EPIGENOMICS
2024, VOL. 16, NOS. 21–22, 1351–1363
https://doi.org/10.1080/17501911.2024.2426450

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

CXXC5 stabilizes DNA methylation patterns in mouse embryonic stem cells

Seung-Gi Jin[a], Jennifer Johnson[a], Zhijun Huang[a], Wei Cui[a], Thomas Dunwell[b] and Gerd P. Pfeifer [a]

[a]Department of Epigenetics, Van Andel Institute, Grand Rapids, MI, USA; [b]GeneFirst, Abingdon Science Park, Oxfordshire, UK

ABSTRACT

Aims: Mammalian genomes encode 12 proteins that contain a CXXC zinc finger domain. Most members of this family are large multi-domain proteins that function in the control of DNA methylation and histone methylation patterns. CXXC5 is a smaller member of the family, along with its closest homologue CXXC4. These two proteins lack known catalytic domains. Here, we have characterized CXXC5 in mouse embryonic stem (ES) cells.

Materials & Methods: We used gene knockouts, RNA sequencing, and DNA methylation analysis by whole-genome bisulfite sequencing.

Results & Conclusions: We show that CXXC5 is a nuclear protein that interacts with 5-methylcytosine oxidases (TET proteins). Removal of CXXC5 from ES cells leads to very few changes in gene expression. CXXC5 extensively colocalizes with TET1 and TET2 at CpG islands. CXXC5 inactivation leads to a substantial reduction of DNA methylation levels that affects all genomic compartments including genic and intergenic regions and CpG island shores. We propose a model in which CXXC5 serves as an anchor for TET proteins at CpG islands. In the absence of CXXC5, the 5-methylcytosine oxidases become dislodged from CpG islands and are enabled to induce genome-scale DNA demethylation. Thus, CXXC5 serves as a stabilizer of DNA methylation patterns.

ARTICLE HISTORY
Received 27 July 2024
Accepted 29 October 2024

KEYWORDS
CXXC5; CXXC4; DNA methylation; TET1; 5-methylcytosine

1. Introduction

Methylation of cytosine in CpG dinucleotides produces 5-methylcytosine (5mC) as an epigenetic mark. This enzymatic reaction is catalyzed by DNA methyltransferases (DNMT proteins). The 5mC residues can be removed by an opposing process of DNA demethylation. Demethylation is initiated by a set of three 5-methylcytosine oxidases in mammals (TET1, TET2, and TET3) that convert the methyl group on 5mC to hydroxymethyl, formyl, and then carboxyl residues [1–3]. The formylcytosines and carboxylcytosines are recognized as unusual or "damaged" bases and are removed from DNA by base excision repair initiated by the TDG DNA glycosylase [2]. This excision repair process includes the incorporation of normal cytosine by a DNA polymerase completing the DNA demethylation reaction.

Although TET proteins may be recruited to their targets by a diverse set of transcription factors, not much is known about general regulators of TET proteins. We recently reported that the chromosomal structural protein SMCHD1 negatively controls the TET protein function [4]. There are two small proteins that are potential-binding partners and regulators of TET enzymes, CXXC4 and CXXC5. These proteins carry their names from the presence of a CXXC-type zinc finger domain in their protein sequence. CXXC domains are often found in much larger proteins, all of which are epigenetic regulators [5]. Functionally, the CXXC domains recognize and bind to unmethylated CpG-rich DNA sequences as found most frequently at CpG islands, but in some cases, these domains bind even more strongly to CpG sequences that harbor 5-carboxylcytosine [6]. Well-characterized CXXC domain-containing proteins include DNMT1, TET1, TET3, histone methyltransferases MLL3 and MLL4, the histone demethylases KDM2A and KDM2B, and the CpG island-binding protein CFP1 [5]. It is thought that CXXC domains are used for recruitment of these large proteins to unmethylated, CpG-rich genomic target sites.

CXXC4, also referred to as IDAX, was initially characterized as an interaction partner of the Disheveled (DVL1) protein and regulator of Wnt signaling [7]. Later, CXXC4 was shown to be a binding partner of TET2 [8]. While TET2 itself lacks a CXXC domain, partnering with CXXC4 may allow TET2 to find CpG-rich genomic target sites. Furthermore, the CXXC4 protein was described as a negative regulator of TET2 by functioning as an adapter to induce TET2 protein degradation [8]. Interestingly, the genomic locus of *TET2* has the *CXXC4* gene as a neighbor at a close distance of 690 kb in the mouse genome and about 450 kb in the human genome suggesting that an ancestral *TET* gene was split and rearranged into these two separate genes during evolution.

CXXC5, also known as a retinoic acid-inducible factor (RINF), has previously been characterized as a regulator of WNT signaling and transforming growth factor beta (TGF-β) signaling [9–11], mTOR/TSC1 signaling [12] and was shown to play a role in the interferon response [13]. Several reports have characterized CXXC5 as a transcription factor that either activates [10,14–16] or represses [12,17,18] genes. One study suggested that CXXC5 interacts with the methyl-CpG binding protein MECP2 indicating a potential DNA methylation-dependent repressor function [19]. These apparent

CONTACT Gerd P. Pfeifer gerd.pfeifer@vai.org Department of Epigenetics, Van Andel Institute, Grand Rapids, MI, USA

Supplemental data for this article can be accessed online at https://doi.org/10.1080/17501911.2024.2426450

Article highlights

- CXXC5 is a nuclear protein that interacts with TET proteins.
- We created ES cells with a single deficiency of CXXC5 and double deficiency of CXXC4 and CXXC5.
- There are minimal changes in gene expression when CXXC5 is inactivated in ES cells.
- As reported previously, CXXC5-deficient ES cells have a subtle defect during differentiation into neurons.
- Using ChIP sequencing, we identified over 12,000 CXXC5 peaks in ES cells. The majority of CXXC5 peaks colocalize with TET1 and/or TET2.
- Inactivation of CXXC5 in ES cells leads to a global reduction of DNA methylation affecting all compartments of the genome including genes, intergenic regions, and CpG island shores.
- We propose a model in which CXXC5 serves as an anchor for TET proteins at CpG islands.
- CXXC5 stabilizes DNA methylation patterns.

inconsistencies and the various proposed functions of CXXC5 have not been reconciled. The *CXXC5* gene is regulated by various cytokines and transcription factors [9], but it has not been clear how the specificity of downstream gene regulation by CXXC5 is attained.

Previous studies showed that TET1 and TET3 have cell-type specific isoforms based upon the presence or absence of a CXXC domain at the N-terminus of the proteins [6,20,21] (Figure 1(a)), suggesting a potential CXXC domain-dependent regulation of DNA methylation in mammalian cells. Here, we focused on a potential epigenetic regulatory function of CXXC5 based on its potential link to TET proteins. We used mouse ES cells for the inactivation of CXXC5 and determined the effects of CXXC5 loss on gene expression and its interplay with TET proteins to regulate DNA methylation patterns.

2. Materials and methods

2.1. Cell culture

The J1 mouse embryonic stem cells (mESCs) (ATCC; SCRC-1010, Manassas, VA) derived from the inner cell mass of a male agouti 129S4/SvJae embryo were grown on gelatin-

