nature

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41586-018-0350-5

In the format provided by the authors and unedited.

Glucose-regulated phosphorylation of TET2 by AMPK reveals a pathway linking diabetes to cancer

Di Wu[1,2,7], Di Hu[1,2,7], Hao Chen[1,3,7], Guoming Shi[1,2,7], Irfete S. Fetahu[2], Feizhen Wu[1,2], Kimberlie Rabidou[2], Rui Fang[2], Li Tan[1], Shuyun Xu[4], Hang Liu[1], Christian Argueta[2], Lei Zhang[5], Fei Mao[6], Guoquan Yan[5], Jiajia Chen[5], Zhaoru Dong[1], Ruitu Lv[1], Yufei Xu[2], Mei Wang[2], Yong Ye[1], Shike Zhang[2], Danielle Duquette[2], Songmei Geng[4], Clark Yin[2], Christine Guo Lian[4], George F. Murphy[4], Gail K. Adler[2], Rajesh Garg[2], Lydia Lynch[2], Pengyuan Yang[5], Yiming Li[6], Fei Lan[1], Jia Fan[1], Yang Shi[1,3] & Yujiang Geno Shi[1,2]*

[1]Key Laboratory of Medical Epigenetics and Metabolism, Institute of Clinical Science of Zhongshan Hospital and Institutes of Biomedical Sciences, Fudan University, Shanghai, China. [2]Division of Endocrinology, Diabetes and Hypertension, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [3]Division of Newborn Medicine and Program in Epigenetics, Department of Medicine, Boston Children's Hospital, Boston, MA, USA. [4]Program in Dermatopathology, Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [5]Department of Chemistry and Institutes of Biomedical Science, Shanghai Medical School, Fudan University, Shanghai, China. [6]Division of Endocrinology and Metabolism, Huashan Hospital, Fudan University, Shanghai, China. [7]These authors contributed equally: Di Wu, Di Hu, Hao Chen, Guoming Shi. *e-mail: yujiang_shi@hms.harvard.edu

Supplementary Information

Supplementary Discussion

Dynamics of 5hmC in response to glucose

In Extended Data Figure 1i-j, results were shown for the short term and long term study of 5hmC dynamics in A2058-TET2WT cells. In Extend Data Figure 1i, the cells were starved overnight and were then transitioned into high-g culturing condition for a 4 day period. The initial, acute high-g shock after the starvation resulted in an early phase (within 30 min) of increased 5hmC levels, which is consistent with the recent study by Yang *et al*[32]. However, sustained exposure (after 4 hours) to high-g resulted in a decline of 5hmC levels back to baseline and the level remained at the baseline thereafter. In Extended Data Figure 1j, cells began culturing (with no starvation) under high-g (red line) or normal-g (blue line) on day 0. Cells cultured in high-g exhibited constant basal low levels of 5hmC over 13 days. However, cells cultured in normal-g showed a gradual increase of 5hmC after 4 days. The 5hmC level steadily accumulated throughout the later phase and reached a plateau at day 13. Importantly, we also observed a rapid return to basal 5hmC levels when these cells were transitioned from normal-g to high-g condition (blue line, days 13 to 15). Extended Data Figure 1i shows that substrate level effect on 5hmC is very transient, and the sustained increase of 5hmC in Extended Data Figure 1j is likely due to the effects on the enzyme itself.

We should note that the long latency for 5hmC accumulation after the switch from high-g to normal-g is intriguing and may require further investigation to determine the underlying reason(s). One possible factor is that it would take time for the cells to reach the metabolic threshold for reactivating AMPK. Another contributing factor could be that the dilution effect of active cell division overshadows the accumulation of 5hmC, therefore resulting in a slow accumulation of net 5hmC during the initial days after the glucose switch. Also, it is possible there are other factors involved that remain to be discovered.

The inverse correlation between DNA 5hmC levels and glucose

The data in Extended Data Figure 1-3 collectively show an inverse correlation between DNA 5hmC levels and hyperglycemic condition. Moreover, DhMR analysis and gene expression profiling illustrate a TET2-dependent, genome-wide reprogramming of the hydroxymethylome and transcriptome under the influence of glucose, both of which are strongly associated with cancer pathways. Lastly, these data not only imply that a functional, full-length TET2 is required to mediate the reversible changes of 5hmC in response to extracellular glucose availability, but also suggest that TET2 protein is less stable in high-g, raising an intriguing question of how glucose levels impact the regulation of TET2 protein stability and 5hmC level.

