Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the glycosyltransferase OGT [$ #]

Christina Bauer[1], Klaus Göbel[1], Nagarjuna Nagaraj[2], Christian Colantuoni[1], Mengxi Wang[1], Udo Müller[1], Elisabeth Kremmer[3], Andrea Rottach[1]*, Heinrich Leonhardt[1,4]*

[1] Ludwig-Maximilians University Munich, Biocenter, Planegg-Martinsried, Germany
[2] Max Planck Institute for Biochemistry, Martinsried, Germany
[3] Helmholtz Center Munich, Institute for Molecular Immunology, München-Großhadern, Germany
[4] Center for Integrated Protein Science Munich (CIPSM)

Running title: Phosphorylation and O-GlcNAcylation of TET proteins

To whom correspondence should be addressed: Prof. Dr. Heinrich Leonhardt, Ludwig-Maximilians University Munich, Department Biology II, Großhadernerstr. 2, 81925 Planegg-Martinsried, Germany
Tel: +49 (0) 89 / 2180 - 74229, Fax: +49 (0) 89 / 2180 - 74236, h.leonhardt@lmu.de

Keywords: epigenetics, 5-hydroxymethylcytosine (5-hmC), post-translational modification (PTM), O-linked N-acetylglucosamine (O-GlcNAc), phosphorylation, dioxygenase, TET proteins, OGT

CAPSULE

Background: TET proteins oxidize 5-methyl-cytosine and contribute to active DNA demethylation.
Results: OGT modifies TET proteins with N-acetylglucosamine and thereby reduces TET phosphorylation.
Conclusion: TET proteins are subjected to a dynamic interplay of post-translational modifications (PTMs) at low complexity regions.
Significance: This first map of TET phosphorylation and O-GlcNAcylation sites at amino acid resolution provides a valuable resource for future studies of TET regulation.

ABSTRACT

TET proteins oxidize 5-methylcytosine (mC) to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC) and 5-carboxylcytosine (caC) and thus provide a possible means for active DNA demethylation in mammals. Although their catalytic mechanism is well characterized and the catalytic dioxygenase domain is highly conserved, the function of the regulatory regions — the N-terminus and the low complexity insert between the two parts of the dioxygenase domains — is only poorly understood. Here, we demonstrate that TET proteins are subject to a variety of PTMs that mostly occur at these regulatory regions. We mapped TET modification sites at amino acid resolution and show for the first time that TET1, TET2, and TET3 are highly phosphorylated. The glycosyltransferase OGT, which we identified as a strong interactor of all three TET proteins, catalyzes the addition of an N-acetylglucosamine (GlcNAc) group to serine and threonine residues of TET proteins and thereby decreases both the number of phosphorylation sites as well as the site occupancy. Interestingly, the different TET proteins display unique PTM patterns and some modifications occur in distinct combinations. In summary, our results provide a novel potential mechanism for TET protein regulation based on a dynamic interplay of phosphorylation and O-GlcNAcylation at the N-terminus and the low complexity insert region. Our data suggest strong crosstalk between the modification sites that could allow rapid adaption of TET protein localization, activity, or targeting due to changing environmental conditions as well as in response to external stimuli.

INTRODUCTION

A major epigenetic mechanism of gene regulation in higher eukaryotes is methylation of DNA at the carbon 5 atom of cytosines (1,2). Recently, the family of TET (ten-eleven-translocation)[5] proteins has been shown to successively oxidize mC to hmC, fC and caC (3-6), providing novel insights

into the dynamics of DNA modifications. TET proteins are also active on genomic thymine residues, leading to the generation of 5-hydroxyuracil (hmU) (7). In gnathostomata, there are three TET proteins: TET1, TET2, and TET3 (8), that show distinct expression patterns and functions in different tissues or during development (9-13). TET1 and TET2 are highly expressed in mouse embryonic stem cells (mESCs) and are associated with oxidation of transcription start sites or gene bodies, respectively (14). TET3 is upregulated in the oocyte and oxidizes the silenced paternal pronuclear DNA (10,15). High levels of TET proteins and genomic hmC are described for neuronal tissues (11,16-18). In several patients with myeloid malignancies, mutations of *TET2* correlate with decreased hmC levels and altered gene expression patterns (19-22).

The activity of TET proteins directly depends on two co-factors: Fe(II) and 2-oxoglutarate (2-OG) (3,8). Interestingly, gain-of-function mutations of the enzymes responsible for 2-OG synthesis, IDH1 and IDH2, have been associated with tumorigenesis, in particular glioblastomata and acute myeloid leukemia (AML) (20,23,24). These mutations lead to the synthesis of 2-hydroxyglutarate (2-HG), a potent inhibitor of 2-OG dependent dioxygenases such as TET proteins (24,25). Since IDH1 and IDH2 are enzymes of the Krebs cycle, these findings represent a direct link of TET protein activity to metabolism, especially since low hmC levels are not only found in AML patients with *TET2* loss-of-function mutations, but also with *IDH2* gain-of-function mutations (20). Besides 2-HG, ascorbate has also been shown to influence cytosine oxidation by TET proteins (26-28). In summary, TET protein activity appears to be modulated by several small molecules, either inhibitory such as 2-HG or stimulating like ascorbate.

TET proteins are not only influenced by certain metabolites, but also by interacting proteins. TET1 forms complexes with heterochromatin-associated proteins like HDAC1, HDAC2, SIN3A or EZH2 (29). All three TET proteins interact with a variety of factors of the base-excision repair pathway, including PARP1, LIG3, or XRCC1, and also with several DNA glycosylases such as TDG, NEIL1, or MDB4 (30). Another known interactor of TET proteins is the glycosyltransferase OGT (31-36) that represents an additional interesting connection with metabolism. OGT catalyzes the addition of a GlcNAc group to serine or threonine residues of target proteins (37). Its activity is dependent on the availability of a variety of metabolic molecules like glucose, ATP, glutamine and acetyl-CoA (38). The association of OGT with TET proteins has been reported to influence histone modifications and gene expression (31,36), TET1 protein stability (33) and activity (34) and, for TET3, also on subcellular localization (35).

Taken together, TET protein activity is widely studied in the context of development, tumorigenesis and metabolic conditions. However, only very little is known about the structure and function of the non-catalytic domains of TET proteins. In this study, we show that TET proteins are subject to a large number of PTMs, predominantly occurring at the two low complexity regions, which display only little sequence conservation: the N-terminus and the insert region that separates the two parts of the catalytic dioxygenase domain and is predicted to be unstructured (8). We describe that TET proteins are phosphorylated and that this phosphorylation can be suppressed via O-GlcNAcylation by the glycosyltransferase OGT. Detailed mapping of modification sites to the protein sequence shows that mostly N-terminus and insert region of TET proteins are subjected to PTMs and that their regulation depends on a dynamic interplay of different PTMs.

EXPERIMENTAL PROCEDURES

Antibody generation

A His-tagged protein fragment from the insert region of each TET protein (Figure 1a) was expressed in *E. coli* BL21 (DE3) cells (Novagen, Darmstadt, Germany) and purified with the TALON™ Superflow Metal Affinity Resin system (Clontech, Saint Germain, France) under native conditions as described previously (39). For TET1, amino acids 1682-1914, for TET2 amino acids 1332-1779, and for TET3 amino acids 976-1521, were used as antigens. Approximately 100 µg of each antigen were injected both intraperitoneally (i.p.) and subcutaneously (s.c.) into Lou/C rats using CPG2006 (TIB MOLBIOL, Berlin, Germany) as adjuvant. After eight weeks, immune response was boosted i.p. and s.c. three days

before fusion. Fusion of the myeloma cell line P3X63-Ag8.653 with the rat immune spleen cells was performed using polyethylene glycol 1500 (PEG 1500, Roche, Mannheim, Germany). After fusion, the cells were cultured in 96 well plates using RPMI1640 with 20 % fetal calf serum, Penicillin/streptomycin, pyruvate, nonessential amino acids (PAA, Linz, Austria) supplemented by aminopterin (Sigma, St. Louis, MO). Hybridoma supernatants were tested in a solid-phase immunoassay. Microtiter plates were coated over night with His-tagged TET antigens at a concentration of 3-5 µg/ml in 0.1 M sodium carbonate buffer, pH = 9.6. After blocking with non-fat milk (Frema, Neuform, Zarrentin, Germany), hybridoma supernatants were added. Bound rat mAbs were detected with a cocktail of biotinylated mouse mAbs against the rat IgG heavy chains, avoiding IgM mAbs (α-IgG1, α-IgG2a, α-IgG2b (ATCC, Manassas, VA), α-IgG2c (Ascenion, Munich, Germany)). The biotinylated mAbs were visualized with peroxidase-labelled avidin (Alexis, San Diego, CA) and o-phenylenediamine as chromogen in the peroxidase reaction. Anti-TET1 5D6 (rat IgG2a), anti-TET1 5D8 (rat IgG2a), anti-TET1 2H9 (rat IgG2a), anti-TET1 4H7 (rat IgG2a), anti-TET2 9F7 (rat IgG2a), anti-TET3 11B6 (rat IgG2a) and anti-TET3 23B9 (rat IgG2a) were stably subcloned and further characterized (Figure 1b).

ESC culture, co-IP and MS/MS analysis

mESCs (J1) were cultured as described previously (9). Endogenous TET1 and TET2 proteins were pulled out via monoclonal antibodies (5D6, 5D8, and 9F7) coupled to protein G sepharose beads as described in (39). After co-immunoprecipitation (co-IP), protein samples were digested on beads with trypsin according to standard protocols. Peptide mixtures were analyzed using electrospray tandem mass spectrometry. Experiments were performed with an LTQ Orbitrap mass spectrometer (Thermo Scientific, Waltham, MA). Spectra were analyzed with the Mascot™ Software (Matrix Science, Boston, MA).

