Cell Reports, Volume *20*

Supplemental Information

Kmt5a Controls Hepatic Metabolic Pathways by Facilitating RNA Pol II Release from Promoter-Proximal Regions

Kostas C. Nikolaou, Panagiotis Moulos, Vangelis Harokopos, George Chalepakis, and Iannis Talianidis

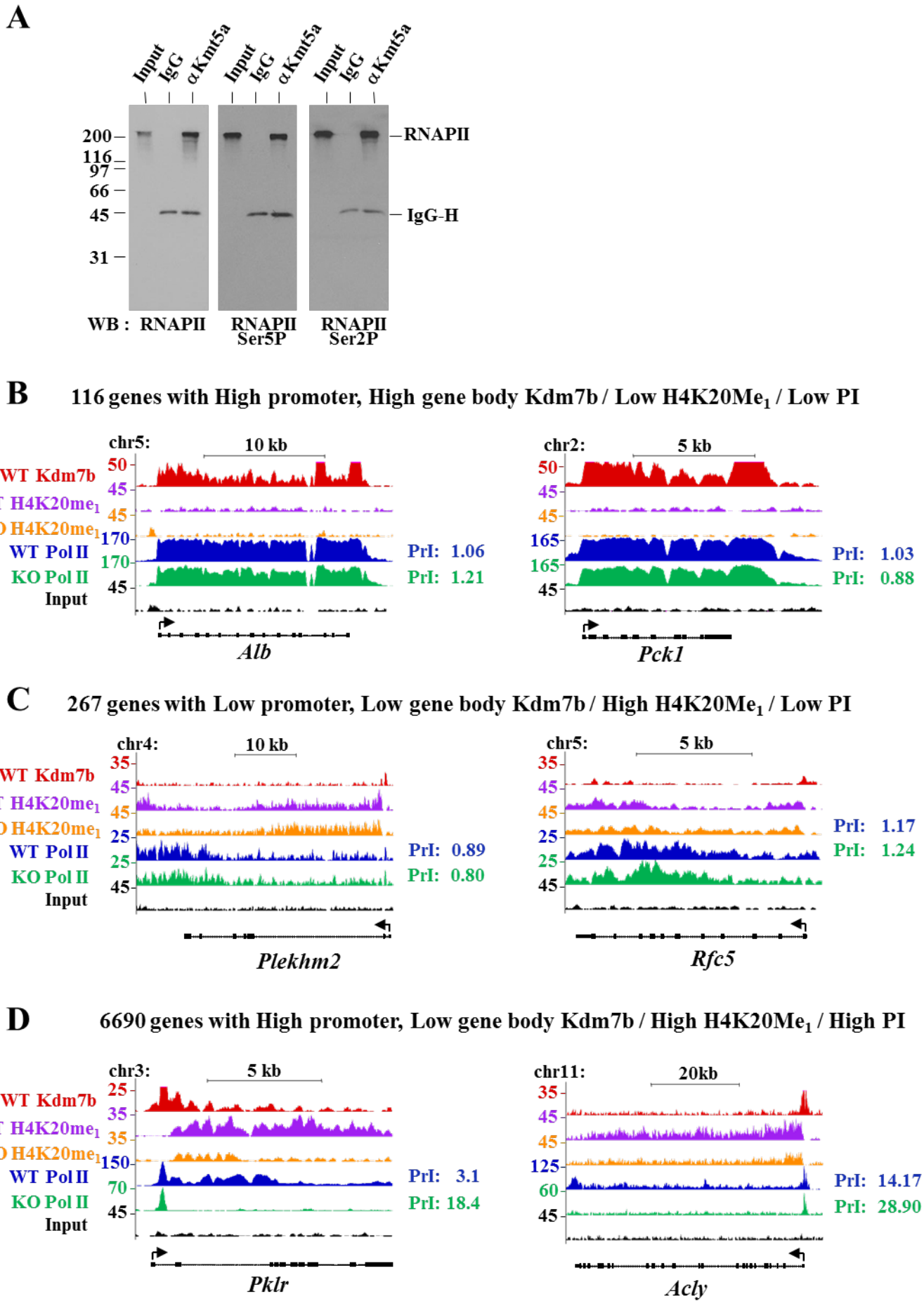
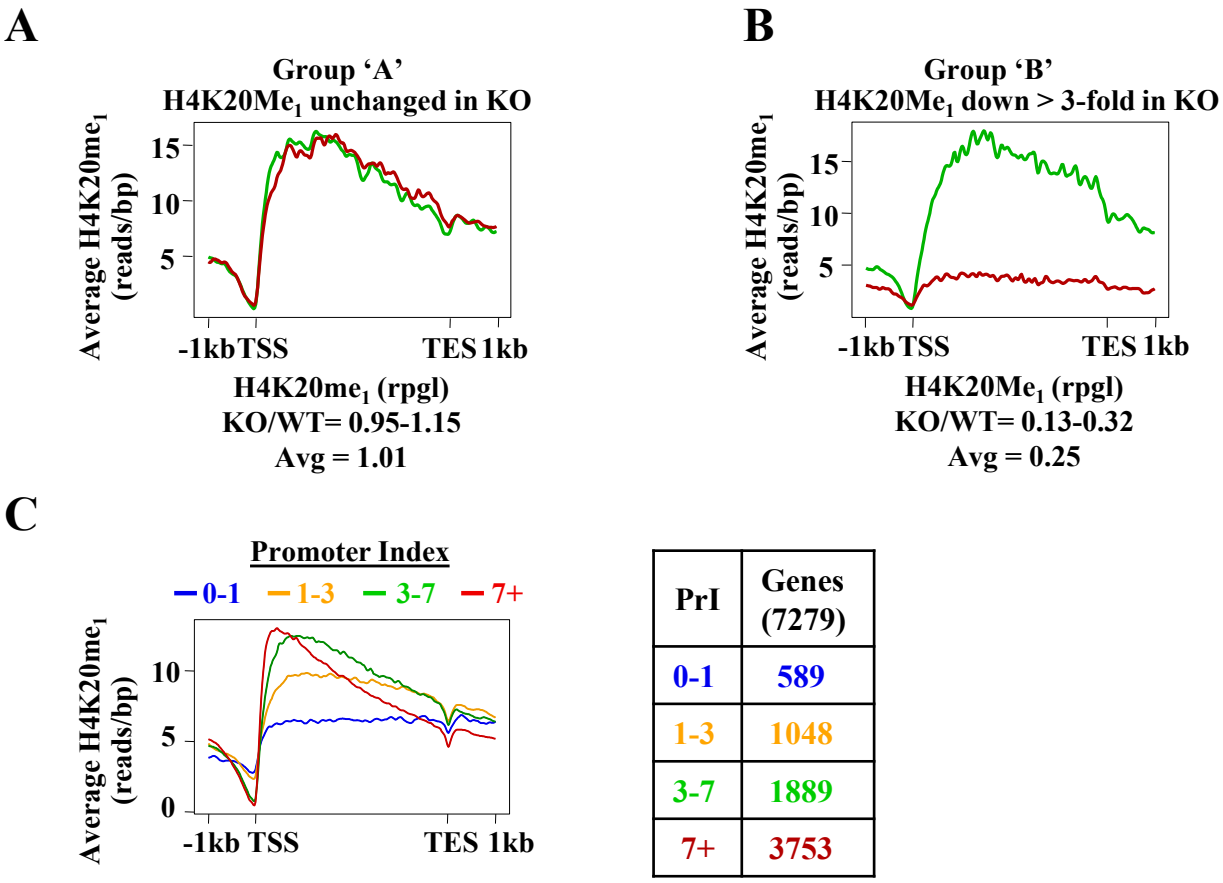

