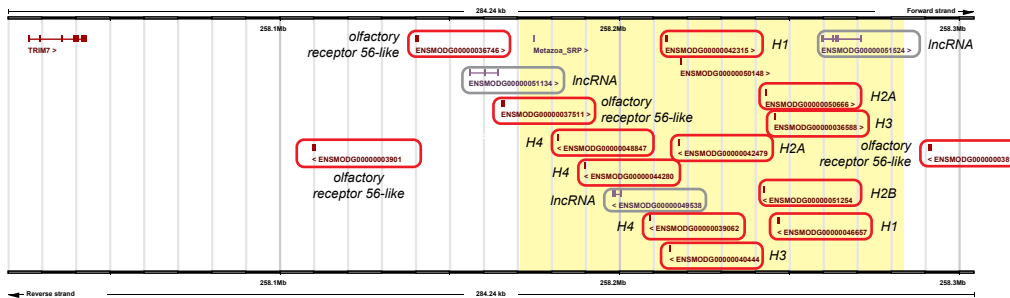
d, Phylogenetic Analysis Using Maximum Likelihood (PAML) of *H3* genes showing ratios of nonsynonymous to synonymous mutation rates (d_n/d_s) ($-\log_{10}$ -transformed). Separate analyses were done for *H3.1*, *H3.2* and *H3.3* genes in selected species, as well as for protein-coding *H3-4* orthologs and their pseudogenes. More detailed results of statistical testing are provided in Supplementary Data 4. As comparative references, d_n/d_s values of several *H2A*- and *H2B*-encoding genes were added from published studies^{7,8}. Color coding of genes is the same as in panel c.

Locations of *H3* genes in opossum

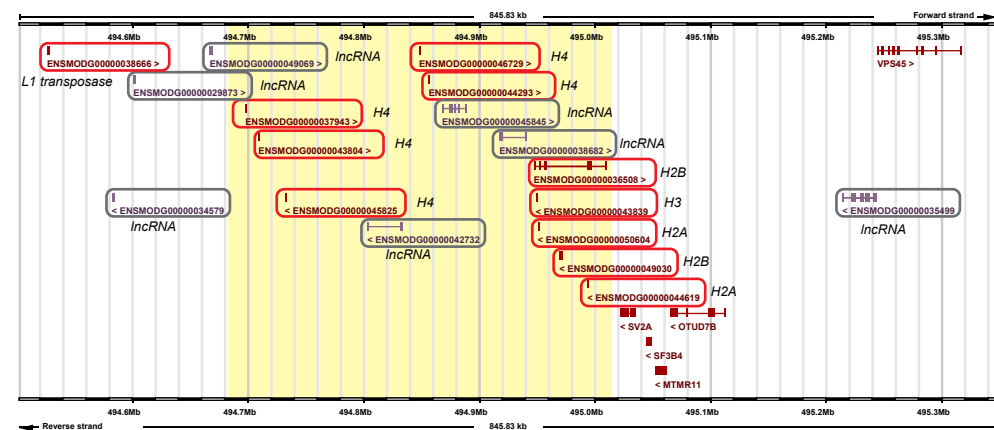
Cluster 1



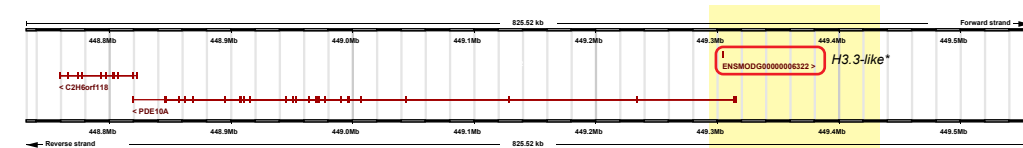
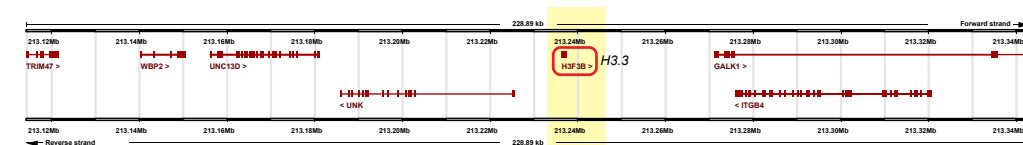
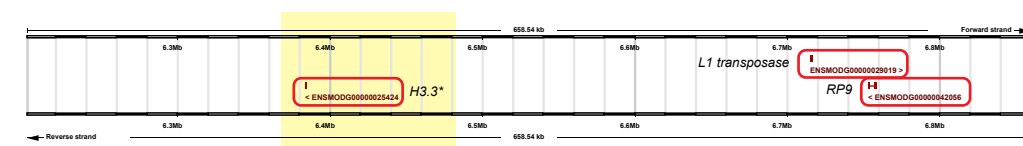
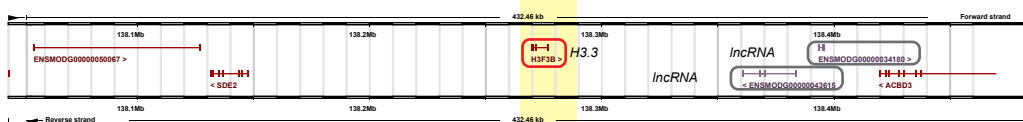
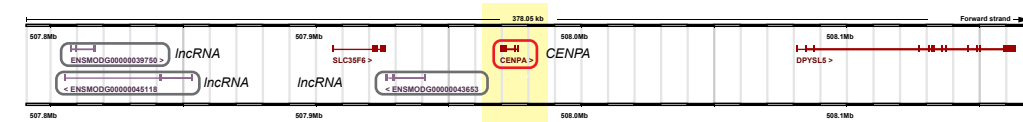
Cluster 2



Cluster 3



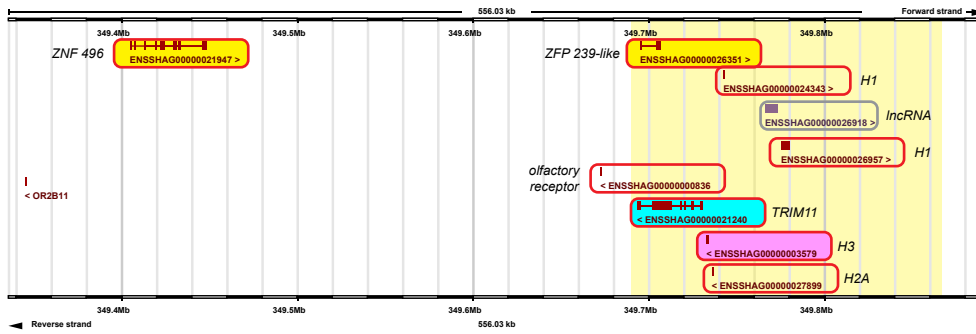
Standalone *H3* genes



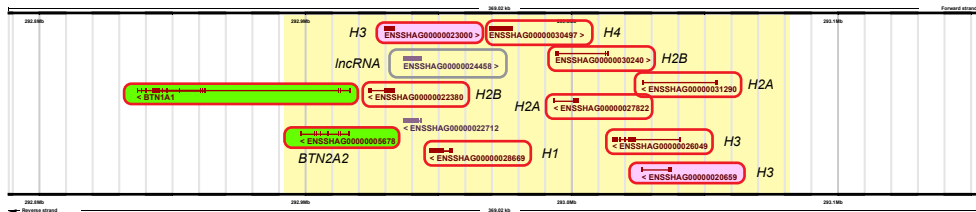
Supplementary Figure 4. Locations of *H3* genes in opossum

The opossum genome (ENSEMBL Release 112) harbors three large histone clusters with replication-dependent *H2A*, *H2B*, *H3* and *H4* genes as well as four *H3.3* loci and one *CENPA* locus. In the Ensembl genome browser snapshots of cluster 1 (on chr. 2: 250,981,589 - 251,620,156), *TRIM11* and *H3.2* (the putative *H3-4* ortholog) are highlighted in cyan and purple, respectively. Neighboring *ZFP239* and *ZFP496* are yellow and *BTN2A2* and *BTN1A1* are green. For each location, the respective coordinates are shown on the right.

Tasmanian devil

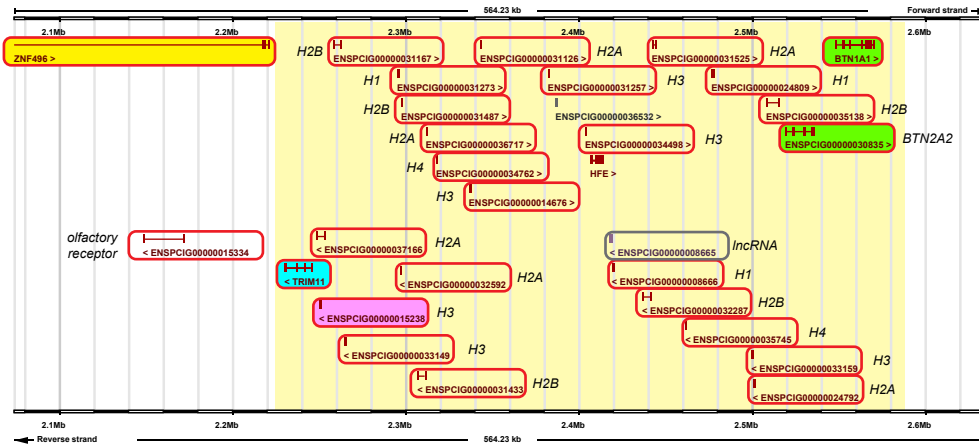


chr1:
349,335,689-349,891,722



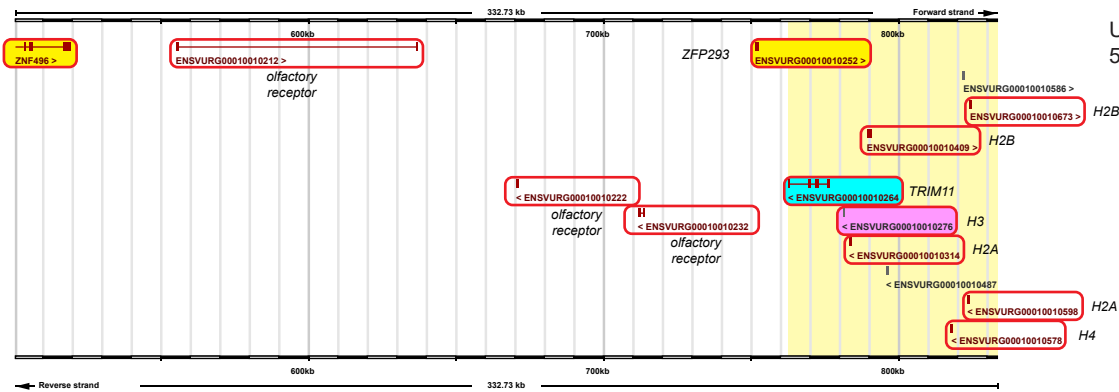
chr4:
292,788,488-293,157,511

Koala

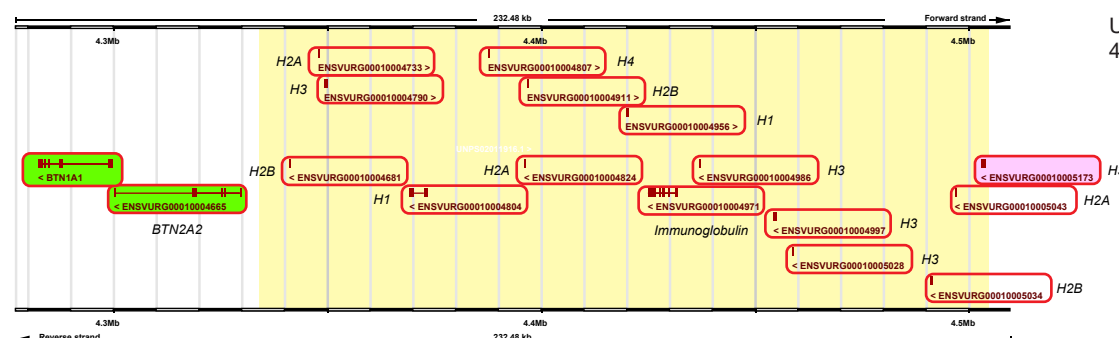


MSTS01000168.1:
2073915-2638140

Common wombat



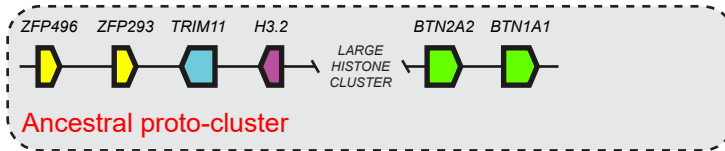
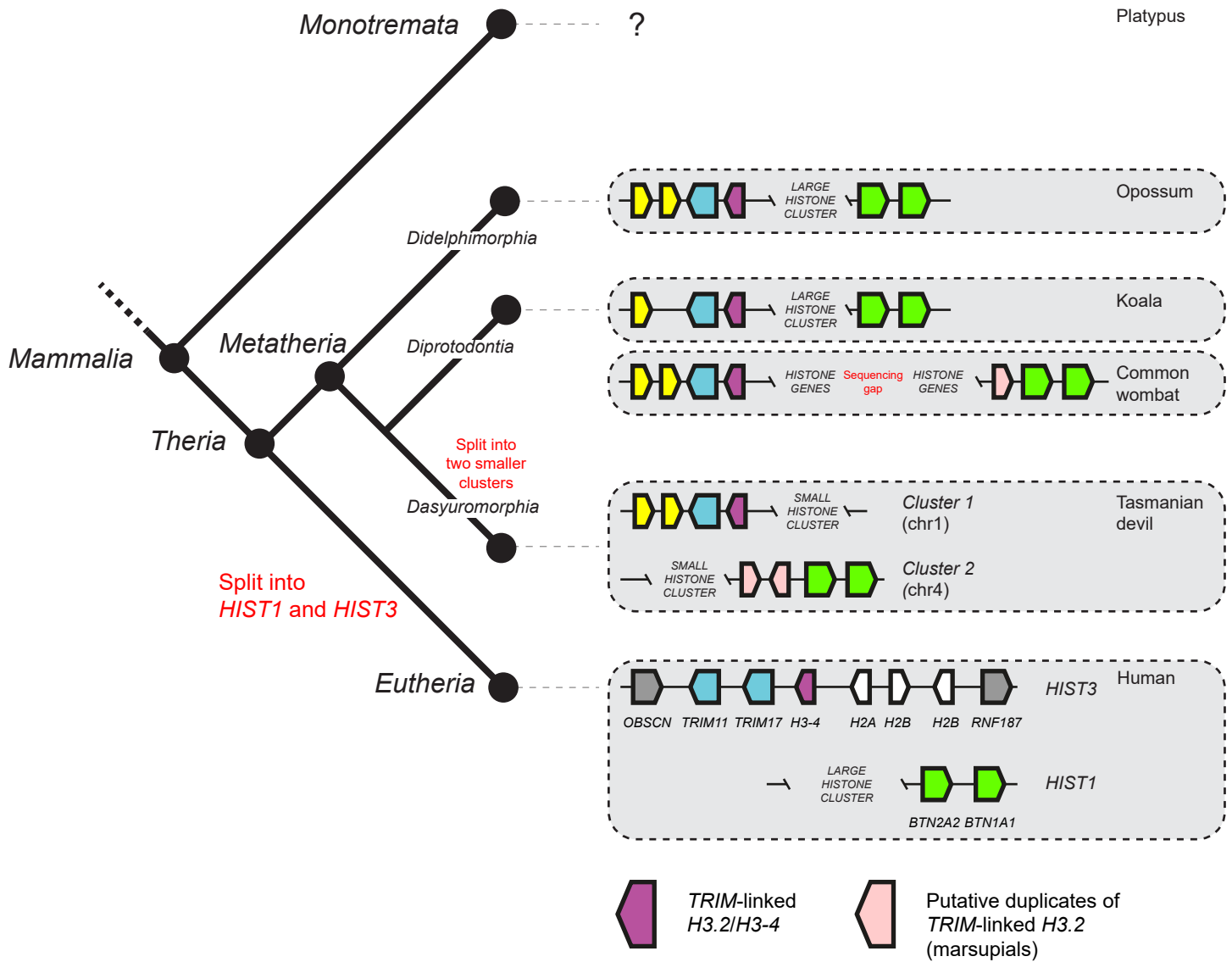
UNPS02014910.1:
500794-833521