Expression constructs

Expression constructs for GFP-TET1, GFP-TET2, GFP-TET3, GFP and Cherry (Ch) were described previously (40-42). To generate the Ch-OGT construct, the coding sequence was amplified using cDNA from mouse E14 ESCs as template and subcloned into the pCAG-Cherry-IB vector. Expression constructs for Ch-OGT^{H508A} (subsequently referred to as OGT^{mut}) were generated by overlap extension PCR. All constructs were verified by DNA sequencing (Eurofins, Ebersberg, Germany).

HEK293T culture, co-IP and Western blot analysis

Co-IP followed by Western blot with GFP- and Ch-tagged proteins expressed in HEK293T cells was performed as described previously (30). O-GlcNAc was detected with a mouse monoclonal antibody (RL2, abcam, Cambridge, UK) and an Alexa647N-conjugated secondary antibody (Sigma, St. Louis, MO).

Sample preparation for mass spectrometric analysis

All experiments were performed in biological triplicates. GFP-tagged TET proteins and/or Ch-tagged OGT or OGT^{mut} were expressed in HEK293T cells. Cell lysis with RIPA buffer and IP with the GFP-Trap® (Chromotek, Martinsried, Germany) was performed as described previously (30). After IP, samples on beads were rinsed two times with wash buffer (20 mM TrisHCl, pH = 7.5, 300 mM NaCl, 0.5 mM EDTA) and two times with IP buffer (20 mM TrisHCl, pH = 7.5, 150 mM NaCl, 0.5 mM EDTA).

100 µl of denaturation buffer (6 M GdnHCl, 10 mM Tris(2-carboxyethyl)phosphine and 40 mM chloroacetamide in 100 M Tris pH = 8.5) was added to the beads and heated at 70 °C for 5 min. The samples were then subjected to sonication in Diagenode bioruptor plus (UCD-300-TO) at maximum power settings, for 10 cycles consisting of 30 s pulse and 30 s pause. Following sonication, the samples were diluted 1:10 with digestion buffer (25 mM Tris pH = 8.5 containing 10 % acetonitrile) and mixed by vortexing prior to enzyme digestion. Each sample was digested with 1 µg of endoproteinase Lys-C (Wako Chemicals, Neuss, Germany) for 4 h with subsequent digestion using 1 µg of trypsin (Promega, Madison, WI) under gentle rotation at 37 °C. After digestion, the samples were placed in a speed-vac for 10 minutes to remove acetonitrile from the sample before StageTip purification using SDB-XC material (43). Peptides were then eluted from

the StageTip, placed in speed-vac to reduce the sample volume to approximately 6 µl and 5 µl of the sample was injected on the column for MS/MS analysis.

Liquid chromatography tandem mass spectrometry and data analysis

Samples were loaded on a column (15 cm length and 75 µm inner diameter (New Objective, Woburn, MA)) packed with 3 µm Reprosil C18 beads (Dr Maisch GmbH, Ammerbuch-Entringen, Germany) using the auto sampler of Thermo Easy n-LC (Thermo Scientific, Waltham, MA) coupled via a nano-electrospray source to a LTQ Orbitrap XL mass spectrometer. Each sample was analyzed using a 2 h reversed phase gradient and using a top 5 method for data dependent acquisition. Full scans were acquired in the Orbitrap after accumulating up to $1x10^6$ charges and the MS/MS of the five most abundant precursors were performed using low energy ion trap CID. MS/MS spectra were recorded using the ion trap by radial ejection.

All raw files were analyzed using the MaxQuant (44) computational proteomics platform (version 1.4.1.6). Peak lists were searched with an initial mass deviation of 7 ppm and fragment ion deviation of 0.5 Th. Carbamidomethylation was used as fixed modification and oxidation of methionine, phosphorylation of serine, threonine and tyrosine, O-linked GlcNAc of serine and threonine, ubiquitination (diglycine motif) of lysine and acetylation of the protein N-terminus were used as variable modification. All unmodified and oxidized methionine and N-acetylation containing peptides were used for protein quantification. The MaxQuant software quantifies the different versions of modified peptides in a label-free fashion. Briefly, the occupancy reflects the extracted signal differences between modified and unmodified peptide and also includes the protein ratios between samples. The different forms of modified peptides, e.g. peptides with single, double and triple O-GlcNAc sites, are individually quantified and listed separately in the output (Supplemental table S4). Details on label-free quantification of modification sites are provided elsewhere (45).

MaxQuant output data were further analyzed with the Perseus software 1.5.0.15 (44). Only modifications that were detected in at least two out of three biological replicates in at least one experimental set-up were included in the analysis. PTMs that were detected in non-unique peptides were also excluded. Significance was tested using a two-tailed paired student's t-test.

RESULTS

Characterization of TET antibodies

The three TET proteins share a common domain architecture: the C-terminal catalytic dioxygenase domain is split into two parts separated by a low complexity insert region and is preceded by an extension enriched in cysteines (8). All three TET proteins have a large N-terminal part that is mostly uncharacterized so far, except for a CXXC-type zinc finger at the N-terminus of TET1 and TET3 (8,40,47). Murine TET3 exists in two isoforms: one with the zinc finger and one without (41). The cysteine-rich region and the split-dioxygenase domain are conserved among the three murine TET proteins whereas N-terminus and insert region display only little sequence similarity (Figure 1a). The three-dimensional structure of mammalian TET proteins remains unresolved with the exception of the cysteine-rich and dioxygenase domains of TET2 (48), leaving structure and function of the N-terminus and the low complexity insert unknown.

We generated antibodies against murine TET1, TET2, and TET3 using protein fragments derived from the insert region of the catalytic domains as antigens. The rat monoclonal antibodies were tested for their applicability in Western blot, IP, and immunofluorescence (IF) (Figure 1b). Figure 1c-e) shows exemplary data from the antibody characterization process of selected clones. For antibody testing, GFP-tagged TET proteins were expressed in HEK293T cells and detected in the cell lysate by Western blot using TET antibodies and a GFP antibody as positive control (Figure 1c and data not shown). For IP, antibodies were coupled to protein G beads, incubated with the cell lysates, and analyzed by anti-GFP Western blot for efficient pull-down of the respective TET protein (Figure 1d and data not shown). mESCs were used to test the suitability of the obtained antibodies for IF. Antibodies preselected for specificity in Western blot analyses that showed a clear nuclear staining were judged as applicable in IF (Figure 1e and data not shown).

TET proteins interact with and are O-GlcNAcylated by OGT

As a first step towards understanding the regulation of TET proteins, we screened for interaction partners in mESCs. Since TET1 and TET2 are contitutively expressed in mESCs, clones 5D6-anti-TET1, 5D8-anti-TET1 and 9F7-anti-TET2 were used to pull down endogenous TET1 and TET2. Subsequent LC-MS/MS analysis revealed that both TET1 and TET2 interact with the glycosyltransferase OGT. In accordance with this result, co-IP analysis of GFP-TET1, TET2, and TET3 expressed in HEK293T cells highly enriched OGT in the pull-down (Figure 2a).

Having observed the interaction between TET proteins and OGT, we examined whether TET proteins are modified by OGT and screened for O-GlcNAcylation, the modification that is transferred to the OH-group of serine or threonine residues of target proteins by OGT (38,49). To this end, we specifically enriched GFP-tagged TET proteins co-expressed with either OGT or its catalytically inactive point mutant OGT^{mut} with the GFP-Trap® and probed the subsequent Western blot with an anti-GlcNAc antibody. All three TET proteins were found to be increasingly O-GlcNAcylated dependent on the coexpression of catalytically active OGT (Figure 2b).

O-GlcNAcylation reduces phosphorylation of TET proteins

To identify OGT-dependent O-GlcNAcylation sites on TET proteins, we performed mass spectrometric analysis of semi-purified proteins. We therefore expressed GFP-tagged TET1, TET2, and TET3 in HEK293T cells, either with OGT, OGT^{mut}, or without interactor. After pull-down with the GFP-Trap® and stringent washing steps, the samples were analyzed by LC-MS/MS. An overall sequence coverage of about 50 % was achieved for TET1, about 60 % for TET2 and about 65 % for TET3 (Supplemental data S1, Supplemental table S4). For data analysis, only sites were considered that were detected in at least two out of three biological replicates. Figure 3a shows an exemplary MS/MS spectrum of a O-GlcNAcylated TET1 peptide. Without co-expression of interactor, only few residues on TET proteins are found to be O-GlcNAcylated at low site occupancy. Coexpression of OGT leads to a strong increase in both number of O-GlcNAcylation sites and site occupancy for TET2 and TET3. The difference in number of O-GlcNAc sites is either due to *de novo* modification by OGT or because the site occupancy without OGT co-expression is below detection limit. For TET1, however, the O-GlcNAc pattern is relatively heterogeneous and only few O-GlcNAc sites can be detected. This heterogeneity is also illustrated by the fact that residues 1327 and 327, which are O-GlcNAcylated in the TET1 samples, are only detected to be modified in one out of three replicates in the TET1/OGT samples. Although Ch-OGT^{mut} is supposed to be catalytically inactive, co-expression leads to a small increase in O-GlcNAcylation and represents a distinct state from basal levels (Figure 4 and Tables 1-3).

Since O-GlcNAcylation occurs at serine or threonine residues of the target protein, we also screened for another post-translational modification that can occur at these amino acids, namely phosphorylation. Interestingly, high phosphorylation of TET1 and TET2 and, to a lesser extent, of TET3, was observed. Phosphorylation of all TET proteins decreased significantly upon co-expression of OGT, regarding both site occupancy and number of detected phosphorylation sites (Figure 4 and Tables 1-3). An example of the MS/MS spectrum of a phosphorylated TET1 peptide is shown in Figure 3b. The MS/MS spectra of all modified TET peptides are provided as Supplemental data S2 and S3.