A
Input
IgG
αKmt5a
200
116
97
66
45
31
RNAPII
IgG-H
WB : RNAPII
RNAPII Ser5P
RNAPII Ser2P
B
116 genes with High promoter, High gene body Kdm7b / Low H4K20Me1 / Low PI
chr5:
10 kb
WT Kdm7b
WT H4K20me1
KO H4K20me1
WT Pol II
KO Pol II
Input
PrI: 1.06
PrI: 1.21
Alb
chr2:
5 kb
PrI: 1.03
PrI: 0.88
Pck1
C
267 genes with Low promoter, Low gene body Kdm7b / High H4K20Me1 / Low PI
chr4:
10 kb
PrI: 0.89
PrI: 0.80
Plekhm2
chr5:
5 kb
PrI: 1.17
PrI: 1.24
Rfc5
D
6690 genes with High promoter, Low gene body Kdm7b / High H4K20Me1 / High PI
chr3:
5 kb
PrI: 3.1
PrI: 18.4
Pklr
chr11:
20kb
PrI: 14.17
PrI: 28.90
Acly

Figure S1. Related to Figure 1.

(A) Interaction between Kmt5a and RNA Polymerase-II (RNAPII) in mouse liver. Liver nuclear extracts from P45 wild type mice were prepared and used for immunoprecipitation with anti-Kmt5a antibody. The presence of different forms of RNAPII in the immunoprecipitates was evaluated by western blots assays using anti-RNAPII, recognizing all forms or anti-RNAPII-Ser5P, recognizing the Ser5 phosphorylated form of the C-terminal domain or anti-RNAPII-Ser2P recognizing the Ser2 phosphorylated form of the C-terminal domain of the enzyme.

(B, C and E) UCSC genome browser tracks showing Kdm7b occupancy, H4K20Me$_1$ and RNAPII distribution on representative genes from different categories in the livers of P45 wild type (WT) and *Kmt5a*$^{\Delta HepA}$ (KO) mice. Active, RNAPII-occupied genes (7279) were categorized according to Phf8 occupancy, H4K20Me$_1$ levels and Promoter Index (Read densities at TSS area divided by read densities at gene bodies; PrI). The number of genes fulfilling the indicated triple criteria is indicated at the lines above the graphs. 226 genes did not belong to any of the indicated group. The Promoter Index values of the individual genes are indicated at right.



PrI	Genes (7279)
0-1	589
1-3	1048
3-7	1889
7+	3753

Figure S2. Related to Figures 1 and 2.

(A and B) The distribution of $H4K20Me_1$ in the group of genes displaying low (Group A) and high (Group B) $H4K20Me_1$ methylation turnover. The graphs show average coverage of ChIP-seq reads obtained with an antibody specifically recognizing $H4K20Me_1$.

(C) Active, RNAPII-occupied genes (7279) were categorized according to their Promoter Index (PrI). Average coverage of $H4K20Me_1$ ChIP-seq reads is shown for each category. The number of genes in each category is indicated in the table at right.

A

Biological Process	Enrichment	FDR
Metabolic process	0.0222	1.41E-09
Oxidation-reduction process	0.0275	3.34E-05
Phosphorylation	0.0261	4.68E-05
Lipid metabolic process	0.0440	0.0001
Response to nutrient	0.2105	0.0005
Epithelial cell development	0.3750	0.0012
Positive regulation of glycolysis	0.3333	0.0013
Cholesterol biosynthetic process	0.1538	0.0014
Regulation of transcription from RNA polymerase II promoter	0.0343	0.0019
Lipid biosynthetic process	0.0555	0.0029
Transport	0.0141	0.0046
Cellular response to insulin stimulus	0.1000	0.0049
Ion transport	0.0217	0.0051
Isoprenoid biosynthetic process	0.1875	0.0055
Male gonad development	0.0869	0.0069
Negative regulation of signal transduction	0.0833	0.0072
Mammary gland alveolus development	0.1578	0.0076

B

KEGG Pathway description	Enrichment	FDR
Insulin signaling pathway	0.0592	7.23E-05
Drug metabolism - other enzymes	0.1071	9.91E-05
Pantothenate and CoA biosynthesis	0.1875	0.0037
Type II diabetes mellitus	0.0801	0.0045
Circadian rhythm - mammal	0.1363	0.0050
Valine, leucine and isoleucine degradation	0.0816	0.0052
Fatty acid biosynthesis	0.3333	0.0060
Maturity onset diabetes of the young	0.1153	0.0062
Purine metabolism	0.0363	0.0066
Bile secretion	0.0563	0.0072
Butanoate metabolism	0.1001	0.0076
Propanoate metabolism	0.0967	0.0076
Pyruvate metabolism	0.0731	0.0147
Fatty acid metabolism	0.0666	0.0179
Terpenoid backbone biosynthesis	0.1428	0.0214
Steroid hormone biosynthesis	0.0566	0.0250
Glycolysis / Gluconeogenesis	0.0508	0.0319
Oxidative phosphorylation	0.0305	0.0444

Figure S3. Related to Figure 3.

(A and B) Top Biological Process terms as derived from the respective Gene Ontology enrichment analysis (A) and the top list of biochemical pathways as derived from the respective KEGG Pathway enrichment analysis (B) of the 217 genes showing decreased H4K20Me$_1$ coverage and decreased mRNA levels in the livers of P45 *Kmt5a*$^{\Delta HepA}$ mice. The Biological Processes are ranked by the False Discovery Rate (FDR) values and the number of the support genes that are connected to the respective process according to the Gene Ontology hierarchy.