Supplementary Acknowledgement

This work was in part supported by a grant from the Key Project of the National Science Foundation of China to J. Xu (81130047), the National Natural Science Foundation of China (grants 2009CB825602, 2009CB825603, and 81272295 to the Epigenetics Laboratory at the Institutes of Biomedical Sciences, Fudan University Medical School), National Institutes of Health awards R01-HL-087060 (G.K.A.), K24-HL-103845 (G.K.A.), UL1 TR000170 (Harvard Catalyst, Harvard Clinical and Translational Science Center) and Brigham and Women's Hospital Center for Clinical Investigation. Y.S. is an American Cancer Society Research Professor.

Methods

Antibodies and Reagents

Anti-5hmC antibody (#39769) was purchased from Active Motif. Anti-FLAG antibody (M2) was purchased from Sigma Aldrich. Anti-AMPK (#2532), anti-AMPKpThr172 (#2535) and anti-pACC (#3661) were purchased from Cell Signaling Technology (CST). Anti-lamin A/C (#ab8984) and anti-p53 (#ab179477) were purchased from Abcam. Anti-TET2 (#NBP2-32104) was purchased from Novus Biologicals. Anti-TET3 (#GTX628893) was purchased from GeneTex. Anti-actin (#AC004) was purchased from ABclonal Biotechnology. Anti-tubulin (#66240) was purchased from Proteintech. Anti-TET2pS99 antibody was homemade, raised against GGIKRTVpSEPSLSGL peptide (TET2 aa 92-106), and purified through subsequent positive and negative selection using the GGIKRTVpSEPSLSGL phosphor peptide

and the unphosphorylated GGIKRTVSEPSLSGL peptide, respectively. Metformin (#D6198-25g) was purchased from Sangon Biotech. Cycloheximide (CHX), anti-FLAG antibody conjugated beads, Calpain protease inhibitor and Streptozotocin were all purchased from Sigma. AMPK activator A769662 (#S2697) was purchased from Selleckchem. Cell Proliferation Assay (MTS) kit was purchased from Promega. Lipofectamine2000 reagent was purchased from Thermo Fisher Scientific. DNeasy Blood & Tissue Kit (#69504) was purchased from QIAGEN.

Constructs

Human TET1 (NM_030625.2), TET2 (NM_001127208.2) and TET3 (NM_001287491.1) were amplified from HEK293T cells and subcloned into pOZN and pPB vectors as previously described[33]. To generate the TET2 catalytically inactive mutant (TET2M), the TET2 iron-binding site H1382RD1384 was mutated to Y1382RA1384 using the QuickChange Site-Directed Mutagenesis Kit (Stratagene). The TET2S99A, TET2S1205A, TET2SLF and TET2S99D mutants were generated using the same kit per the manufacturer's instructions. Mock, TET2WT, and TET2M containing retroviruses were generated and the virus-infected A2058 cells were selected as previously described[34]. To obtain pure monoclonal stable cell lines, the selected cells were serially diluted and the expression of TET2WT, TET2S99A, TET2S99D and TET2M was verified by RT-qPCR at the mRNA level and by Western blot at the protein level.

Cell Culture

Dulbecco's Modified Eagle Medium (DMEM) high glucose (4.5 g/L) and normal glucose (1 g/L) were used with 10% (v/v) fetal calf serum (Biological Industries) and 1% penicillin/streptomycin for all the cell lines except for TF-1 cells. TF-1 cells were cultured in RPMI 1640, 10% fetal bovine, serum (FBS), 2 ng/ml Human Granulocyte Macrophage Colony Stimulating Factor (GM-CSF). Cells were grown in high glucose or normal glucose medium at 37˚C in 5% CO_2. For long-term (≥ 7 days) high glucose or normal glucose treatment, cells were split periodically (see below) into 10 cm plates (10% confluence after splitting) and reached 60%-70% confluence at collection. For PBMCs, cells were grown in high glucose or normal glucose for two days, and then collected for following Western blot or DNA extraction. *In vitro* metformin treatments were done in high glucose DMEM. It should be noted that, although consistent with the accepted practice in the field[35,36], the metformin dosage used in these *in vitro* experiments (up to 5 mM) are higher than the circulating metformin concentration (~20 μM) in humans taking the drug[35,36]. Cell lines used have been tested to be free of mycoplasma contamination by VENORGEM mycoplasma detection kit (Sigma Aldrich).