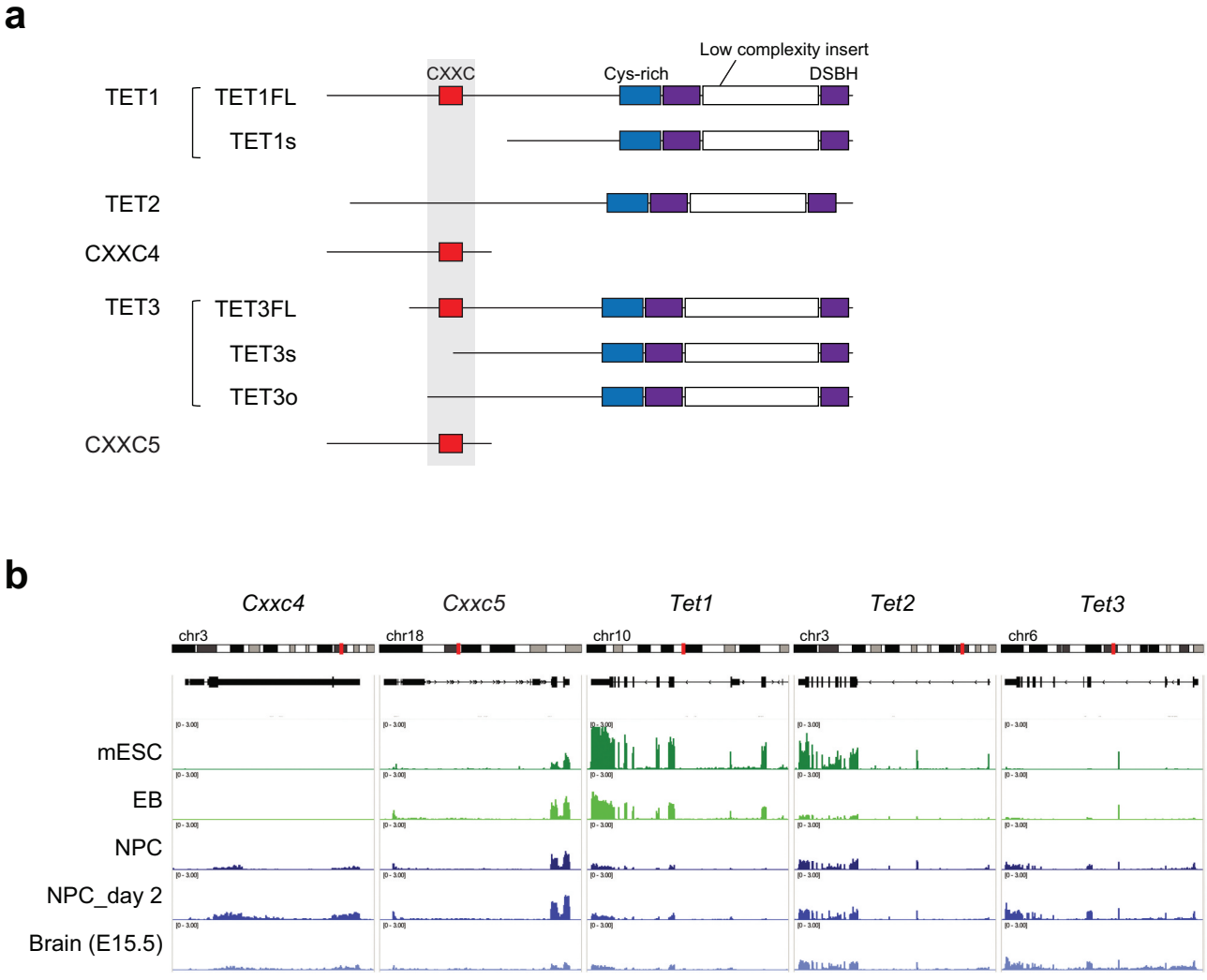

Figure 1. Protein domains, isoforms and expression pattern of the TET family and CXXC4/CXXC5 proteins.

(a) Protein domain structures and isoforms of mammalian TET1, TET2, and TET3, and mammalian CXXC4 and CXXC5. The cysteine-rich domain and the double-stranded β-helix (DSBH) domain together form the core catalytic domain of TET proteins. The CXXC domains that recognize unmethylated CpG motifs are indicated in red color. Splitting and inversion within a putative ancestral *TET2* gene has led to the formation of the separate *TET2* and *CXXC4* genes. (b) RNA-seq data, IGV browser view of the expression profiles of the *Cxxc4*, *Cxxc5*, and *Tet1/2/3* genes during mouse ES cell differentiation toward the neuronal lineage obtained from mouse ES cells (mESC), embryonic body (EB), neuronal progenitor cells (NPC), NPC cultured in N2 media for 2 days (NPC_day 2), and E15.5 mouse brain.

coated plates without feeder cell layers in knockout DMEM (Invitrogen; Carlsbad, CA) supplemented with 15% (v/v) ES-qualified fetal bovine serum (FBS), 1× non-essential amino acids, 2 mm L-glutamine, 100 μM β-mercaptoethanol, and 1,000 U/ml LIF (ESGRO; Millipore, USA) at 37°C in a 5% CO_2 standard incubator. To induce neuronal differentiation as previously described [6], mESCs were cultured on bacterial dishes (Greiner Bio-one) in a medium lacking LIF and with 10% FBS to induce the formation of embryoid body (EB) for 4 days followed by additional incubation in a cultural medium with 5 μM retinoic acid (RA) for 4 days. The RA-treated cells were trypsinized and plated on PORN/laminin-coated plates at a density of 1.5×10^7 cells per 100 mm plate. The cells were further cultured for 2 days in an N2 medium (DMEM/F12 mixture, 1:1 with 1× N2 supplement) and 7 days in B27 media.

293T and Hela cells were maintained as a monolayer in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum at 37°C in a 5% CO_2 standard incubator. Plasmid transfections were carried out using BioT transfection reagent (Bioland; Paramount, CA) according to the manufacturer's instructions.

2.2. Generation of gene knockout (KO) mice and mESC clones

Cxxc4$^{\pm}$ (C57BL/6NJ-Cxxc4<em1J>/J) mice were obtained from the JAX Laboratory and were bred to homozygosity. *Cxxc5*$^{-/-}$ mice were established by CRISPR/Cas9 mediated gene knockout. Animal breeding and experiments were performed under the approved Institutional Animal Care and Use Committee protocol (approval number AUP 18-01-001) of the Van Andel Institute Vivarium. The sgRNA oligonucleotide targeting the mouse *Cxxc5* gene (Table S1) was subcloned into the peSpCas9 (ver. 1.1) vector (Addgene 718,714; a gift from Feng Zhang) [22]. Cas9 mRNA and sgRNAs were generated using mMESSAGE T7 ULTRA kit and MEGAshortscript T7 kit (ThermoFisher Scientific) and microinjected into mouse embryos at the Van Andel Institute's Transgenic Core Facility. The tails of pups from founder mice were used for genotyping to identify genetically engineered mice. Both *Cxxc4* and *Cxxc5* KO mice were backcrossed to wild type C57BL/6 mice for at least 10 generations to eliminate off-target mutations.

To generate CXXC5 KO mESC clones, the same *Cxxc5* targeting CRISPR/Cas9 vector as described above was directly transfected into J1 mESCs and selected by GFP expression. The GFP-selected clones were screened by Sanger sequencing at sgRNA-targeted positions. CXXC4 KO mESC clones were also generated by using CRISPR/Cas9 mediated gene knockout. The *Cxxc4* targeting sgRNA sequence (Table S1) was subcloned into the peSpCas9 vector with a puromycin selection marker, and the targeting vector was directly transfected into wildtype J1 mESCs for CXXC4 single knockout or into two CXXC5 KO mESC clones established above for the generation of double knockout (DKO) mESC clones. Puromycin selected KO clones were screened by Sanger sequencing at sgRNA-targeted positions. The sequencing primers are shown in Table S1. The protein levels of CXXC5 and CXXC4 were determined by Western blot assay as described below.

2.3. Immuno-dot blot analysis

Genomic DNAs (gDNAs) from the mES cells were isolated by using the Quick-DNA MiniPrep Plus kit (Zymo Research; Irvine, CA) followed by the RNase A treatment for 30 min at 37°C. The gDNAs were fragmented by brief sonication and purified using QIAquick PCR purification kits (Qiagen; Valencia, CA). The gDNAs were denatured in a TE buffer by heating for 10 min at 98°C, then immediately chilled on ice for 10 min, and spotted onto a charged nylon-based membrane using a Bio-Dot apparatus (96-well; Bio-Rad). The membrane was treated by UV crosslinking and blocked with 5% nonfat milk in 0.15% Tween-20/PBS (PBST) for 2 h at room temperature, then incubated with anti-5hmC antibody (Active Motif, rabbit polyclonal, 1:8,000) for 1 hour at room temperature. After washing with PBST, the membrane was incubated with peroxidase-conjugated anti-rabbit IgG (Active Motif, 1:15,000) secondary antibody for 1 hour at room temperature. After washing, the signal was visualized by using the ECL Prime detection reagent (GE Healthcare).

2.4. Western blot analysis

Nuclear or cytoplasmic extracts from mES cells were prepared by using NE-PER Nuclear and Cytoplasmic Extraction Reagents (ThermoFisher Scientific) according to the manufacturer's instructions. The extracts were then resuspended in an RIPA buffer (25 mm Tris-HCl pH 7.6, 150 mm NaCl, 1% NP-40, 1% sodium deoxycholate, and 0.1% SDS) and separated by 4%–15% SDS-polyacrylamide gels, and then blotted onto PVDF membranes (Bio-Rad). The membranes were blocked with blocking buffer (5% nonfat milk, 0.05% Tween-20 in PBS) for 1 h at room temperature with gentle shaking and then incubated with affinity purified anti-CXXC5 antibody (in house; 1:2,000) raised against residues 134–195 of mouse CXXC5, anti-CXXC5 (Cell Signaling, #84546; 1:1,000), anti-CXXC4 (Active Motif, #61679; 1:1,000), anti-TET2 (Cell Signaling, #18950), anti- HDAC1 (Abcam, ab31263; 1:2,000), or anti-vinculin (Cell Signaling, #13901; 1:2,000) antibodies for 1 h at room temperature. After washing with PBST, the membranes were incubated with peroxidase-conjugated anti-rabbit IgG or anti-mouse IgG (Active Motif, 1:20,000) secondary antibodies for 1 h at room temperature. Bands were detected using ECL Prime detection reagent (GE Healthcare).