UNPS02011916.1:
4,277,130-4,509,606

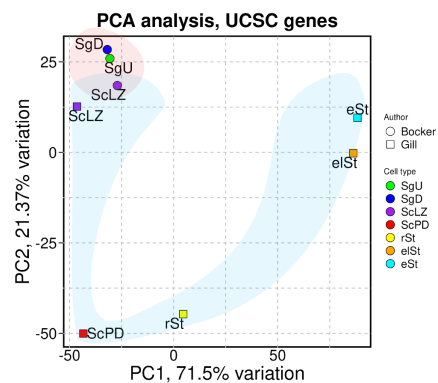
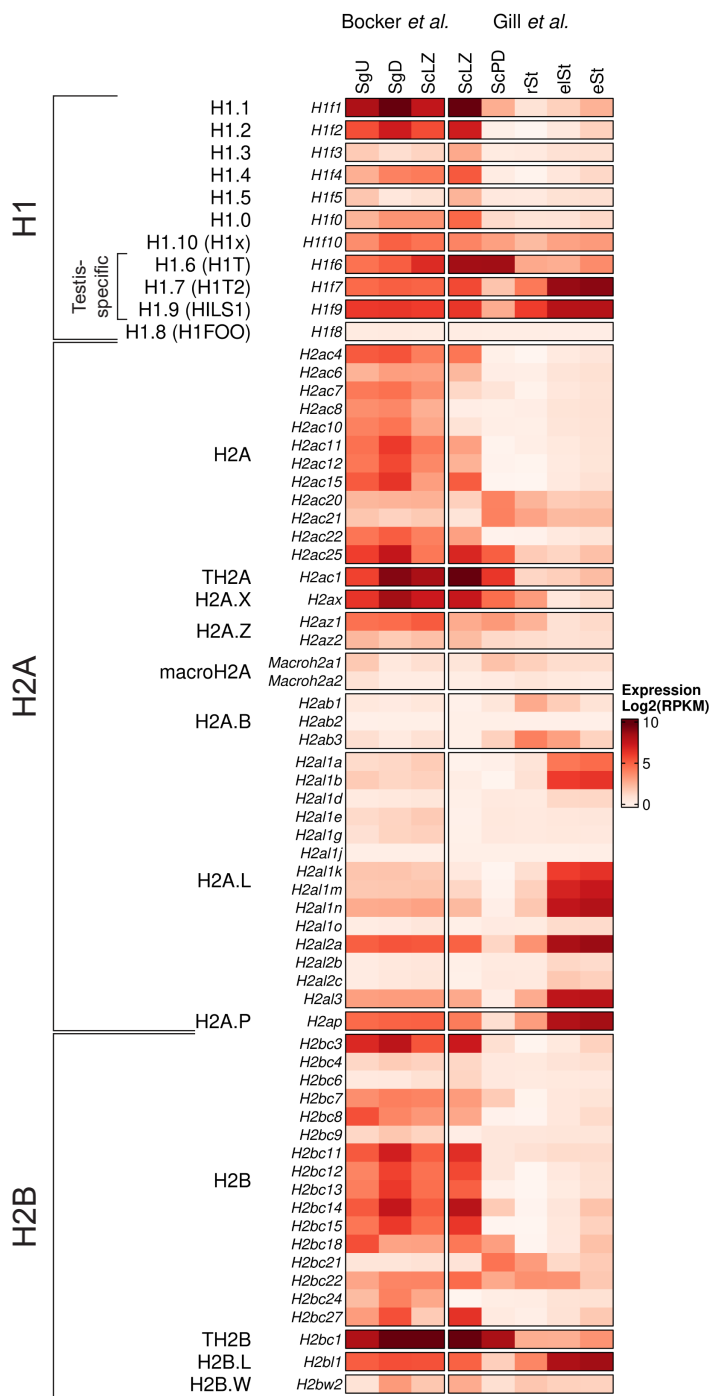
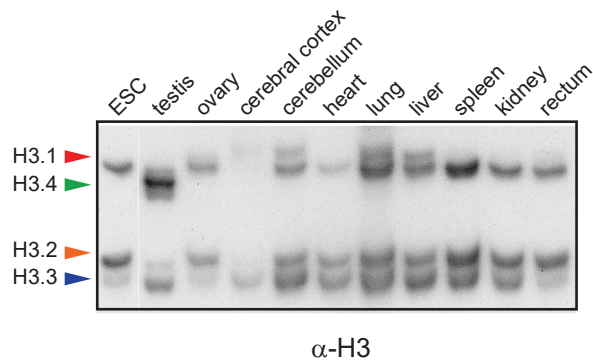
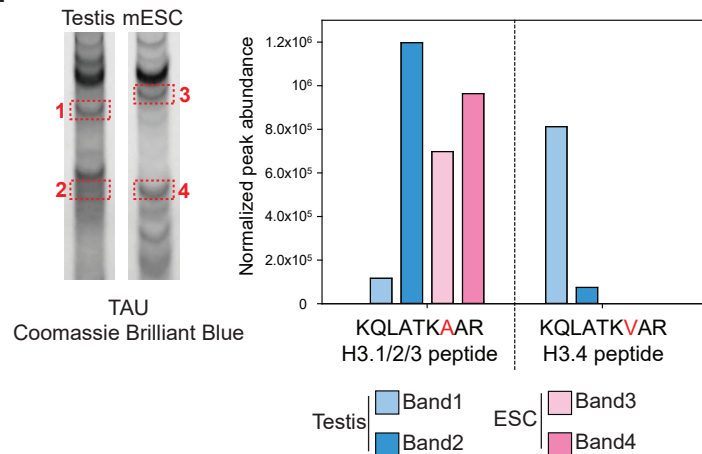
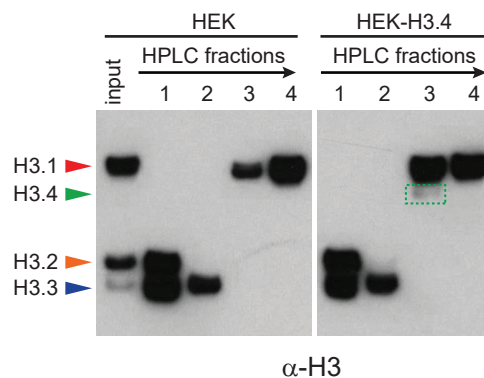
Supplementary Figure 5. Genomic locations of putative *H3-4* orthologs in Tasmanian devil, Koala and Common wombat

Ensembl genome browser snapshots of histone clusters harboring *H3.2* genes that are putative *H3-4* orthologs in Tasmanian devil, Koala, and Common wombat. The color annotation of genes is identical to the one shown in Supplementary Fig. 4. *H3.2* genes with sequence similarity to *H3.2* genes that are putative *H3-4* orthologs but that are not located next to *TRIM11* genes are displayed in light-purple. For each location, the respective coordinates are shown on the right.



Supplementary Figure 6. Evolution of the ancestral *TRIM11*-linked histone cluster and its derivatives in marsupials and eutherian species

Top: composition of an ancestral histone protocluster proposed to exist in the common ancestor of marsupials and eutherians. Schematic views of histone cluster derivatives existing in four marsupial species belonging to the Monotremata and Metatheria lineages and in Eutherians (human) are indicated below. In Common wombat, due to missing sequencing information, we were unable to establish whether the shown histone cluster is intact or split in two. In Tasmanian devil, two smaller histone clusters were detected on chr1 and chr4. In Wallaby, we could not obtain data due to poor assembly. Cladogram structure is inferred from TimeTree (<https://timetree.org>), nodes and branches are not scaled.

a**b****c****d****e**

Supplementary Figure 7. The expression of H3 variants in mouse tissues

a, Principal component analysis (PCA) comparing RNA-seq expression of UCSC-annotated genes in different cell types^{57,58}. The RNA-seq libraries of the two independent studies recapitulate the respective developmental trajectories during spermatogenesis.

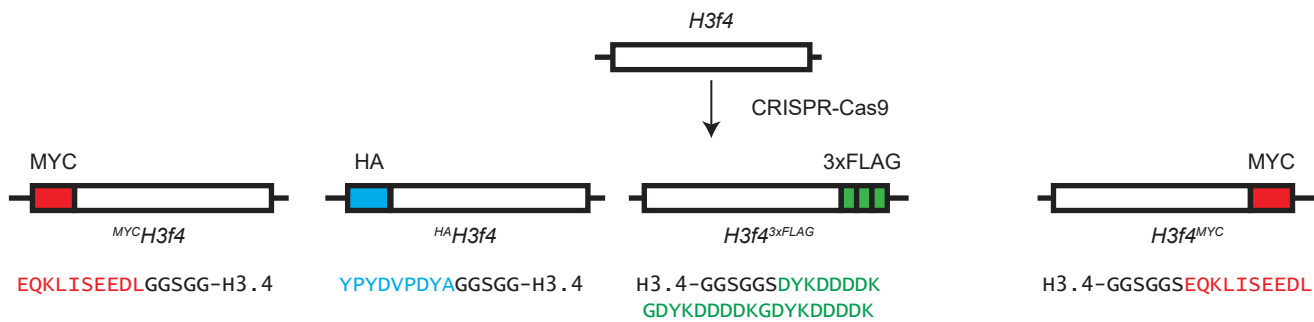
b, Heatmap showing mRNA expression H1, H2A and H2B encoding genes during spermatogenesis. The names of histone genes are according to the current nomenclature³⁴. Abbreviations: SgU, SgD, ScLZ, ScPD, rSt, elSt and eSt refer to populations of FACS-sorted undifferentiated and differentiating spermatogonia, leptotene/zygotene and pachytene/diplotene stage spermatocytes and round, elongating and elongated spermatids.

c, Western blot showing H3.1, H3.4, H3.2 and H3.3 proteins separated by TAU gel electrophoresis in extracts of mouse embryonic stem cells and various tissues isolated from adult mice. A panH3 antibody recognizing the C-terminus of H3 as used to detect all H3 variants.

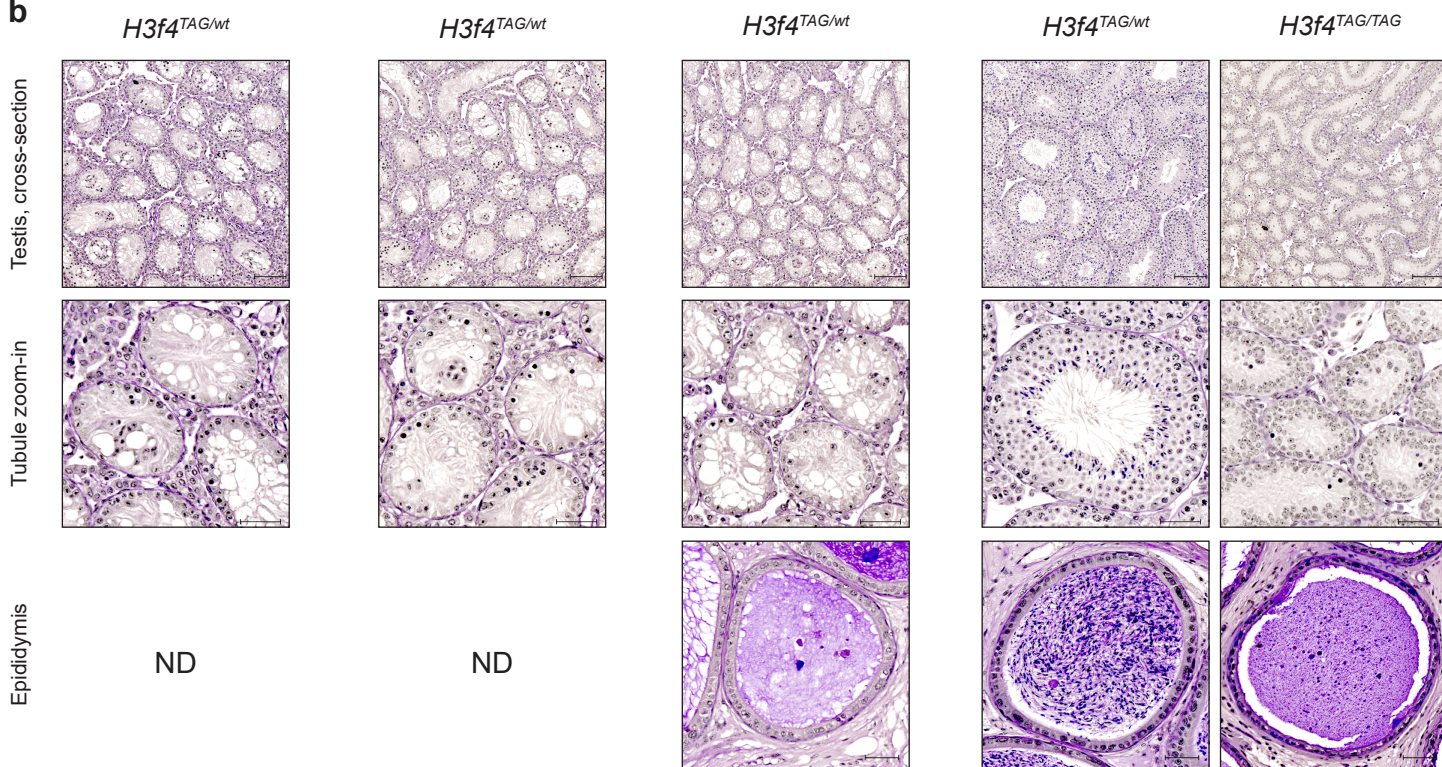
d, Mass spectrometry analysis of histones isolated from testes and mESCs. We separated histones by TAU gel electrophoresis and stained them with Coomassie Brilliant Blue. We excised histone protein containing bands and subjected them to mass spectrometry. We quantified normalized peak abundances corresponding to the H3.1/H3.2/H3.3 (KQLATKA₂₄AR, left) common peptide and the H3.4 (KQLATV₂₄AR, right) specific peptide in the four excised bands. The diagnostic amino acids of the two peptides are labeled in red. Band 1 contains: H3.4, band 2: H3.3, band 3: H3.1 and band 4: H3.3.

e, Proteins blots showing the migration characteristics of H3.4 relative to other H3 proteins separated by TAU gel electrophoresis. We prepared histone extracts from control HEK293 cells or HEK293 cells expressing H3.4. We obtained different H3 protein containing fractions by reverse phase HPLC separation and subjected them to TAU gel electrophoresis. Fraction 3 contains the H3.4 protein.

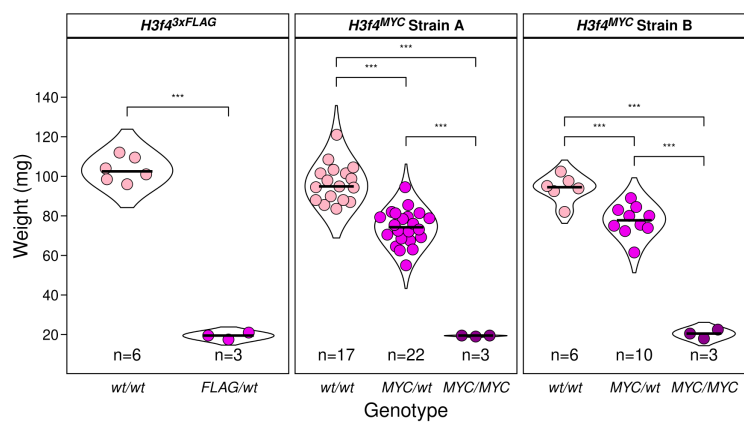
a



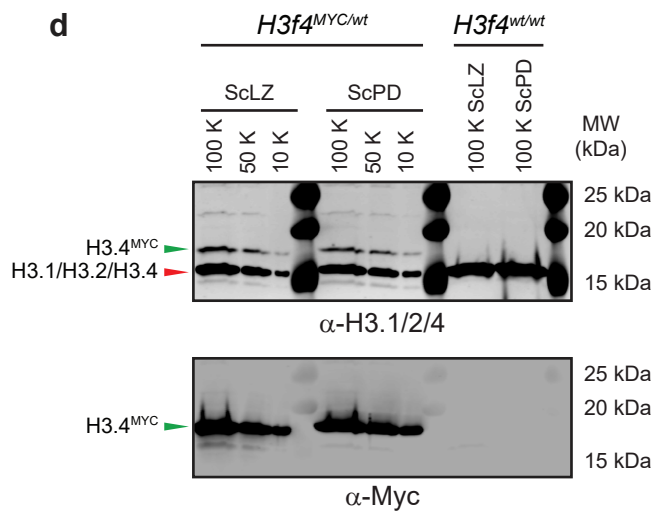
b



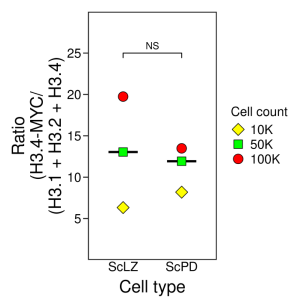
C



d



e



Supplementary Figure 8. CRISPR/Cas9 approaches to tag the endogenous *H3f4* gene

a, Schematic overview of four different transgenic mouse lines generated by CRISPR/Cas9-mediated tagging of the endogenous *H3f4* gene. The sequences of protein tags are displayed in different colors and linker sequences in black. Sequences of sgRNAs and repair templates are provided in Supplementary Data 6 and 7, respectively.

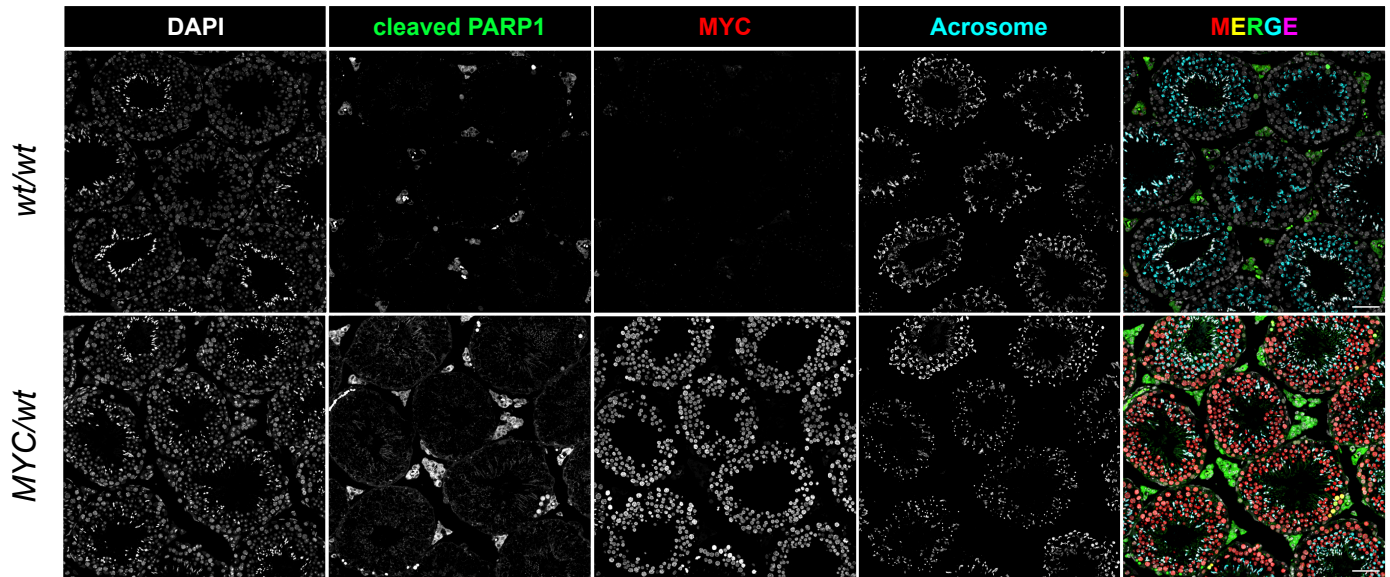
b, PAS-haematoxylin-stained histological cross-sections of testis and epididymal tissues obtained from mice bearing tagged *H3f4* genes. Scale bars: 100 μ m (testis, cross-section overview) and 50 μ m (tubule zoom-in and epididymis). ND: not determined.

c, Violin plots showing testicular weights of control, heterozygous and homozygous transgenic mice of the *H3f4*^{3xFLAG} and two independent *H3f4*^{MYC} strains. In the violin plots, each dot represents the mean weight of two testes taken from each mouse. The number (n) of mice per genotype are indicated. Two-sided t-test was performed. When three genotypes were compared, Bonferroni correction was used. *** $P \leq 0.001$.