5hmC dynamics in high-g and normal-g treatments

Cells were initially cultured in high-g for 2 days after thawing from a frozen vial, and then switched to fresh high-g or normal-g to begin respective treatments on day 0. Cells were cultured in the respective medium for 13 days, with splitting on days 2, 4,

7, and 10. DNA was collected for 5hmC analysis on days 0, 2, 4, 7, 10, and 13. After the normal-g to high-g switch, DNA was collected at 1 h, 2 h, 8 h, 24 h, and 48 h. To study the effect of high-g shock on 5hmC during the first few hours (the short-term high-g curve), cells were starved in no glucose DMEM overnight before being treated with high-g. On day 0 (the beginning of the short-term curve), cells were changed to high-g medium. DNA was collected for 5hmC analysis at the following time points: 0 min, 5 min, 10 min, 30 min, 2 h, 4 h, 8 h, 24 h, and day 4.

Transfection and infection

Three different small hairpin RNAs (shRNAs) purchased from 3DMAX were designed to knock down AMPKα2 or TET2 expression in A2058-TET2WT and TF-1 cells. AMPKα2 shRNA-1: CGCAGTTTAGATGTTGTTGGA, shRNA-2: GTGGCTTATCATCTTATCATT. TET2 shRNA: ACACCCTCTCAAGATTGTTTA. The shRNA constructs together with packing plasmids (psPAX2 and pMD2.G) were transiently transfected into 293T cells using Lipofectamine2000 according to the manufacturer's protocol. Conditioned medium containing lentiviral particles was harvested 72 h after transfection and was then used to infect recipient A2058-TET2WT or TF-1 cells. Lentivirus-infected A2058-TET2WT or TF-1 cells were selected with 2 μg/ml puromycin and collected for future assays.

Genomic DNA Extraction from human blood samples

Genomic DNA was purified from blood samples of 28 normal donors and 29 diabetic patients by DNeasy Blood & Tissue Kit (QIAGEN) according to the manufacturer's protocol. Partners HealthCare Institutional Review Board approved the protocol, and all participants provided written informed consent. The clinical information is listed in the following table. The univariate regression analysis was done by SPSS.

		Control		Diabetes	
Variable	Units	Mean	SD	Mean	SD
Age	years	26.5	7.3	50.7	8.7
BMI	Kg/m^2	23.4	3.3	37.7	7.3
Systolic	mmHg	114.7	8.4	129.3	16.8
Diastolic	mmHg	67.1	9.9	76.8	8.7
Age of Diabetes DX	Years			40.2	11
Duration of DM	years			10.5	6.2
WBC	x10E3/uL	5.7	1.4	6.7	1.6
Random Glucose, Serum	mg/dL	86	10.6	254.7	118.7
BUN	mg/dL	13.6	4	15.2	4.5
Creatinine, Serum	mg/dL	1	0.2	0.9	0.2
eGFR	mL/min/1.73	95.1	26.4	75.9	23.4
Sodium, Serum	mmol/L	139.1	2.1	137.1	2.9
Potassium, Serum	mmol/L	4.2	0.3	4.2	0.3
AST	IU/L	23.9	19.9	26.8	20.7
ALT	IU/L	20.9	14.1	32.6	22.9

Genomic DNA Extraction from cultured cells and Dot Blot

Cultured cells were harvested and re-suspended in digestion buffer (100 mM Tris pH 8.0, 100 mM NaCl, 25 mM EDTA, 0.5% SDS) at 50˚C for 12 to 18 hours. Equal volumes of phenol-chloroform were added to the lysate and vortexed. The upper aqueous phase was separated by centrifugation and transferred to new tubes. 1/10