2.5. Protein purification and electrophoretic mobility shift assay (EMSA)

For the bacterial expression vectors, the wild type and mutant mouse *Cxxc4* and *Cxxc5* sequences were cloned into the GST pGEX-5X-1 expression vector (GE Healthcare) and expressed in BL21 (DE3) competent cells. The cells were induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to 0.25 mm and further incubated for 4 h at 37°C with shaking. The GST-tagged proteins were purified on Glutathione

Sepharose 4B beads (GE Healthcare) and eluted using an elution buffer (50 mm Tris-HCl pH 8.5, 150 mm NaCl, 20 mm reduced glutathione, 0.1% Triton X-100), and then stored at −80°C in storage buffer containing 50 mm HEPES pH 7.4, 150 mm NaCl, 2 mm DTT, and 50% glycerol.

For protein-DNA binding analysis, the 35 bp double-stranded probe (sense strand; 5′-TATATTACGCGTATATCGCGTATTTCG CGTTATAA-3′) were ^{32}P-end-labeled with [γ-^{32}P]ATP. The protein–DNA binding reaction was prepared in the following order: H_2O for a final volume of 20 µl, 2 µl 10× binding buffer (60 mm Tris-HCl pH 7.5, 1.5 M NaCl, 60 mm $MgCl_2$, and 30% glycerol), 1 µl of fresh 20 mm DTT, 1 µl (10 fmoles) probe, and approximately 400 ng recombinant proteins. The reaction mixtures were incubated for 45 min at room temperature and fractionated by electrophoresis on 6% non-denaturing polyacrylamide gels, and then visualized by autoradiography.

2.6. Immunofluorescence (IF) analysis

To examine the cellular localization of CXXC5 in cells, the HA-tagged CXXC5 expression plasmids were constructed by insertion of the cDNAs encoding full-length, N-terminal (amino acids 1–244), or C-terminal region (amino acids 245–317) into the pCMV-HA vector. The 2 µg of plasmids were transfected into HeLa cells cultured on a coverslip in a 6-well plate. 48 h after transfection, the cells were fixed in methanol and permeabilized with 0.2% Triton X-100 in PBS for 10 min. After blocking with 2.5% goat serum (Millipore Sigma) for 1 h at room temperature, the cells were incubated with the primary anti-HA antibody in 1% BSA in PBST (PBS containing 0.2% Tween 20) for 1 h followed by incubation with the secondary antibody for 1 h in 1% BSA in PBST. The cells were then mounted with ProLong Gold anti-fade reagent with DAPI (Invitrogen). Fluorescence images were acquired using an I×81 inverted fluorescence microscope (Olympus).

2.7. Plasmid constructions and protein interaction by co-IP

For mammalian expression vectors, the full-length human *CXXC5* and *CXXC4* sequences were cloned into pEF1α-DEST51 Gateway vector with V5-tag (Invitrogen 430,106). The full-length mouse TET2 plasmid (FH-Tet2-pEF) with N-terminal FLAG tag was obtained from Addgene (Addgene 41,710; a gift from Anjana Rao). For the co-IP of exogenously expressed proteins, 10 µg of plasmids were transfected into 293T cells on a 10-cm dish using BioT transfection regent (Bioland). The cells were harvested 48 h after transfection, and then nuclear lysates were prepared by using NE-PER Nuclear and Cytoplasmic Extraction Reagents (ThermoFisher Scientific) according to the manufacturer's instructions. For immunoprecipitations, the nuclear lysates were incubated with 2 µg of the appropriate antibody and 20 µl of Dynabeads Protein G (Invitrogen) overnight, then washed with ice-cold washing buffer containing 10 mm Tris-HCl pH 7.4, 250 mm NaCl, 0.125% NP-40, 2.5 mm EDTA, and protease inhibitors. The samples were boiled with SDS-PAGE loading buffer, followed by SDS-PAGE and Western blot analysis.

2.8. Chromatin immunoprecipitation (ChIP)-seq and data analysis

Ten million mESCs per ChIP reaction were fixed with freshly prepared cross-linking solution (1% formaldehyde, 50 mm HEPES-KOH pH 7.5, 1 mm EDTA, 0.5 mm EGTA, and 100 mm NaCl) for 8 min at room temperature, with gentle shaking, and then the fixation was stopped by adding final 125 mm of glycine and an additional incubation for 5 min. For nuclear isolation, the fixed cells were washed twice with PBS and resuspended in a lysis buffer (50 mm HEPES-KOH pH 7.9, 140 mm NaCl, 1 mm EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, and protease inhibitors) for 10 min on ice. The isolated nuclei were pelleted and washed twice with a washing buffer (10 mm Tris-HCl pH 8.1, 1 mm EDTA, 0.5 mm EGTA, 200 mm NaCl, and protease inhibitors). The nuclear pellets were resuspended in 1 ml of cold shearing buffer (10 mm Tris-HCl pH 8.1, 1 mm EDTA, 0.1% SDS, and protease inhibitors) and sheared to 100–300 bp fragments using a Covaris S220 sonicator (Covaris; Woburn, MA) with the following conditions: time for 12.5 min, 5% duty cycle, 140 Watts power, and 200 cycles per burst. The sheared chromatin was mixed after adding 1% (v/v) Triton X-100 and 150 mm NaCl (final concentrations) and cleared by centrifugation at 14,000 rpm for 10 min at 4°C. The chromatin was incubated with 2.5 µg of affinity purified anti-CXXC5 antibody (in house) overnight at 4°C with rotation and then immunoprecipitated using Dynabeads protein G (ThermoFisher). The libraries for Illumina sequencing of ChIP DNA were prepared using TruSeq Sample Prep kit (Illumina; San Diego, CA) according to the manufacturer's instructions, and then samples were sequenced on an Illumina HiSeq instrument.

For CXXC5 ChIP-seq analysis, the reads were trimmed using Trim Galore (version 0.4.0) with default parameters and aligned to the mouse reference genome (mm10) with Bowtie2 (version 2.3.5). SAMtools (version 2.0.5) was used to remove PCR duplicates and for sorting data into BAM files, then the data were used for peak calling with MACS2 (version 2.2.9.1). To visualize the data, bigwig files were generated using bamCoverage (version 3.5.4) and loaded into IGV viewer (version 2.8.9) for genome browser snapshots of ChIP-seq enrichment. For overlap analysis of ChIP-seq peaks, the intersection function of bedtools (version 2.30.0) provided on the Galaxy web platform was used.

Published data for TET1 mapping in mESCs were obtained from the GEO database (GSE24841 for TET1-N and TET1-C and GSE77453 for TET1FL) [20]. TET2 ChIP-seq data from mESCs were downloaded from the published GEO database (GSE115964) [23]. The downloaded raw data were reanalyzed as described above.

2.9. RNA-seq and data analysis

To determine the changes of gene expression in CXXC4 and CXXC5 DKO cell clones compared to wildtype cells, total RNAs were isolated from the cells with PureLink RNA Mini Kit

(Invitrogen). RNA quality was evaluated with Bioanalyzer 2100 (Agilent Technologies). The RNA-seq libraries were prepared with the KAPA RNA HyperPrep kit (KAPA Biosystems) according to the manufacturer's instructions. The libraries were then submitted to the Van Andel Institute's Genomic Core to perform paired-end sequencing on an Illumina instrument.

The reads were trimmed with Trim_Galore (version 0.4.0). Then the reads were then aligned to the mouse reference genome (mm10) and counted with HISAT2 (version 2.2.1) provided on the Galaxy web platform. To determine differentially expressed genes from the counted sample files with two replicates for each sample, DEseq2 (version 1.40.2) provided on the Galaxy web platform was used using the following parameters, $q < 0.05$ and fold change > 2.

Gene ontology (GO) analysis of differentially expressed gene sets was performed using GSEA functional annotation analysis.

2.10. WGBS and data analysis

Genomic DNAs were isolated from the cells using Quick-DNA MiniPrep Plus kit (Zymo Research) according to the manufacturer's instructions. Then, the gDNA samples with two biological replicates were submitted to the Van Andel Institute's Genomic Core to perform bisulfite conversion and WGBS library preparation. Paired-end sequencing of 150 bp reads was performed on an Illumina NovaSeq 6000 instrument.

WGBS data analysis was performed with our standard procedures [4,24]. Briefly, 150 bp paired-end whole-genome bisulfite reads were trimmed using TrimGalore (v 0.6.7) with the following parameters: length 50, clip_R1 10, clip_R2 18, three_prime_clip_R1 10, and three_prime_clip_R2 10. The trimmed reads were then aligned to the mouse reference genome (mm10) using Bismark [25] (v0.19.0) and Bowtie2 [26] (v2.3.3.1) with the following parameters: X 1000, nucleotide_coverage, and bowtie2. Duplicate reads were removed using the deduplicate_bismark script provided with Bismark. CpG methylation values were called by using the bismark_methylation_extractor script provided with Bismark and the following parameters: no_overlap, comprehensive, merge_non_CpG, and cytosine_report. DMRseq [27] (0.99.0) was used to identify DMRs. Briefly, the sequencing depth for CpGs with at least three covering reads for each sample were considered for DMR calling, and a single CpG coefficient cutoff of 0.05 was used for candidate regions. Significant DMRs were identified using a $q < 0.05$ and a minimum of 10% difference. The methylation of CpG tracks for the IGV browser was generated with bigwig tools.