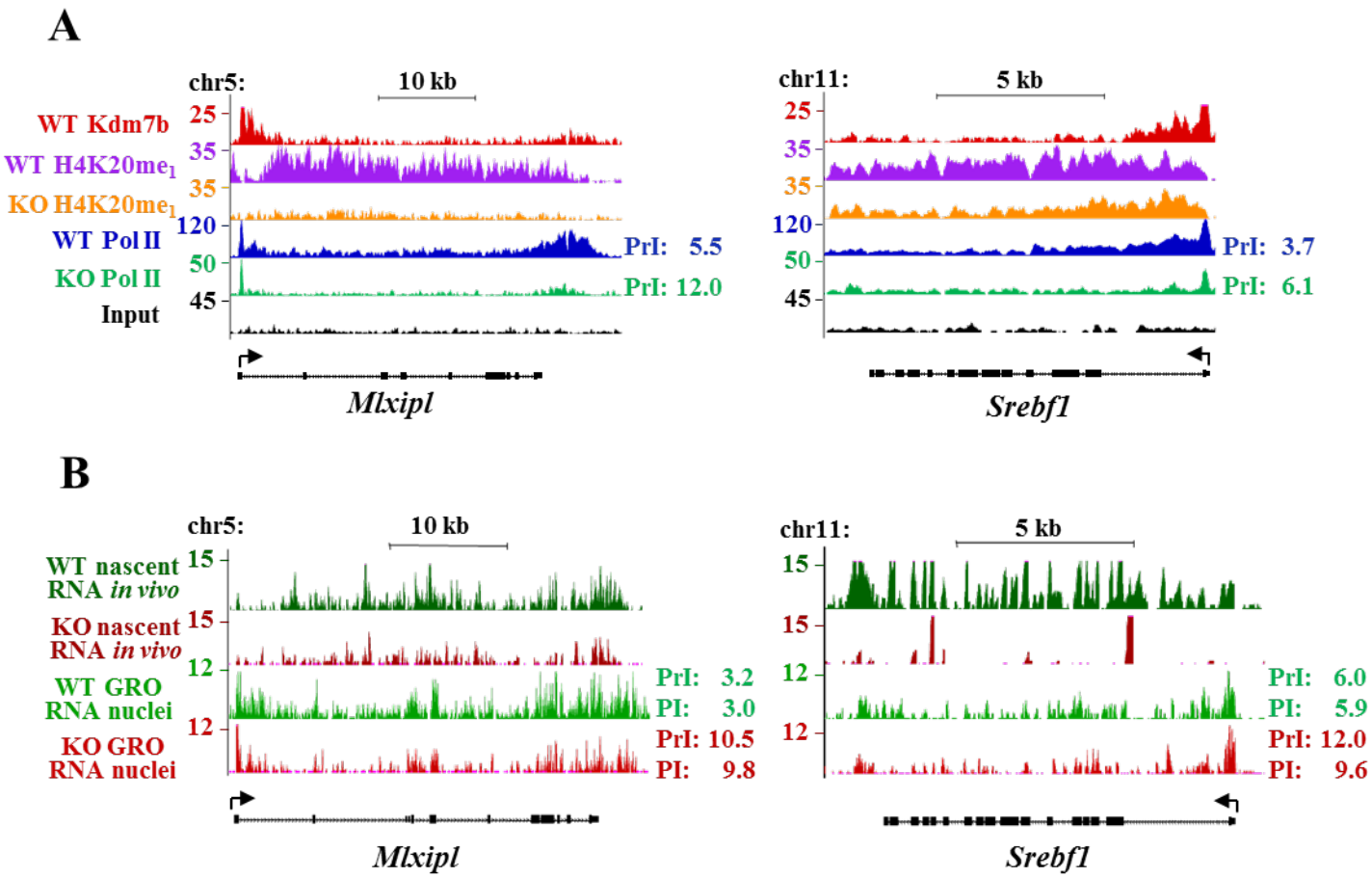


Figure S4. Related to Figure 4.

(A) UCSC genome browser tracks showing Kdm7b occupancy, H4K20Me$_1$ and RNAPII distribution on genes encoding Mlxipl and Srebf1 transcription factors in the livers of P45 wild type (WT) and *Kmt5a*$^{\Delta HepA}$ (KO) mice. The Promoter Index (PrI) value of the individual genes is indicated at right.

(B) UCSC genome browser tracks showing the distribution on Mlxipl and Srebf1 genes of in vivo EU-labeled nascent RNA reads and reads obtained from Global Run-On sequencing in isolated nuclei (GRO-seq) in the livers of P45 wild type (WT) and *Kmt5a*$^{\Delta HepA}$ (KO) mice. The Promoter+Pausing Index (PrI) value (-100 to +100 relative to TSS region reads divided by the reads in downstream gene body region) and the Pausing Index (PI) value (TSS to +50 nt reads per downstream gene body reads) are indicated at right.

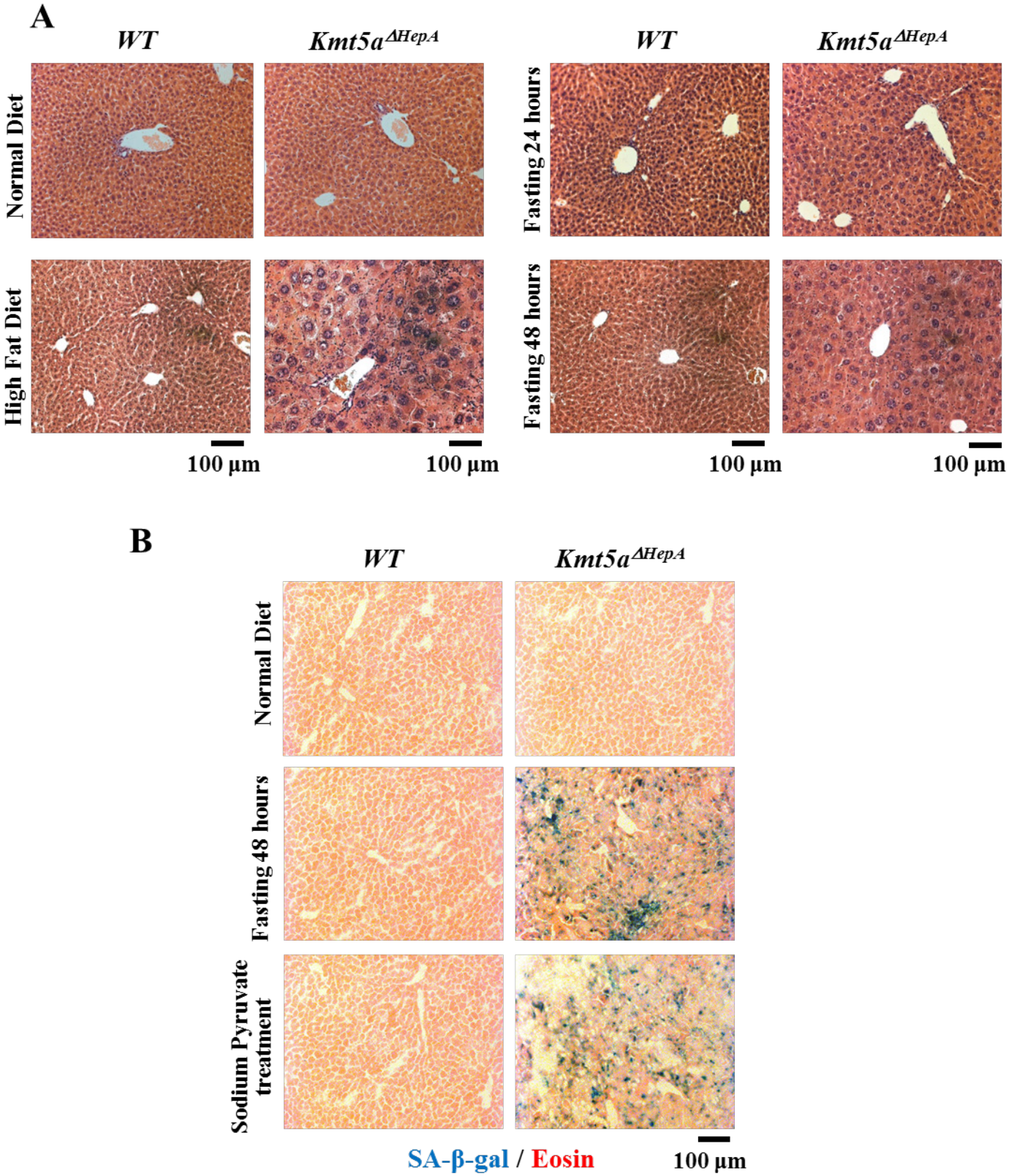


Figure S5. Related to Figure 7.