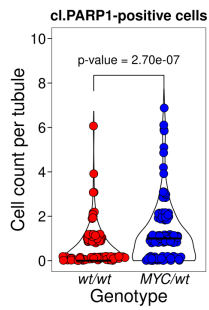
d, SDS-PAGE Western blot analysis of H3.4^{MYC} expression in FACS-purified ScLZ and ScPD cells of wildtype *H3f4*^{wt/wt} and heterozygous *H3f4*^{MYC/wt} mice. Blots were probed with antibodies recognizing H3.1/H3.2/H3.4 (upper panel) and MYC (lower panel) epitopes. Bands corresponding to the tagged H3.4^{MYC} and untagged H3.1, H3.2 and H3.4 proteins are indicated by arrows.

e, ImageJ quantification of bands in panel d. Ratios of H3.4^{MYC} (upper band) relative to the band corresponding to the untagged H3 proteins (lower). Separate data points correspond to lanes with 100K, 50K and 10K cells. Two-sided t-test was performed to compare the values of ScLZ and ScPD bands. NS – not significant.

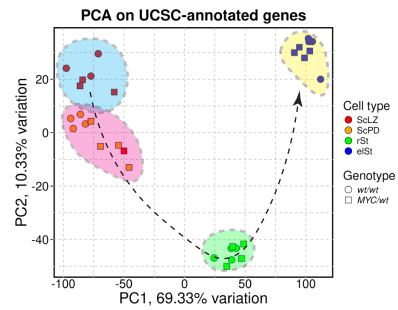
a



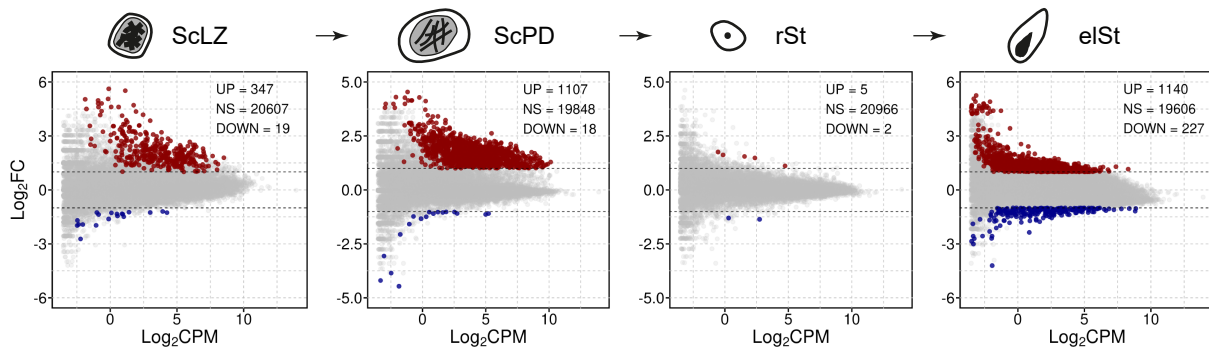
b



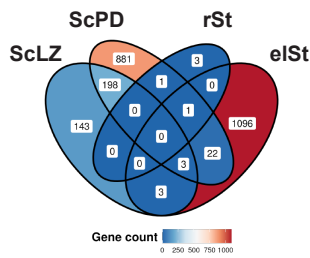
c



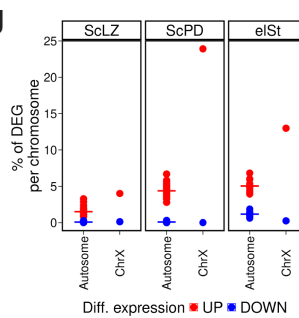
d



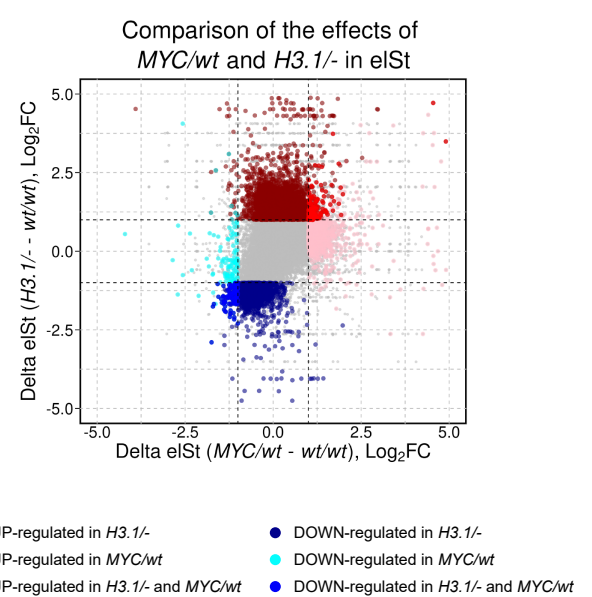
e



g

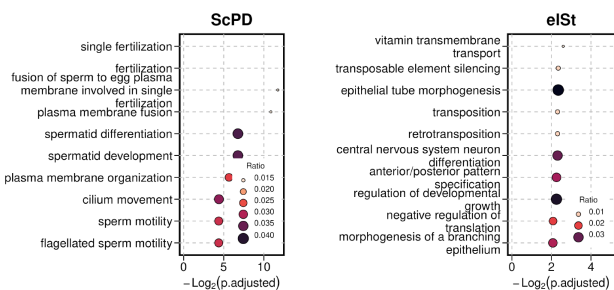


h



f

GO: Biological Process



Supplementary Figure 9. Cell death and transcriptional mis-regulation in germ cells expressing MYC-tagged H3.4

a, Immunofluorescence images showing cleaved PARP1 and MYC signals in sections of seminiferous tubules of control ($H3f4^{wt/wt}$) and heterozygous ($H3f4^{MYC/wt}$) mice. Scale bar: 50 μ m.

b, Quantification of cells positive for cleaved PARP1, as displayed in panel a. The number of tubules analyzed: 159 for $H3f4^{wt/wt}$ and 154 for $H3f4^{MYC/wt}$. *p*-value of paired t-test comparison is displayed.

c, Principal component analysis (PCA) comparing RNA-seq expression of UCSC-annotated genes among of biological replicates and cell types from $H3f4^{wt/wt}$ and $H3f4^{MYC/wt}$ animals.

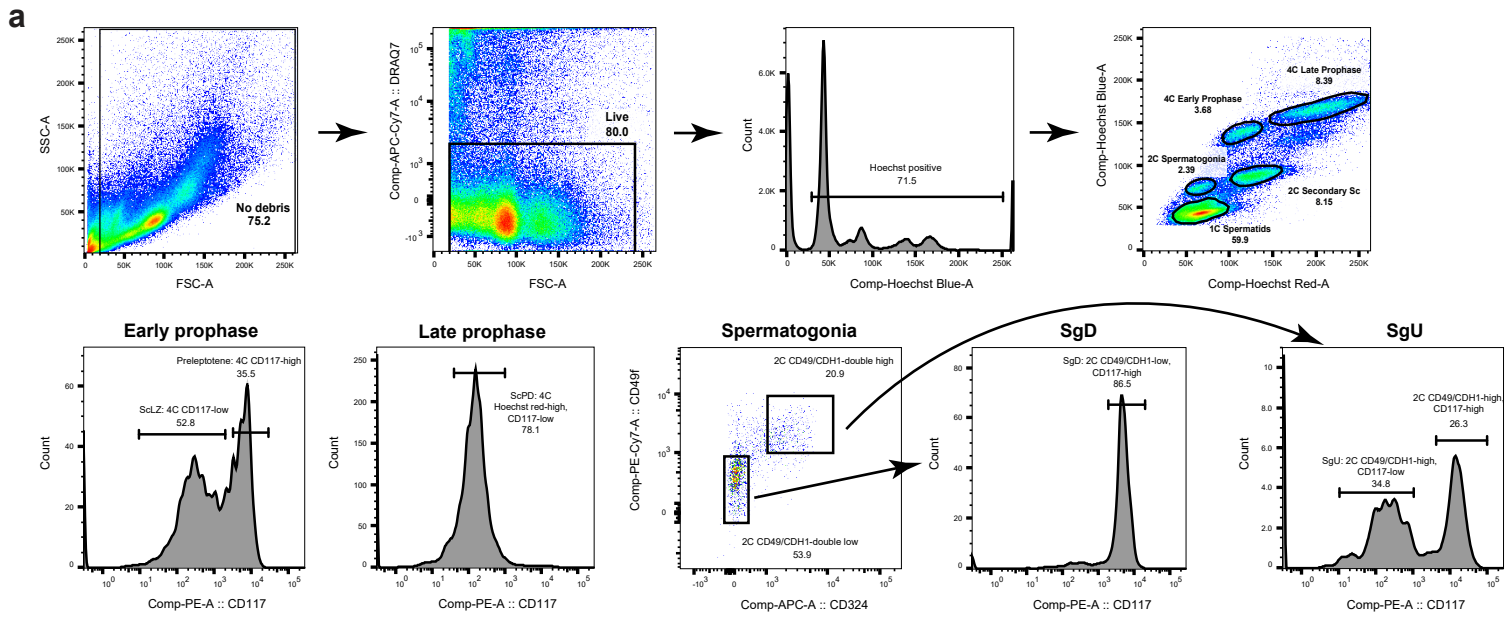
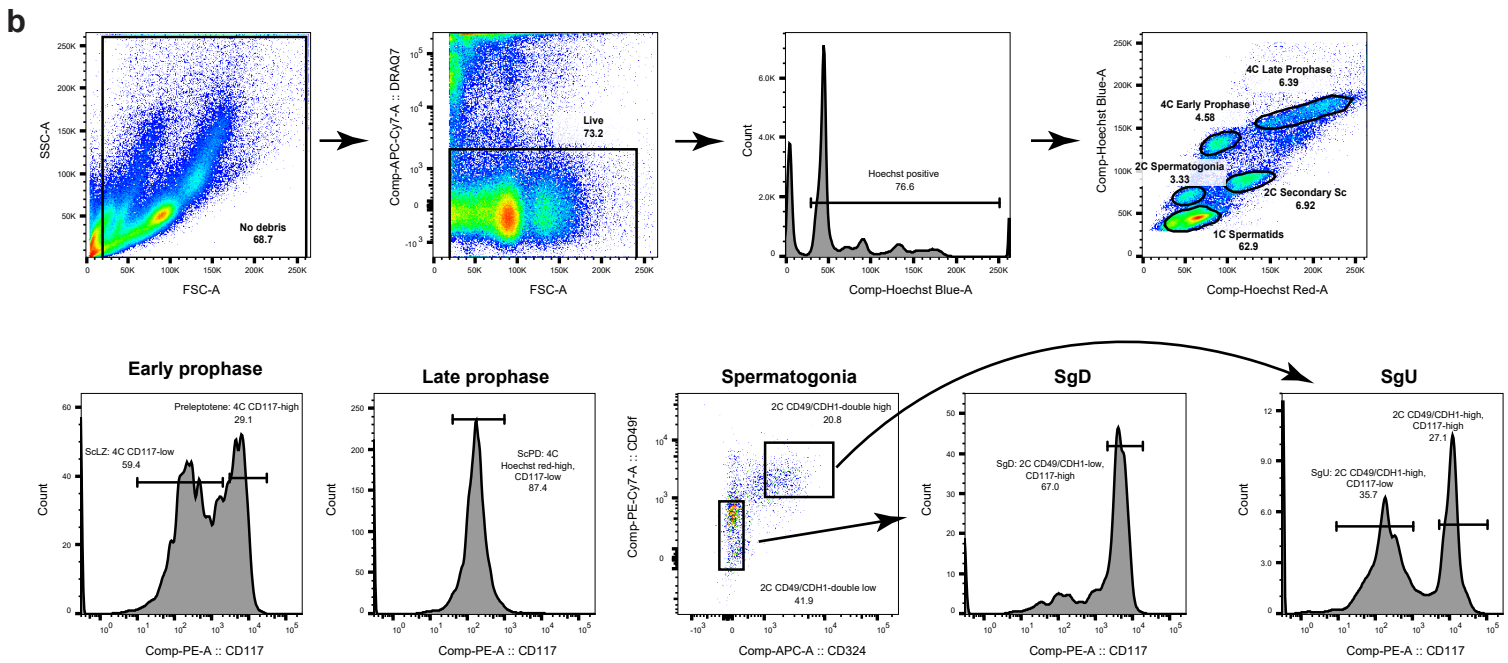
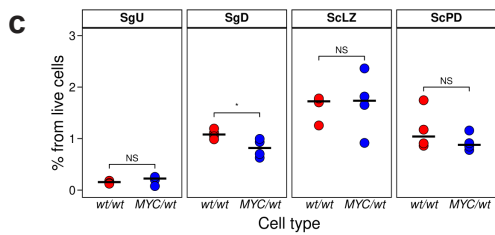
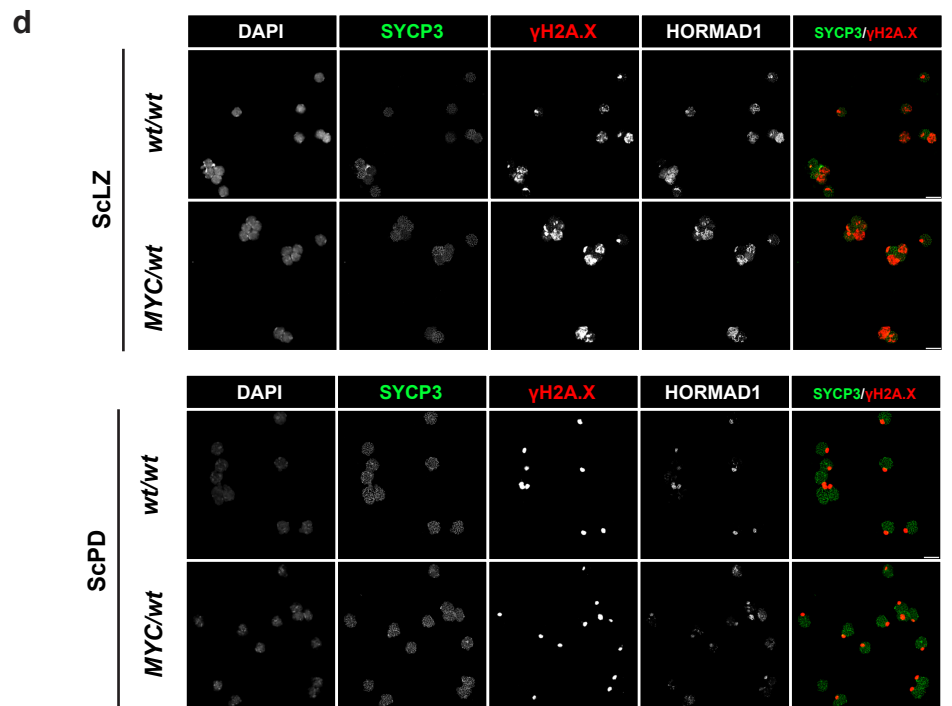
d, Scatter plots showing gene expression log2 fold changes in FACS-isolated ScLZ, ScPD, rSt and eSt in heterozygous ($H3f4^{MYC/wt}$) cells plotted against average gene expression levels in control ($H3f4^{wt/wt}$) cells. Data for full-length genic counts is shown. The numbers of significantly upregulated (red) or downregulated (blue) and non-significant (grey) genes are displayed on the top right (logFC ≥ 1 , max. FDR = 0.05).

e, Venn diagram showing the overlap between genes upregulated in ScLZ, ScPD, rSt and eSt cells of $H3f4^{MYC/wt}$ versus $H3f4^{wt/wt}$ males.

f, Results of gene ontology (GO: Biological process) analysis of genes upregulated in ScPD and eSt cells. For ScLZ and rSt populations, we were unable to identify terms overrepresented in upregulated genes. A one-sided over representation hypergeometric test was used, followed by Benjamini–Hochberg correction for multiple testing.

g, Dot plot showing the percentage of up- and downregulated genes for each chromosome in ScLZ, ScPD and eSt cells. The bars represent the mean values for the selected DEGs. Data for rSt population is not shown due to low number of DEGs. Each dot has been derived from the edgeR estimation of DEGs (panel d) based on 4 biological replicates (n=4) for $H3f4^{wt/wt}$ and $H3f4^{MYC/wt}$ animals.

h, Scatter plot showing differential expression (log2FC) in $H3f4^{MYC/wt}$ versus $H3f4^{wt/wt}$ eSts (FDR ≤ 0.05) (x axis) compared to differential expression (log2FC) in $H3f4^{H3.1/-}$ versus $H3f4^{wt/wt}$ eSts (FDR ≤ 0.05) (y axis).