volume of sodium acetate and 2 volumes of ethanol were added to let the genomic DNA precipitate. DNA was washed using 70% ethanol and dissolved in TE buffer (10 mM Tris pH 8.0, 1 mM EDTA). RNase A was then added into the DNA solution and incubated at 37˚C overnight, after which the phenol-chloroform extraction was repeated. For dot blot assay, 1μg DNA was first sonicated and leveled to 1000 ng per 60 μl. Leveled DNA was then loaded into a 96-well plate followed by two additional serial dilutions (final volume was 30 μl per well). 20 μl of buffer (1 M NaOH, 25 mM EDTA) was added and samples were incubated at 95˚C for 10 minutes. 50 μl of ammonium acetate (2 M, pH 7.0) was then added and cooled on ice for another 10 minutes. Meanwhile, the nitrocellulose membrane was incubated in double-distilled H_2O and then in 6×SSC buffer (0.9 M NaCl, 90 mM $Na_3C_6H_5O_7$) for 20 minutes. After setting up the dot blot apparatus (BIO-RAD), the membrane was rehydrated with TE buffer. Then, the denatured DNA was applied and washed by 2×SSC buffer (0.3 M NaCl, 30 mM $Na_3C_6H_5O_7$). The membrane was then cross-linked by UV light for 20 minutes and subjected to the Western blot protocol for the remaining steps.

Co-immunoprecipitation

Cells were lysed with lysis buffer (20 mM TrisCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 0.5% NP-40, protease inhibitors (including calpain inhibitor) and phosphatase inhibitor cocktail (Roche Applied Science)). The extract was spun at 14,000 rpm for 15 min at 4°C. Two percent was kept for input, while the rest was incubated with the FLAG antibody for 1 hr at 4°C. Protein A/G beads (Millipore) were then added for

overnight incubation at 4°C. The beads were washed five times with lysis buffer, and the bound proteins were boiled in SDS sample buffer and blotted with indicated antibodies.

Full-length TET2 Recombinant Protein Expression and Purification

Genes for full length TET2 WT and mutants were subcloned into a modified pFastBac-HTA vector with a FLAG affinity tag fused to the N terminus. Bacmids were generated in DH10Bac cells following the Bac-to-Bac baculovirus expression system protocol (Invitrogen). Baculovirus was generated and amplified in Sf-9 insect cells. For protein expression and purification, Sf-9 insect cells were grown in Sf-900™II medium supplemented with FBS. TET2 protein was expressed at 27°C for 72 hours using the previously generated viruses. Infected cells were harvested by centrifugation at 2,000g for 15 min and homogenized in ice-cold lysis buffer (25 mM Tris pH 8.0, 300 mM NaCl, and 0.5 mM phenylmethanesulfonyl-fluoride (PMSF)). The cells were disrupted using a cell homogenizer and the insoluble fraction was precipitated by ultracentrifugation (20,000g) for 30 min at 4°C. FLAG M2 Agarose beads (Sigma-Aldrich) were added into supernatant, incubated for 2 hours and washed four times with lysis buffer. Elution was performed twice with buffer containing 50 mM Tris, pH 8.0, 300 mM NaCl and 100 μg/ml FLAG peptide.

***In Vitro* Kinase Phosphorylation Assay**

Active recombinant human AMPK (α2, β1, γ1) holoenzyme (14-902) was purchased from EMD Millipore. The recombinant FLAG-TET2 (WT/S99A/SLF/S1205A) proteins purified from Sf-9 insect cells were incubated with recombinant AMPK holoenzyme in a 20 μl reaction volume containing 50 mM Tris, 10 mM $MgCl_2$, 5 mM DTT and 200 μM ATP. For the radioactivity-based assay, 0.1mCi/ml γ-^{32}P labeled ATP was added to the reaction. The reactions were incubated at 30°C for 1 hr and stopped by addition of SDS sample loading buffer.

To identify AMPK-phosphorylated TET2 sites and their relative phosphorylation intensity, treated samples were processed by LC-MS/MS as described below in phosphorylation modification and analysis sections.

Alternatively, the samples were also run on SDS–PAGE gels and immunoblotted with TET2 pSer99 specific antibody.

For the classical radiometric assay using γ-^{32}P labeled ATP, 20μg of FLAG-TET2 protein was immobilized on FLAG beads and the kinase assay reaction was set up as described above. The reaction was terminated by washing the beads four times with buffer (50 mM Tris-HCl (pH 8.0), 300 mM NaCl). The washed beads were then subjected to measurement by Liquid Scintillation Analyzer. Similar assay reactions with AMPK kinase and FLAG beads were used as a blank for background subtraction. The calculation is based on the specific activity [γ-^{32}P]ATP (800 c.p.m. per pmol) used and the calculated molar amounts of TET2 added to the assay tube. As shown in the Ext Fig.4d, the stoichiometry is around 0.38 mol phosphates per mol of

TET2 WT, while it decreases dramatically after mutating Ser99 into Ala (TET2S99A) or disrupting AMPK recognizing consensus sequence (TET2SLF).