(A) Hematoxylin and eosin staining of liver sections from P45 wild type (*WT*) and in *Kmt5a*$^{\Delta HepA}$ mice under different feeding conditions. The mice were fed with Normal chow diet (Altromin 1324; 19% protein, 5% fat) or High Fat Diet (34% Fat, 23% Protein, 5% fiber by Mucedola s.r.l.) or fasted for 24 hours or 48 hours, as indicated.
(B) Representative SA-β-gal stainings of liver sections from mice fed with normal chow diet or fasted for 48 hours or treated with 2g/kg Na-pyruvate for 2 hours.

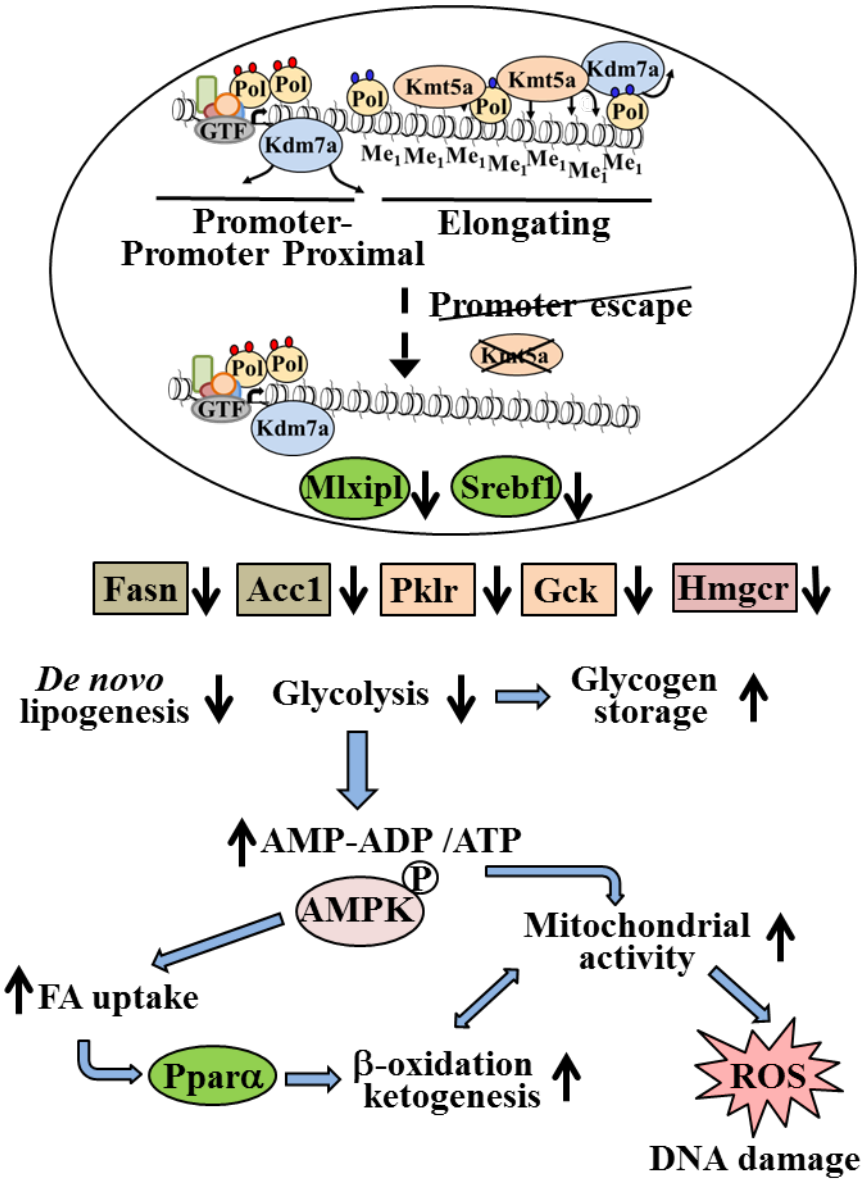


Figure S6. Related to Figures 6 and 7.

Schematic presentation of how alterations of H4K20Me$_1$ in gene bodies lead to ROS-mediated DNA damage in *Kmt5a*$^{\Delta HepA}$ mice.

Table S1 Related to Figure 4. Sequences of PCR primers used in this study.

CD36	NM_001159558.1	5' CAA AGT TGC CAT AAT TGA GTC C 3' 5' CGA ACA CAG CGT AGA TAG ACC T 3'
CYP7B1	NM_007825.4	5' GGG CTA AAT CAC AGT CGG ATG T 3' 5' GAG TGC ACC AAA TCA CGG AG 3'
CYP8B1	NM_010012.3	5' GTA CGC TTC CTC TAT CGC CTG A 3' 5' TGG AGG GAT GGC GTC TTA TG 3'
SCD1	NM_009127.4	5' GGG CAG TTC TGA GGT GAT TAG AG 3' 5' TGG CAG TGG GTA GGT AGT CTT G 3'
ACACA	NM_133360.2	5' GGG GAC AGG TTG TCA CTA TGG G 3' 5' AAA CAC CAG CAC CGA GAC TGA A 3'
ACLY	NM_001199296.1	5' GGC TTC ATT GGG CAC TAC CTT G 3' 5' GTA GGG CTC CTG GCT CAG TTA CA 3'
ACOT1	NM_012006.2	5' TTG GGT AGT GAG TGT TTG CAT GT 3' 5' CAG CCC TTG AAT CAG CAC TAT TT 3'
MLXIPL	NM_021455.4	5' TTG GGG ATG GAG AAA AAC CTA GT 3' 5' GAA GTT GTG AGG AAA CAG ACC G 3'
CPT1a	NM_013495.2	5' GGA AGT GAG GGG CTT GGT AGT C 3' 5' TGG CTG CCT GTG TCA GTA TGC 3'
FASN	NM_007988.3	5' GGA CTT GGG TGC TGA CTA CAA C 3' 5' CCT CCC GTA CAC TCA CTC GT 3'
FGF21	NM_020013.4	5' GAG CAT GGT AGA GCC TTT ACA 3' 5' CGT AAT AAA TAA GTG GAA ACC CAA 3'
HMGCS2	NM_008256.4	5' GAT TCA ACT AGG GGA TGA TCT C 3' 5' CAC TTT CAG GAA TGG GTT ATC T 3'
HMGCR	NM_008255.2	5' TAA TTC ATT TGG TGC TGG TCT AT 3' 5' GCC TCT CTG AAC AAA GAC TCA C 3'
PKLR	NM_013631.2	5' GCG ACC TGG CTC TGG CTA TAC 3' 5' GGC AGA TGT GGG ACT ATG GG 3'
PPARα	NM_011144.6	5' CCT TCT ACG CTC CCG ACC CA 3' 5' CCA TGT CCA TAA ATC GGC ACC A 3'
SREBF1	NM_011480.4	5' GGA CTT CCC ATC TGT TGT AAG G 3' 5' TGT GTT TCT GGA TAT ACG GGG TTA 3'
GAPDH	NM_008084	5' CCA ATG TGT CCG TCG TGG ATC T 3' 5' GTT GAA GTC GCA GGA GAC AAC C 3'
GCK	NM_010292.5	5' GGC CAC AAA CAT TCC AGA 3' 5' GCT GCC ATC CTG CTA ACA 3'
CYP7A1	NM_007824.2	5' CCG TTT CTG TGC AGC CTA AGT 3' 5' TGG AAT TGT AGA CAT AAC GCA AAG A 3'
Kmt5a	NM_030241.3	5' AAA GGC AGG GGT GTG ATC GCT 3' 5' AGT CTC CCT AGG CGG TTC GTT 3'