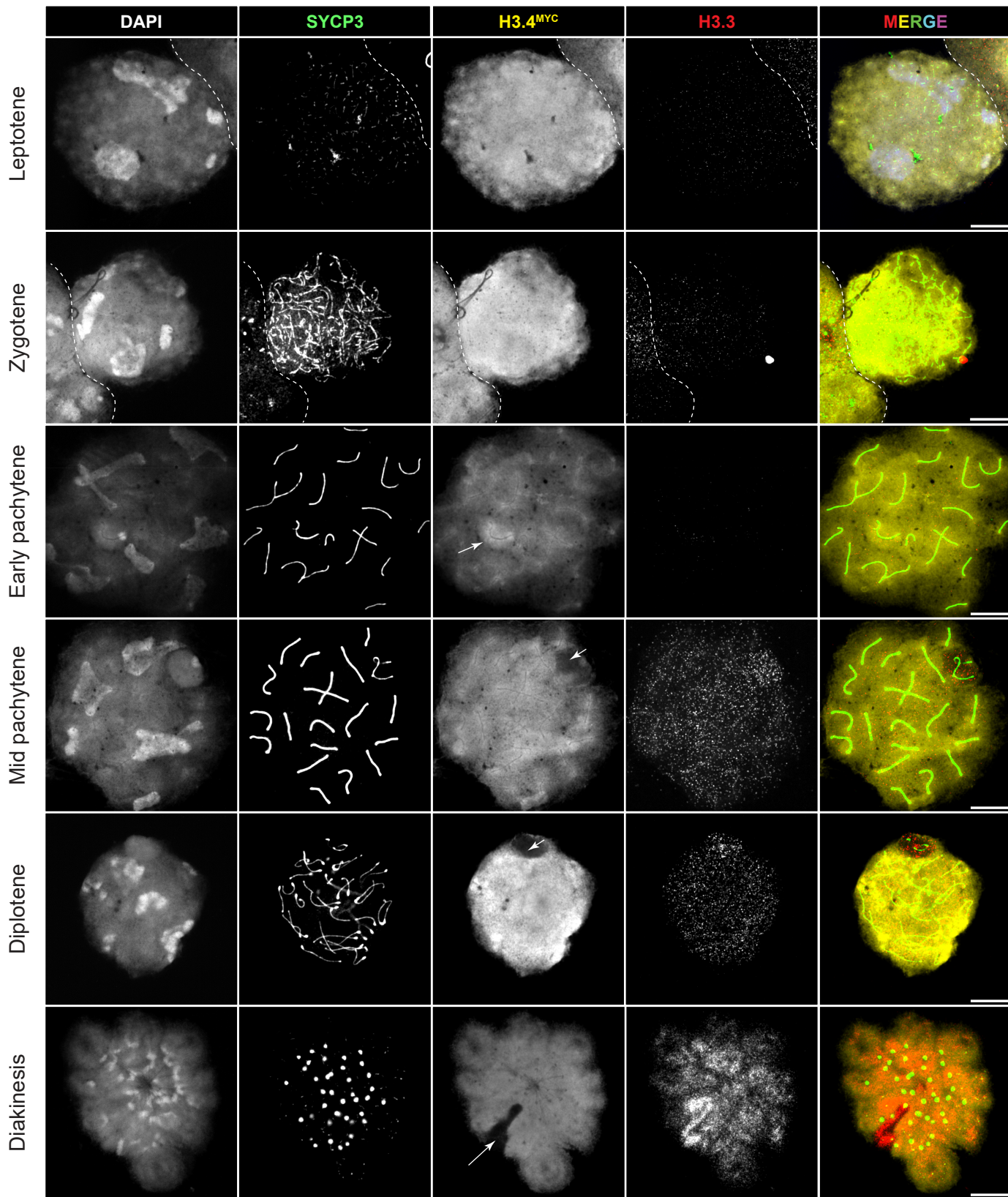
Supplementary Figure 10. Gating parameters used for the FACS isolation of spermatogonia and spermatocytes

a-b, Representative FACS sorting profiles of germ cells with gating parameters from *H3f4^{wt/wt}* control (a) and *H3f4^{MYC/wt}* tagged (b) mice. FSC (forward size scatter) versus SSC (side size scatter) plot indicating the gate excluding debris. We next gated for DRAQ7-negative cells in the FCS versus DRAQ7 plot to select for living cells. The Hoechst-blue histogram indicates the gate with stained cells. The Hoechst-Red vs Hoechst-Blue plot indicates gates with populations of spermatogonia, spermatocytes and haploid spermatids. To obtain leptotene and zygotene spermatocytes (ScLZ), we gated for CD117 (C-KIT) – negative population in the CD117 histogram. The CD324 (E-cadherin) vs CD49f (Integrin alpha 6) plot shows the gates for undifferentiated spermatogonia (SgU, CD324-high, CD49f-high) and differentiating spermatogonia (SgD, CD324-low, CD49f-low). To sort for SgD, we gated for CD117-positive population and to sort for SgU, we gated for CD117-negative population.

c, Dot plot showing the percentages of sorted SgU, SgD, ScLZ and ScPD cells relative to all viable Hoechst-positive cells from *H3f4^{wt/wt}* and *H3f4^{MYC/wt}* animals. Two-sided t-test was performed to compare between genotypes. * $p < 0.05$, NS – not significant.

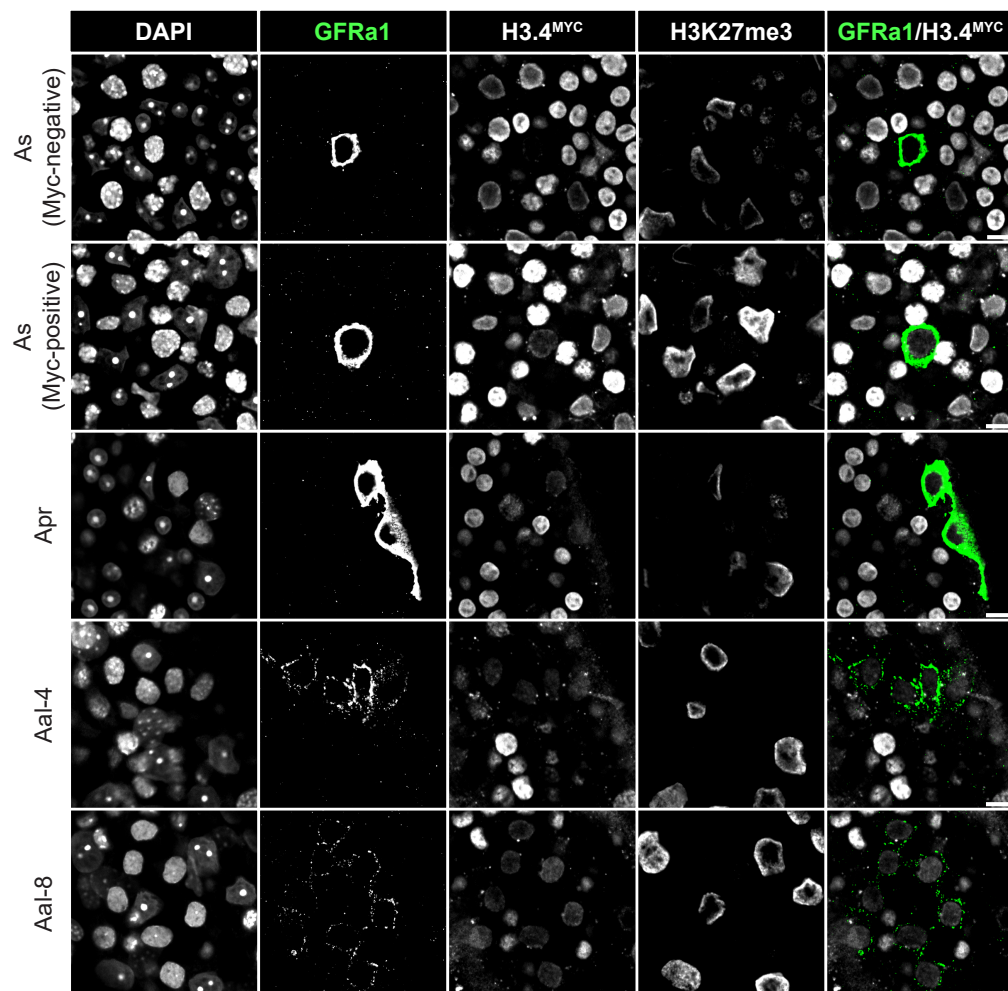
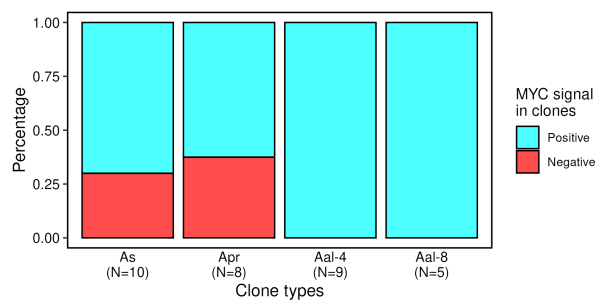
d, Purity control of FACS-sorted populations of ScLZ and ScPD spermatocytes. Representative immunofluorescence staining of spread cells with antibodies against synaptonemal complex protein 3 (SYCP3), phosphorylated H2A.X (γ H2A.X) marking damaged DNA and HORMAD1 marking unsynapsed chromosome axes. Scale bar: 20 μ m.

e, Stacked bar plot showing the relative percentages of cell types within the target ScLZ and ScPD populations of sorted cells. Data from two biological replicates were pooled, numbers on the top represent the total number of cells analyzed. NA – cell type not possible to identify. We observed no differences in meiotic cell type purities between genotypes.



Supplementary Figure 11. H3.4 is evicted from the X and Y chromosomes in mid-pachytene spermatocytes

Meiotic spreads showing temporal dynamics of H3.4^{MYC} and H3.3 proteins during progression through meiotic prophase I. The X and Y chromosomes undergoing eviction of H3.4 are indicated by white arrows. Scale bar: 10 μ m.

a**b**

Supplementary Figure 12. H3.4^{MYC} is expressed in undifferentiated spermatogonia

a, Whole-mount immunofluorescence images showing H3.4^{MYC} expression in different clones of GRFα1-positive spermatogonia with different chain length: A_s (single), A_{pr} (paired), A_{al-4} (chain of 4 aligned cells) and A_{al-8} (chain of 8 aligned cells). For A_s, an example of a clone positive or negative for H3.4^{MYC} are displayed. Scale bar: 10 μm.

b, Stacked bar plot showing the proportion of GRFα1-positive spermatogonia clones (A_s, A_{pr}, A_{al-4} and A_{al-8}) positive or negative for H3.4^{MYC} staining. N corresponds to the number of clones assessed.

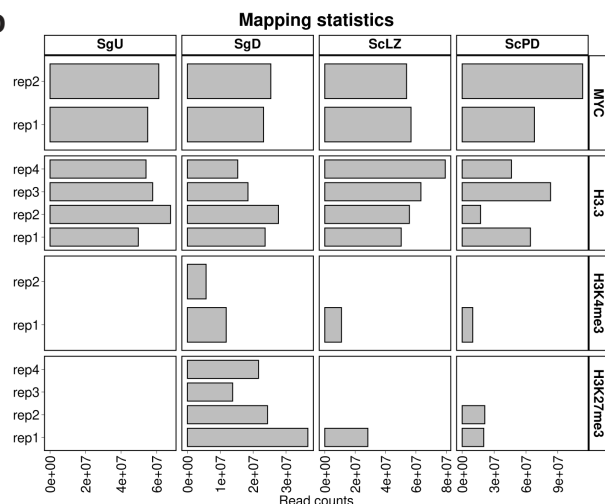
a

ULLI-NChIP

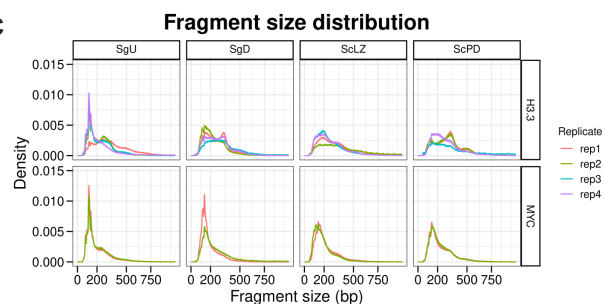
Celltype	Replicate	Genotype	ChIP	Type	Library Name
SgU	rep1	wt/wt	H3.3	PE	SgU_H3.3_rep1
SgU	rep2	wt/wt	H3.3	PE	SgU_H3.3_rep2
SgU	rep3	MYC/wt	H3.3	PE	SgU_H3.3_rep3
SgU	rep4	MYC/wt	H3.3	PE	SgU_H3.3_rep4
SgD	rep1	wt/wt	H3.3	PE	SgD_H3.3_rep1
SgD	rep2	wt/wt	H3.3	PE	SgD_H3.3_rep2
SgD	rep3	MYC/wt	H3.3	PE	SgD_H3.3_rep3
SgD	rep4	MYC/wt	H3.3	PE	SgD_H3.3_rep4
ScLZ	rep1	wt/wt	H3.3	PE	ScLZ_H3.3_rep1
ScLZ	rep2	wt/wt	H3.3	PE	ScLZ_H3.3_rep2
ScLZ	rep3	MYC/wt	H3.3	PE	ScLZ_H3.3_rep3
ScLZ	rep4	MYC/wt	H3.3	PE	ScLZ_H3.3_rep4
ScPD	rep1	wt/wt	H3.3	PE	ScPD_H3.3_rep1
ScPD	rep2	wt/wt	H3.3	PE	ScPD_H3.3_rep2
ScPD	rep3	MYC/wt	H3.3	PE	ScPD_H3.3_rep3
ScPD	rep4	MYC/wt	H3.3	PE	ScPD_H3.3_rep4
SgU	rep1	MYC/wt	MYC	PE	SgU_MYC_rep1
SgU	rep2	MYC/wt	MYC	PE	SgU_MYC_rep2
SgD	rep1	MYC/wt	MYC	PE	SgD_MYC_rep1
SgD	rep2	MYC/wt	MYC	PE	SgD_MYC_rep2
ScLZ	rep1	MYC/wt	MYC	PE	ScLZ_MYC_rep1
ScLZ	rep2	MYC/wt	MYC	PE	ScLZ_MYC_rep2
ScPD	rep1	MYC/wt	MYC	PE	ScPD_MYC_rep1
ScPD	rep2	MYC/wt	MYC	PE	ScPD_MYC_rep2
SgD	rep1	wt/wt	H3K4me3	SE	SgD_H3K4me3_rep1 ¹
SgD	rep2	wt/wt	H3K4me3	SE	SgD_H3K4me3_rep2 ¹
ScLZ	rep1	wt/wt	H3K4me3	SE	ScLZ_H3K4me3_rep1
ScPD	rep1	wt/wt	H3K4me3	SE	ScPD_H3K4me3_rep1
SgD	rep1	wt/wt	H3K27me3	SE	SgD_H3K27me3_rep1 ¹
SgD	rep2	wt/wt	H3K27me3	SE	SgD_H3K27me3_rep2 ¹
SgD	rep3	wt/wt	H3K27me3	SE	SgD_H3K27me3_rep3 ¹
SgD	rep4	wt/wt	H3K27me3	SE	SgD_H3K27me3_rep4 ¹
ScLZ	rep1	wt/wt	H3K27me3	SE	ScLZ_H3K27me3_rep1 ¹
ScPD	rep1	wt/wt	H3K27me3	SE	ScPD_H3K27me3_rep1 ¹
ScPD	rep2	wt/wt	H3K27me3	SE	ScPD_H3K27me3_rep2 ¹

¹Bocker *et al.*, (2025) Nat Commun

b



c



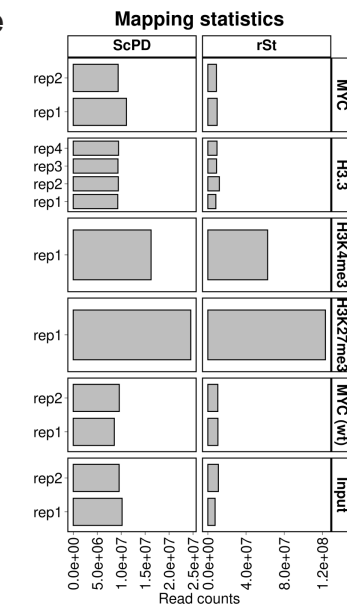
d

conventional ChIP

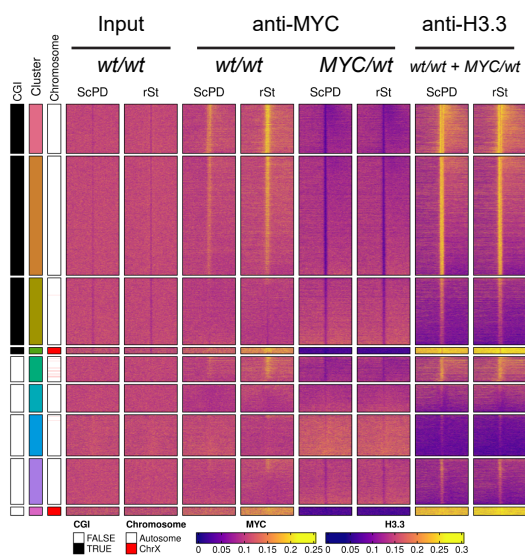
Cell type	Replicate	Genotype	ChIP	Type	Library Name
ScPD	rep1	wt/wt	H3.3	SE	ScPD_H3.3_rep1
ScPD	rep2	wt/wt	H3.3	SE	ScPD_H3.3_rep2
ScPD	rep3	MYC/wt	H3.3	SE	ScPD_H3.3_rep3
ScPD	rep4	MYC/wt	H3.3	SE	ScPD_H3.3_rep4
rSt	rep1	wt/wt	H3.3	SE	rSt_H3.3_rep1
rSt	rep2	wt/wt	H3.3	SE	rSt_H3.3_rep2
rSt	rep3	MYC/wt	H3.3	SE	rSt_H3.3_rep3
rSt	rep4	MYC/wt	H3.3	SE	rSt_H3.3_rep4
ScPD	rep1	MYC/wt	MYC	SE	ScPD_MYC_rep1
ScPD	rep2	MYC/wt	MYC	SE	ScPD_MYC_rep2
rSt	rep1	MYC/wt	MYC	SE	rSt_MYC_rep1
rSt	rep2	MYC/wt	MYC	SE	rSt_MYC_rep2
ScPD	rep1	wt/wt	H3K4me3	SE	ScPD_H3K4me3_rep1
rSt	rep1	wt/wt	H3K4me3	SE	rSt_H3K4me3_rep1 ²
ScPD	rep1	wt/wt	H3K27me3	SE	ScPD_H3K27me3_rep1
rSt	rep1	wt/wt	H3K27me3	SE	rSt_H3K27me3_rep1 ²
ScPD	rep1	wt/wt	MYC	SE	ScPD_MYC
ScPD	rep2	wt/wt	MYC	SE	ScPD_MYC
rSt	rep1	wt/wt	MYC	SE	rSt_MYC
rSt	rep2	wt/wt	MYC	SE	rSt_MYC
ScPD	rep1	wt/wt	Input	SE	ScPD_Input_rep1
ScPD	rep2	wt/wt	Input	SE	ScPD_Input_rep2
rSt	rep1	wt/wt	Input	SE	rSt_Input_rep1
rSt	rep2	wt/wt	Input	SE	rSt_Input_rep2

²Erkek *et al.*, (2013) NSMB

e



g



Supplementary Figure 13. Quality control of ChIP-seq libraries used in this study

a, Names of ChIP-seq libraries generated in this study using an ultra-low native ULI-NChIP method or previously published (Bocker *et al.*⁵⁷). Libraries related to different antibodies are highlighted in different shades of gray.

b, Mapping statistics of libraries displayed in panel a.