PTMs occur mostly at the N-terminus and the low complexity insert of TETs

To date, the domains of TET proteins are largely uncharacterized except for the conserved catalytic dioxygenase domain and the CXXC-type zinc finger at the N-terminus of TET1 (8,40,48). Mapping of the detected O-GlcNAc and phosphorylation sites to the TET protein sequence reveals that mostly the N-terminus and the low complexity insert, which separates the two parts of the dioxygenase domain, are subjected to PTMs (Figure 5). Remarkably, O-GlcNAcylation and phosphorylation rarely occur at the exact same residue, although O-GlcNAcylation suppresses phosphorylation. Furthermore, the three TET proteins carry different modification patterns: Whereas TET1 is mostly modified at the N-terminus and the very C-terminal part and is

hardly glycosylated, TET2 and TET3 show strong O-GlcNAcylation at the low complexity insert region. The first 350 amino acids of TET3 remain free of PTMs. The observed pattern is not due to differences in sequence coverage as the detected peptides are homogenously distributed over the whole protein sequence (Supplemental data S1).
Interestingly, some of the modifications are detected on the same peptides, indicating that they occur together at the same molecule. For example, TET2-S23 phosphorylation can be found with S15 phosphorylation and phosphorylation of S376 only occurs when S374 is O-GlcNAcylated, but not when it is phosphorylated (Table 2). For TET3, a variety of PTM combinations can be observed for residues 360-368 and 1071-1077. Phosphorylation at S362, for example, exists either alone or in combination with S360 O-GlcNAcylation and S368 phosphorylation. Phospho-S362 also co-occurs with O-GlcNAc-S361. If S362 is O-GlcNAcylated, however, no further modifications on this peptide was observed (Table 3, Figure 4). Apparently, some residues such as TET3-S362 serve as O-GlcNAc/phosphorylation switches that can either promote or suppress neighboring PTMs. These data indicate a strong crosstalk between O-GlcNAcylation and phosphorylation at different residues. Modifications on TET1, on the other hand, appear more isolated and no peptide bearing more than one modification was detected (Table 1). In summary, we detect many interdependent modification sites on TET proteins suggesting that TET1, TET2, and TET3 are dynamically regulated by PTMs.

DISCUSSION

Since oxidation of mC to hmC, fC and caC by TET proteins represents a potential mechanism for active DNA demethylation in higher vertebrates (3-5), these proteins are intensively investigated. Here, we provide evidence that all three TET proteins are subject to O-GlcNAcylation through OGT. This finding is in accordance with previous studies, showing that TET1 and TET2 interact with OGT in ESCs and are O-GlcNAcylated (33,34). TET3 has also been described to associate with OGT (32,35) and to alter its subcellular localization dependent on glucose metabolism and O-GlcNAcylation (35). OGT not only directly modifies TET proteins, but the interaction also promotes histone modifications such as H3K4me3 or H2BS112GlcNAc (31,36). TET1 has been shown to associate with the repressive SIN3A complex (50) and TET2 and TET3 with the SET1/COMPASS complex (31).
We show that per default, TET proteins are phosphorylated. Basal O-GlcNAc levels are low, but increase upon OGT expression. Simultaneously, the phosphorylation levels decrease. This finding indentifies regulation of the phosphorylation signal as a novel function for TET O-GlcNAcylation. Interestingly, the underlying mechanism of this observation seems not to be direct competition for the serine or threonine residue that is to be modified, but rather proximal site competition as neighboring residues are interdependent (51). O-GlcNAcylation and phosphorylation of TET proteins occur at distinct amino acids and often, several modifications of the same type appear in close proximity in "modification islands", e.g. O-GlcNAcylation at S1252/S1256/ S1263 of TET3, or phosphorylation at S15/S23/S39 of TET2. It is important to notice that only few and more isolated O-GlcNAcylation sites are detected on TET1 in contrast to TET2 and TET3 and that glycosylation of TET1 is less conserved within biological replicates. We also do not observe O-GlcNAcylation of TET1 at T535 which has been described as a major TET1 glycosylation site before (33,52). Taken together, O-GlcNAcylation of TET1 seems to be very dynamic. This hypothesis is also supported by the fact that Myers *et al* detect TET1-T535 O-GlcNAcylation in only one out of three replicates (52), similar to our observation of heterogeneous TET1 glycosylation patterns.

To distinguish between mere interaction of TET proteins with OGT and catalytic activity of OGT on TET proteins, we used a catalytically inactive point mutant of OGT as a control. Interestingly, O-GlcNAcylation of TET proteins is slightly increased by OGT^{mut}. This might be due to residual activity of the mutant (53) or, more likely, due to recruitment of endogenous active OGT via trimerization of the TPR domain (54). Nevertheless, this supposed heterotrimer seems to target the same residues as 91 % of all detected O-GlcNAc sites in the OGT^{mut} samples are also modified in the OGT samples. Regarding phosphorylation, co-expression of OGT^{mut} also represents an intermediate state between the basal

state, i.e. only TET expression, and co-expression of active OGT. 74 % of all detected phosphorylation sites in the basal state also appear during co-expression of OGT^{mut}, arguing against steric hindrance of the kinase by the inactive enzyme as a mechanism for reduced phosphorylation.

In this study, we have investigated TET protein PTMs dependent on OGT levels. The observed effect that O-GlcNAcylation of TET proteins reduces phosphorylation is of particular interest since protein O-GlcNAc levels are influenced by a variety of factors, such as different subcellular localization of OGT and nutrient availability, and seem to be tightly regulated (38). For example, O-GlcNAcylation of TET3 can be enhanced when cells are cultured in high glucose medium, leading to nuclear export of TET3 (35). Furthermore, not only OGT activity but also OGT expression levels are tightly controlled in living cells. During chondrocyte differentiation, for example, OGT is upregulated upon insulin stimulance (55). Moreover, the *Ogt* gene is located on the X chromosome and is subjected to dosage compensation through X chromosome inactivation (56). Taken together, OGT-dependent dephosphorylation represents a novel mechanism on how TET proteins could be regulated in response to changing environmental conditions or during differentiation.

Interestingly, some residues remain stably phosphorylated even at high OGT levels. For TET2, they appear in close proximity to each other and just N-terminal of the cysteine-rich region. This persistence of phosphorylation suggests an important OGT-independent regulatory role of these residues that is of interest for future studies. Nevertheless, the majority of phosphorylation sites are reduced in occupancy upon O-GlcNAcylation. We thus observe two different types of phosphorylation, dependent and independent of O-GlcNAcylation.
The hypothesis of interdependence of PTMs on TET proteins is further strengthened by the fact that some modifications are detected on the same peptides in stable combinations whereas others occur as stand-alone modifications. Certain residues appear to be O-GlcNAc/phosphorylation switches that influence the PTM pattern on the neighboring amino acids. The observed crosstalk of modifications enables a variety of potential regulatory mechanisms that could fine-tune TET activity dependent on different environmental conditions such as nutrient availability.

To date, the domain architecture and three-dimensional structure of TET proteins is only poorly understood. The catalytic domain is highly conserved and homologous to other types of Fe(II)- and 2-OG-dependent dioxygenases that act on nucleic acids (8,57). Recently, the crystal structure of the catalytic part of TET2 has provided insights into the reaction mechanism (48). However, the large N-terminus and the low complexity insert, that is characteristic for TET proteins, remain poorly understood — both in terms of structure and of function. So far, no homologous domains have been described except for the CXXC-type zinc finger at the N-terminus, and the insert region is predicted to be largely unstructured (8). In this study, we show that these two regions are subject to many dynamic PTMs. For TET1 and TET3, few modification sites are also found at the very C-terminus of the proteins, but N-terminus and insert region are the major target of O-GlcNAcylation and phosphorylation. In general, the lower the conservation of one region, the more modification sites are detected. The selective modification of these regions might contribute to the regulation of TET protein activity, stability, or targeting. In line, TET1, TET2, and TET3 have been described to colocalize with OGT at transcription start sites and influence gene expression (31,34). The different modifications described in this study might alter binding of TET interaction partners and thus provide a possible explanation for the observed dual role in transcriptional activation and repression (58).

In summary, we provide the first systematic mapping of O-GlcNAcylation and phosphorylation sites on TET proteins at amino acid resolution. The distribution of these PTMs and the described crosstalk open new perspectives on the regulatory role of the so far poorly characterized non-catalytic domains, the N-terminus and the low complexity insert region. The observed O-GlcNAcylation and phosphorylation are linked to metabolic conditions and thus provide a possible

mechanism of TET protein regulation in response to external stimuli.

REFERENCES

1. Goll, M. G., and Bestor, T. H. (2005) Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* **74**, 481-514
2. Suzuki, M. M., and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**, 465-476
3. Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935
4. Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., and Zhang, Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-1303
5. He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C. X., Zhang, K., He, C., and Xu, G. L. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-1307
6. Pfaffeneder, T., Hackner, B., Truss, M., Munzel, M., Muller, M., Deiml, C. A., Hagemeier, C., and Carell, T. (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl* **50**, 7008-7012
7. Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S. K., Eisen, D., Truss, M., Steinbacher, J., Hackner, B., Kotljarova, O., Schuermann, D., Michalakis, S., Kosmatchev, O., Schiesser, S., Steigenberger, B., Raddaoui, N., Kashiwazaki, G., Muller, U., Spruijt, C. G., Vermeulen, M., Leonhardt, H., Schar, P., Muller, M., and Carell, T. (2014) Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat Chem Biol* **10**, 574-581
8. Iyer, L. M., Tahiliani, M., Rao, A., and Aravind, L. (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8**, 1698-1710
9. Szwagierczak, A., Bultmann, S., Schmidt, C. S., Spada, F., and Leonhardt, H. (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* **38**, e181. 110.1093/nar/gkq1684
10. Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F., Liu, W., Xie, Z. G., Shi, L., He, X., Jin, S. G., Iqbal, K., Shi, Y. G., Deng, Z., Szabo, P. E., Pfeifer, G. P., Li, J., and Xu, G. L. (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606-610
11. Kim, M., Park, Y. K., Kang, T. W., Lee, S. H., Rhee, Y. H., Park, J. L., Kim, H. J., Lee, D., Lee, D., Kim, S. Y., and Kim, Y. S. (2014) Dynamic changes in DNA methylation and hydroxymethylation when hES cells undergo differentiation toward a neuronal lineage. *Hum Mol Genet* **23**, 657-667
12. Langlois, T., da Costa Reis Monte-Mor, B., Lenglet, G., Droin, N., Marty, C., Le Couedic, J. P., Almire, C., Auger, N., Mercher, T., Delhommeau, F., Christensen, J., Helin, K., Debili, N., Fuks, F., Bernard, O. A., Solary, E., Vainchenker, W., and Plo, I. (2014) TET2 Deficiency Inhibits Mesoderm and Hematopoietic Differentiation in Human Embryonic Stem Cells. *Stem Cells* **32**, 2084-2097
13. Rudenko, A., Dawlaty, M. M., Seo, J., Cheng, A. W., Meng, J., Le, T., Faull, K. F., Jaenisch, R., and Tsai, L. H. (2013) Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* **79**, 1109-1122
14. Huang, Y., Chavez, L., Chang, X., Wang, X., Pastor, W. A., Kang, J., Zepeda-Martinez, J. A., Pape, U. J., Jacobsen, S. E., Peters, B., and Rao, A. (2014) Distinct roles of the methylcytosine

oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* **111**, 1361-1366

15. Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C. J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun* **2**, 241. 210.1038/ncomms1240
16. Hahn, M. A., Qiu, R., Wu, X., Li, A. X., Zhang, H., Wang, J., Jui, J., Jin, S. G., Jiang, Y., Pfeifer, G. P., and Lu, Q. (2013) Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell Rep* **3**, 291-300
17. Kriaucionis, S., and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929-930
18. Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J. G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., Diao, J., Wu, F., He, H. H., Cui, Q., Clark, E., Ma, C., Barbara, A., Veenstra, G. J., Xu, G., Kaiser, U. B., Liu, X. S., Sugrue, S. P., He, X., Min, J., Kato, Y., and Shi, Y. G. (2012) Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. *Cell* **151**, 1200-1213
19. Abdel-Wahab, O., Mullally, A., Hedvat, C., Garcia-Manero, G., Patel, J., Wadleigh, M., Malinge, S., Yao, J., Kilpivaara, O., Bhat, R., Huberman, K., Thomas, S., Dolgalev, I., Heguy, A., Paietta, E., Le Beau, M. M., Beran, M., Tallman, M. S., Ebert, B. L., Kantarjian, H. M., Stone, R. M., Gilliland, D. G., Crispino, J. D., and Levine, R. L. (2009) Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood* **114**, 144-147
20. Konstandin, N., Bultmann, S., Szwagierczak, A., Dufour, A., Ksienzyk, B., Schneider, F., Herold, T., Mulaw, M., Kakadia, P. M., Schneider, S., Spiekermann, K., Leonhardt, H., and Bohlander, S. K. (2011) Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia* **25**, 1649-1652
21. Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Masse, A., Kosmider, O., Le Couedic, J. P., Robert, F., Alberdi, A., Lecluse, Y., Plo, I., Dreyfus, F. J., Marzac, C., Casadevall, N., Lacombe, C., Romana, S. P., Dessen, P., Soulier, J., Viguie, F., Fontenay, M., Vainchenker, W., and Bernard, O. A. (2009) Mutation in TET2 in myeloid cancers. *N Engl J Med* **360**, 2289-2301
22. Ko, M., Huang, Y., Jankowska, A. M., Pape, U. J., Tahiliani, M., Bandukwala, H. S., An, J., Lamperti, E. D., Koh, K. P., Ganetzky, R., Liu, X. S., Aravind, L., Agarwal, S., Maciejewski, J. P., and Rao, A. (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839-843
23. Dang, L., Jin, S., and Su, S. M. (2010) IDH mutations in glioma and acute myeloid leukemia. *Trends Mol Med* **16**, 387-397
24. Dang, L., White, D. W., Gross, S., Bennett, B. D., Bittinger, M. A., Driggers, E. M., Fantin, V. R., Jang, H. G., Jin, S., Keenan, M. C., Marks, K. M., Prins, R. M., Ward, P. S., Yen, K. E., Liau, L. M., Rabinowitz, J. D., Cantley, L. C., Thompson, C. B., Vander Heiden, M. G., and Su, S. M. (2009) Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739-744
25. Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S. H., Ito, S., Yang, C., Wang, P., Xiao, M. T., Liu, L. X., Jiang, W. Q., Liu, J., Zhang, J. Y., Wang, B., Frye, S., Zhang, Y., Xu, Y. H., Lei, Q. Y., Guan, K. L., Zhao, S. M., and Xiong, Y. (2011) Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17-30
26. Blaschke, K., Ebata, K. T., Karimi, M. M., Zepeda-Martinez, J. A., Goyal, P., Mahapatra, S., Tam, A., Laird, D. J., Hirst, M., Rao, A., Lorincz, M. C., and Ramalho-Santos, M. (2013) Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222-226
27. Chen, J., Guo, L., Zhang, L., Wu, H., Yang, J., Liu, H., Wang, X., Hu, X., Gu, T., Zhou, Z., Liu, J., Liu, J., Wu, H., Mao, S. Q., Mo, K., Li, Y., Lai, K., Qi, J., Yao, H., Pan, G., Xu, G. L., and Pei,

D. (2013) Vitamin C modulates TET1 function during somatic cell reprogramming. *Nat Genet* **45**, 1504-1509
28. Minor, E. A., Court, B. L., Young, J. I., and Wang, G. (2013) Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. *J Biol Chem* **288**, 13669-13674
29. Cartron, P. F., Nadaradjane, A., Lepape, F., Lalier, L., Gardie, B., and Vallette, F. M. (2013) Identification of TET1 Partners That Control Its DNA-Demethylating Function. *Genes Cancer* **4**, 235-241
30. Muller, U., Bauer, C., Siegl, M., Rottach, A., and Leonhardt, H. (2014) TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res* **42**, 8592-8604
31. Deplus, R., Delatte, B., Schwinn, M. K., Defrance, M., Mendez, J., Murphy, N., Dawson, M. A., Volkmar, M., Putmans, P., Calonne, E., Shih, A. H., Levine, R. L., Bernard, O., Mercher, T., Solary, E., Urh, M., Daniels, D. L., and Fuks, F. (2013) TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J* **32**, 645-655
32. Ito, R., Katsura, S., Shimada, H., Tsuchiya, H., Hada, M., Okumura, T., Sugawara, A., and Yokoyama, A. (2014) TET3-OGT interaction increases the stability and the presence of OGT in chromatin. *Genes Cells* **19**, 52-65
33. Shi, F. T., Kim, H., Lu, W., He, Q., Liu, D., Goodell, M. A., Wan, M., and Songyang, Z. (2013) Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells. *J Biol Chem* **288**, 20776-20784
34. Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K., and Pasini, D. (2013) Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol Cell* **49**, 645-656
35. Zhang, Q., Liu, X., Gao, W., Li, P., Hou, J., Li, J., and Wong, J. (2014) Differential Regulation of the Ten-Eleven Translocation (TET) Family of Dioxygenases by O-Linked beta-N-Acetylglucosamine Transferase (OGT). *J Biol Chem* **289**, 5986-5996
36. Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. (2013) TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* **493**, 561-564
37. Hanover, J. A., Yu, S., Lubas, W. B., Shin, S. H., Ragano-Caracciola, M., Kochran, J., and Love, D. C. (2003) Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc transferase encoded by a single mammalian gene. *Arch Biochem Biophys* **409**, 287-297
38. Harwood, K. R., and Hanover, J. A. (2014) Nutrient-driven O-GlcNAc cycling - think globally but act locally. *J Cell Sci* **127**, 1857-1867
39. Jost, K. L., Rottach, A., Milden, M., Bertulat, B., Becker, A., Wolf, P., Sandoval, J., Petazzi, P., Huertas, D., Esteller, M., Kremmer, E., Leonhardt, H., and Cardoso, M. C. (2011) Generation and characterization of rat and mouse monoclonal antibodies specific for MeCP2 and their use in X-inactivation studies. *PLoS One* **6**, e26499. 26410.21371/journal.pone.0026499
40. Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenoder, S., Wang, M., Qin, W., Soding, J., Spada, F., and Leonhardt, H. (2011) Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS One* **6**, e16627. 16610.11371/journal.pone.0016627
41. Liu, N., Wang, M., Deng, W., Schmidt, C. S., Qin, W., Leonhardt, H., and Spada, F. (2013) Intrinsic and extrinsic connections of Tet3 dioxygenase with CXXC zinc finger modules. *PLoS One* **8**, e62755. 62710.61371/journal.pone.0062755
42. Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W., Bauer, C., Munzel, M., Wagner, M., Muller, M., Khan, F., Eberl, H. C., Mensinga, A., Brinkman, A. B., Lephikov, K., Muller, U., Walter, J., Boelens, R., van Ingen, H., Leonhardt, H., Carell, T., and Vermeulen, M. (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146-1159

43. Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75**, 663-670
44. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372
45. Sharma, K., D'Souza, R. C., Tyanova, S., Schaab, C., Wisniewski, J. R., Cox, J., and Mann, M. (2014) Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep* **8**, 1583-1594
46. Tusher, V. G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**, 5116-5121
47. Zhang, H., Zhang, X., Clark, E., Mulcahey, M., Huang, S., and Shi, Y. G. (2010) TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res* **20**, 1390-1393
48. Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y. G., Zhu, J., Wang, P., and Xu, Y. (2013) Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell* **155**, 1545-1555
49. Hanover, J. A., Krause, M. W., and Love, D. C. (2012) Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation. *Nat Rev Mol Cell Biol* **13**, 312-321
50. Williams, K., Christensen, J., Pedersen, M. T., Johansen, J. V., Cloos, P. A., Rappsilber, J., and Helin, K. (2011) TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343-348
51. Butkinaree, C., Park, K., and Hart, G. W. (2010) O-linked beta-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. *Biochim Biophys Acta* **1800**, 96-106
52. Myers, S. A., Panning, B., and Burlingame, A. L. (2011) Polycomb repressive complex 2 is necessary for the normal site-specific O-GlcNAc distribution in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* **108**, 9490-9495
53. Lazarus, M. B., Nam, Y., Jiang, J., Sliz, P., and Walker, S. (2011) Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564-567
54. Jinek, M., Rehwinkel, J., Lazarus, B. D., Izaurralde, E., Hanover, J. A., and Conti, E. (2004) The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. *Nat Struct Mol Biol* **11**, 1001-1007
55. Andres-Bergos, J., Tardio, L., Larranaga-Vera, A., Gomez, R., Herrero-Beaumont, G., and Largo, R. (2012) The increase in O-linked N-acetylglucosamine protein modification stimulates chondrogenic differentiation both in vitro and in vivo. *J Biol Chem* **287**, 33615-33628
56. Olivier-Van Stichelen, S., and Hanover, J. A. (2014) X-inactivation normalizes O-GlcNAc transferase levels and generates an O-GlcNAc-depleted Barr body. *Front Genet* **5**, 256. 10.3389/fgene.2014.00256
57. Loenarz, C., and Schofield, C. J. (2009) Oxygenase catalyzed 5-methylcytosine hydroxylation. *Chem Biol* **16**, 580-583
58. Wu, H., D'Alessio, A. C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y. E., and Zhang, Y. (2011) Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**, 389-393
59. Sievers, F., and Higgins, D. G. (2014) Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* **1079**, 105-116
60. Rottach, A., Kremmer, E., Nowak, D., Leonhardt, H., and Cardoso, M. C. (2008) Generation and characterization of a rat monoclonal antibody specific for multiple red fluorescent proteins. *Hybridoma (Larchmt)* **27**, 337-343

FOOTNOTES

[$] This article contains Supplemental data S1, S2, and S3, and Supplemental table S4

[#] This work was funded by the Deutsche Forschungsgemeinschaft (DFG, Collaborative Research Centers SFB 646/B10 and SFB 1064/A17). C. Bauer is supported by the International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS). H. Leonhardt is a member of the Nanosystems Initiative Munich (NIM).

* corresponding authors: A. Rottach (a.rottach@lmu.de) and H. Leonhardt (h.leonhardt@lmu.de)

[1] Ludwig-Maximilians University Munich, Biocenter, Planegg-Martinsried, Germany

[2] Max Planck Institute for Biochemistry, Martinsried, Germany

[3] Helmholtz Center Munich, Institute for Molecular Immunology, München-Großhadern, Germany

[4] Center for Integrated Protein Science Munich (CIPSM)

[5] abbreviations: TET (ten-eleven-translocation), O-GlcNAc transferase (OGT), post-translational modifications (PTMs), 5-methylcytosine (mC) , 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), 5-carboxylcytosine (caC), 5-hydroxyuracil (hmU), O-linked N-acetylglucosamine (O-GlcNAc); mouse embryonic stem cells (mESCs); 2-oxoglutarate (2-OG); acute myeloid leukemia (AML); 2-hydroxyglutarate (2-HG); intraperitoneally (i.p.); subcutaneously (s.c.); monoclonal antibodies (mAbs); co-immunoprecipitation (co-IP); Cherry fluorescent protein (Ch)

FIGURE LEGENDS

Figure 1: Generation of monoclonal TET antibodies

a) Schematic representation of the domain architecture of the three murine TET proteins: The catalytic dioxygenase domain (D) is split in two parts, separated by a presumably unstructured low complexity insert (8), and is N-terminally preceded by a cysteine-rich region (Cys). The Fe(II)-binding residues are marked by green asterisks. The N-terminus (NT) of TET1 contains a CXXC-type zinc finger (ZF); TET3 exists in two isoforms, one with zinc finger and one without (41). The mean percent identity of the single domains of TET1, TET2, and TET3 is represented by different shades of grey and was calculated with Clustal2.1 (59).

b) Overview of the generated TET antibodies and their possible applications. IP: immunoprecipitation, WB: Western blot, IF: immunofluorescence, x: antibody not suited for the indicated application.

c) Example of Western blot analysis of two TET antibodies with GFP antibody as positive control. The antibodies detect only their target protein, but not the other two TET proteins. The wt protein from mESC whole cell lysates is also detected specifically (black arrowhead).

d) Example of an IP experiment with the indicated TET3 antibodies. Clone 23B9 efficiently precipitates TET3 compared to clone 11B6. Western blot analysis was performed with a GFP antibody. (I: Input, FT: Flowthrough, B: Bound)

e) IF staining of mESCs with the TET1 antibodies 5D8 or 4H7 and DAPI as DNA counterstain. Whereas clone 5D8 shows a clear nuclear pattern, clone 4H7 displays only a weak and diffuse signal. Confocal imaging was performed with a Leica TCS SP5 confocal laser scanning microscope with a 63x/1.4 NA Plan-Apochromat oil immersion objective. Scale bar: 5 µm

Figure 2: TET proteins interact with OGT and are O-GlcNAcylated

a) The tables depict the number of detected unique peptides in IP experiments followed by LC-MS/MS. Left: IP of endogenous TET1 or TET2 with the indicated antibodies. Negative control: protein G beads without antibody. Right: IP of GFP-coupled TET1, TET2, or TET3 expressed in HEK293T cells. Pull-down of GFP served as negative control.

b) Western blot analysis of TET1, TET2, and TET3 specifically enriched with the GFP-Trap®. Upon co-expression of active OGT, the O-GlcNAcylation signal increases for TET1 and TET3 (black arrowheads) compared to coexpression of catalytically inactive OGT^{mut}. For TET2, protein levels in the OGT^{mut} samples are higher (white arrowhead) whereas the O-GlcNAc signal remains constant suggesting a higher

proportion of O-GlcNAcylated TET2 in the OGT sample. Interaction between TET proteins and OGT is independent of OGT activity. Anti-RED antibody (60) detects the coexpressed Cherry-tagged OGT. (I: Input, B: Bound)

Figure 3: Exemplary MS/MS spectra of modified TET1 peptides
a) MS/MS spectrum of a TET1 peptide modified with O-GlcNAc at the threonine residue (o-). O-GlcNAcylation is characterized by a neutral loss of 203.8 Da as indicated. y ions are depicted in red, b ions in blue. Labeling of neutral losses of H_2O or NH_3 (orange peaks) has been removed for clarity reasons. Fully annotated spectra are provided as Supplemental data S2 and S3.
b) MS/MS spectrum of the same TET1 peptide phosphorylated at the serine residue (ph). Phosphorylated ions show a neutral loss of 97.98 Da as indicated. y ions are depicted in red, b ions in blue. Labeling of neutral losses of H_2O or NH_3 (orange peaks) has been removed for clarity reasons. Fully annotated spectra are provided as Supplemental data S2 and S3.

Figure 4: TET phosphorylation is reduced upon O-GlcNAcylation
Boxplots depict the distribution of O-GlcNAc and phosphorylation occupancy in the three conditions, expression of TET protein only, coexpression of OGT^{mut}, and coexpression of OGT. Missing values have been substituted by an occupancy of 0.005; with 0.001 being the lowest measured occupancy. Mean occupancies of single sites can be found in Tables 1-3. Asterisks indicate results of student's t-test: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns: not significant

Figure 5: N-termini and insert regions of TET proteins are densely modified
Schematic and scaled mapping of all TET phosphorylation and O-GlcNAcylation sites to the protein sequence. Modifications are mostly found in the N-terminus and the insert region and rarely occur at the same residue. Residue numbering refers to the murine protein sequences specified in the Supplemental data S1. Green asterisks mark catalytic Fe(II)-binding residues. Basal O-GlcNAc sites occur without any coexpression of OGT or OGT^{mut}; persistent phosphorylation sites show high occupancy despite increase of O-GlcNAcylation. An example of the PTM crosstalk on TET proteins is shown for TET3-S360/361/362/368. empty arrowhead: 2 cooccurring modifications; full arrowheads: 3 cooccurring modifications; blunt arrows: mutual exclusivity.

TABLES

# aa	LP	Modified sequence		Mean TET1	Mean TET1 + OGTmut	Mean TET1 + OGT
160	1.00	H[...]ATVS(ph)PGTENGEQNR		0.15 ± NaN	0.98 ± NaN	0.34 ± 0.08
177	1.00	CLVEGES(ph)QEITQSCPVFEER		0.63 ± 0.01	0.32 ± 0.15	0.48 ± NaN
253	0.85	NT(o-)SNQLADLSSQVESIK		NaN ± NaN	0.07 ± 0.01	0.37 ± NaN
270	0.84	LS(o-)DPSPNPTGSDHNGFPDSSFR		NaN ± NaN	0.08 ± 0.02	0.67 ± 0.06
320	1.00	FILAGS(ph)QPDVFDTKPQEK		0.60 ± 0.16	0.44 ± 0.03	0.30 ± 0.20
327	1.00	FILAGSQPDVFDT(o-)KPQEK	b	0.32 ± 0.47	0.36 ± 0.15	0.55 ± NaN
556	0.81	A[...]STSS(ph)PPCNSTPPMVER		0.23 ± 0.10	NaN ± NaN	NaN ± NaN
561	0.89	A[...]STSSPPCNS(ph)TPPM(ox)VER		0.20 ± 0.09	0.88 ± NaN	NaN ± NaN
734	0.98	QQTNPS(ph)PTFAQTIR		0.44 ± NaN	0.46 ± 0.33	0.32 ± NaN
736	0.96	QQTNPSPT(ph)FAQTIR		0.67 ± 0.06	NaN ± NaN	NaN ± NaN
794	0.77	DAM(ox)SVTTS(o-)GGECDHLK		NaN ± NaN	0.48 ± NaN	1.00 ± 0.00
854	1.00	DGS(ph)PVQPSLLSLMK		0.73 ± 0.13	0.54 ± 0.07	0.25 ± 0.34
892	0.70	L[...]SESSS(ph)PSKPEK		0.51 ± 0.48	0.27 ± 0.03	0.79 ± NaN
950	1.00	S(ph)PDSFATNQALIK	p	0.68 ± 0.26	0.72 ± 0.20	0.49 ± 0.11
969	0.74	SQGYPSS(ph)PT[...]		0.61 ± 0.03	NaN ± NaN	NaN ± NaN
1327	0.66	REAQT(o-)SSN[...]K	b	0.01 ± 0.00	NaN ± NaN	0.79 ± NaN
1964	0.89	ELHATTSLRS(ph)PK		0.33 ± 0.21	0.17 ± NaN	0.47 ± 0.33
2016	1.00	PADRECPDVS(ph)PEANLSHQIPSR		0.68 ± 0.21	0.37 ± 0.18	0.81 ± NaN
2016	0.56	PADRECPDVS(o-)PEANLSHQIPSR		NaN ± NaN	0.26 ± 0.10	0.55 ± 0.41
2042	0.99	DNVVTVS(ph)PYSLTHVAGPYNR		0.73 ± 0.12	NaN ± NaN	0.38 ± NaN