Supplemental Methods

Antibodies

Antibodies used for this study were from:
Cell Signaling Technology: anti-phospho histone H2A.X #9718; anti-phospho Ser2-Rpb1 CTD #13499; anti-phospho Ser5 Rpb1 CTD #13523; anti-AMPKα #2532; anti-phospho AMPKα Cat#4188.
Santa Cruz Biotrechnology: anti-ChREBP #SC-33764; anti-SREBP1 #SC-8984; anti-GAPDH #SC-32233; anti-TFIIB #SC-225.
Euromedex: anti-RNA Pol II #PB-7G5.
Abcam: anti-CPS1 #ab45956; anti-phospho Ser2-RNA Pol-II #ab5095; anti-PHF8 #ab36068; anti-histone H4 mono methyl K20 #ab9051; anti-LC3B #ab48394

Chromatin Immunoprecipitation (ChIP)

Liver tissue was minced to small pieces in PBS and after addition of formaldehyde to a 1% final concentration immediately was subjected to 10 strokes of dounce homogenization. Cross-linking was continued for 10 min and stopped by the addition of glycine at 0.125 M final concentration. Cross-linked cells were treated with a buffer containing 0.25% Triton-X100, 10 mM EDTA, 0.5 mM EGTA, 20 mM HEPES pH 7.9 and then with a buffer containing 0.15 M NaCl, 1 mM EDTA, 0.5 mM EGTA and 20 mM HEPES pH 7.9 for 10 minutes each. The nuclei were resuspended in 10 volumes of sonication buffer containing 50 mM Hepes pH 7.9, 140 mM NaCl, 1 mM EDTA, 0.1% Na-deoxycholate and 0.5% Sarkosyl and sonicated for 10 minutes in Covaris Sonicator instrument at maximal settings. After centrifugation at 14000 rpm for 15 minutes, the extracted chromatin was supplemented with 0.5 volume of 50 mM Hepes pH 7.9, 140 mM NaCl, 1 mM EDTA, 3% Triton X-100, 0.1% Na-deoxycholate, and incubated overnight with Dynabeads Protein G, that were prebound by 10 μg of the respective antibodies. The beads were washed sequentially first with a buffer containing 50 mM Hepes pH 7.9, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate and 0.1% SDS, then with a buffer containing 50 mM Hepes pH 7.9, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate and 0.1% SDS, followed by a buffer containing 20 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5 % NP-40 and 0.5 % Na-deoxycholate and finally with TE buffer (10 mM Tris pH 8.0, 1 mM EDTA). Immunoprecipitated chromatin was eluted from the beads by two sequential incubations with a buffer containing 50 mM Tris, pH 8.0, 1 mM EDTA, 1 % SDS, 50 mM $NaHCO_3$ at 65 ^{0}C for 15 minutes. After de-crosslinking by incubation at 65^{0}C for 16 hours, RNAs and proteins were removed by incubation with 10 μg/ml RNAse-I for 1 hour at 37 ^{0}C, followed by incubation with 10 μg/ml Proteinase-K for 2 hours at 42 ^{0}C. DNA was extracted by phenol/chlorophorm and precipitated with ethanol.
About 10 ng of the immunoprecipitated DNA and input DNA was used for library preparation following the optimized workflow of Ion Proton system (Life Technologies).

RNA purification and RT-PCR

Total RNA was prepared by homogenizing liver pieces in 10 volumes of Trizol reagent followed by the addition of 2 volumes of chlorophorm and centrifugation at 12000 g for 15 minutes. The aqueous phases were precipitated with ethanol. The RNA samples were digested with 10 units of DNase I for 10 min at 37 ^{0}C, followed by purification with phenol/chlorophorm extraction and ethanol precipitation. For first strand cDNA synthesis 1 μg of total RNA was incubated with 200 units of MMLV reverse transcriptase in a buffer containing 50 mM Tris-HCl pH 8.3, 75mM KCl, 3mM MgCl2, 10mM DTT for 60 minutes at 37 ^{0}C. Quantitative PCR analyses were carried out in STEP-ONE Real time PCR detection system using Fast Start Universal SYBR Green Master.

RNA sequencing

RNA-seq was performed on an Ion Proton^{TM} System as described previously (Sarris et al., 2016). Briefly, approximately 20 μg of total RNAs was used for mRNA isolation with the Dynabeads mRNA DIRECT^{TM} Micro Kit (Life Technologies, Carlsbad, CA, USA). The isolated mRNA was digested with RNase III, hybridized and ligated to Ion Adaptors, reverse transcribed, barcoded and amplified, using the Ion Total RNA-Seq Kit v2 (Life Technologies). Samples were processed on an OneTouch 2 instrument and enriched on a One Touch ES station. Templating was performed using the Ion PITM Template OT2 200 Kit (Life,Technologies) and sequencing with the Ion PITM Sequencing 200 Kit on Ion Proton PITM chips (Life Technologies) according to the manufacturer's protocols.