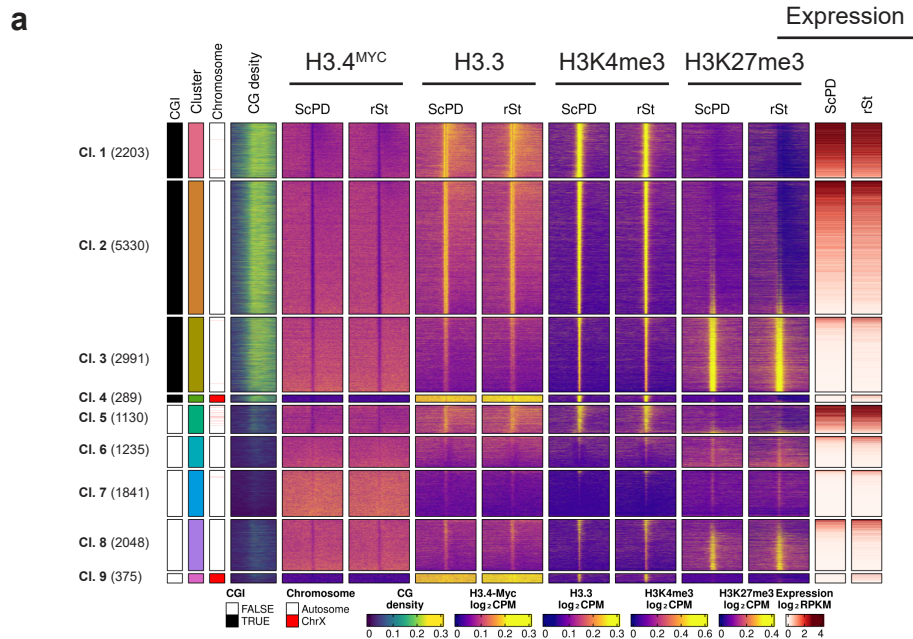
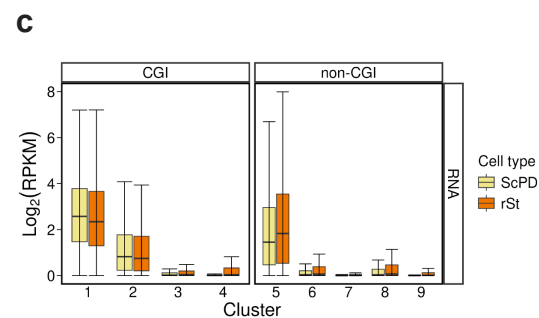
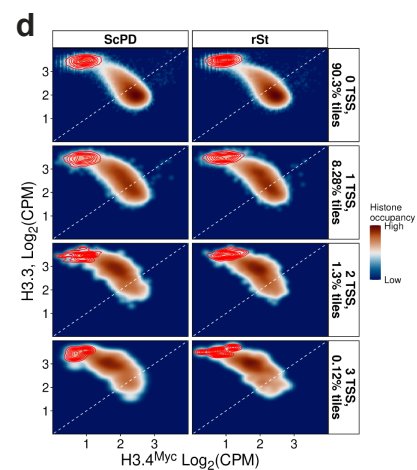
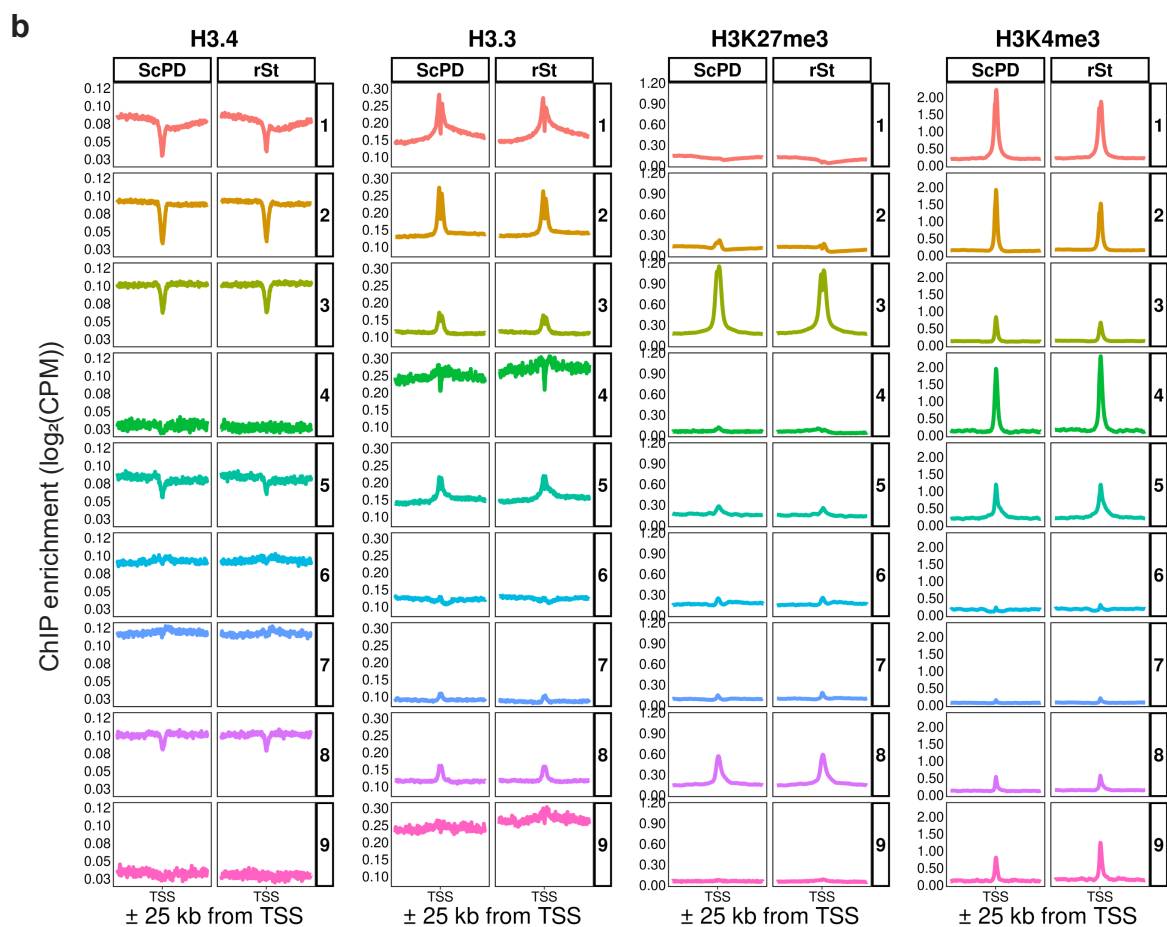
c, Fragment size distribution of reads from libraries in panel a (only for paired-end libraries)

d, Names of ChIP-seq libraries generated in this study using a conventional ChIP method (see Materials and Methods) or previously published (Erkek *et al.*⁶¹). Libraries related to different antibodies are highlighted in different shades of gray.

e, Mapping statistics of libraries displayed in panel d.

g, Heatmap comparing ChIP-signal in input, anti-MYC libraries (in *H3f4^{wt/wt}* and MYC-expressing *H3f4^{MYC/wt}* animals separately) and anti-H3.3 libraries (signal pooled from control *H3f4^{wt/wt}* and MYC-expressing *H3f4^{MYC/wt}* animals). The intensity scale for input and anti-MYC libraries is the same. The order of clusters and genes is the same as in Fig. 3a.

The experiment shows the presence of H3.4^{MYC} in chromatin of ScPD and rSt cells at regions up- and downstream of TSS in a sequence-context specific manner. In cells from transgenic animals we observe the following: At CGI promoter genes, H3.4^{MYC} occupancy is visible in flanking regions surrounding TSS while the signal is reduced at the central CGI regions (see Fig 3a for CGI information). At non-CGI promoter genes, H3.4^{MYC} occupancy is largely uniform across the regions except for genes at which nucleosome turnover is observed. Finally, H3.4^{MYC} occupancy is absent at X-linked genic regions indicative of MSCI-related nucleosome replacement of H3.4 by H3.3. H3.3 occupancy reflects an inverse occupancy pattern as seen for H3.4^{MYC}, with high levels at regions having undergone eviction of H3.4^{MYC}, either at TSS or along the X chromosome. In cells of wildtype control mice, the anti-MYC antibody reveals a non-uniform pattern around TSS of CGI genes and partially along X-linked genic regions, opposite of the patterns observed in transgenic H3.4^{MYC} expressing cells. Since H3.4^{MYC} enrichment in *H3f4^{wt/wt}* mice does not reflect the presence or absence of the DNA-sequence motif of the endogenous MYC protein at promoter regions (data analysis not shown), it may point to cross-reactivity with other epitopes. Together, given the inverse ChIP-signal patterns of H3.4^{MYC} and H3.3 proteins, we consider the H3.4^{MYC} signal as representing genuine H3.4^{MYC} occupancy in transgenic animals.



Supplementary Figure 14. Genome localization of H3 variants in ScPD and rSt cells

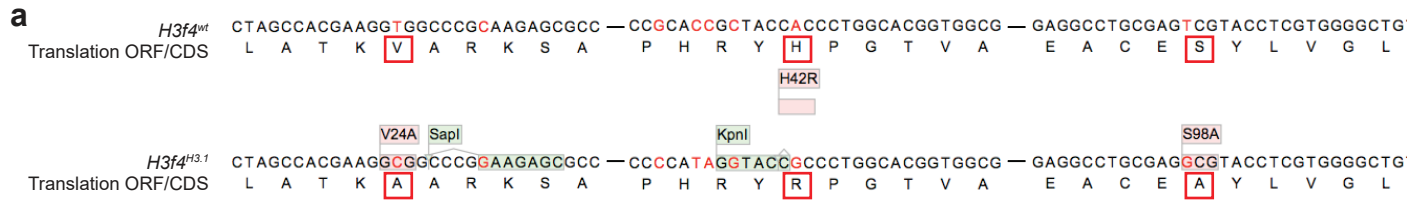
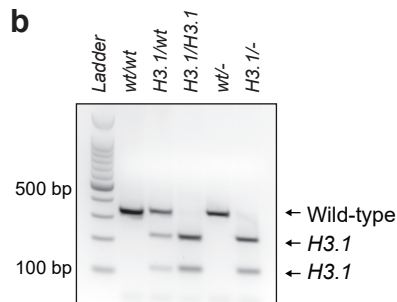
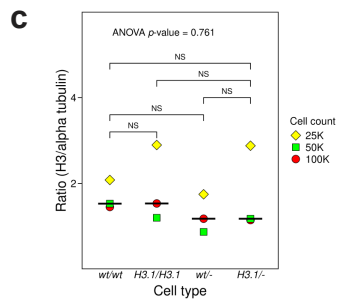
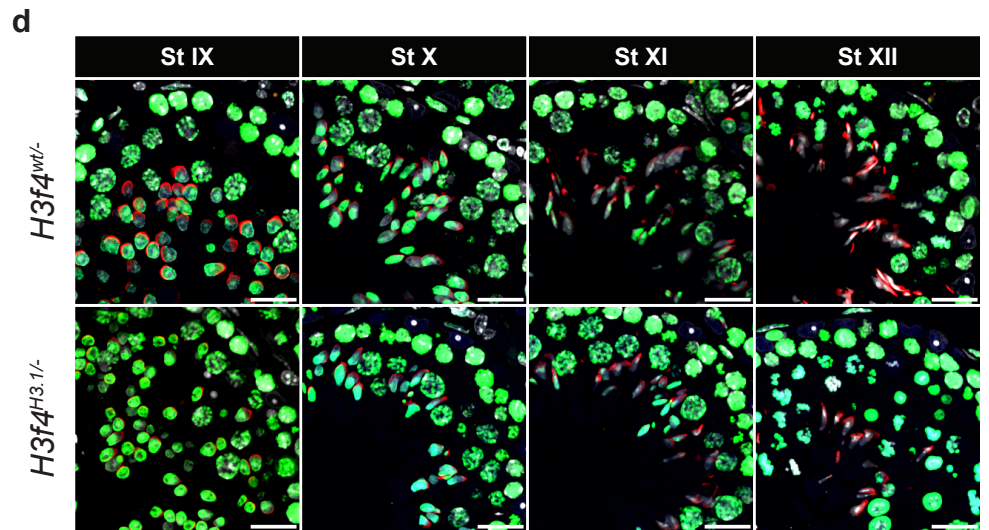
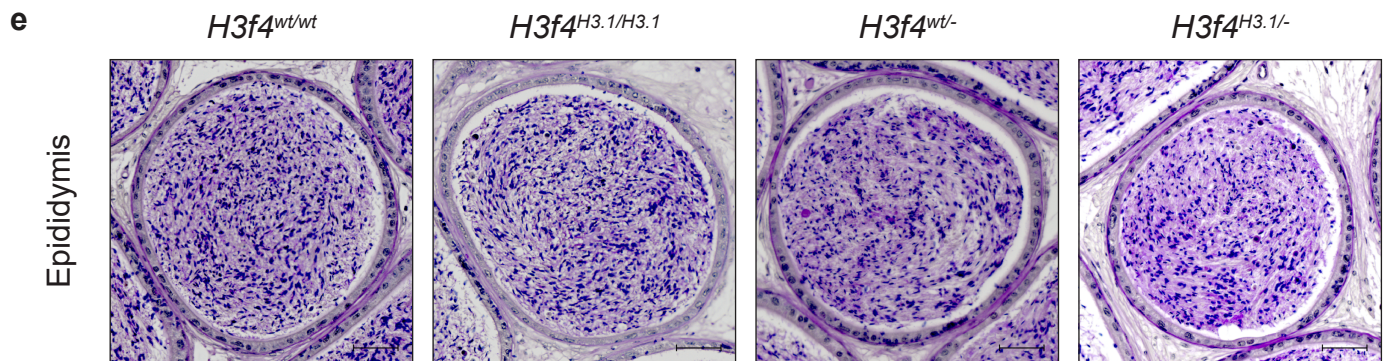
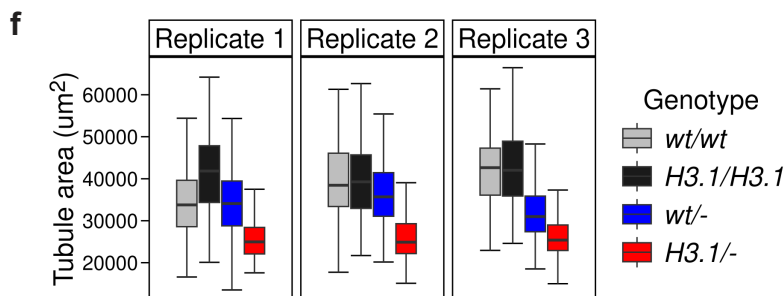
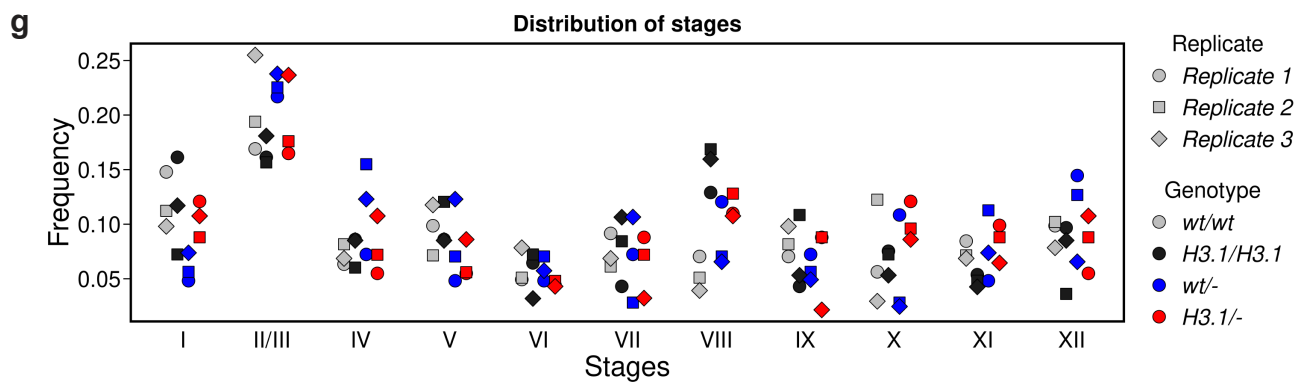
a, Heatmap displaying chromatin states at TSS regions (± 25 kb; log2CPM) and gene expression levels (log2RPKM) of CGI- and nonCGI-containing genes measured in FACS-purified ScPD and rSt cells. Regions were grouped into 9 gene clusters by k-means clustering on chromatin variables performed separately for CGI- and non-CGI-type promoters.

The number of gene promoters is displayed for each cluster. CG density (± 1 kb around TSS), is indicated in percentage. The ChIP-seq for H3K4me3 and H3K27me3 data in rSt were obtained from Erkek *et al.*⁶¹ and RNA expression data in ScPD and rSt from Gill *et al.*⁵⁸. The order of clusters and genes is the same as in Fig. 3a.

b, Metaplots showing quantification of ChIP-seq data at TSS regions (± 25 kb; log2CPM) in clusters 1-9, as displayed in panel a.

c, Boxplot showing quantification of expression of genes (in log2RPKM) belonging to clusters 1-9 in panel a.

d, Density scatter plots displaying log2CPM counts of H3.4^{MYC} (x-axis) and H3.3 (y-axis) on genomic tiles that contain 0, 1, 2 or 3+ transcription units with TSS in ScPD and rSt cells. Red contour lines represent the density of tiles located on the X chromosome. The presence of a transcription unit with TSS was defined by an overlap of a given tile with a region ± 1 kb around TSS.



Supplementary Figure 15. Generation of the *H3f4*^{H3.1} allele by CRISPR/Cas9 gene editing.

a, Schematic representation of the *H3f4*^{wt} wildtype and *H3f4*^{H3.1} modified alleles, with corresponding H3.4 to H3.1-specific amino acid substitutions (in red boxes). Variant positions in DNA sequences encoding H3.4 and H3.1 are indicated in red, while relevant restriction sites are shown in light green. Abbreviations: V24A, Valine 24 to Alanine; H42R, Histidine 42 to Arginine; S98A, Serine 98 to Alanine. CDS, coding sequence; ORF, open reading frame.

b, PCR genotyping results of the *H3.1* mouse model. The following genotypes are shown: *H3f4*^{wt/wt}, *H3f4*^{H3.1/wt}, *H3f4*^{H3.1/H3.1}, *H3f4*^{wt/-} and *H3f4*^{H3.1/-}. Primers and PCR protocols used for genotyping are provided in Supplementary Data 8 and 9, respectively. The PCR products were digested with KpnI into a 338bp restriction fragment for the wildtype allele and two bands of 222 bp and 116 bp for the *H3f4*^{H3.1} allele. A hundred base pair ladder is shown for size comparison.