Table 1: Detected modified peptides of TET1

aa: residue number of modified amino acid; LP: Localization probability; (ph): phosphorylated; (o-): O-GlcNAcylated; (ox): oxidized; LP was calculated with the MaxQuant software (44). Residue numbering refers to the murine protein sequences specified in the Supplemental data S1. Basal O-GlcNAc sites and persistent phosphorylation sites are marked with "b" and "p", respectively. The arithmetic mean and the standard deviation of the occupancy is depicted for each dataset. NaN: not a number / not detected

# aa	LP	Modified sequence		Mean TET2	Mean TET2 + OGTmut	Mean TET2 + OGT
15	1.00	TTHAEGTRLS(ph)PFLIAPPS[...]K		0.57 ± 0.06	0.20 ± 0.01	0.01 ± 0.00
23	1.00	T[...]LS(ph)PFLIAPPS(ph)PIS[...]K		0.66 ± 0.09	0.32 ± 0.17	0.32 ± 0.04
39	0.98	LQNGS(ph)PLAERPHPEVNGDTK		0.45 ± 0.10	NaN ± NaN	NaN ± NaN
95	0.98	RT(o-)VS(o-)EPSLSGLHPNK		NaN ± NaN	0.06 ± 0.00	0.26 ± 0.17
97	1.00	TVS(ph)EPSLSGLHPNK		0.53 ± 0.26	0.49 ± 0.30	0.29 ± 0.02
97	0.97	RT(o-)VS(o-)EPSLSGLHPNK		NaN ± NaN	0.01 ± NaN	0.29 ± 0.06
165	1.00	S[...]TSTTQESSGADAFPT(o-)R		NaN ± NaN	0.74 ± 0.06	0.98 ± 0.02
317	0.98	SALDIGPS(o-)RAENK		NaN ± NaN	NaN ± NaN	0.48 ± 0.03
374	0.82	DS(ph)ISPTTVTPPSQSLLAPR		NaN ± NaN	NaN ± NaN	0.34 ± 0.49
374	0.99	DS(o-)IS(ph)PTTVTPPSQSLLAPR		NaN ± NaN	0.19 ± 0.22	0.49 ± 0.37
376	0.99	DS(o-)IS(ph)PTTVTPPSQSLLAPR		1.00 ± NaN	0.44 ± 0.38	0.30 ± 0.20
464	1.00	T[...]LPEQHQNDCGS(ph)PS(ph)PEK		0.79 ± 0.03	NaN ± NaN	NaN ± NaN
466	1.00	T[...]LPEQHQNDCGS(ph)PS(ph)PEK		0.79 ± 0.03	NaN ± NaN	NaN ± NaN
514	0.89	QT(o-)QGSVQAAPGWIELK		NaN ± NaN	0.09 ± 0.03	0.59 ± 0.23
545	0.94	DIS(o-)LHSVLHSQT[...]M(ox)SSK		NaN ± NaN	0.46 ± 0.10	0.78 ± 0.02
552	0.87	DIS(o-)LHSVLHS(o-)QT[...]MSSK		NaN ± NaN	0.13 ± NaN	0.76 ± 0.20
561	0.95	DIS[...]VNQMS(o-)S(o-)K		NaN ± NaN	0.07 ± 0.00	0.79 ± 0.17
562	0.97	DIS[...]VNQMS(o-)S(o-)K	b	0.01 ± 0.01	0.42 ± 0.06	0.80 ± 0.17
565	0.98	QS(o-)TGNVNM(ox)PGGFQR		NaN ± NaN	NaN ± NaN	0.41 ± 0.04
603	1.00	AQMYQVQVNQGPS(ph)PG[...]K		0.41 ± 0.17	0.06 ± NaN	0.14 ± 0.05
625	0.96	ALYQECIPRT(o-)DPSS[...]R	b	0.05 ± 0.01	0.73 ± 0.13	0.98 ± 0.02
746	0.98	VEESFCVGNQYS(o-)K		NaN ± NaN	0.23 ± 0.16	0.83 ± 0.06
778	0.92	ILT(o-)PNSSNLQILPSNDTHPACER		0.09 ± NaN	0.31 ± 0.01	0.64 ± 0.05
807	1.00	EQALHPVGS(o-)K		NaN ± NaN	0.01 ± NaN	0.58 ± 0.12
889	1.00	ALPVPEQGGSQTQT(ph)PPQK	p	0.57 ± 0.23	0.78 ± 0.30	0.57 ± 0.05
944	1.00	YPLS(ph)PPQENMSSR	p	0.43 ± 0.14	0.46 ± 0.22	0.67 ± 0.11
951	0.97	PSSYRYPLSPPQENMS(ph)SR	p	0.24 ± NaN	0.32 ± NaN	0.51 ± 0.68
1437	0.63	QM(ox)T(o-)AQPQLS[...]R		NaN ± NaN	NaN ± NaN	0.67 ± 0.06
1443	0.98	QMTAQPQLS(o-)GPVIR		NaN ± NaN	0.05 ± 0.05	0.52 ± 0.42
1613	0.87	D[...]PPIHT(o-)LHQQTFGDSPSK		NaN ± NaN	NaN ± NaN	0.09 ± 0.10
1622	0.74	Y[...]TLHQQTFGDS(ph)PSK		0.45 ± 0.15	0.07 ± 0.04	0.76 ± NaN
1640	0.76	DAFT(o-)TNSTLKPN[...]K	b	0.05 ± 0.01	0.50 ± 0.20	0.84 ± 0.03
1672	1.00	M(ox)DSHFM(ox)GAAS(o-)R		NaN ± NaN	NaN ± NaN	0.93 ± 0.01
1749	1.00	TASAQELLYSLTGSS(ph)QEK		0.31 ± 0.02	0.07 ± 0.05	0.27 ± 0.02

Table 2: Detected modified peptides of TET2

aa: residue number of modified amino acid; LP: Localization probability; (ph): phosphorylated; (o-): O-GlcNAcylated; (ox): oxidized; grey: multiple modifications occuring on one peptide; LP was calculated with the MaxQuant software (44). Residue numbering refers to the murine protein sequences specified in the Supplemental data S1. Basal O-GlcNAc sites and persistent phosphorylation sites are marked with "b" and "p", respectively. The arithmetic mean and the standard deviation of the occupancy is depicted for each dataset. NaN: not a number / not detected

# aa	LP	Modified sequence		Mean TET3	Mean TET3 + OGTmut	Mean TET3 + OGT
360	0.93	VEAPS(o-)SS(ph)PAPVPS(ph)PISQR		NaN ± NaN	0.10 ± 0.07	0.91 ± 0.09
361	0.79	VEAPSS(o-)S(ph)PAPVPSPISQR	b	0.02 ± 0.01	0.52 ± 0.13	NaN ± NaN
362	1.00	VEAPSSS(ph)PAPVPSPISQR		0.03 ± 0.03	0.38 ± 0.03	0.01 ± NaN
362	0.67	VEAPSSS(o-)PAPVPSPISQR		NaN ± NaN	0.96 ± 0.02	NaN ± NaN
368	1.00	VEAPS(o-)SS(ph)PAPVPS(ph)PISQR		0.34 ± NaN	0.55 ± 0.40	0.28 ± NaN
478	1.00	S(ph)RDM(ox)QPLFLPVR		0.46 ± 0.13	0.38 ± 0.13	0.66 ± NaN
557	0.83	S(ph)PSPM(ox)VALQSGST[...]R		0.23 ± NaN	0.22 ± 0.27	NaN ± NaN
557	0.76	S(o-)PSPM(ox)VALQSGST[...]R		NaN ± NaN	NaN ± NaN	0.44 ± 0.30
1008	0.83	VS(o-)SGAIQVLTAFPR		NaN ± NaN	0.91 ± 0.01	0.36 ± 0.51
1071	0.97	QEALELAGVT(o-)T(o-)DPGLSLK		NaN ± NaN	NaN ± NaN	0.96 ± 0.01
1072	0.89	QEALELAGVT(o-)T(o-)DPGLSLK		NaN ± NaN	NaN ± NaN	0.96 ± 0.01
1077	0.99	QEALELAGVTT(o-)DPGLS(o-)LK		NaN ± NaN	0.50 ± 0.40	0.53 ± 0.31
1105	0.89	YS(o-)GNAVVESYSVLGS[...]R		NaN ± NaN	0.40 ± 0.07	0.73 ± 0.11
1252	0.94	VPQLHPAS(o-)RDPSPFAQSSSCYNR		NaN ± NaN	0.42 ± NaN	0.95 ± 0.03
1256	0.62	VPQLHPASRDPS(o-)PFAQSSSCYNR	b	0.04 ± 0.01	NaN ± NaN	0.98 ± 0.01
1263	0.84	VPQLHPASRDPSPFAQSSS(o-)CYNR		NaN ± NaN	0.48 ± NaN	0.98 ± 0.03
1282	0.88	QEPIDPLTQAES(o-)IPR		NaN ± NaN	0.30 ± 0.06	0.91 ± 0.09
1293	1.00	T(o-)PLPEAS[...]SGGPSMSPK	b	0.01 ± 0.00	0.53 ± 0.06	0.99 ± 0.01
1318	1.00	TPLPEAS[...]SGGPSM(ox)S(ph)PK		0.43 ± 0.07	0.00 ± NaN	NaN ± NaN
1351	0.61	LNSFGAS(ph)CLTPSHFPES [...]R		0.45 ± 0.39	NaN ± NaN	NaN ± NaN
1404	0.76	FGNGTSALTGPSLT(o-)EK	b	0.02 ± 0.02	NaN ± NaN	0.74 ± 0.15
1412	1.00	PWGM(ox)GT(o-)GDFNPALK		NaN ± NaN	0.06 ± 0.00	0.65 ± 0.14
1651	0.72	Q[...]SAVT(o-)VSSYAYTK		NaN ± NaN	NaN ± NaN	0.18 ± 0.05
1653	0.71	Q[...]SAVTVS(o-)SYAYTK		NaN ± NaN	0.43 ± 0.19	0.73 ± 0.05
1654	0.77	Q[...]SAVTVS(o-)S(o-)YAYTK		NaN ± NaN	NaN ± NaN	0.24 ± 0.05
1658	0.99	G[...]TDSAVTVSSYAYT(o-)K	b	0.14 ± 0.03	0.12 ± 0.05	0.75 ± 0.14

