nature research

Corresponding author(s): Tony Kouzarides, George Vassiliou

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.	No sample size calculations were performed. Sample sizes were determined based on previous experience of the specific experimental setup.

2. Data exclusions

Describe any data exclusions.	No Data were excluded

3. Replication

Describe whether the experimental findings were reliably reproduced.	Experiments were successfully replicated the stated number of times in each figure legend.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.	No randomization

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.	No blinding was used

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

The softwares used for the study are all publicly available.

>CRISPR Screens:
Enrichment and depletion of guides and genes were analyzed using MAGeCK statistical package; gRNAs were designed using WTSI Genome Editing website (http://www.sanger.ac.uk/htgt/wge/); Efficiency of genome editing in the pool of sgRNA-targeted cells was evaluated by Tracking of Indels by Decomposition (TIDE); Flow cytometry data were analysed with FlowJo.

>ChIP-seq:
Read quality was assessed using FastQC. Ribosomal contamination was removed by first mapping the reads to hg38 rDNA sequences using bwa and reads were mapped to the hg38 genome using bwa; Multiple reads mapping to a single genomic locus were removed using samtools rmdup. Reads were filtered to remove those with mapping quality less than 20 using samtools. Genome coverage bedgraph files were generated using genomeCoverageBed from the bedtools suite. Coverage files were converted to bigwig format using bedGraphToBigWig. Peaks were called against input sample using MACS2. Read count for each replicate was calculated using the GenomicRanges package in R. Genomic annotation of ChIP peaks and reads and gene set operations were performed taking advantage of the R packages ChIPseeker and VennDiagram, respectively. HOMER tool suite was used for DNA motif discovery.

>RNA sequencing experiments (M6A-IP, Ribosome Profiling, RNA-seq):
Read quality was assessed using FastQC. Ribosomal contamination was removed by first mapping the reads to hg38 rDNA sequences using bwa and reads were mapped to the hg38 genome using Tophat2; Reads were filtered to remove those with mapping quality less than 20 using samtools. Genome coverage bedgraph files were generated using genomeCoverageBed from the bedtools suite. Coverage files were converted to bigwig format using bedGraphToBigWig. Transcript assembly was performed using cufflinks and a single transcript database was generated using cuffmerge.

>m6A RNA-IP:
Statistical analysis of differentially methylated peaks was performed using the R package MeTDiff. Metagene plots were generated by RNAModR package (https://github.com/mevers/RNAModR). For evaluating statistical significance of [GAG]n motif enrichment, individual motif occurrences were searched throughout human transcriptome with FIMO program (MEME suite).

>Ribosome Profiling:
Statistical analysis of differentially translated genes in CTRL and METTL3 KD cells was performed using the R package xtail. In order to estimate the offset value relative to the read start required to localize the position of P-sites we employed the function “psite” of the plastid Python library.

>RNA-seq:
Gene counts were calculated at the transcript level for the combined transcript database from cuffmerge using summarizeOverlaps from the GenomicAlignments package in R. Differential gene expression analysis was conducted using DESeq2. Gene set enrichment analysis was performed using R package GAGE.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	All materials are readily available.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	For the ChIP experiments the following antibodies were used: anti-METTL3 from Bethyl Laboratories (A301-568A), anti-METTL3 from Bethyl Laboratories (A301-567A), rabbit polyclonal anti-METTL14 from Abcam (ab98166), anti H3k4me3 from Abcam (ab8580) and IgG Isotype Control (ab171870). Western blot experiment were performed using the following antibodies: anti-METTL3 from Bethyl Laboratories (A301-568A), anti-METTL14 from Abcam (ab98166), anti-Histone H3 from Active motif (39763), anti-WDR5 from Abcam (ab178410), anti-CEBPZ from Abcam (ab176579), anti-SP1 from Abcam (ab13370) and anti-SP2 from Abcam (ab137238), anti-ACTIN from Abcam (ab8227). For each genomic experiment the antibody lot number is provided on the GEO submission.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	All human cancer cell lines were obtained from the Sanger Institute Cancer Cell Collection.
b. Describe the method of cell line authentication used.	Cell lines were checked for morphology by microscope, as indicated by ATCC.
c. Report whether the cell lines were tested for mycoplasma contamination.	Cells were routinely tested for mycoplasma contamination by PCR.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	None of the cell lines used is listed within the ICLAC dataset.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	1 x 10^5 cells were transplanted into female 6-week old Rag2-/- IL2RG-/- mice by tail-vein injection. At day 14 post-transplant, the tumour burdens of the animals were detected using IVIS Lumina II (Caliper) with Living Image version 4.3.1 software (PerkinElmer). Animals were culled when the tumour burden was 108 photons per second or higher. Diseased mice welfare was assessed blindly by qualified animal technicians from the Sanger mouse facility. All animal studies were carried out in accordance with the Animals (Scientific Procedures) Act 1986, UK and approved by the Ethics Committee at the Sanger Institute.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	The study does not involve human research participants.