GRO-seq assays in isolated nuclei

Nuclei from mouse livers were isolated by dounce homogenization of liver pieces in a buffer containing 10 mM Tris pH 7.5, 2 mM $MgCl_2$, 3 mM $CaCl_2$, 2 U/ml RNaseOUT. After filtering through a cell strainer (100 μm, BD Biosciences), the lysates were centrifuged at 400xg for 10 min, and resuspended in the above buffer supplemented with 10% glycerol and 1% NP40. After incubation in ice for 5 min and two successive washes with the same buffer, the nuclei were resuspended in 50 mM Tris pH 8.3, 40% glycerol, 5 mM $MgCl_2$, 0.1 mM EDTA.
$2x10^7$ nuclei were mixed with an equal volume of run-on buffer (10 mM Tris pH 8.0, 5 mM $MgCl_2$, 1 mM DTT, 300 mM KCl, 200 U/ml RNaseOUT, 1% Sarkosyl, 500 μM ATP, GTP, CTP and Br-UTP) and incubated for 5 min at 30°C. RNA was extracted with Trizol reagent. Purified RNA was treated with 10 Units DNaseI (Invitrogen) for 10 min at 37

^{0}C, extracted with acid-phenol and precipitated with ethanol. The RNA was fragmented by hydrolysis in 0.2N NaOH in ice for 10 min. Fragmented RNA was extracted with acid-phenol and precipitated with ethanol. Next, 10 µg anti-BrdU antibody was conjugated to 50 µl Dynabeads by incubation in blocking buffer, containing 0.25x SSPE, 38 mM NaCl, 1 mM EDTA, 0.05% Tween-20, 0.1% PVP, and 0.1% BSA for 1h and the nuclear RNA was incubated with beads for 1 hour, followed by two washes in binding buffer (0.25x SSPE, 38 mM NaCl, 1 mM EDTA, 0.05% Tween-20), two washes in low salt buffer (0.2x SSPE, 1 mM EDTA, 0.05% Tween-20), one wash in high salt buffer (0.2x SSPE, 1 mM EDTA, 0.05% Tween-20, 138 mM NaCl), and two washes in TET buffer (TE pH 7.4, 0.05% Tween-20). BrU-labeled RNA was extracted from the beads using Trizol Reagent and treated with 5 Units RNA 5' Pyrophosphohydrolase (Rpph) for 1h at 37 ^{0}C. The 5' end repaired RNA was extracted using acid phenol followed by ethanol precipitation and treated with 10 units T4 polynucleotide kinase without ATP for 30 min and then with ATP for an additional 30 min at 37 ^{0}C.
The RNAs were then directly hybridized with Ion adaptor Mix V2, by successive incubations at 65°C for 10 min and at 30°C for 5 min, followed by ligation reaction at 30°C 70 min. The samples were then reverse transcribed in a mixture containing Ion RT Primer V2 and Superscript III enzyme. The cDNAs were then amplified by Platinum High Fidelity PCR Supermix, Ion Xpress RNA 3' Barcode primer and Ion Xpress RNAseq BC Barcode primer 1-24.
Due to the T4 polynucleotide kinase step above "strandedness" detection provided by the adaptor primers is not possible in these libraries.

***In vivo* nascent RNA-seq assay (EU-seq)**
This assay is based on short *in vivo* labeling of nascent RNA with the nucleoside analogue 5-ethynyl uridine. EU-labeled nuclear RNA is subjected to fragmentation, followed by biotinylation and purification using streptavidin beads. This procedure, similar to the recently described TT-seq method (Schwalb et al., 2016), uniformly maps nascent transcripts along genes and also maps transient RNA downstream of the polyadenylation sites.
Mice were injected intraperitoneally with 25 mg/kg 5-ethynyl uridine (EU). Mice were sacrificed 30 minutes following EU treatment and liver nuclei were isolated as above. Nuclear RNA was isolated by Trizol extraction and subjected to fragmentation by hydrolysis in 0.2N NaOH in ice for 10 min, followed by biotinylation using Click-iT Nascent RNA Capture Kit according to manufacturer's instructions. Biotinylated RNA was incubated with Dynabeads MyOne Streptavidin T1 magnetic beads for 30 min at room temperature. The beads were washed 5 times with Click-iT reaction wash buffer and biotinylated RNA was isolated by extraction with Trizol. The RNAs were directly processed to hybridizations with Ion adaptor Mix V2, ligations and cDNA synthesis and library preparation as above, without prior removal of 5'Cap structures by RppH treatment.

Short read mapping
The RNA-Seq FASTQ files obtained from the Ion Proton sequencing procedure were mapped on the UCSC mm9 reference genome using a two phase mapping procedure. Firstly, the short reads were mapped using tophat2 (Kim et al., 2013), with default settings and using additional transcript annotation data for the mm9 genome from Illumina iGenomes (http://cufflinks.cbcb.umd.edu/igenomes.html). Next, the reads which remained unmapped were submitted to a second round of mapping using Bowtie2 (Langmead and Salzberg, 2012) against the mm9 transcriptome with the *--local* and *--very-sensitive* local switches turned on. The ChIP-Seq FASTQ files were mapped to the mm9 genome using Bowtie2 with standard parameters. The resulting ChIP-Seq BAM files were processed so as all Ion Torrent Proton reads would have a length of 180bp. The GRO-Seq FASTQ files were mapped using Bowtie2 with standard parameters in two rounds: in the first mapping round, GRO-Seq samples were aligned against the Mus musculus clone RP23-225M6 45S pre-ribosomal RNA gene (https://www.ncbi.nlm.nih.gov/nuccore/22297525) to remove ribosomal RNA contamination which led to the removal of 3 to 10% of short reads. The remaining were aligned against the mm9 genome. All the resulting BAM files were visualized in the UCSC Genome Browser using bedtools and tools provided by the UCSC Genome Browser toolkit.

Reference mouse gene annotation
Prior to any subsequent analyses (gene differential expression, examination of binding genomic profiles etc.) and in order to avoid ambiguities when for example determining the genomic location distribution of H4K20me1 profiles due to potential alternative isoform expression in the mouse strains used, the following data-driven procedure was followed: firstly, a combined mouse gene (UCSC mm9/GRCm37) annotation was constructed from several sources (UCSC, Ensembl, NCBI Entrez) using the Bioconductor package *OrganismDbi*. At the same time, a combined BAM file of RNA PolII short reads in WT was constructed from the two RNA PollII sequenced biological replicates. Then, genomic areas +/- 1kb from the Transcription Start Site (TSS, referred as *promoters*) and +/- 0.5kb from the Transcription End Site (TES, referred as *downstream regions*) were retrieved from the combined annotation described above and the RNA PolII occupancy inside these regions was measured. Based on these occupancy data, a final unambiguous set of 19583 dominant transcripts was constructed by resolving ambiguities (e.g. different isoforms of the same gene) by keeping those whose promoters showed the greatest RNA PolII occupancy as well as the downstream region.

ChIP-seq analyses

H4K20me1 enriched regions calling

The two biological replicates of H4K20Me$_1$ WT ChIP were analyzed with SICER (Xu et al., 2014) in order to detect genomic regions enriched in H4K20Me$_1$ signal. SICER was applied for each replicate with the calling procedure which uses control files (Input DNA for the H4K20me1 samples) with the following parameters: *redundancy threshold=1*, *window size=400*, *fragment size=180*, *effective genome fraction=0.67*, *gap size=1200*, *fdr=0.05*. Prior to applying SICER and in order to compensate for differences in library sizes between ChIP and Input DNA sequencing samples, the total number of reads of each sample was equalized by uniformly downsampling reads relatively to the sample with the lower number of reads. The latter were used for calling H4K20Me$_1$ enriched regions. At the same time, combined H4K20Me$_1$ tracks were created for WT, Kmt5a$^{\Delta HepA}$ and Input DNA by merging the respective duplicates and normalizing as above. These tracks were used for further quality control of the SICER enriched region calling as well as quantifications of the difference between H4K20Me$_1$ signals in WT and Kmt5a$^{\Delta HepA}$. The final set of H4K20Me$_1$ enriched regions was determined by merging the regions called by SICER for each individual replicate and retaining only those demonstrating a fold enrichment greater than 0.58 in log2 scale or 1.5 in natural scale (where fold enrichment is the ratio of reads under an enriched region in the H4K20Me$_1$ ChIP sample to the respective number of reads in the Input DNA sample).