Table 3: Detected modified peptides of TET3

aa: residue number of modified amino acid; LP: Localization probability; (ph): phosphorylated; (o-): O-GlcNAcylated; (ox): oxidized; grey: multiple modifications occuring on one peptide; LP was calculated with the MaxQuant software (44). Residue numbering refers to the murine protein sequences specified in the Supplemental data S1. Basal O-GlcNAc sites and persistent phosphorylation sites are marked with "b" and "p", respectively. The arithmetic mean and the standard deviation of the occupancy is depicted for each dataset. NaN: not a number / not detected

Figure 1

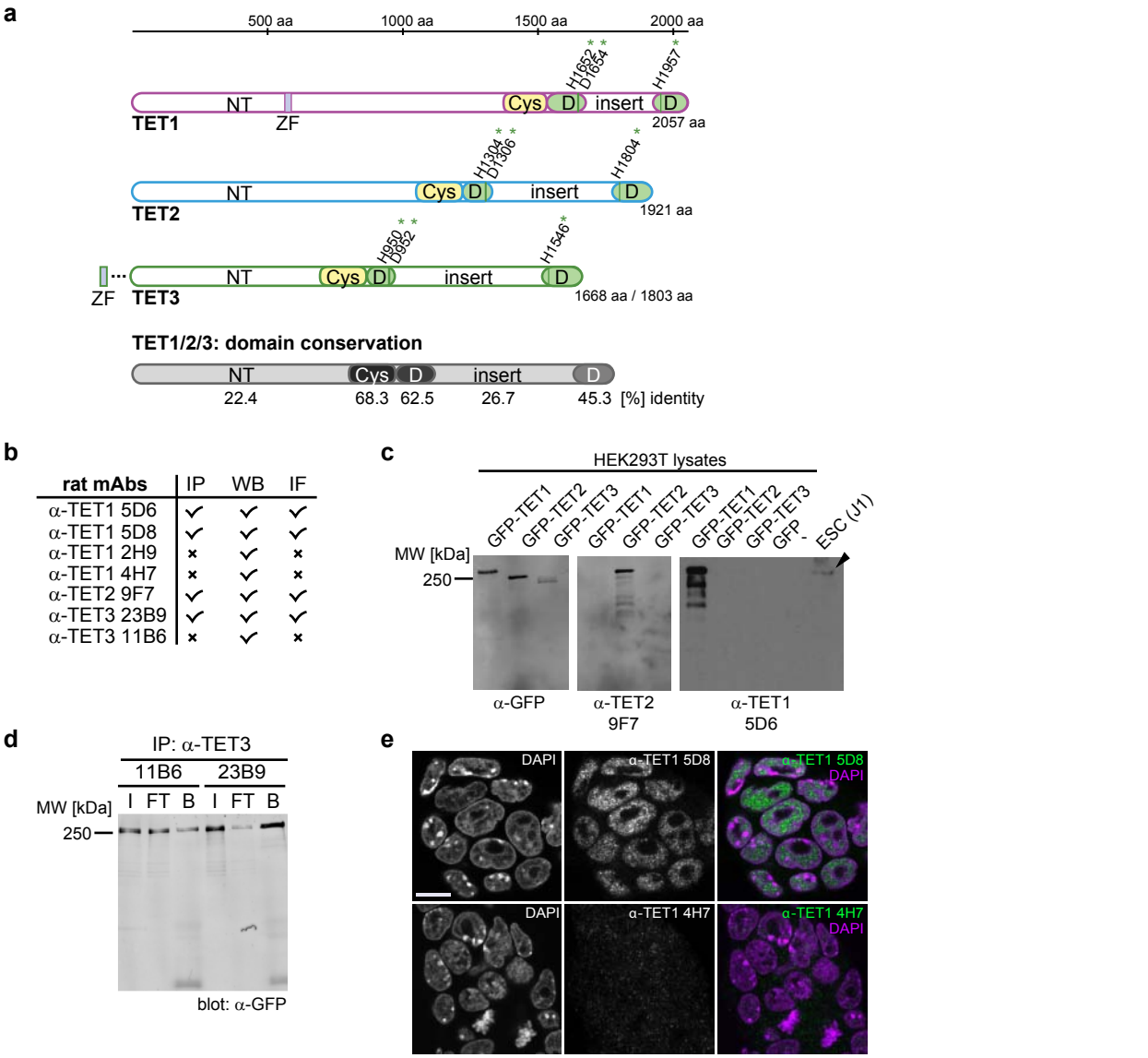

a
500 aa
1000 aa
1500 aa
2000 aa
H1652
D1654
H1957
NT
Cys
D
insert
D
TET1
ZF
2057 aa
H1304
D1306
H1804
NT
Cys
D
insert
D
TET2
1921 aa
H1850
D1852
H1546
NT
Cys
D
insert
D
ZF
TET3
1668 aa / 1803 aa
TET1/2/3: domain conservation
NT
Cys
D
insert
D
22.4
68.3
62.5
26.7
45.3
[%] identity
b
rat mAbs
IP
WB
IF
α-TET1 5D6
α-TET1 5D8
α-TET1 2H9
α-TET1 4H7
α-TET2 9F7
α-TET3 23B9
α-TET3 11B6
c
HEK293T lysates
GFP-TET1
GFP-TET2
GFP-TET3
GFP-TET1
GFP-TET2
GFP-TET3
GFP-TET1
GFP-TET2
GFP-TET3
GFP -
ESC (J1)
MW [kDa]
250
α-GFP
α-TET2
9F7
α-TET1
5D6
d
IP: α-TET3
11B6
23B9
I FT B I FT B
MW [kDa]
250
blot: α-GFP
e
DAPI
α-TET1 5D8
α-TET1 5D8
DAPI
DAPI
α-TET1 4H7
α-TET1 4H7
DAPI

Figure 2

a

ESCs	α-TET1 5D6	α-TET1 5D8	α-TET2 9F7	neg. ctrl
TET1	3	5	0	0
TET2	0	0	2	0
TET3	0	0	0	0
OGT	3	7	1	0

	α-GFP			
HEK293T	GFP-TET1	GFP-TET2	GFP-TET3	GFP
TET1	90	5	4	4
TET2	3	96	18	2
TET3	2	4	63	1
OGT	38	43	50	6

b

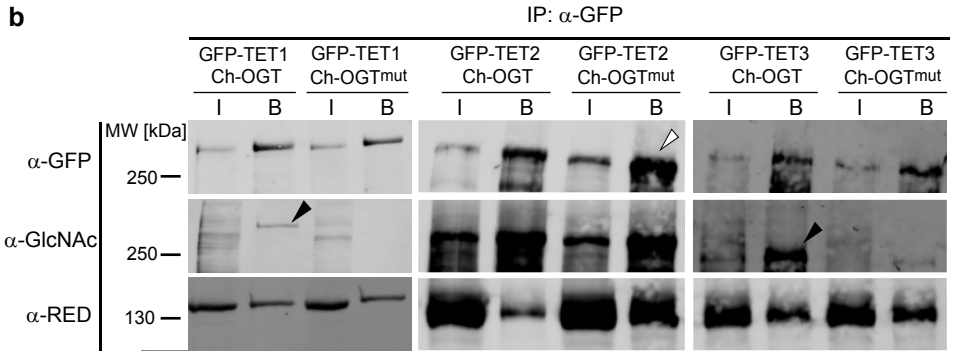
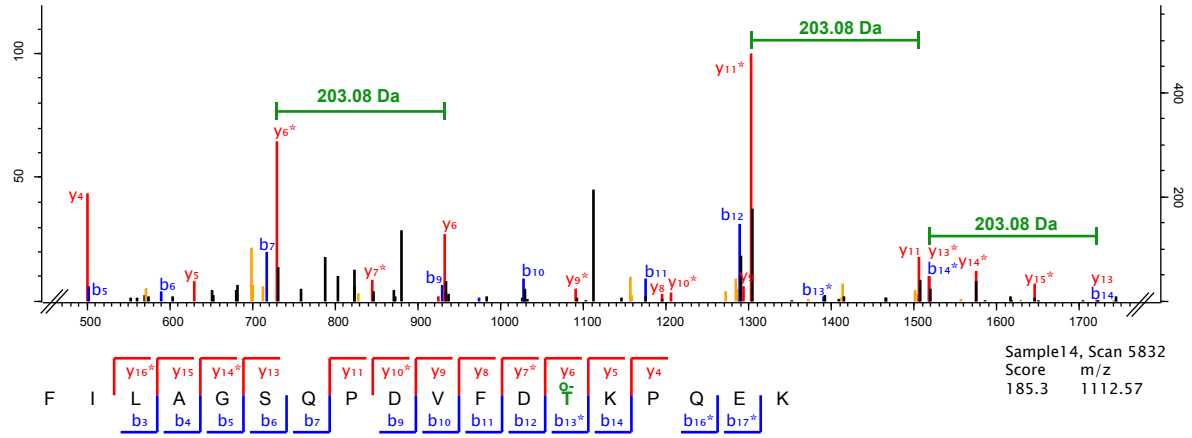