3. Results

3.1. Expression and biochemical features of CXXC4 and CXXC5

CXXC4 and CXXC5 are two non-catalytic proteins that have been considered as partners and regulators of the TET proteins. TET proteins are 5-methylcytosine oxidases involved in DNA demethylation [28,29]. Inactivation of TET genes and proteins leads to genome-wide changes in DNA methylation as first reported over a decade ago [30–35].

Currently, there is limited information on the contribution of CXXC4 or CXXC5 to the establishment or maintenance of DNA methylation. In mouse ES cells, only CXXC5 but not CXXC4 is expressed at detectable levels, both at the RNA level (Figure 1(b)) and at the protein level [10,14–16], as confirmed by us (see below). *Cxxc4* expression becomes activated upon differentiation of ES cells toward the neuronal lineage (Figure 1(b)). *Tet1* and *Tet2*, but not *Tet3*, are highly expressed in ES cells and become downregulated during neuronal differentiation at a time when *Tet3* is upregulated (Figure 1(b)). It was also reported that full-length *Tet1* with the N-terminal CXXC domain is the only expressed *Tet1* isoform in mES cells [20,21]. Thus, in undifferentiated ES cells, *Tet1* and *Tet2* are the most highly expressed, but full-length *Tet1* and *Cxxc5* are the only CXXC domain-containing TET/CXXC factors in these cells.

To verify a possible DNA binding function, using electrophoretic mobility shift assays, we first tested if CXXC4 or CXXC5 proteins can bind to oligonucleotides containing unmodified CpG dinucleotides by comparing wildtype proteins and cysteine-mutant proteins that have mutations in the CXXC domain as shown in Figure 2(a). We confirmed that both CXXC4 and CXXC5 have the capacity to bind to unmethylated CpG sequences (Figure 2(b, c)). Mutations in the CXXC domains of these proteins abolished the binding. We then used co-immunoprecipitation (co-IP) to test the interaction of CXXC4 and CXXC5 with the 5mC oxidase TET2 (Figure 2(d, e)) after transfection of the tagged protein expressing plasmids into 293T cells. Using reciprocal co-IP, we show that both CXXC5 (Figure 2(d)) and CXXC4 (Figure 2(e)) interact with the TET2 protein in cells.

Previous studies have suggested that CXXC5 plays a role in WNT signalingfor example, by binding to the DVL1 protein [7]. These processes are expected to take place in the cytoplasm. To assess the cellular localization of CXXC5, we performed cell fractionation experiments. Figure S1A shows that CXXC5 is clearly a nuclear protein with almost no presence in the cytoplasm in mouse ESCs. Separate transfections of full-length CXXC5 and its N-terminal and C-terminal regions into HeLa cells show localization of the full-length and C-terminal protein fragments in the nucleus but cytoplasmic localization of the N-terminal fragment (Fig. S1B). The DNA-binding CXXC domain is in the C-terminal region of CXXC5.

3.2. Inactivation of CXXC4 and CXXC5

To study CXXC5, we inactivated this gene using CRISPR/Cas9 technology in mouse ES cells (Fig. S2A). Starting from the derived CXXC5 knockout cells and to rule out a possible compensation effect of altered CXXC4 expression in the single knockouts, we also created CXXC4/CXXC5 double knockout (DKO) ES cells (Fig. S2A), two independent clones each. The knockouts were confirmed by Western blots (Fig. S2B, S2C) and by extensive Sanger sequencing. For this purpose, we established an in-house anti-CXXC5 antibody and used a commercial antibody from Cell Signaling Technology in parallel, with similar results (Fig. S2B). Since CXXC4 is not expressed in ES cells, we differentiated them with retinoic

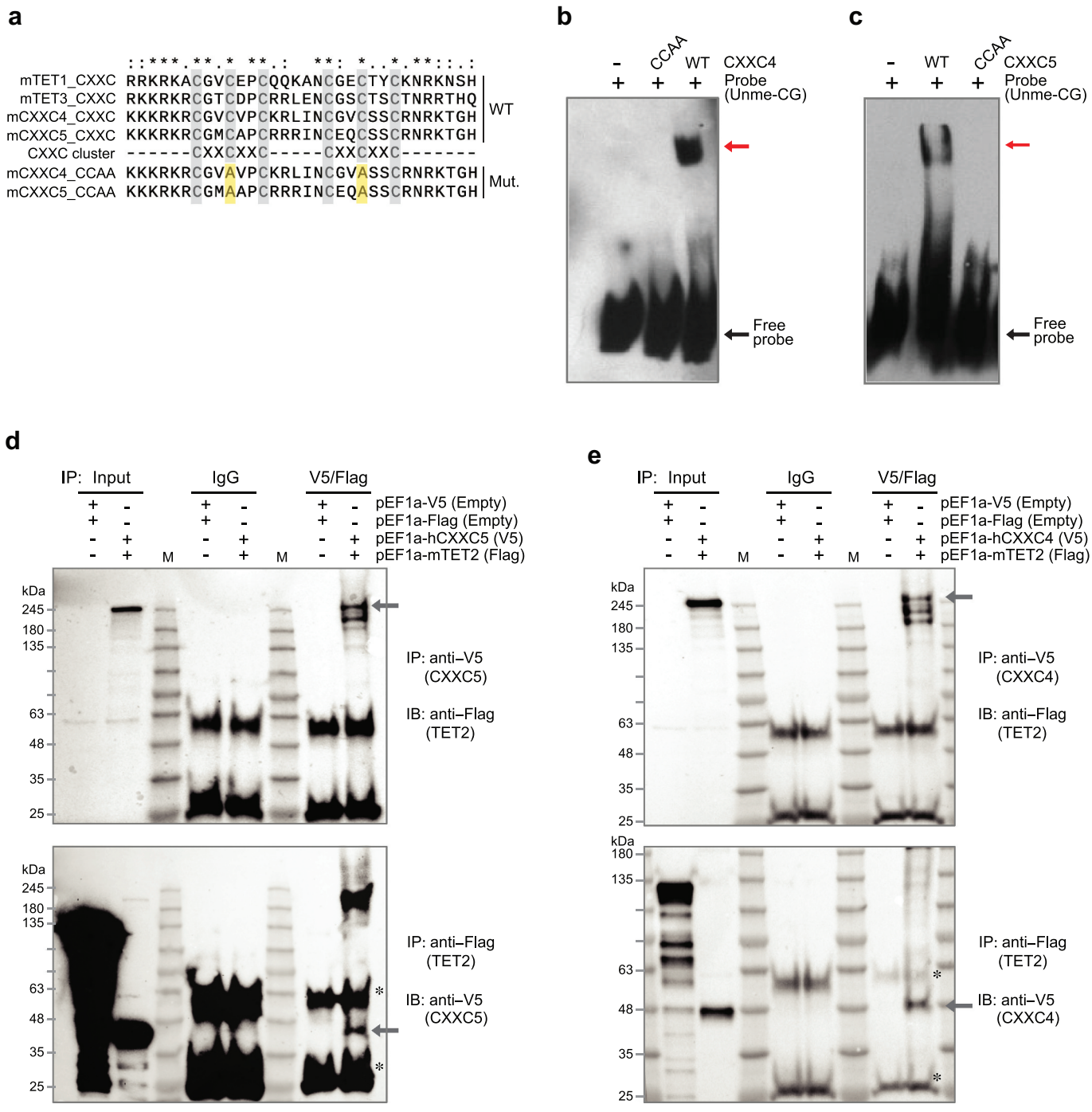

Figure 2. DNA binding of CXXC4 and CXXC5 proteins and interactions with TET2.

(a) Sequence alignment for the CXXC domains of mouse CXXC4, CXXC5, TET1, and TET3. The positions of cysteine within the CXXC domains are indicated by gray shading. CXXC4 and CXXC5 cysteine-mutant proteins are shown with yellow shading below the wildtype sequences. (b) Electrophoretic mobility shift assay of wildtype and mutant CXXC4 (CCAA) using an unmethylated CpG-rich DNA target. (c) Electrophoretic mobility shift assay of wildtype and mutant CXXC5 (CCAA) using an unmethylated CpG-rich DNA target. Red arrows indicate shifted bands. (d) Co-immunopreciptation of CXXC5 and TET2. We used reverse co-IP and Western blotting to demonstrate the interaction of the two proteins after transfection into 293T cells. (e) Co-immunoprecipitation of CXXC4 and TET2. Reverse co-IP was used. Asterisks indicate nonspecific bands from IgG heavy and light chains.

acid before Western blot to confirm CXXC4 inactivation (Fig. S2C).