H4K20Me$_1$, H3K4Me$_3$, RNA PolII and Kdm7b data processing

Prior to any subsequent analyses (differential expression, examination of binding genomic profiles etc.), all genes/transcripts derived from the procedure described in the section 'Reference mouse gene annotation' with length less than 1kb were excluded, resulting in a set of 18042 genes/transcripts.

In order to determine the location distribution of the identified H4K20Me$_1$-enriched regions in relation to their distance from the closest genes, facilities from the R/Bioconductor platform were used in a custom script. In addition, the absolute quantifications of H4K20Me$_1$ signals over annotated genes in WT and Kmt5a$^{\Delta HepA}$, were determined by calculating the number of H4K20Me$_1$ Reads Per Window Length (RPWL) normalized by the gene length (taken to be the window length) in H4K20Me$_1$ regions characterized as enriched by SICER. The library sizes were also accounted for the normalization process described in the 'ChIP-Seq analysis' section. A gene was considered to present H4K20Me$_1$ signal if its RPWL was above 0.03 and to present differential H4K20Me$_1$ signal if the RPWL ratio between Kmt5a$^{\Delta HepA}$ and WT was greater than 0.58 in log2 scale or 1.5 in natural scale.

Absolute quantifications for RNA PolII and Ser2 were also determined by calculating the number of sequenced ChIP Reads Per Window Length (RPWL) normalized by the gene length (taken to be the window length) in gene bodies. The Pausing Index (PI) was calculated as the ratio of average normalized coverage (read density) in a window of +/-250bp from the TSS to the average normalized coverage in the rest of the gene body. For the PI calculations, only RNA PolII active genes were used, which were those presenting an absolute quantification greater than 0.01 in WT. H3K4Me$_3$ data were retrieved from Gene Expression Omnibus (GSE44571) and the four sets of H3K4Me$_3$ data in GSE44571 were merged into one.

GRO-seq and Nascent RNA-seq analyses

The two biological replicates of WT and Kmt5a$^{\Delta HepA}$ for Br-UTP-labeled or EU-labeled nascent RNA-seq were merged into one, to create one sample for each condition. Normalization between the two conditions was performed as follows: firstly, the reads overlapping exons were counted and assembled to create a read count value for each gene. Subsequently, the standard Reads Per Kilobase per Million reads (RPKM) score was calculated for each gene, using the sum of exon lengths as the actual gene length in the calculations, resulting in a matrix of RPKM scores, where each row represented a gene while each column represented a sample. Then, genes presenting an RPKM value less than the median RPKM value of all genes and genes non-expressed at all (RPKM=0) in one of WT or Kmt5a$^{\Delta HepA}$ conditions were filtered out. The remaining genes were used to for a simple regression model: $y = ax + b$, where y corresponds to Kmt5a$^{\Delta HepA}$ gene expression values and x to WT gene expression values. The fitted slope a was used for normalization of the coverage profiles, which were calculated after removing duplicate reads from each merged sample and down-sampling the largest library to the size of the lowest to adjust for differences in library sizes.

Coverage profiles

All average coverage profiles were calculated and visualized with the Bioconductor package *recoup* (https://bioconductor.org/packages/release/bioc/html/recoup.html). Briefly, for coverage profiles across gene bodies, as the latter have different lengths, the gene bodies were split in 200 bins of dynamic length each (so that all genes are split in the same number of bins). Then, the ChIP-seq signal was calculated as normalized (from normalized BAM files, see 'ChIP-seq analysis' above) coverage per base-pair for each bin and averaged per bin, resulting in 200 data points for each gene body, comprising the binding pattern for each gene. For coverage profiles across the TSS, the ChIP-seq signal was calculated as normalized coverage per base-pair.

RNA-seq data analyses

The resulting RNA-Seq BAM files were analyzed with the Bioconductor package *metaseqR* (Moulos and Hatzis, 2015). Briefly, the raw BAM files, one for each RNA-Seq sample, were summarized to an exon read counts table, using the

Bioconductor package GenomicRanges (Lawrence et al., 2013) and the Ensembl version 67 (NCBIM37 release, corresponding to mm9). In the resulting read counts table, each row represented one exon, each column one RNA-Seq sample and each cell, the corresponding read counts associated with each row and column. The exon read counts were filtered for artifacts that could affect the subsequent normalization and statistical testing procedures as follows: if an annotated gene had up to 5 exons, read presence was required in at least 2 of the exons, else if an annotated gene had more than 5 exons, then read presence was required in at least $0.2x\lceil E \rceil$ exons, where $\lceil . \rceil$ is the ceiling mathematical function. The final read counts for each gene model were calculated as the sums of their exon reads, creating a gene counts table where each row corresponded to an Ensembl gene model and each column corresponded to an RNA-Seq sample. The gene counts table was normalized for inherent systematic or experimental biases using the Bioconductor package *DESeq* after removing genes that had zero counts over all the RNA-Seq samples. Prior to the statistical testing procedure, the gene read counts were filtered for possible artifacts that could affect the subsequent statistical testing procedures. Genes presenting any of the following were excluded from further analysis: i) genes/transcripts with length less than 500, ii) genes/transcripts with read counts below the median read counts of the total normalized count distribution. Similar expression thresholds (e.g. the median of the count distribution) have been previously used in the literature, where the authors use the median RPKM value instead of normalized counts). The resulting gene counts table was subjected to differential expression analysis for the contrasts Kmt5a-KO versus WT using the statistical test from the Bioconductor package *DESeq*. Genes presenting a *DESeq* p-value less than 0.05 and fold change (for each contrast) greater than 0.58 or less than -0.58 in log2 scale corresponding to 1.5 times up and down in natural scale respectively were considered differentially expressed.

GSEA, GO and pathway analysis

Gene set enrichment analysis was performed with GSEA (Broad Institute) using default gene sets as well as mouse specific gene sets from the Bioconductor package *gskb* (https://bioconductor.org/packages/release/data/experiment/html/gskb.html). Gene Ontology (GO) enrichment and biochemical pathway analysis was performed using GeneCodis (Tabas et al., 2012). The results of GeneCodis for GO and pathway analysis were further filtered in order to remove very general (e.g. 'metabolic_process') or very specific functions (e.g. functions representing single-gene leaves in the GO hierarchy) based on the enrichment score reported by GeneCodis.

H&E staining and immunohistochemistry

Liver tissue was fixed in 4% Formaldehyde/PBS overnight, followed by tissue dehydration and paraffin infiltration. Preparation of tissue sections and staining with hematoxylin and eosin was performed as described in (Tatarakis et al. 2008; Elkouris et al. 2016).