Figure 3

a TET1 T327-GlcNAc



b TET1 S320-ph

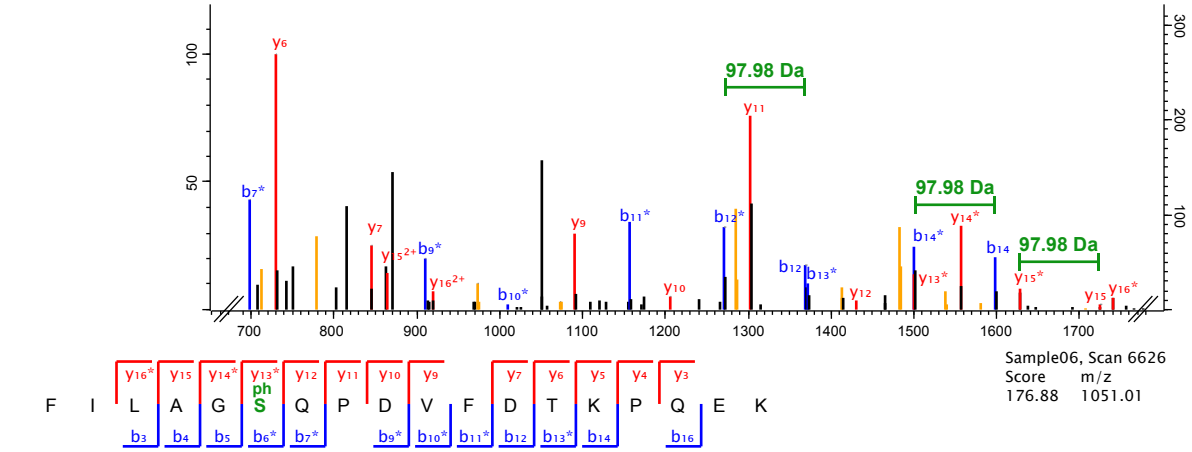


Figure 4

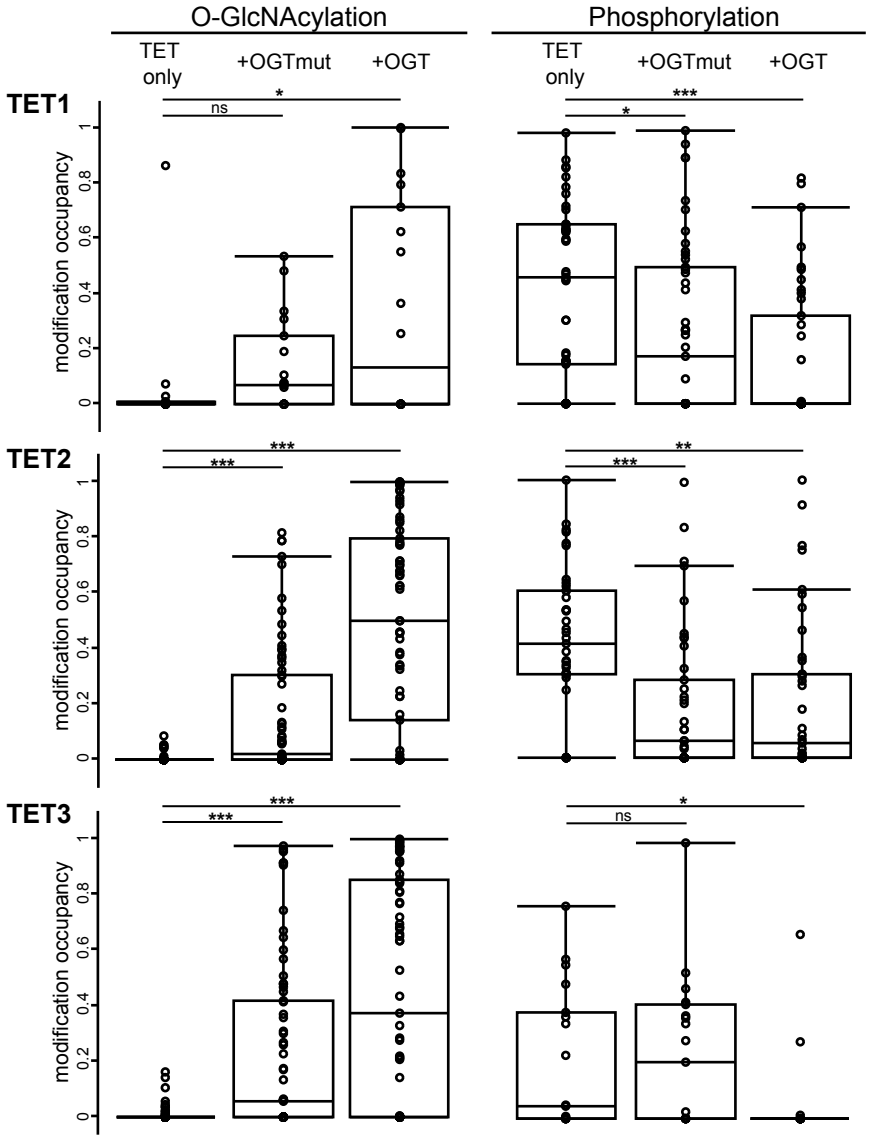
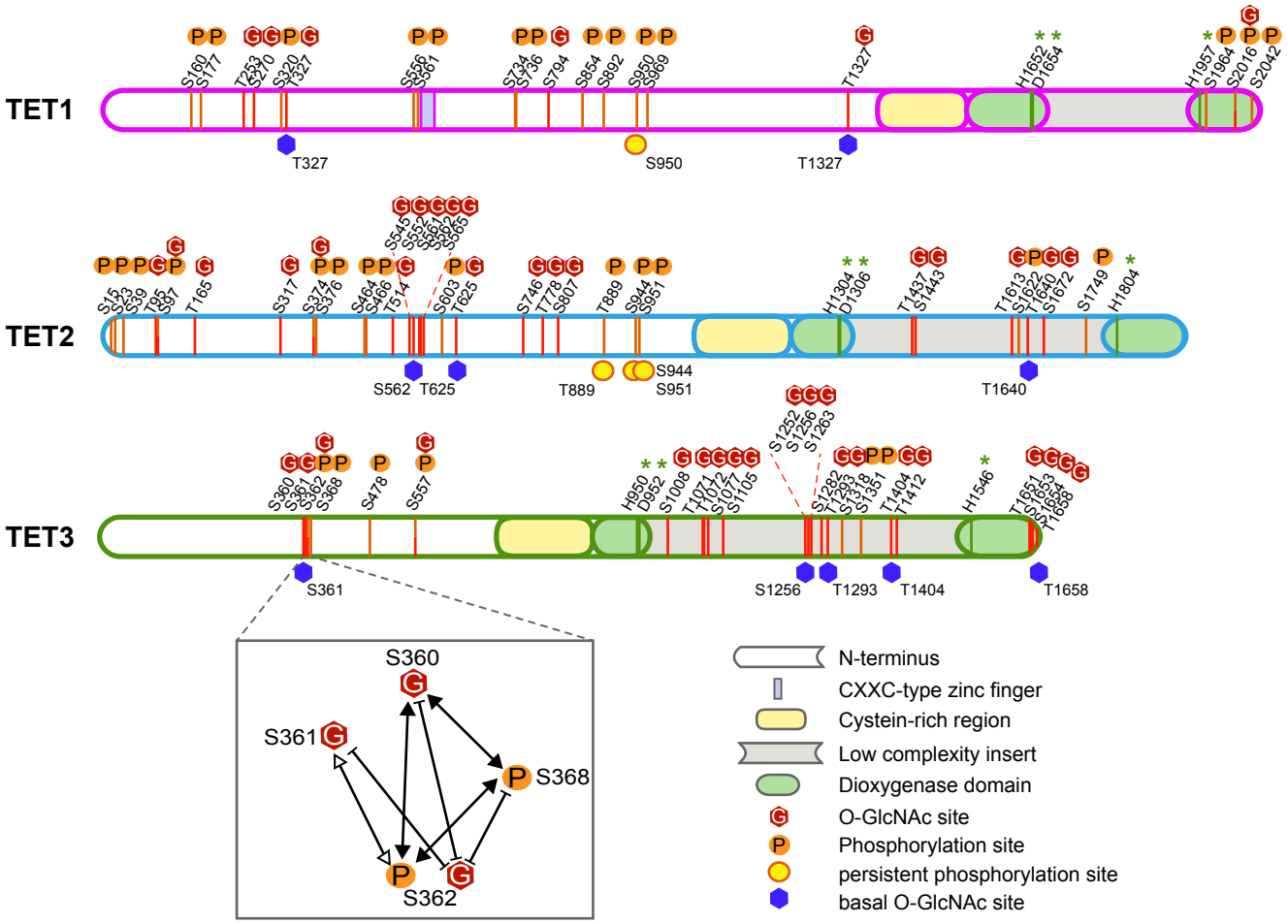

O-GlcNAcylation
Phosphorylation
TET only
+OGTmut
+OGT
TET1
TET2
TET3
modification occupancy
0
0.2
0.4
0.6
0.8
1
ns
*
**
***

Figure 5


TET1
T327
S950
T1327
TET2
S562
T625
T889
S944
S951
T1640
TET3
S361
S1256
T1293
T1404
T1658
S360
S361
S362
S368
N-terminus
CXXC-type zinc finger
Cystein-rich region
Low complexity insert
Dioxygenase domain
O-GlcNAc site
Phosphorylation site
persistent phosphorylation site
basal O-GlcNAc site