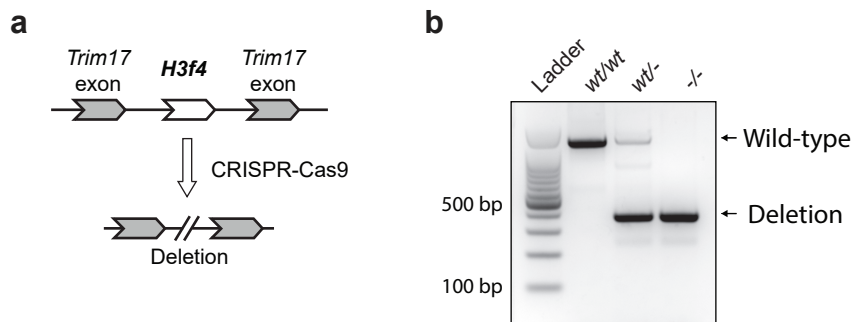
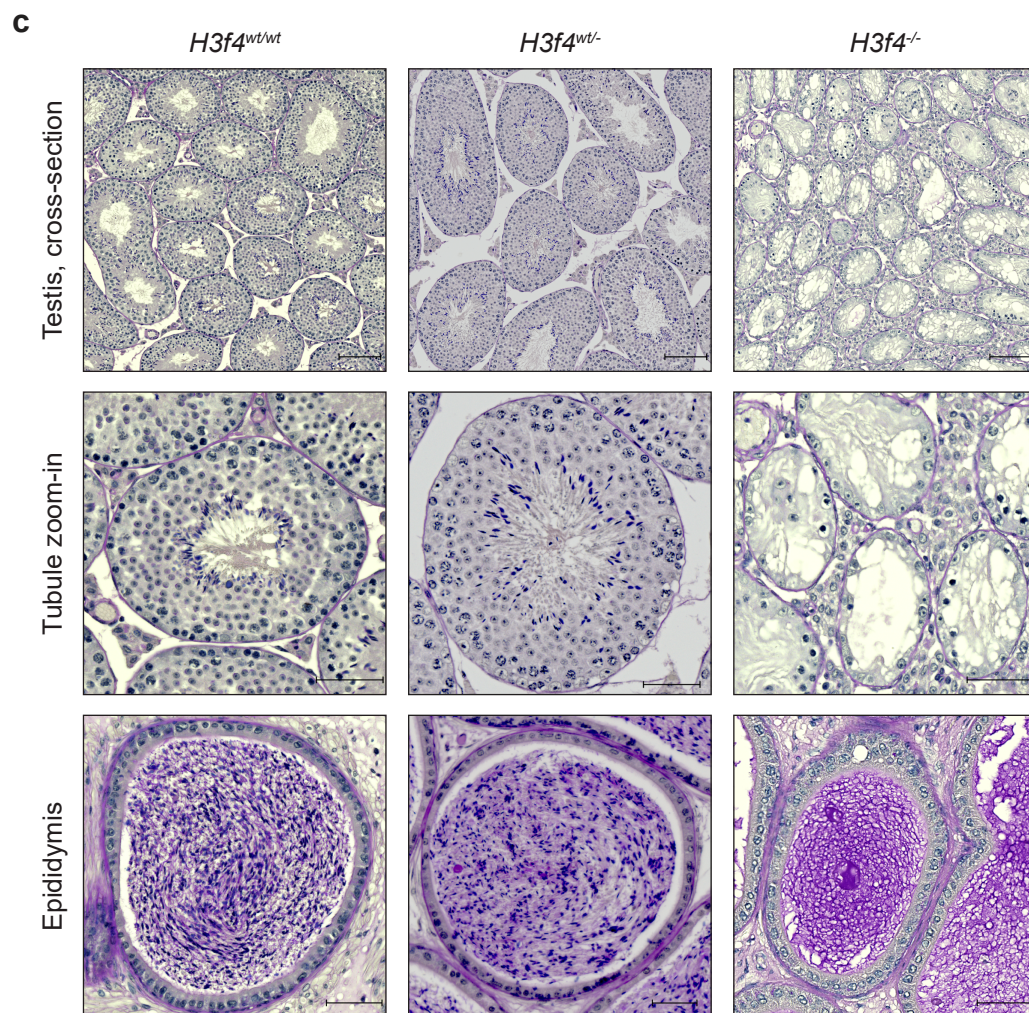
c, ImageJ quantification of Western blot data shown in Fig. 4c and calculated as the ratio of intensities of bands detected by the pan-H3 antibody over the bands detected by alpha tubulin antibody. Two-sided t-test was performed for each comparison group. NS – not significant. The *p*-value obtained from an ANOVA test is provided at the top.

d, Representative immunofluorescence images of stage IX to XII seminiferous tubules of *H3f4*^{wt/-} and *H3f4*^{H3.1/-} animals, co-stained for DAPI (grey), an antibody that recognizes H3.1, H3.2 and H3.4 (#ab34)⁶² (green) and acrosome (red). Scale bar: 20 μm.

e, PAS-haematoxylin staining of caudal epididymal tissues obtained from *H3f4*^{wt/wt}, *H3f4*^{H3.1/H3.1}, *H3f4*^{wt/-} and *H3f4*^{H3.1/-} mice. Scale bar: 50 μm.

f, Boxplots showing the area of segmented seminiferous tubules used in the quantification analysis shown in Fig. 4i. Data for three biological replicates per genotype are shown.

g, Distribution of developmental stages of the seminiferous tubule cycle in testes of 3 mice per genotype. Cells in stage V and X tubules were used for automated cell quantification analysis as shown in Fig. 4i. Staging was manually performed as described before⁶⁶. Each dot represents the frequency of a given stage per biological replicate.

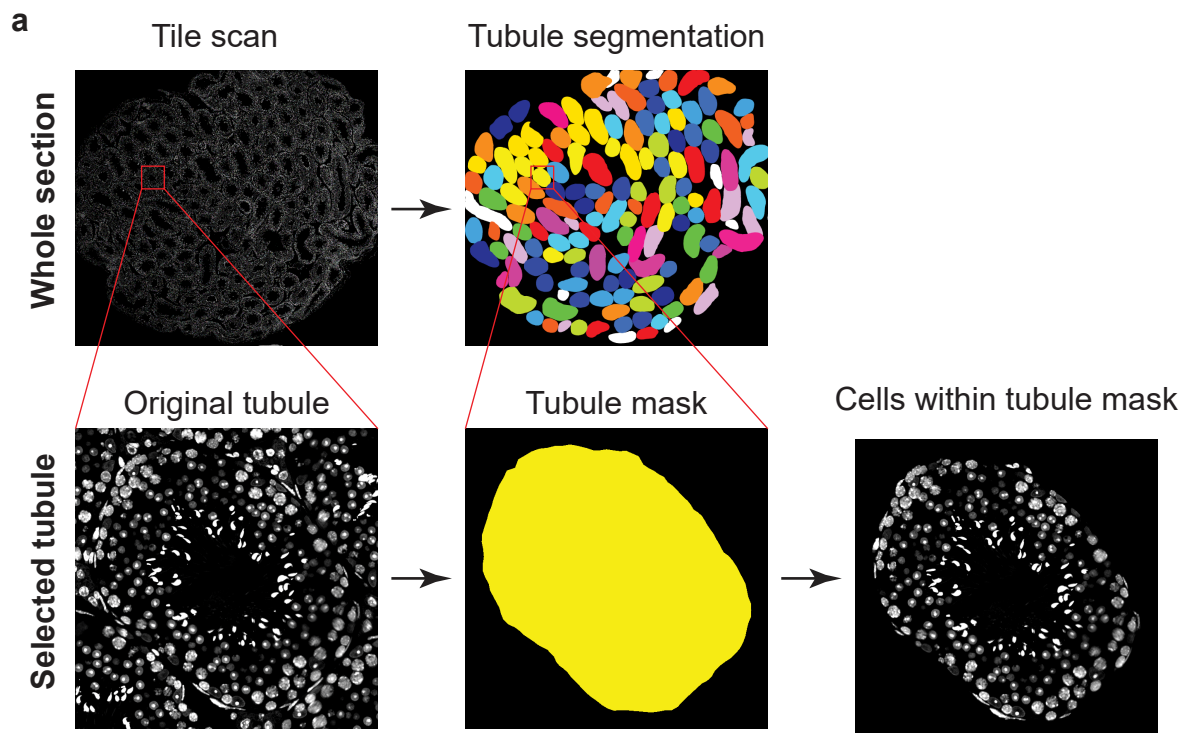
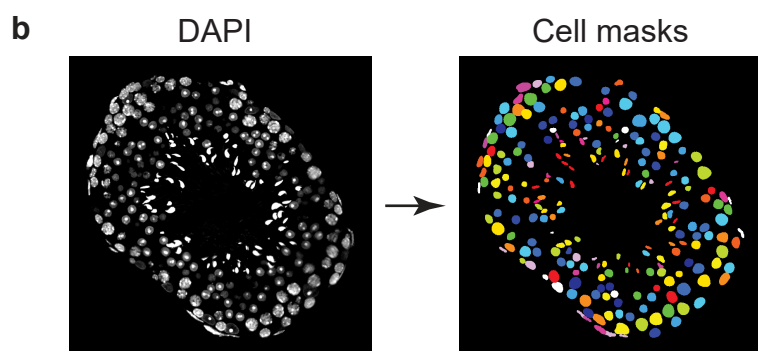
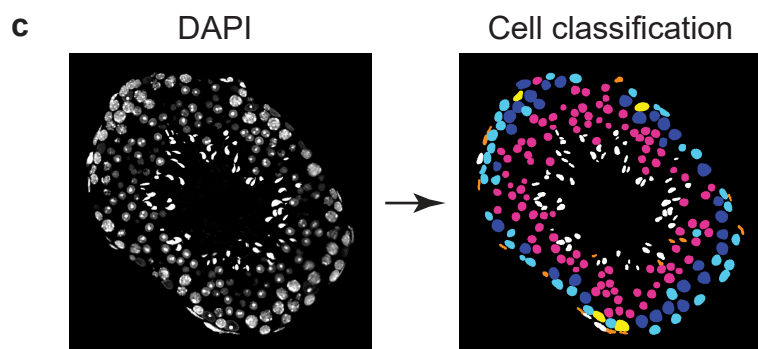
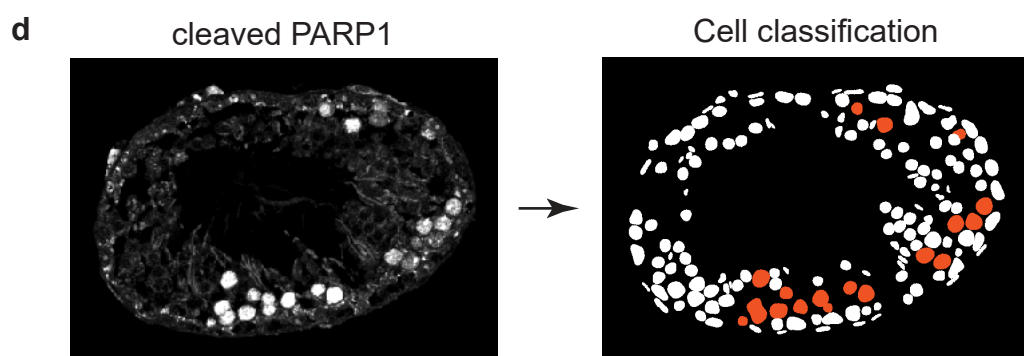


Supplementary Figure 16. Deletion of the *H3f4* gene leads to infertility.

a, CRISPR/Cas9 strategy to delete the *H3f4* gene. Sequences of sgRNAs are indicated in Supplementary Data 6.

b, Genotyping results of mice with wildtype (*H3f4*^{wt/wt}), heterozygous (*H3f4*^{wt/-}) and *H3f4* knock-out (*H3f4*^{-/-}) genotypes. PCR using primers (Supplementary Datas 8 and 9) that bind up- and downstream of the deletion shows fragment sizes of 1180bp (wildtype allele) and 330bp (deleted allele). Ladder in multiples of hundred base pairs is shown for size comparison.

c, PAS-haematoxylin-stained histological cross-sections of testes and epididymal tissues obtained from *H3f4*^{wt/wt}, *H3f4*^{wt/-} and *H3f4*^{-/-} mice. Scale bars: 100 µm (cross-section overview) and 50 µm (tubule zoom-in and epididymis).



Supplementary Figure 17. The pipeline for automated quantification of male germ cells

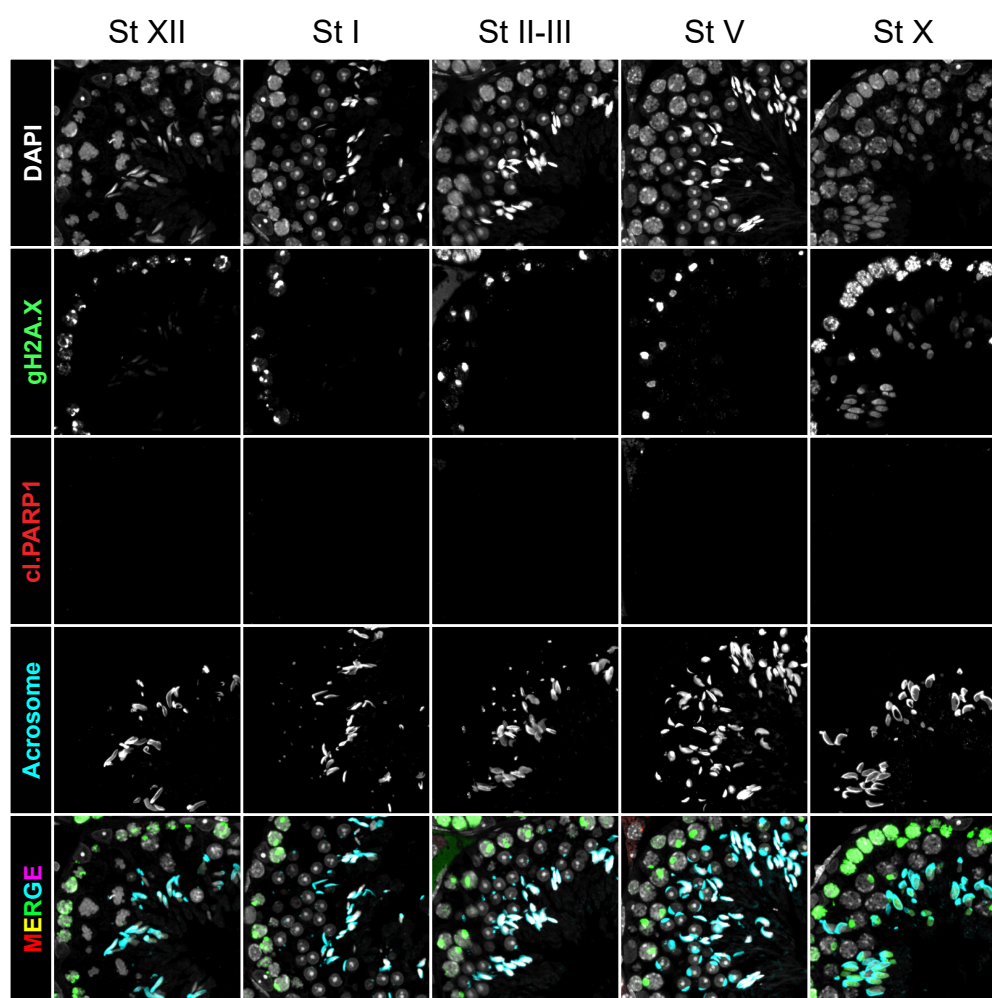
a, Example of tubule segmentation on an 8-binned stitched scan (DAPI channel used as an input). Raw DAPI and tubule masks are shown for the whole scan (upper panels) and for a selected tubule example (bottom).

b, Example of StarDist-based cell segmentation after manual curation of cell shapes. DAPI images were used as input information for cell segmentation.

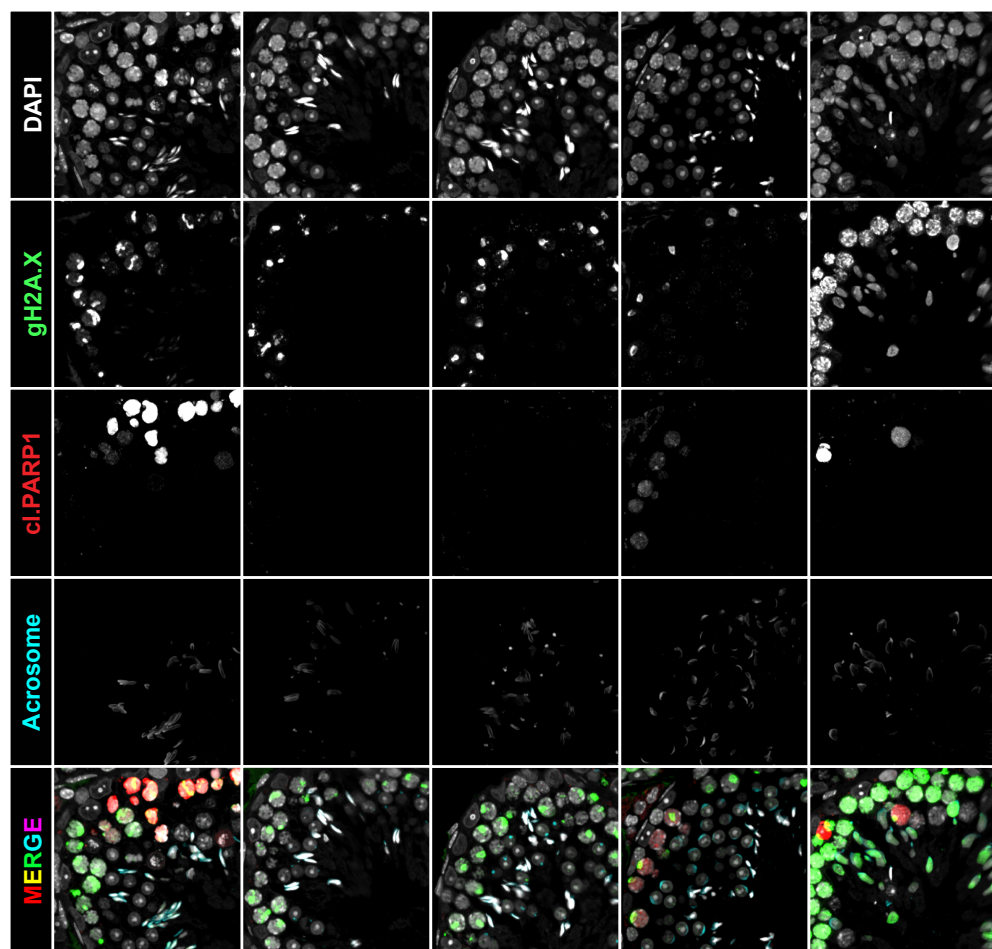
c, Example of a multi-type cell classification results performed by Ilastik software¹¹⁶ on a single channel (DAPI) channel. Ground truth annotation was used to assign classes to cells in a learning phase. Different colors represent cell different cell types.

d, Example of a single-type cell classification results performed by Ilastik software¹¹⁶ on a single channel (cleaved PARP1) channel. Ground truth annotation was used to assign classes to cells in a learning phase.