We obtained a *Cxxc4* knockout mouse line from JAX Laboratories. To create a *Cxxc5* knockout mouse, we used CRISPR/Cas9 microinjection into mouse embryos to target near the initiation site of the coding region as shown in Fig. S2A. Progeny mice were confirmed by genotyping and subjected to further breeding and backcrossing. Neither the *Cxxc4* knockout nor the *Cxxc5* knockout mice had a recognizable phenotype. We also created *Cxxc4-Cxxc5* double knockout mice, and these mice did not have any major phenotype either, consistent with a previous study [36]. From then on, we focused on the ES cells and on the function of CXXC5.

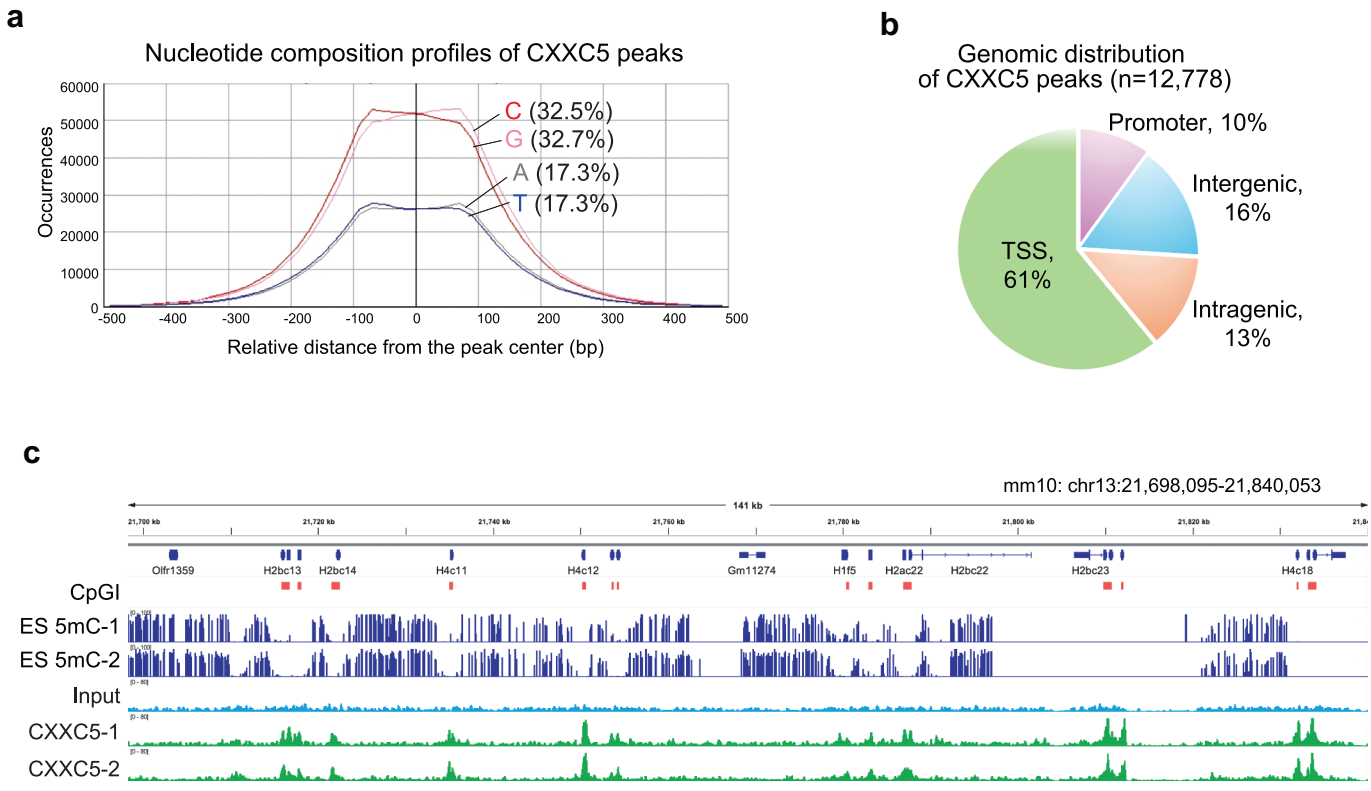

Figure 3. Localization of CXXC5 in the ES cell genome.

(a) Nucleotide composition profiles of CXXC5 peaks obtained by ChIP-seq indicate that CXXC5 binds preferentially to G/C-rich DNA sequences. (b) Genomic distribution of CXXC5 peaks. (c) CXXC5 binds to unmethylated DNA regions. The DNA methylation patterns were obtained by WGBS analysis. CpG islands are shown as red bars.

3.3. Role of CXXC5 in gene expression in mouse ES cells

During induced neural differentiation, the expressions of *Tet1* and *Tet2* are biased toward undifferentiated mouse ES cells, and the transcript levels of *Tet3* and *Cxxc4* are detectable after the neural progenitor cell (NPC) stage, while *Cxxc5* expression is continuously observed as shown in Figure 1(b). To gain insights into the role of CXXC5 in gene regulation, we performed RNA-seq on the single and double knockouts at 2 time points: undifferentiated ES cells and immature neurons (ND9, Fig. S3A). RNAs were isolated from undifferentiated and differentiated wild type and knockout ES cells after embryoid body formation and retinoic acid (RA) treatment using the protocol outlined in Figure S3A.

When analyzing the RNA-seq data in undifferentiated ES cells, we obtained very few differences in gene expression when comparing knockout to wildtype cells. Twenty-two genes were upregulated, and nine genes were downregulated in the DKO in undifferentiated cells (Fig. S3B). This was surprising given the fact that CXXC5 is expressed in ES cells and is a DNA binding protein. However, 9 days after differentiation, when immature neurons are formed with little overt differences in wild type and knockout cells, we observed almost 1,000 genes as up- or down-regulated between wild type cells and the double knockouts (Fig. S3C). We also observed these dramatic changes in gene expression in CXXC5 single KO ES cells after differentiation, similar to that shown in a previous study (Ravichandran et al., 2019). To understand these dramatic variations in gene expression, we performed gene ontology analysis and revealed that the terms for the down-regulated genes primarily reflect functions involved in neuron differentiation (Fig. S3D). Using DisGeNet, we found the following links to neurodevelopmental and other brain disorders among the top 20 most highly dysregulated genes: *ASCL1*, mental retardation; *MLC1*, schizophrenia; *POU4F1*, cerebellar ataxia; *SLC1A2*, schizophrenia, epilepsy; *SLC14A2*, depression; *ZIC1*, Dandy-Walker syndrome.

Additionally, RNA-seq showed a defect in the induction of specific neuronal differentiation genes such as *Pax6*, *Nes*, and *Sox1*, while pluripotency markers such as *Pou5f1*, *Zfp42,* and *Nanog* were still expressed at substantial levels in the RA-induced DKO ES cells (Fig. S3D).

Together, these data suggest that the ES cells lacking CXXC5 or both CXXC4 and CXXC5 have a subtle defect in neuronal differentiation. A similar phenotype has been observed in previous work [14,36]. This defect seems relatively mild because no major morphological differences were observed in the differentiating cells and in mice.

3.4. Genomic localization of CXXC5

To gain a better understanding into the genomic targets of CXXC5, we performed chromatin immunoprecipitation sequencing (ChIP-seq). We observed 12,777 CXXC5 peaks in ES cells, a number much greater than that reported in a previous study [14]. As expected for a CXXC-domain protein, 78% of the CXXC5 peaks localized to CpG islands and the

nucleotide composition profile shows CXXC5 mapped to G/C-rich DNA with about 65% of C or G occurrence (Figure 3(a)). The overall genomic distribution of CXXC5 showed that 61% of the peaks were found at transcription start sites (TSS), which are often embedded within CpG islands, an additional 10% were at extended promoters (extending to −2.4 kb upstream to 0.5 kb downstream of the TSS), 13% were intragenic and only 16% were intergenic (Figure 3(b)). Next, we determined the complete DNA methylation pattern in our ES cell lines using whole-genome bisulfite sequencing (WGBS). Comparison of the DNA methylation data with the CXXC5 peak distribution showed that the CXXC5 peaks were generally localized in methylation-free regions, many of them being annotated as CpG islands (Figure 3(c), small red bars).

To compare the distribution of CXXC5 with that of TET proteins, we used published data for TET1 including those that have been obtained with both an N-terminal and a C-terminal TET1-specific antibody [20,21] and data from published mapping of TET2 [23]. Inspection of genome browser windows revealed a substantial overlap between our mapped CXXC5 peaks and those mapped for TET1 or TET2 in ES cells (Figure 4(a); Figure S4). These proteins overlapped at CpG islands and in other methylation-free genomic regions.