HPLC-MS-based quantitative measurement of 5mC/5hmC levels

HPLC-MS/MS analysis was performed using MassHunter System (Agilent). Briefly, extracted genomic DNA was hydrolyzed and dephosphorylated. 10 μg DNA was added to hydrolysis buffer (50 mM $ZnSO_4$, 1 M NaOAc pH 5.3) containing 0.5 unit of Nuclease P1 (Sigma Aldrich) and incubated at 37˚C for 16 hours. Dephosphorylation buffer (1.5 M Tris pH 8.3, 3 M NaCl, 0.5 M $MgCl_2$) and 0.5 unit of CIP (NEB) were added and incubated for 2 hours at 37˚C. The digested samples were then subjected to HPLC-MS/MS analysis via manufacturer's protocol. The mass spectrometer was optimized and set up in selected reaction monitoring (SRM) scan mode for monitoring the [M+H+] of 5hmC (258.1→142.1), deoxyguanosine(268.1→152) and deoxycytidine (228.1→112.1).

MeDIP-Seq, hMeDIP-Seq and validation

Genomic DNA of A2058-TET2WT under different treatments was purified, sonicated, and ligated with Illumina barcode adapters. Adaptor-ligated DNA (5 μg) was denatured and incubated with 3 μg of 5-hmC antibody or 10 μg of 5-mC antibody (Eurogentec) at 4°C overnight. Antibody-DNA complexes were captured by protein A/G beads. The immunoprecipitated DNA was purified and sequenced following the Illumina standard protocols. Read sequences were aligned to the human genome

(hg19) using bowtie (v1.0.0) with parameter,"-m 1", to allow unique mapping. The duplicated reads were removed from the mapped reads. The significantly enriched regions were identified by the MACS2 (v2.0.10). In order to compare the 5mC or 5hmC difference between normal and high glucose conditions, the peaks were merged in both normal glucoses and high glucoses using the merge command in bedtools (v2.17.0) package. The 5mC and 5hmC densities normalization was processed in two steps. First the densities were normalized based on their corresponding input reads, and then normalized into 20 million reads. After normalization, the normalized 5mC and 5hmC densities difference between the two conditions were identified on the merged peaks. Locus specific hMeDIP was performed using Active Motif's hMeDIP kit (#55010) following the manufacturer's instructions. Primer sequences for genes of interest are listed below. Data were calculated following the percent input method.

Primer	Sequence
CENPA-F	GTTTTGGAGACGGAGTCAGG
CENPA-R	AGACCATCCTGGCTAACACG
CHEK1-F	TGCCTATCCCTGAGTTCCAG
CHEK1-R	CGGGGAGAGCCAAAATAAAT
COBLL1-F	CCAGCCAACTTAGAATGTCCA
COBLL1-R	GCACTCTTTCCCCTGTACGA
CREB3L2-F	TTGCCCAGTCTTGGGTACTT
CREB3L2-R	ACAGGGAATCGCAATGAGAG
GABPB1-F	GAATGCAATGGCGCTATCTT
GABPB1-R	CCGATCACCTGAGGTCAAGA
GAPDH-F	ACACGCTCCCCTGACTTG
GAPDH-R	AGAAAGGTGGGAGCCTCAGT
MCM7-F	CTCTTGTACGGCATCAGCAA
MCM7-R	GCAAAGTGAATGTGGCTCTG
TAC1-F	AGCCTCAGCAGTTCTTTGGA
TAC1-R	GGAAGGAGTGAGCAGACAGG