H3f4^{wt/-}



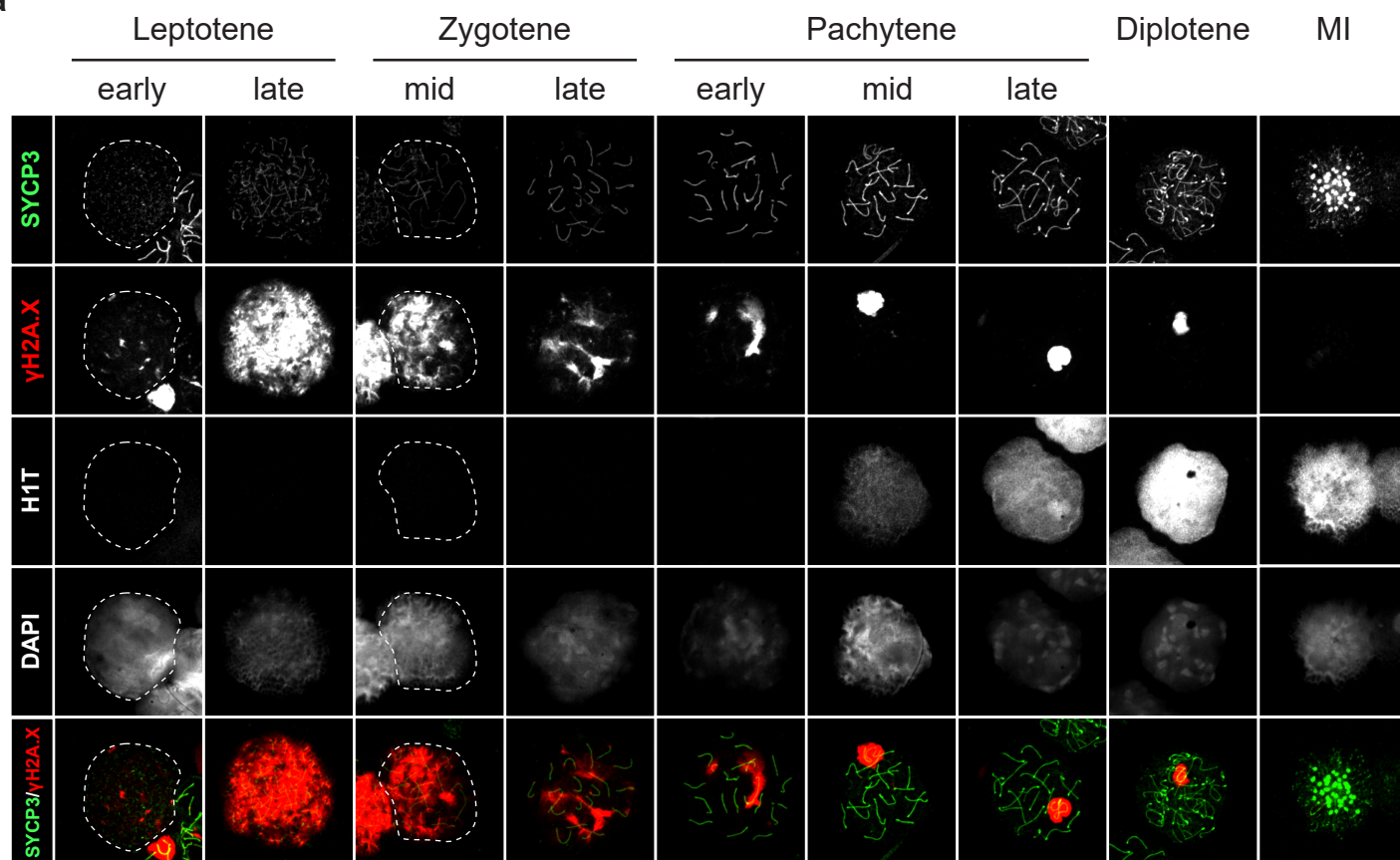
H3f4^{H3.1/-}



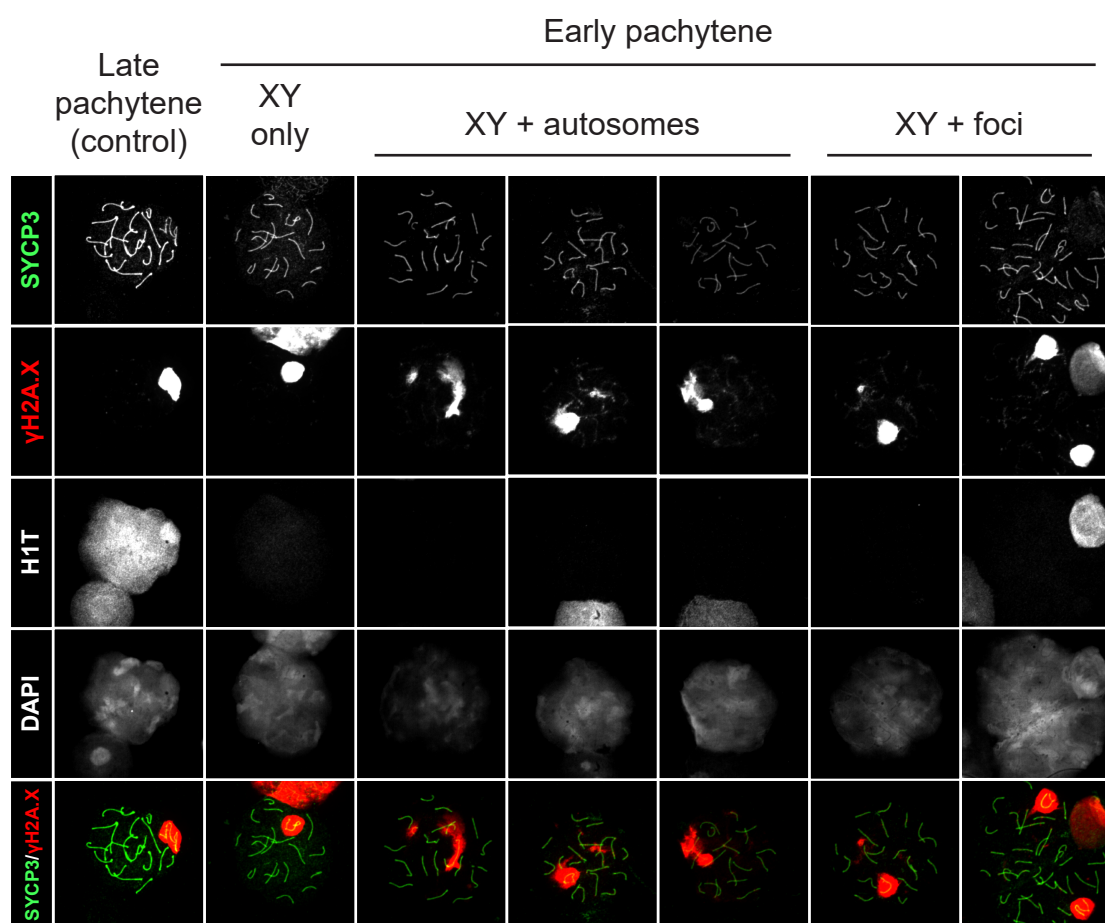
Supplementary Figure 18. Examples of cleaved PARP1 signal in testicular tubules in *H3f4*^{wt/-} and *H3f4*^{H3.1/-} animals

Representative zoom-up examples of stage I, II-III, V, X and XII seminiferous tubules, related to Fig. 5a. Scale bar: 20 μ m.

a



b

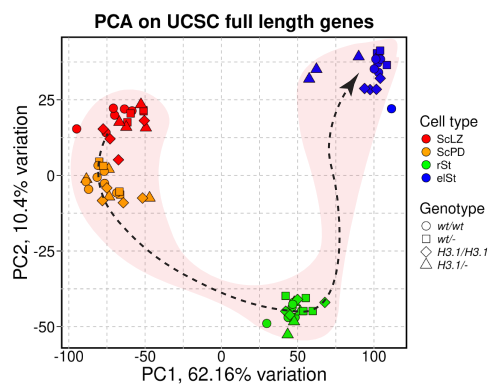


Supplementary Figure 19. Synapsis and double strand break formation and repair during meiotic prophase in *H3f4^{wt/-}* control mice

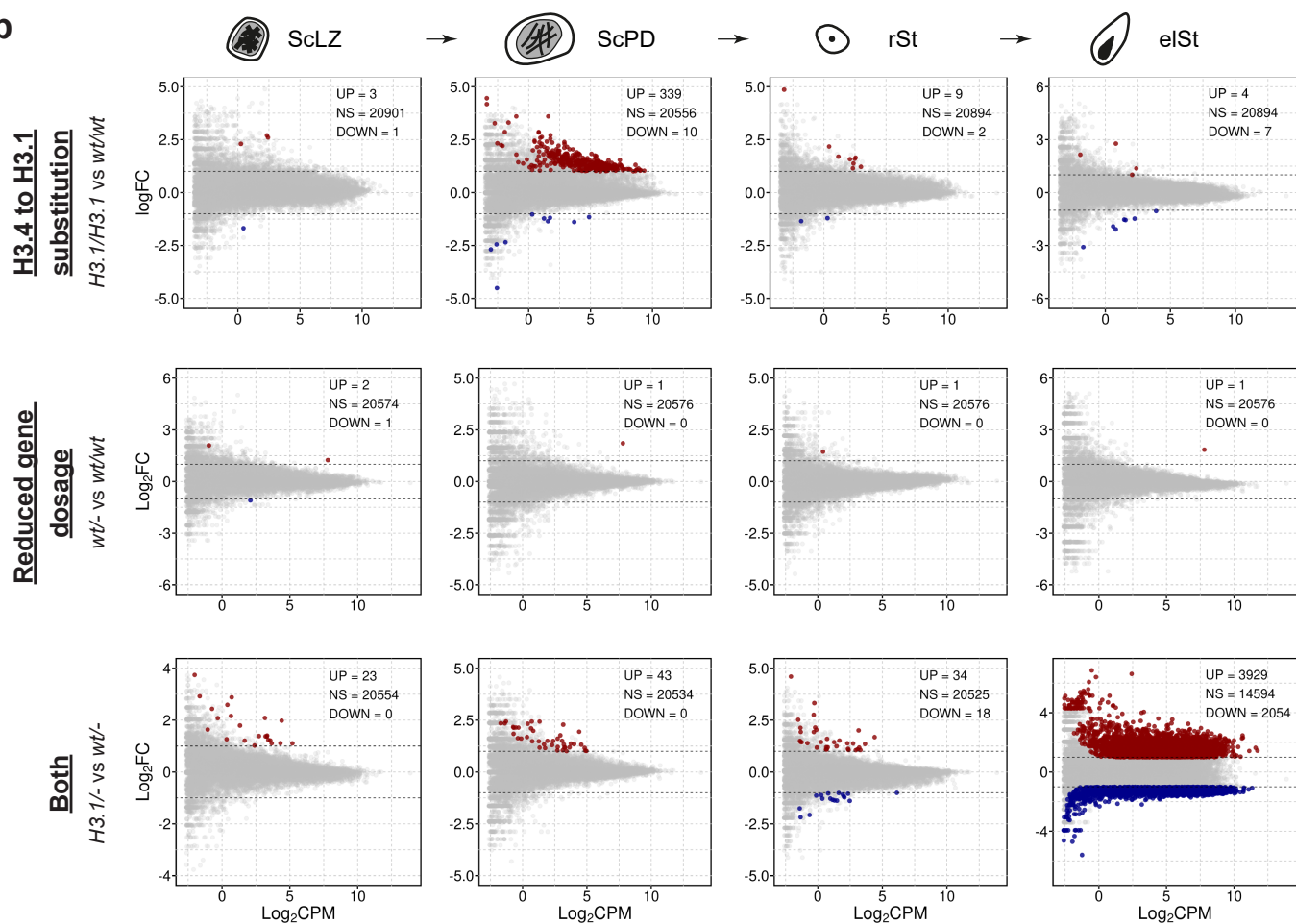
a, Examples of SYCP3 and γ H2A.X signals in spread of spermatocytes during their developmental time course of prophase I in *H3f4^{wt/-}* animals. To assess the developmental progression of spermatocytes through pachytene substages, we used immunodetection of H1T histone variant incorporation into meiotic chromatin. Scale bar: Scale bar: 10 μ m.

b, Examples of SYCP3 and γ H2A.X signals in spread of early pachytene spermatocyte nuclei (lacking H1T expression) in *H3f4^{wt/-}* animals. Scale bar: 10 μ m.

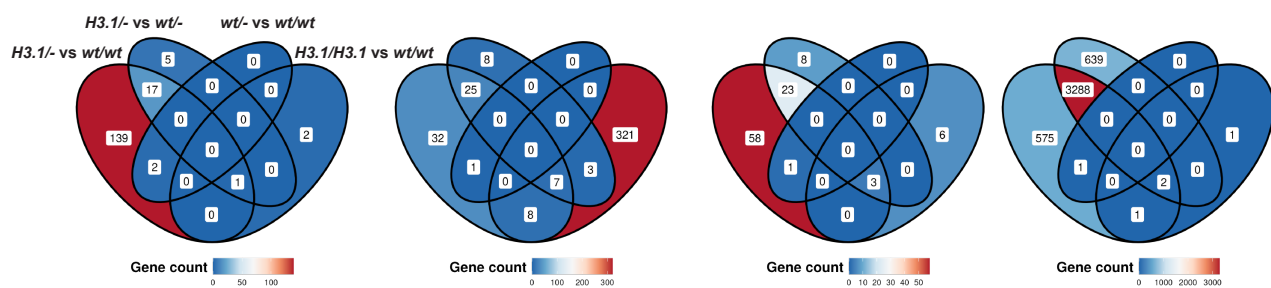
a



b



c

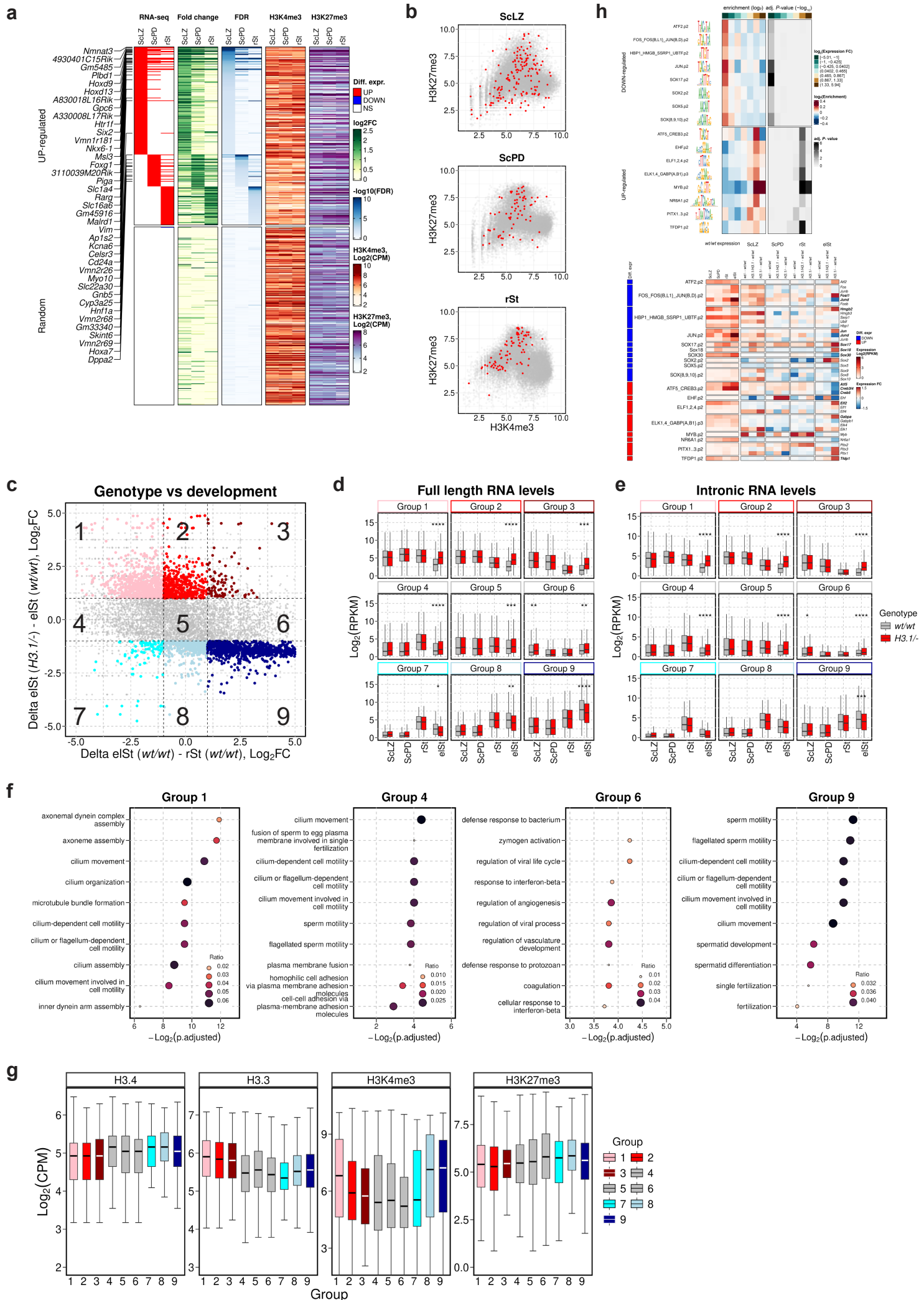


Supplementary Figure 20. Transcriptional impact of H3.4 to H3.1 substitution and/or of reduced *H3f4* gene dosage.

a, Principal component analysis (PCA) comparing RNA-seq expression of UCSC-annotated genes among biological replicates and genotypes in samples that were used in the study.

b, Scatter plots showing differential gene expression (log2FC) in *H3f4*^{H3.1/-} versus *H3f4*^{wt/-} FACS-isolated ScLZ, ScPD, rSt and eSt cells plotted against gene expression levels in *H3f4*^{wt/-} cells (top row), in *H3f4*^{H3.1/H3.1} against *H3f4*^{wt/wt} (middle row) and in *H3f4*^{wt/-} against *H3f4*^{wt/wt} (bottom row). Numbers of significantly upregulated (red), downregulated (blue) and non-significant (grey) genes are displayed (logFC ≥ 1, FDR ≤ 0.05).

c, Venn diagrams showing the overlap between genes upregulated in the contrasts shown in Fig. 6a and Supplementary Fig. 20b.