Of the 12,777 CXXC5 peaks, 85% overlapped with the TET1 peaks obtained with the N-terminal TET1 antibody and 56% overlapped with the TET1 peaks obtained with the C-terminal TET1 antibody (Figure 4(b)). The greater overlap of CXXC5 with the peaks obtained with the N-terminal TET1 antibody is perhaps seen because CXXC5, in analogy with CXXC4 [8], is expected to bind to the C-terminus of the TET proteins thus occluding the antibody. When comparing the positions of TET2, 79% of CXXC5 peaks overlapped with the TET2 peaks obtained for the endogenous TET2 protein and 80% of CXXC5 peaks overlapped with Flag-tagged TET2 peaks (Figure 4(c)). When we compared CXXC5 to both TET proteins, we found that 70% of CXXC5 peaks overlapped with both TET1 and TET2 (Figure 4(d)) indicating a large degree of colocalization of CXXC5 with TET proteins in ES cells.

3.5. Control of DNA methylation patterns by CXXC5

Given the interaction and colocalization of CXXC5 with the TET proteins, we then went on to investigate if the loss of CXXC5 and CXXC4 would lead to a change in DNA methylation patterns. We performed WGBS on two wild-type samples and two CXXC5-CXXC4 double knockout ES cell clones (Figure 5). When integrating DNA methylation levels at all quantifiable CpG sites of the mouse genome, we found that the DKO cells had globally reduced DNA methylation levels. The average genome-wide DNA methylation levels were reduced from about 73.5% in wildtype ES cells to nearly 65.5% in the DKO cells, a relative reduction of methylation levels by about 11% (Figure 5(a)). Given the knowledge that WGBS scores the sum of 5mC and 5-hydroxymethylcytosine [37], it may be argued that the reduction in the WGBS signal is due to the sole loss of 5-hydroxymethylcytosine. However, the levels of 5hmC are much smaller than 8% in mouse ES cells [38]. Furthermore, we measured the levels of 5hmC in wildtype and CXXC5 knockout ES cells using semiquantitative dot blots and did not see any difference (Fig. S5). To address differentially methylated regions (DMRs), we used the DMRseq pipeline with the following criteria: $q < 0.05$ and a minimum 10% difference. We observed almost 98,000 DMRs. Of these DMRs, over 99.9% were hypomethylation DMRs (Figure 5(b)). Figure 5(c) shows a genome browser view of the entire chromosome 11 with black color indicating methylation in wildtype and orange color showing methylation in knockout cells. The dual-color overlay indicates that DNA hypomethylation occurs at the chromosome-wide level. Supplementary Figure S6 shows several additional examples including chromosome 8, 10, and 12. When assessing genomic features of interest, we noticed that hypomethylation in the DKO affects all genome compartments tested, including genic regions and DNA regions around enhancers (Figure 5(d)). Because the global hypomethylation affects thousands of genes, this methylation difference is unlikely to be responsible for upregulation of the few differentially expressed genes. On the other hand, hypermethylation of the CpG islands, where TET1, TET2, and CXXC5 are normally located, was not observed.

4. Discussion

CXXC4 and CXXC5 are two small proteins that do not have a known enzymatic activity and have only partially been characterized. Here, we show that CXXC5 is a nuclear protein that interacts with and extensively colocalizes with TET1 and TET2 in mouse ES cells. CXXC4 has been described previously as a protein interacting with TET2 [8]. In our study, we confirmed this interaction for both CXXC4 and CXXC5 (Figure 2(d, e)). We mapped the localization of CXXC5 and determined that this protein strongly prefers to bind to CpG islands (approximately 10,000 peaks). There is a subset of CpG islands that appear to lack CXXC5 peaks. It is still possible that those CpG islands have weak CXXC5 peaks that escape detection by our peak finding algorithm or, alternatively, they may have a higher frequency of CpG methylation which should be incompatible with CXXC5 binding. CXXC5 appears to execute a general function at CpG islands, which suggests that the protein likely does not operate as a sequence-specific transcription factor as previously suggested [12,15–18]. Furthermore, the very small list of differentially expressed genes in both the CXXC5 single knockout and the CXXC4-CXXC5 double knockout is more consistent with a structural role of CXXC5 at CpG islands rather than this protein functioning as a gene-specific transcriptional activator or repressor.

In this study, we mostly worked with the CXXC4-CXXC5 double knockout to avoid any compensatory effects of these highly similar proteins. However, the differences we observed in ES cells are likely due to CXXC5 because CXXC4 is not expressed at detectable levels in undifferentiated ES cells. Nonetheless, we performed RNA-seq in CXXC5 single knockouts and similarly observed very few changes in gene expression (not shown). We also analyzed DNA methylation patterns in the CXXC5 single knockout and obtained a reduction of DNA methylation at the global genome scale comparable to the DKO (data not shown). It remains to be determined if CXXC4 alone can influence DNA methylation and gene

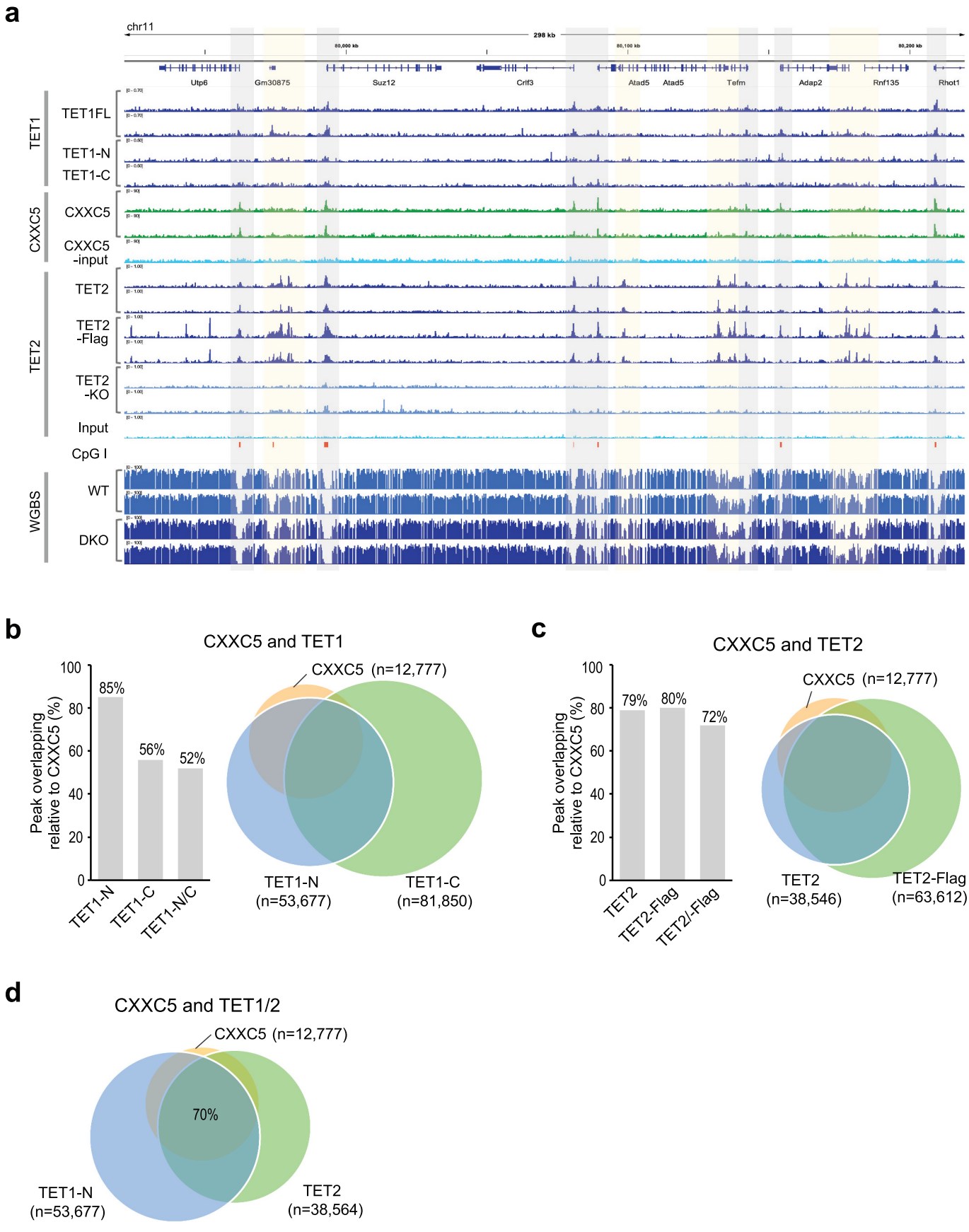

Figure 4. Co-localization of CXXC5 with TET1 and TET2 in the ES cell genome.

(a) Genome browser views of mapping TET1 (TET1FL, TET1-N; N-terminal region including CXXC domain, and TET1-C; C-terminal region including catalytic domain), CXXC5, and TET2 (endogenous and FLAG-tagged protein). The WGBS DNA methylation tracks are shown at the bottom in blue. CpG islands are represented as small red bars. (b,c) Venn diagrams illustrate the overlap between CXXC5 and TET1 or TET2 peaks, and bar graphs show that the percentages of the peaks overlap. (d) Venn diagram illustrates the overlap between CXXC5 and TET1-N and TET2 (endogenous) peaks. The percentage of the peak overlap is 70%.