Animal Study

Four-week-old athymic BALB/c nude mice (female) were purchased from Fudan University Animal Center (Shanghai, China) and maintained under pathogen-free conditions with approved experimental protocols. Experiments were performed in accordance with the institutional ethical guidelines and were approved by the institutional review board of Fudan University. None of the experiments exceeded the limits (2000 mm3) which were permitted by IACUC of Fudan University. Nude mice were randomly divided into 2 main groups: diabetic and non-diabetic as described in Extended Data Figure 9. Each group was further divided into four subgroups of n=6: A2058-TET2WT +/- Met and Mock +/- Met. First, the diabetic group received a single dose of streptozotocin (1% w/v solution in fresh cold sodium citrate buffer, pH 4.5) injection into the abdominal cavity, with the dosage ranging from 180-200 mg/kg of body weight. Glucose measurements were performed three days after streptozotocin treatment using blood collected from the tail vein. Eight days after streptozotocin injection, mice subjected to metformin treatment started taking the drug dissolved in drinking water (500 mg/L) daily until the end of study. All the mice received tumor cell injections seven days after the commencement of metformin treatment. A2058-TET2WT or Mock cell suspensions (0.2 ml, 5×10^6 cells in PBS) were injected subcutaneously. After injection, the mice were inspected daily for any bleeding or wounds. One week after injection, tumor sizes were measured every 2-3 days.

Immunohistochemical Staining

Immunohistochemical studies were performed on 5 μm sections of formalin-fixed, paraffin-embedded tissue. Slides were first deparaffinized, and rehydrated. Antigen retrieval was carried out in pressure cooker (119.5°C 30 sec, 90°C 10 sec) with 0.01M citrate buffer at pH 6.0 (DAKO Target Retrieval solution). Sample section slides were rinsed in dH_2O, following blocking for endogenous peroxidase activity with 1.5% H_2O_2/ETOH for 10 minutes, after which they were washed with PBS. Sections were blocked for 1 hour with normal goat serum followed by incubation with rabbit anti-5hmC (1:6000) and anti-pAMPK (1:100) overnight at 4°C. After three washes with TBS-0.05% Tween-20 buffer, sections were incubated with horseradish peroxidase conjugated goat-anti-rabbit-IgG for 1 hour. The sections were incubated with horseradish peroxide (Dako) and signals were visualized with NovaRED (pAMPK) and diaminobenzidine (5hmC), followed by a hematoxylin counterstain. Positive staining was defined as dark brown and red staining pattern. All of the immunostained sections were reviewed and scored independently by two pathologists in a blinded manner without knowledge of the clinicopathological information, based on the H-score method, which considers the staining intensity together with the percentage of cells staining positively. For H-score method, 10 fields were chosen randomly at ×200 magnification. The staining intensity in the cells was scored as 0, 1, 2 and 3 corresponding to the negative, weak, intermediate and strong staining, respectively. In each field the total number of cells and cells stained at each intensity was counted. The H-score was calculated using the following formula: (% of cells stained at

intensity category 1×1) + (% of cells stained at intensity category 2×2) + (% of cells stained at intensity category 3×3). H-scores varied from 0 to 300 where 300 represented 100% of cells strongly stained.

RNA Extraction and RT-qPCR

Cells were rinsed with ice cold PBS, following RNA isolation using TRIZOL according to the manufacturer's protocol. cDNA was synthesized using the SuperScript III First-Strand Kit (Invitrogen). Relative gene expression was determined by the Livak method using β-ACTIN as a housekeeping gene. Primers used in the RT-qPCR are listed below:

TET1 F	GCTATACACAGAGCTCACAG
TET1 R	GCCAAAAGAGAATGAAGCTCC
TET2 F	CTTTCCTCCCTGGAGAACAGCTC
TET2 R	TGCTGGGACTGCTGCATGACT
TET3 F	GTTCCTGGAGCATGTACTTC
TET3 R	CTTCCTCTTTGGGATTGTCC
ACTIN F	ACTCGTCATACTCCTGCT
ACTIN R	GAAACTACCTTCAACTCC
CHEK1 F	TTGGGCTATCAATGGAAGAAA
CHEK1 R	CCCTTAGAAAGCCGGAAGTC
CENPA F	AGCACACACCTCTTGATAAGGA
CENPA R	CACACCACGAGTGAATTTAACAC
MCM7 F	CAGGAGCAAGACCTCTGAGC
MCM7 R	ACTGCTTCTTCCCGAGTTCA
MKI67 F	CCTGTACGGCTAAAACATGGA
MKI67 R	TGACTTCCTTCCATTCTGAAGAC
KIF14 F	CAGAGTTGGCTAGAATTGGGAAA
KIF14 R	TCCATTAGTGTGAGCCGTAGA
NESTIN F	AGAGGGGAATTCCTGGAG
NESTIN R	CTGAGGACCAGGACTCTCTA
CD133 F	GAGCGCAAAGACTACCTGAAGA
CD133 R	CGACTCTAGCTCGATGCTCTTG
ALDH2 F	CGAGGTCTTCTGCAACCAG
ALDH2 R	TCCTTGTCCCCTTCAGCTAC