Supplementary Figure 21. Major transcriptional changes in elongating spermatids of *H3f4^{H3.1/-}* mice

a, Heatmap showing ChIP-seq enrichments of H3K4me3 and H3K27me3 in control samples at promoters of genes (TSS ± 1 kb; log₂(CPM)) upregulated in ScLZ, ScPD and rSt cells of *H3f4^{H3.1/-}* animals (Fig. 6a). For reference, a randomly selected set of genes is shown below. The EdgeR results for RNA-seq data (significance, fold-change and FDR) are shown for each gene. Names of genes commonly upregulated in at least two cell populations are indicated on the left. H3K27me3 data in ScLZ and ScPD was obtained from Bocker *et al.*⁵⁷, H3K4me3 and H3K27me3 in rSt was obtained from Erkek *et al.*⁶¹

b, Scatter plots showing the enrichment of H3K27me3 versus H3K4me3 at promoters (TSS ± 1 kb; log₂(CPM)) of all genes (grey) and of the genes upregulated in ScLZ, ScPD and rSt cells (red) (Fig. 6a).

c, Scatter plot showing differential expression (log₂FC) of non-CGI type promoter genes in *H3f4^{wt/wt}* elSts versus *H3f4^{wt/wt}* rSts (FDR ≤ 0.05) (x axis) compared to differential expression (log₂FC) in *H3f4^{H3.1/-}* versus *H3f4^{wt/wt}* elSts (FDR ≤ 0.05) (y axis). Genes up- and downregulated in *H3f4^{H3.1/-}* elSts are highlighted in red and blue, respectively. DEGs are separated into nine groups, based on the $|\log_2\text{FC}| = 1$ cutoff in both contrasts (dashed lines).

d-e, Boxplots showing absolute quantification of exonic and intronic gene expression in groups 1-9 defined in panel d, comparing *H3f4^{H3.1/-}* to *H3f4^{wt/wt}* cells. Paired t-tests were performed between indicated genotypes: * $P < 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$.

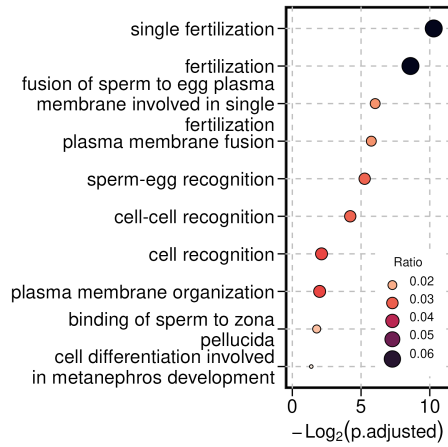
f, GO term search (GO: Biological Process) for genes in groups 1, 4, 6 and 9 (panel c). A one-sided over representation hypergeometric test was used, followed by Benjamini–Hochberg correction for multiple testing.

g, Boxplots displaying ChIP-seq enrichments of different chromatin signatures (H3.4^{MYC}, H3.3, H3K4me3 and H3K27me3 at TSS ± 1 kb) in wildtype rSts for groups 1-9 as defined in panel d.

h, Top: heatmap showing enrichments and adjusted p-values of transcription factor (TF) motifs (log₂) among up- and downregulated genes in *H3f4^{H3.1/-}* elSt cells. Bottom: expression (log₂(RPKM)) of corresponding TF genes in *H3f4^{wt/wt}* and of differential expression (log₂FC) in 3 substitution genotypes of ScLZ, ScPD, rSt and elSt cells. Color annotation shows whether the corresponding TF motif was found among the up- or downregulated genes in *H3f4^{H3.1/-}* elSt cells.

a

GO: Biological Process

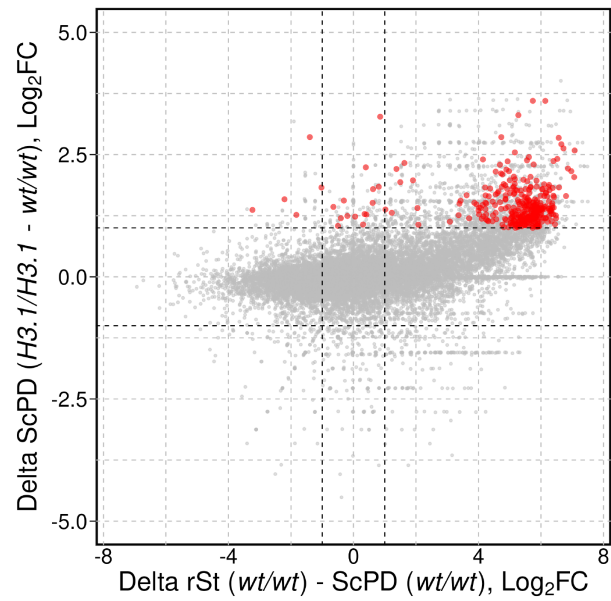


Examples of genes:

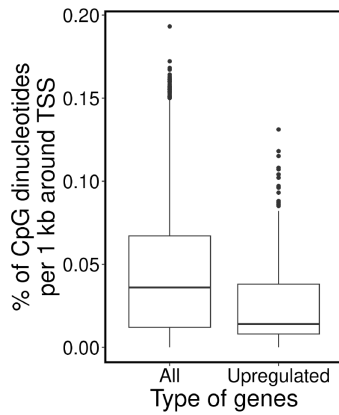
Single fertilization:

Dcst2, Spaca6, Fer115, Irag2, Zan, Prss37, Adam24, Trim36, Spam1, Lyzl4, Plcz1, Spaca7, Izumo1, Cylc2, Dcst1, Crisp4, Spaca4, Catsper1, Foxl2, Fam170b, Spaca3, Iqcf1

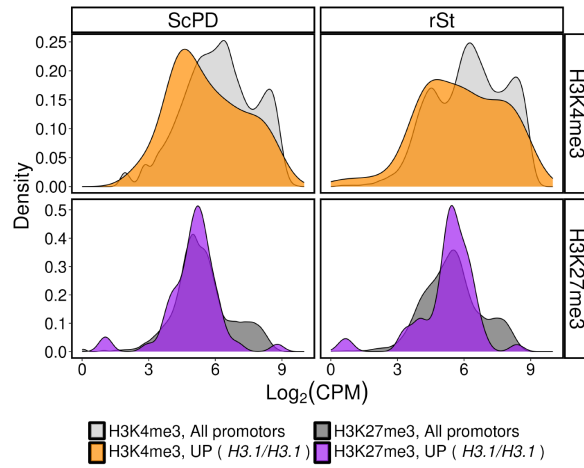
b



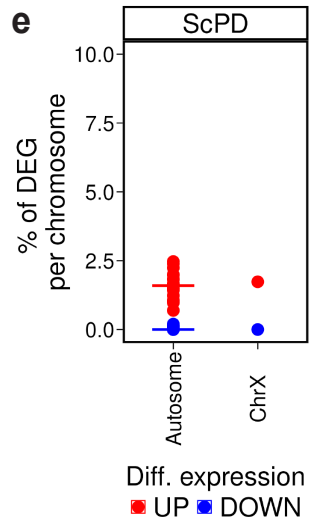
c



d



e



Supplementary Figure 22. Gene ontology and chromatin features of genes differentially expressed in ScPD spermatocytes of *H3f4^{H3.1/H3.1}* animals

a, GO term analysis (GO: Biological process) of genes upregulated in *H3f4^{H3.1/H3.1}* versus *H3f4^{wt/wt}* ScPD spermatocytes. Examples of genes within selected terms are displayed below.

A one-sided over representation hypergeometric test was used, followed by Benjamini–Hochberg correction for multiple testing.

b, Scatter plot showing differential expression (log2FC) of genes in *H3f4^{wt/wt}* rSts versus *H3f4^{wt/wt}* ScPD (FDR ≤ 0.05) (x axis) compared to differential expression (log2FC) in *H3f4^{H3.1/H3.1}* versus *H3f4^{wt/wt}* ScPD (FDR ≤ 0.05) (y axis). Genes upregulated in *H3f4^{H3.1/H3.1}* ScPD are highlighted in red (logFC ≥ 1, FDR ≤ 0.05).

c, Density plots showing the distribution of H3K4me3 and H3K27me3 enrichments at TSS (±1 kb) for promoters of genes upregulated in *H3f4^{H3.1/H3.1}* ScPD (colored fill) and for promoters of all genes (grey fill), as measured in indicated wildtype cell types.

d, Dot plot showing percentages of up- and downregulated genes for each chromosome in *H3f4^{H3.1/H3.1}* ScPD cells. The bars represent the mean values for the selected DEGs. Each dot has been derived from the edgeR estimation of DEGs (panel d) based on 3 biological replicates (n=3) for *H3f4^{wt/wt}* and four biological replicates (n=4) for *H3f4^{H3.1/H3.1}* animals.

a

Valine 24 to Alanine

H3f4^{wt} CTAGCCACGAAGG**T**GGCCCG**C**AAGAGCGCC
Translation ORF/CDS L A T K V A R K S A

H3f4^{V24A} CTAGCCACGAAG**G**CAGCCCG**G**AAGAGCGCC
Translation ORF/CDS L A T K A A R K S A

Histidine 42 to Arginine

H3f4^{wt} CCGCACCGCTAC**C**ACCCCTGGCACGGT**G**GGC
Translation ORF/CDS P H R Y H P G T V A

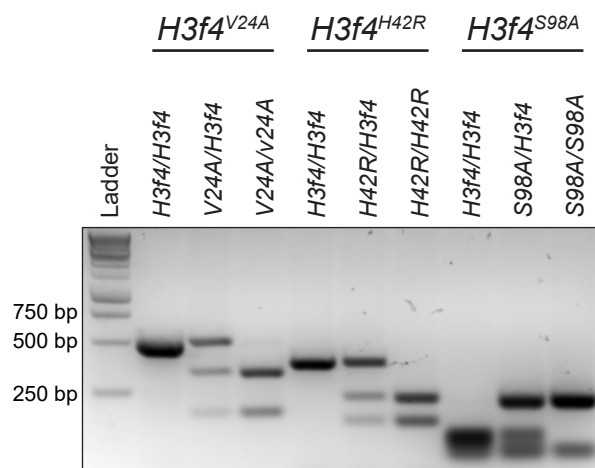
H3f4^{H42R} CCGCACCGCTAC**A**GGCCTGGCACGGT**A**GCG
Translation ORF/CDS P H R Y R P G T V A

Serine 98 to Alanine

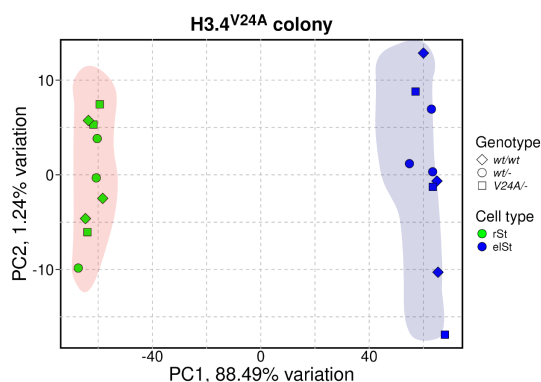
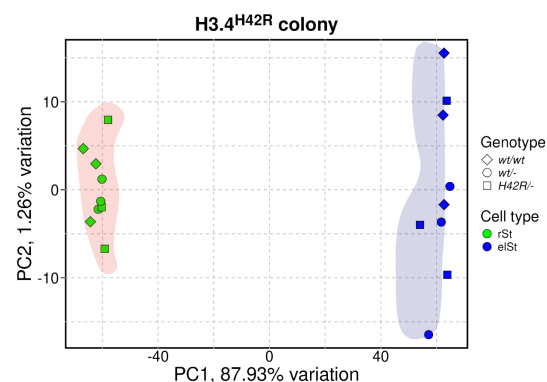
H3f4^{wt} GAGGCCTGCGAG**T**CGTACCT**C**GTGGGCTG
Translation ORF/CDS E A C E S Y L V G L

H3f4^{S98A} GAGGCCTGCGAG**G**CGTACCT**A**GTAGGGCTG
Translation ORF/CDS E A C E A Y L V G L

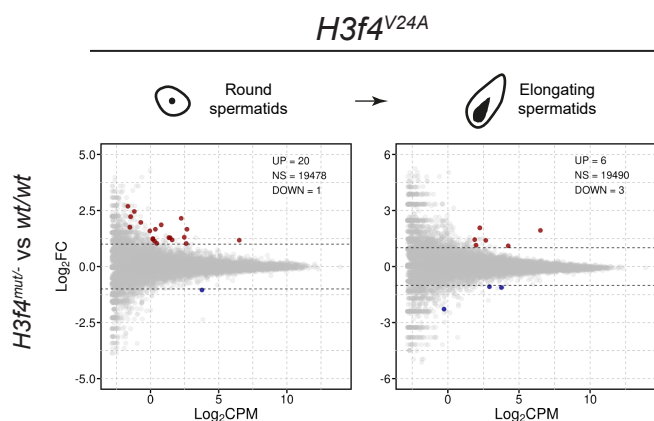
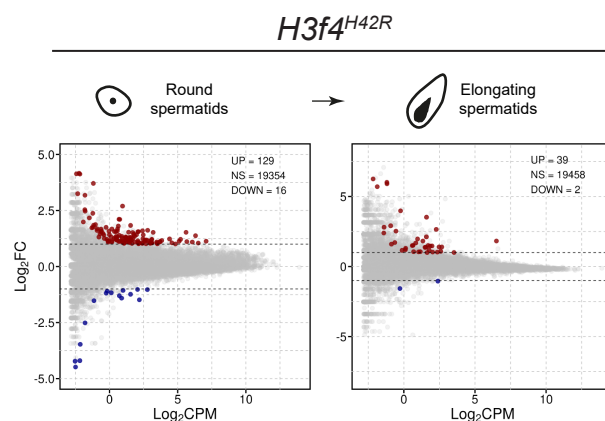
b



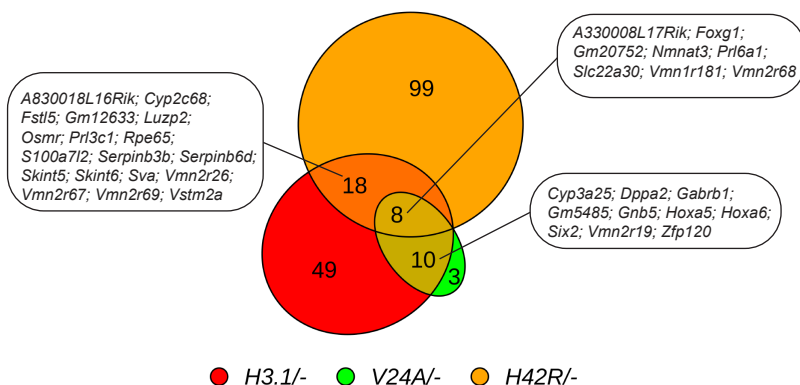
c



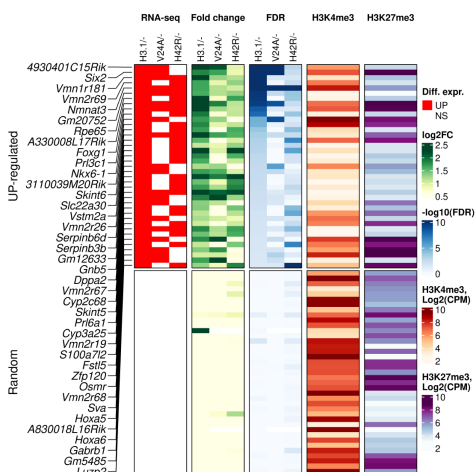
d



e



f



Supplementary Figure 23. Individual contributions of H3.4-specific amino acids to efficiency of spermatogenesis.

a, Schematic representation of *H3f4* point substitution alleles. Variant positions in DNA sequence between wildtype *H3f4* (up) and point mutation alleles (down) are indicated in red. For each mouse model, the DNA sequence of the mutated position within *H3f4* is shown and the translation is shown below. Changes in amino acid sequence are indicated in light red and relevant restriction sites are shown in light green. Changes in DNA sequence that are not annotated are silent mutations that were added to prevent cutting of the sgRNA after successful recombination. Abbreviations: CDS, coding sequence; ORF, open reading frame.

b, Genotyping results of wildtype, heterozygous and homozygous mice from colonies with H3.1-specific point mutations. Fragments were PCR amplified (Supplementary Datas 8 and 9) and digested with the following restriction enzymes: Sapl for *H3f4*^{V24A}, Stul for *H3f4*^{H42R}, HinfI for *H3f4*^{S98A} allele. Fragment sizes after nuclease digestion: Sapl: 516 bp (*H3f4*^{wt/wt}), 340 bp + 176 bp (*H3f4*^{V24A/V24A}); Stul: 384 bp + 27 bp (shorter band not shown) (*H3f4*^{wt/wt}), 227 bp + 157 bp + 27 bp (shorter band not shown) (*H3f4*^{H42R/H42R}); HinfI: 125 bp + 107 bp + 76 bp (*H3f4*^{wt/wt}), 230 bp + 76 bp (*H3f4*^{S98A/S98A}). Ladder with indicated size (bp) is shown for comparison on the left.

c, Principal component analysis (PCA) comparing RNA-expression of USCS-annotated genes among FACS-isolated rSt and eSt cells of *H3f4*^{wt/wt}, *H3f4*^{wt/-}, *H3f4*^{V24A/-} and *H3f4*^{H42R/-} genotypes.

d, Scatter plots showing differential gene expression (log2FC) in *H3f4*^{V24A/-} (left) or *H3f4*^{H42R/-} (right) versus *H3f4*^{wt/wt} FACS-isolated rSt and eSt cells plotted against gene expression levels in control *H3f4*^{wt/wt} cells. Numbers of significantly upregulated (red), downregulated (blue) and non-significant (grey) genes are displayed (logFC ≥ 1, FDR ≤ 0.05).

e, Euler diagram showing overlaps in DEGs between rSt cells of three different mutants: *H3f4*^{H3.1/-} (Fig. 6a), *H3f4*^{V24A/-} and *H3f4*^{H42R/-} (Supplementary Fig. 23d).

f, Heatmap showing ChIP-seq enrichments of H3K4me3 and H3K27me3 in control rSt cells at promoters of genes (TSS ±1 kb; log2(CPM)) upregulated in *H3f4*^{H3.1/-}, *H3f4*^{V24A/-} and *H3f4*^{H42R/-} rSt cells (Fig. 6a). For reference, a randomly selected set of genes is shown below. The EdgeR results for RNA-seq data (significance, fold-change and FDR) are shown for each gene. Names of genes commonly upregulated in at least two cell populations are indicated on the left. H3K4me3 and H3K27me3 in rSt was obtained from Erkek *et al*⁶¹.