nature research

Corresponding author(s): Tony Kouzarides, George Vassiliou

☐ Initial submission ☐ Revised version ☒ Final submission

ChIP-seq Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data deposition

1. For all ChIP-seq data:

☒ a. Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ b. Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

2. Provide all necessary reviewer access links. *The entry may remain private before publication.*	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE94613
3. Provide a list of all files available in the database submission.	For each file we provide both raw sequencing data in SRA format and track-level information in bigwig format. Peak information is provided in the Supplementary Tables for easier access (equivalent to bed files).
4. If available, provide a link to an anonymized genome browser session (e.g. UCSC).	N/A

Methodological details

5. Describe the experimental replicates.	For the METTL3, METTL14, H3K4me3 and IgG ChIP-Seq experiment 5, 5, 2 and 3 independent biological replicates were used, respectively.
6. Describe the sequencing depth for each experiment.	Total read number for each single replicate was on average around 30M, of which 85% unique mapping.
7. Describe the antibodies used for the ChIP-seq experiments.	IP antibody: H3K4me3; vendor: Abcam; catalog num: ab8580; lot num: GR240214-4; validation: http://www.abcam.com/histone-h3-tri-methyl-k4-antibody-chip-grade-ab8580.html IP antibody: METTL3; vendor: Bethyl Laboratories; catalog num: A301-568A; lot num: 1; IP antibody: METTL14; vendor: Abcam; catalog num: ab98166; lot num: GR277753-1;
8. Describe the peak calling parameters.	Peaks were called against input sample using MACS2 using default parameters. Peaks from all replicates were merged to give a master list of potential binding loci per condition, and read count (normalised by overall read depth of the library) for each replicate was calculated using the GenomicRanges package in R. Peaks were treated as potential binding loci if all replicates showed normalised score greater than 1 and did not overlap a peak called in the IgG. Genomic annotation of ChIP peaks and reads and gene set operations were performed taking advantage of the R packages ChIPseeker and VennDiagram, respectively.
9. Describe the methods used to ensure data quality.	In our analysis, METTL3 called peaks all had by definition FDR < 5% and 36% of them displayed fold enrichment greater than 5 (99% had FE greater than 2.5%)
10. Describe the software used to collect and analyze the ChIP-seq data.	Reads were trimmed to remove the TRUseq adapter using trim_galore with parameters '-q 0 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC --phred33 --fastqc'. Read quality was assessed using FastQC. Ribosomal

contamination was removed by first mapping the reads to hg38 rDNA sequences using bwa36 (default parameters), and removing all reads that mapped. Trimmed non-ribosomal reads were mapped to the hg38 genome using bwa with parameters '-n 3 -k 2 -R 300 -t 4'. Multiple reads mapping to a single genomic locus were treated as PCR duplicates, and were removed using samtools rmdup. Mapped reads were filtered to remove reads mapping to more than one unique genomic locus (multihits) by keeping only reads with flag XT:A:U in the output bam file from bwa. Reads were further filtered to remove those with mapping quality less than 20 using samtools. Genome coverage bedgraph files were generated using genomeCoverageBed from the bedtools suite of tools. Coverage files were converted to bigwig format using bedGraphToBigWig. HOMER tool suite was used for DNA motif discovery coupled with the hypergeometric enrichment calculations (or binomial) to determine motif enrichment.

nature research

Corresponding author(s): Tony Kouzarides, George Vassiliou

☐ Initial submission ☐ Revised version ☒ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.	Cells were detached by trypsinization, washed and resuspended in PBS at RT. For immunodetection experiments, cells were stained with anti-mouse CD11b PE/Cy5 (Biolegend, cat. no. 101210) and anti-human CD11b PE (eBiosciences, cat. no. 9012-0118).
6. Identify the instrument used for data collection.	All flow cytometry data were collected on a LSRFortessa (BD) except for sorting experiment, which were performed on SH800S Cell Sorter (Sony).
7. Describe the software used to collect and analyze the flow cytometry data.	Data were acquired with the default control softwares of the machines employed, exported as .FCS files and analyzed with FlowJo software (FLOWJO, LLC).
8. Describe the abundance of the relevant cell populations within post-sort fractions.	The purity of the relevant cells was over 98% in the post-sort fraction as assessed by flow cytometry.
9. Describe the gating strategy used.	Cells were gated according to physical parameters in order to discard cell debris (FSC/SSC) and cell clumps (Width vs. Area). Dead cells were exclused by selecting the DAPI or PI- negative cell populations. Fluorescent cells were gated with a threshold capturing the 0.5% upper tail of a negative control population.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