BRMS1 F	CTCAGGGAATCTCCGTCGT
BRMS1 R	AGAATCACCCTCTGCTTCCA
RASA1 F	GGCGTTCTTCTGCTATCGTT
RASA1 R	TGGACCAGAATACGAGGAGG
RIN1 F	GCTCCCTCAGCCTCTGG
RIN1 R	CCACTGCTGGGATCCTGATA
EIF5A F	GGGTCGAGTCAGTGCGTT
EIF5A R	CATTGGGAAGGTGGCTGA
CDK5 F	GGCTTCAGGTCCCTGTGTAG
CDK5 R	ATGGTGACCTCGATCCTGAG

Soft Agar Assay

Briefly, 6% agar was prepared in sterile water and boiled twice for 30 seconds to dissolve and sterilize the agar. 1 volume of 6% agar was then mixed with 9 volumes of warmed complete DMEM medium. A 3 ml mixture was placed into each well of a 6-well plate and stored at 4˚C for 1 hour or overnight to solidify. Before use, plates with 0.6% agarose were warmed in 37˚C. 3% agar with low melt agarose was prepared in sterile water and microwaved until dissolved. Cells were mixed with warmed medium to make a 5×10^3 cells/well mixture, and 1/9 volume of 3% melted agar was added and mixed immediately by inversion. The mixture was dispensed on top of the 0.6% base gel. The plates solidified in less than 30 minutes and were then kept in a cell incubator. Colonies were counted after 4 weeks.

Cell Proliferation Assay

For TF-1 cells, the cell suspensions were prepared to a final concentrations of 2×10^4 cells/ml in RPMI 1640 medium with indicated GM-CSF concentration (0, 0.001, 0.01, 0.1, 1 stand for relative the ratios of the standard GM-CSF concentration (2ng/ml) in

Extended Data Figure 8d). Viable cell number was determined by trypan blue exclusion assay every 24 hours. For A2058-TET2WT, A2058-TET2S99A and Mock cells, cell suspensions were prepared to a final concentrations of 2.5×10^4 cells/ml in DMEM medium +/- metformin (5 mM). 100 μl of the cell suspension was dispensed into a 96-well plate and incubated for 24-72 hours at 37˚C in a humidified, 5% CO_2 atmosphere. At each measurement time point, the medium was replaced with 20 μl MTS + 100 μl DMEM (120 μl total) mixture per well and incubated for 2 hours. The absorbance was recorded at 490 nm using a microplate reader (BioTek).

Liquid Chromatography-Mass Spectrometry for Phosphorylation Modification

After in-gel digestion of proteins, LC-MS/MS was used to analyze the extracted peptides for protein identification. The experiments were performed on a Nano Aquity UPLC system (Waters Corporation, Milford, MA) connected to a quadrupole-Orbitrap mass spectrometer (Q-Exactive, Thermo Fisher Scientific, Bremen, Germany) equipped with an online nano-electrospray ion source (HaoChuang Biotech, Zhejiang, China).The peptide samples were resuspended in 12 μl of solvent A (A: water with 0.1% formic acid; B: ACN with 0.1% formic acid), and then separated on the analytical column (Acclaim PepMap C18, 75μm × 50cm) with a linear gradient, from 2% B to 45% B in 75 minutes. The column flow rate was maintained at 300 nl/min and the column temperature was maintained at 40˚C.

The Q-Exactive mass spectrometer was operated in the data-dependent mode to switch automatically between MS and MS/MS acquisition. Survey full-scan MS spectra (m/z 350-1200) were acquired with a mass resolution of 70K, followed by twenty sequential high energy collisional dissociation (HCD) MS/MS scans with a resolution of 17.5K. Single charge state was rejected and dynamic exclusion was used with one microscan and 30 seconds exclusion duration.