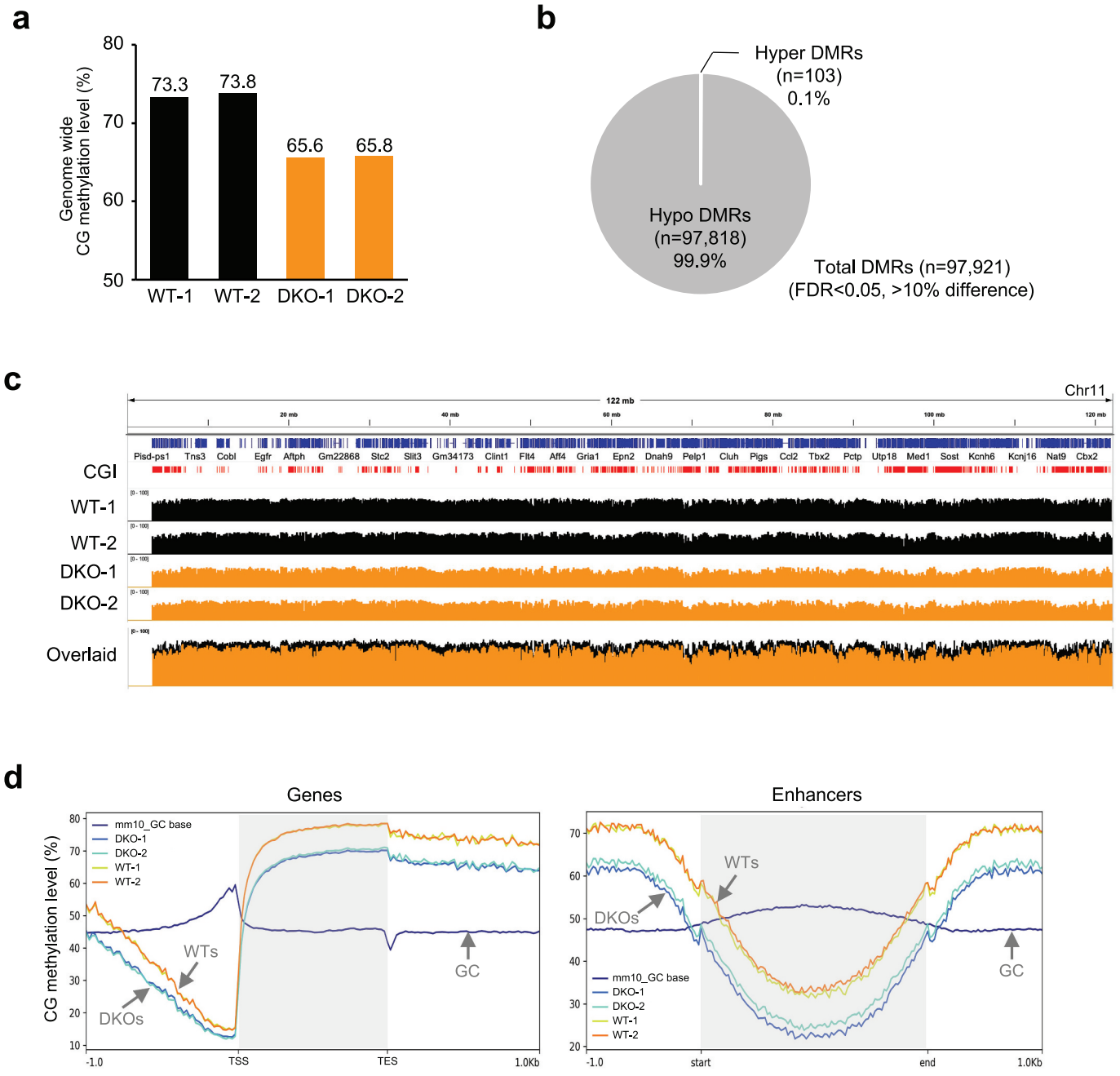

Figure 5. CXXC5 deficiency leads to a global loss of DNA methylation.

(a) Genome-wide DNA methylation levels were calculated from WGBS data using all analyzed CpG sequences. The methylation percentages are represented as black bars for wildtype and as orange bars for DKO ES cells. (b) Analysis of differentially methylated regions (DMRs). Of all DMRs, 99.9% were hypomethylation DMRs. (c) WGBS analysis of DNA methylation on chromosome 11 in wildtype and CXXC4/CXXC5 DKO ES cells. An overlaid view of the tracks is visualized using an IGV viewer and shown at the bottom. CpG islands are represented by a small red bar track (CGI). (d) DNA methylation at defined genomic features in WT and DKO ES cells. The regions analyzed were genes from the transcription start sites (TSS) to the transcription end sites (TES) and enhancers. Wildtype samples are shown in orange and yellow colors; knockout samples are in blue and green. The GC-content tracks in dark blue color are shown for comparison.

expression. However, this will be difficult to establish because CXXC5 is expressed ubiquitously in mouse and human tissues and could compensate for the loss of CXXC4.

CXXC5 is important for maintaining higher DNA methylation levels in ES cells. Cells lacking CXXC5 have a moderate differentiation defect in vitro (Fig. S3) [14,36]. However, the mild phenotypes of *Cxxc5* single knockout and *Cxxc4/5* double knockout mice would suggest that there are pathways in vivo that can compensate for the lack of CXXC5. It is currently unknown how this compensation is achieved.

Interestingly, there is a genome-wide loss of DNA methylation when CXXC5 is missing from ES cells. This is different from some studies obtained with TET1 and TET2 deficiency. One such study reported a general reduction in 5hmC and an increase in 5mC across all chromosomes in the *Tet1-Tet2* DKO embryos [31] and also in TET-TKO cells [30]. However, another publication examining DNA methylation in TET-TKO ES cells showed an average CpG methylation level a few percent lower in TET-TKO ESCs (64.7%) than in wild-type cells (69.0%) [33]. Yet another report suggested that the TET triple-knockout human embryonic stem cells display

bivalent promoter hypermethylation [34]. In other cell types, DNA hypomethylation has also been observed in TET-deficient genomes [32]. These various results are difficult to explain and may have been affected by the different methodologies used for DNA methylation analysis, but the data may also be strongly cell-type dependent.

Another possibility that also applies to our study of CXXC4 and CXXC5 knockouts is that the permanent removal of a DNA methylation regulator (TET, CXXC4/5) leads to compensatory effects of long-term loss of these regulators. To address this issue, rapid protein degradation systems should be applied in future experiments to deplete TET proteins or CXXC5 in order to have more information on how immediate loss of these factors impacts the genomic features of DNA methylation patterns.

Although the mechanism of reduced DNA methylation still needs to be elucidated further, we hypothesize that the loss of methylation in the absence of CXXC5 occurs because of an interplay between CXXC5 and the TET proteins. Both TET1 and TET2, which are expressed in ES cells and CXXC5 are colocalized extensively (Figure 4 and Figure S4). There is also an interaction between CXXC4 or CXXC5 and TET proteins upon co-transfection, suggesting at least a physical proximity of these proteins (Figure 2(d, e)). Our data are consistent with a model in which CXXC5 functions as an anchor for TET1 and TET2 at CpG islands (Fig. S7). Upon loss of CXXC5, TET proteins are at least partially detached from their CpG island-binding sites and become available for 5-methylcytosine oxidation and DNA demethylation genome-wide. Unfortunately, this model cannot currently be tested directly by performing TET1/2 ChIP-sequencing in CXXC5 wild type and knockout ES cells. We employed several commercially available anti-TET1 or anti-TET2 antibodies, even those that have been used in the literature. None of them worked properly in ChIP-seq when we used *TET* knockout cells as stringent negative controls. Therefore, this model can only be substantiated further in future studies. An alternative model is that CXXC5 is a negative regulator of TET protein stability, perhaps analogous to what has been reported for CXXC4/IDAX which causes caspase-mediated degradation of TET2 when bound to DNA targets [8]. While the levels of *Tet2* and *Cxxc4* are indeed negatively correlated in some settings, such as in undifferentiated ES cells, this would not seem to apply to levels of *Tet1/2* versus those of *Cxxc5*, which is expressed invariably during the differentiation process (Figure 1(b)). Thus, we consider this scenario to be less likely. Indeed, levels of the TET2 protein were not increased in the CXXC4/CXXC5 DKO cells as would be expected if CXXC5 functions as a TET2 degrader (Fig. S8). mRNA levels of *Tet1*, *Tet2*, *Tet3, Dnmt1*, *Dnmt3a*, and Dnmt3b *were* not changes in the DKO cells either. A very similar percent reduction of 5mC was observed , recently, in cells inactivated for the protein O-linked N-acetylglucosamine transferase (OGT), which also binds TET proteins [39], and which was proposed as a negative regulator of genome-wide TET activity [40].