For immunofluorescence staining, freshly isolated liver tissues were embedded in Optimal Cutting Temperature (OCT) embedding medium and were frozen in liquid nitrogen. Frozen sections (5μm thick) were fixed in 4% formaldehyde, blocked in 1% BSA/0,1% Triton X-100 for 1 hr and then incubated with the primary antibodies at room temperature for 2 hr or at 4^0C overnight. After incubation with AlexaFluor 568, (Molecular Probes) secondary antibodies for 1 hr at room temperature and counterstaining with DAPI, fluorescence images were observed using a Zeiss Axioscope 2 Plus microscope.

Accumulation of Reactive Oxygen Species (ROS) was measured as described (Nikolaou et al. 2012), by staining with 5-(and-6)-chloromethyl-20,70-dichlorodihydrofluoresceindiacetate, acetyl ester (CM-H_2DCFDA).

Periodic acid Schiff (PAS) staining was performed in formalin fixed paraffin-embedded liver sections. The sections were treated with 0.5% Periodic acid for 5min and stained in Schiff reagent (Sigma) for 10 min. After washings with PBS, the sections were counterstained with hematoxylin. Liver glycogen levels were measured by using Glycogen Assay II kit (Abcam).

Senescence-associated beta-galactosidase (SA-β-gal) activity was performed in frozen liver sections. After fixation in 0.5% glutaraldehyde for 20 min, the sections were washed with PBS and stained for 3 hours with 5mM of $K_3Fe(CN)_6$ and $K_4Fe(CN)_6$ supplemented with 1mg/ml X-Gal, pH=6. Sections were counterstained with eosin and observed with light microscopy.

Electron microscopy

Liver tissues were fixed for 2 h at room temperature in 0.08 M sodium cacodylate buffer, containing 2% of each glutaraldehyde and paraformaldehyde, followed by 1 h post-fixation with 1% osmium tetroxide. Following 1% uranyl acetate treatment for 20 min, samples were dehydrated with serial ethanol treatment and subsequently embedded in LR White resin/propylene oxide (Polysciences). Approximately 100 nm thin sections on copper grids were observed at 80 kV with a JEOL JM2100 transmission electron microscope.

Glucose tolerance (GTT), Insulin sensitivity (IST) and Pyruvate tolerance (PTT) test

Baseline blood glucose levels were measured from mice fasted for 5 hours (0 min time point). Then, 1g/kg glucose (for GTT) or 0.5 units/kg insulin (for IST) or 2g/kg sodium pyruvate (for PTT) was administered via intraperitoneal injection and blood glucose levels were measured 15, 30, 60, 120 minutes following the treatments. Blood samples

were collected from the tail vein. Glucose concentration was measured using AlphaTRAK glucose monitoring system (Abbott Laboratories).

Serum metabolite and enzyme measurements
Total serum cholesterol, triglyceride, free fatty acid and β-hydroxybutyrate levels were determined by using Cholesterol-cholesteryl Ester quantification kit (Abcam), the Triglyceride Quantification Assay kit (Abcam), the Free Fatty Acid Quantification Assay kit (Abcam) and the beta-Hydroxybutyrate Assay kit (Abcam), respectively. Serum alanine aminotransferase (ALT) activity was determined by using the ALT assay kit from Diasys.

Colorimetric ADP and ATP assays and Cox-I and Cox-IV activity
ADP/ATP levels were determined with ADP/ATP colorimetric assay kits (Biovision). Complex I and IV enzyme activities were measured by using enzyme activity dipstick assay kits (Abcam).

Whole Cell extract preparation
Freshly dissected liver pieces were rinsed in PBS and supplemented with 10 volumes of modified RIPA Buffer containing 50 mM Tris pH7,5, 1% NP40, 0,25% Na-Deoxycholate, 150 mM NaCl, 1mM EDTA pH8, 10% Glycerol, 1mM PMSF and 2 ug/ml Aprotinin. After homogenization with Polytron tissumizer for 10 seconds, the extracted tissue was incubated at 4^0C for 20 min with constant agitation. Extracted proteins were recovered after centrifugation at 14000 rpm for 20 minutes and stored at -80^0C

Nuclear Extracts preparation
Nuclear extracts from liver tissues were prepared as described in (Tatarakis et al., 2008). Briefly, liver tissue was minced to small pieces in a buffer containing 0.32 M sucrose, 15 mM Hepes pH 7.9, 60 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 0.5 % BSA, 0.5 mM spermidine, 0.15 mM spermine and 0.5 mM DTT, supplemented with protease inhibitor coctail (Roche). After 10 strokes of homogenization with Teflon pestle the nuclei were layered over equal volume of a buffer containing 30% sucrose, 15 mM Hepes pH 7.9, 60 mM KCL, 2 mM EDTA, 0.5 mM EGTA, 0.5 mM spermidine, 0.15 mM spermine and 0.5 mM DTT and centrifuged for 15 min at 3000 rpm. Pelleted nuclei were washed with PBS, resuspended in 10 volumes of modified RIPA buffer containing 50 mM Tris pH 7.5, 1% NP40, 0.25% deoxycholate, 150 mM mM NaCl, 1 mM EDTA, 10% glycerol and protease inhibitor coctail (Roche) and incubated at 4 ^{0}C for 20 min with constant agitation. After centrifugation at 14000 rpm, the extracts were collected.

Immunoprecipitation and Western blot assays
Immunoprecipitation or western blot analysis as described (Elkouris et al., 2016). Briefly, nuclear extracts containing ~ 500 μg total protein were mixed with primary antibodies and incubated with 50 μl bed volume Protein-G magnetic Dynabeads for 5 hours at 4 ^{0}C with constant shaking. The beads were washed 4 times with excess RIPA buffer.
For western blot analysis extracted or immunoprecipitated proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes (BIORAD). The membranes were blocked by incubation in TBST buffer containing 20 mM Tris, pH 7.5, 150 mM NaCl and 0.1% Tween 20, supplemented with 5% non-fat dry milk for 1 hour at room temperature. Incubations with primary and HRP-conjugated secondary antibodies were performed in TBST buffer for 1 hour each. After 4 washing steps with excess TBST buffer, the membranes were developed using ECL chemiluminescent kit from Thermo Fisher and exposed to X-ray films.

Supplemental References

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. Science *352*, 1225-1228.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14,* R36

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Moulos, P., and Hatzis, P. (2015). Systematic integration of RNA-Seq statistical algorithms for accurate detection of differential gene expression patterns. Nucleic Acids Res *43*, e25

Tabas-Madrid, D., Nogales-Cadenas, R., Pascual-Montano, A. (2012). GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. Nucleic Acids Res. *40,* W478-83.

Xu, S., Grullon, S., Peng, W. (2014). Spatial Clustering for Identification of ChIP-Enriched Regions (SICER) to Map Regions of Histone Methylation Patterns in Embryonic Stem Cells. Methods Mol. Biol. *1150,* 97-111.