Database Searching and Phosphopeptide Site Analysis

Thermo Scientific Proteome Discoverer software version 1.4 with the MASCOT v2.3.2 search engine was used for all searches of the Human Swiss-Prot database (Release 2014-04-10, with 20264 sequences). The MS/MS spectra was searched with a precursor-ion mass tolerance of 10 ppm and fragment-ion mass tolerance of 0.02 Da; full tryptic specificity was applied; 2 missed cleavages were allowed; dynamic modifications were set for phosphorylated Ser, Thr and Tyr (+80); peptides with at least 7 amino acids were retained. The target-decoy based strategy was applied to control peptide level false discovery rates (FDR) lower than 1%.The localization of phosphorylation in peptide sequences was calculated by the phosphoRS 3.0 algorithm, and probabilities above 99% were accepted as truly phosphorylated sites.

Relative Quantification of Phosphorylation Sites

For relative quantification of phosphorylated peptide signal levels, an isotope-free (label-free) method was used by first integrating the total ion counts (TIC) for each

MS/MS sequencing event during a data-dependant acquisition. For each targeted phosphorylation site, a ratio of phosphorylated peptide signal (TIC of phosphorylated form) to the total peptide signal (TIC of non-phosphorylated plus phosphorylated forms) for both *in vitro* kinase assay and high/normal glucose treated samples were calculated according to the following equation:

$$RRRRRRRRRR = \frac{TTTTTT_{PPPP_4}}{TTTTTT_{PPPP_4} + TTTTTT_{nnRRnnPPPP_4}}$$

Microarray processing and clustering

Microarray experimental design and processing were performed in compliance with the MIAME standard. Gene expression profile was determined using Affymetrix GeneChip Human Genome U133 Plus 2.0 Array. Raw microarray data were analyzed using Affymetrix Expression Console Software and customized R scripts for quality assessment and quality control, background correction, and normalization across arrays. Standard Affymetrix quality metrics were also assessed, such as 3'/5' ratios, background, scaling factor, control probes, GAPDH, and percent present (PP) calls. We did average calculation of probe signal level for probes in the same Ensembl annotated gene. To find genes that were regulated by glucose through TET2, genes that were differentially expressed (DE) in normal-g and high-g in A2058-TET2WT cells were first selected into a list using dChip software. Genes that did not have DE in normal-g and high-g in both A2058-TET2M and Mock cells were then excluded from the list. Unsupervised hierarchical clustering was done by dChip on both

samples and genes, using Euclidean Distance and Average Linkage, P-value < 0.001. Disease functional ontology pathway analyses were performed by FunDO.

Quantification of Western Blots and Dot Blots

Western blots and dot blots were quantified using ImageJ and Bio-Rad Image Lab software. Quantification was normalized to appropriate internal references, e.g. β-Actin/β-Tubulin/Lamin A/C (for Western blots), control DNA 5hmC (for dot blot), or FLAG-TET2 (for IP). For CHX half-life experiments, the maximum was scaled to 1 by dividing all normalized time points by the normalized control.

Statistical Analysis

Samples were compared using two-tailed, unpaired Student's t-test, unless otherwise stated. Error bars were represented by SEM +/- (for animals) or SD +/- (for others) as indicated.

Data Availability

The microarray and hMeDIP-seq datasets generated and analyzed during the current study are available in the Gene Expression Omnibus (GEO) with accession number GSE86376.

Code Availability

Custom R codes can be available upon request.

References

32 Yang, H. *et al.* TET-catalyzed 5-methylcytosine hydroxylation is dynamically regulated by metabolites. *Cell research* **24**, 1017-1020, doi:10.1038/cr.2014.81 (2014).

33 Lian, C. G. *et al.* Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell* **150**, 1135-1146, doi:10.1016/j.cell.2012.07.033 (2012).

34 Nakatani, Y. & Ogryzko, V. Immunoaffinity purification of mammalian protein complexes. *Methods Enzymol* **370**, 430-444, doi:10.1016/S0076-6879(03)70037-8 (2003).

35 Dowling, R. J., Niraula, S., Stambolic, V. & Goodwin, P. J. Metformin in cancer: translational challenges. *Journal of molecular endocrinology* **48**, R31-43, doi:10.1530/JME-12-0007 (2012).

36 Barbieri, F. *et al.* In vitro and in vivo antiproliferative activity of metformin on stem-like cells isolated from spontaneous canine mammary carcinomas: translational implications for human tumors. *BMC cancer* **15**, 228, doi:10.1186/s12885-015-1235-8 (2015).