CXXC5 is downregulated in several human cancer types including glioblastoma, kidney cancer, and lung cancer (TCGA database). The *CXXC5* gene was identified as a fusion partner with *MN1* in meningiomas [41]. Many cancer genomes are characterized by genome-wide DNA hypomethylation [42]. It remains to be determined if dysregulation of CXXC5 contributes to the DNA hypomethylation phenotype of human cancer.

5. Conclusions

In this study, we show that the protein CXXC5 extensively colocalizes with the 5-methylcytosine oxidases TET1 and TET2. Loss of CXXC5 leads to a substantial reduction of DNA methylation levels that affects all genome compartments. We propose a model in which CXXC5 aids in recruitment of TET proteins to their genomic target regions. While this study was conducted on embryonic stem cells, future work needs to establish if a similar mechanism may operate in somatic cells.

Acknowledgments

The authors are grateful for the core facility services at the Van Andel Institute.

Author contributions

Seung-Gi Jin: study design, execution, acquisition of data, analysis, and interpretation.
Jennifer Johnson: acquisition of data.
Zhijun Huang: analysis of data.
Wei Cui: acquisition of data.
Thomas Dunwell: acquisition of data.
Gerd P. Pfeifer: conception, study design, analysis, and interpretation of data.

All authors have drafted or written, or substantially revised or critically reviewed the article.

"Gerd P. Pfeifer is a member of the Epigenomics Editorial Board. They were not involved in any editorial decisions related to the publication of this article, and all author details were blinded to the article's peer reviewers, as per the journal's double-anonymized peer review policy."

Disclosure statement

No potential conflict of interest was reported by the author(s).

Ethical declaration

Animal breeding and experiments were performed under the approved Institutional Animal Care and Use Committee protocol (approval number AUP 18-01-001) of the Van Andel Institute Vivarium.

Funding

Work described in this manuscript was funded by the National Institutes of Health [CA234595] to Gerd P. Pfeifer. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability statement

The sequencing data for RNA-seq, ChIP-seq, and WGBS have been deposited into the GEO database (accession numbers GSE268884, GSE268885, and GSE268886). The data that support the findings of this study are available from the corresponding author, Gerd P. Pfeifer, upon reasonable request.

ORCID

Gerd P. Pfeifer http://orcid.org/0000-0002-5080-9604

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Tahiliani M, Koh KP, Shen Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324(5929):930–935. doi: 10.1126/science.1170116
   •• **This work describes the initial identification of a 5-methylcytosine oxidase.**
2. He YF, Li BZ, Li Z, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011;333(6047):1303–1307. doi: 10.1126/science.1210944
   •• **This study describes the completion of the 5-methylcytosine oxidation process and removal of oxidized bases by base excision repair.**
3. Ito S, Shen L, Dai Q, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011;333:1300–1303. doi: 10.1126/science.1210597
4. Huang Z, Yu J, Cui W, et al. The chromosomal protein SMCHD1 regulates DNA methylation and the 2c-like state of embryonic stem cells by antagonizing TET proteins. Sci Adv. 2021;7(4). doi: 10.1126/sciadv.abb9149
   • **Identification of SMCHD1 as a negative regulator of TET proteins.**
5. Long HK, Blackledge NP, Klose RJ. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochem Soc Trans. 2013;41:727–740. doi: 10.1042/BST20130028
6. Jin SG, Zhang ZM, Dunwell TL, et al. Tet3 reads 5-carboxylcytosine through its CXXC domain and is a potential guardian against neurodegeneration. Cell Rep. 2016;14(3):493–505. doi: 10.1016/j.celrep.2015.12.044
   • **This publication describes the identification of three isoforms of TET3 and shows preferential binding of the TET3 CXXC domain to 5-carboxylcytosine**.
7. Hino S, Kishida S, Michiue T, et al. Inhibition of the wnt signaling pathway by Idax, a novel Dvl-binding protein. Mol Cell Biol. 2001;21:330–342. doi: 10.1128/MCB.21.1.330-342.2001
8. Ko M, An J, Bandukwala HS, et al. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. Nature. 2013;497:122–126. doi: 10.1038/nature12052
   • **In this study, CXXC4/IDAX was identified as an interactor and regulator of TET2**.
9. Yan X, Wu J, Jiang Q, et al. CXXC5 suppresses hepatocellular carcinoma by promoting tgf-β-induced cell cycle arrest and apoptosis. J Mol Cell Biol. 2018;10(1):48–59. doi: 10.1093/jmcb/mjx042
10. Sun Y, Yan K, Wang Y, et al. Context-dependent tumor-suppressive BMP signaling in diffuse intrinsic pontine glioma regulates stemness through epigenetic regulation of CXXC5. Nat Cancer. 2022;3:1105–1122. doi: 10.1038/s43018-022-00408-8
11. Andersson T, Sodersten E, Duckworth JK, et al. CXXC5 is a novel BMP4-regulated modulator of wnt signaling in neural stem cells. J Biol Chem. 2009;284:3672–3681. doi: 10.1074/jbc.M808119200
12. Wang W, Zhang Z, Zhao M, et al. Zinc-finger protein CXXC5 promotes breast carcinogenesis by regulating the TSC1/mTOR signaling pathway. J Biol Chem. 2023;299:102812. doi: 10.1016/j.jbc.2022.102812
13. Ma S, Wan X, Deng Z, et al. Epigenetic regulator CXXC5 recruits DNA demethylase Tet2 to regulate TLR7/9-elicited IFN response in pDCs. J Exp Med. 2017;214:1471–1491. doi: 10.1084/jem.20161149
14. Ravichandran M, Lei R, Tang Q, et al. Rinf regulates pluripotency network genes and tet enzymes in embryonic stem cells. Cell Rep. 2019;28(8):1993–2003.e5. doi: 10.1016/j.celrep.2019.07.080
15. Kim HY, Yang DH, Shin SW, et al. CXXC5 is a transcriptional activator of flk-1 and mediates bone morphogenic protein-induced endothelial cell differentiation and vessel formation. Faseb J. 2014;28:615–626. doi: 10.1096/fj.13-236216
16. Kim MY, Kim HY, Hong J, et al. CXXC5 plays a role as a transcription activator for myelin genes on oligodendrocyte differentiation. Glia. 2016;64:350–362. doi: 10.1002/glia.22932
17. Tsuchiya Y, Naito T, Tenno M, et al. ThPOK represses CXXC5, which induces methylation of histone H3 lysine 9 in Cd40lg promoter by association with SUV39H1: implications in repression of CD40L expression in CD8+ cytotoxic T cells. J Leukoc Biol. 2016;100:327–338. doi: 10.1189/jlb.1A0915-396RR
18. Aras S, Pak O, Sommer N, et al. Oxygen-dependent expression of cytochrome c oxidase subunit 4-2 gene expression is mediated by transcription factors RBPJ, CXXC5 and CHCHD2. Nucleic Acids Res. 2013;41:2255–2266. doi: 10.1093/nar/gks1454
19. Ayaz G, Turan G, Olgun CE, et al. A prelude to the proximity interaction mapping of CXXC5. Sci Rep. 2021;11(1):17587. doi: 10.1038/s41598-021-97060-6
20. Williams K, Christensen J, Pedersen MT, et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature. 2011;473:343–348. doi: 10.1038/nature10066
   • **This work reports the genome-wide characterization of TET1 and of 5-hydroxymethylcytosine in ES cells.**
21. Zhang W, Xia W, Wang Q, et al. Isoform switch of TET1 regulates DNA demethylation and mouse development. Mol Cell. 2016;64:1062–1073. doi: 10.1016/j.molcel.2016.10.030
   • **Several isoforms of TET1 were identified and described.**
22. Slaymaker IM, Gao L, Zetsche B, et al. Rationally engineered Cas9 nucleases with improved specificity. Science. 2016;351:84–88. doi: 10.1126/science.aad5227
23. Rasmussen KD, Berest I, Kebetaler S, et al. TET2 binding to enhancers facilitates transcription factor recruitment in hematopoietic cells. Genome Res. 2019;29(4):564–575. doi: 10.1101/gr.239277.118
24. Cui W, Huang Z, Jin SG, et al. Deficiency of the Polycomb protein RYBP and TET methylcytosine oxidases promotes extensive CpG island hypermethylation and malignant transformation. Cancer Res. 2023;83:2480–2495. doi: 10.1158/0008-5472.CAN-23-0269
25. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. Bioinformatics. 2011;27:1571–1572. doi: 10.1093/bioinformatics/btr167
26. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9:357–359. doi: 10.1038/nmeth.1923
27. Korthauer K, Chakraborty S, Benjamini Y, et al. Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. Biostatistics. 2019;20:367–383. doi: 10.1093/biostatistics/kxy007
28. Wu X, Zhang Y. Tet-mediated active DNA demethylation: mechanism, function and beyond. Nat Rev Genet. 2017;18:517–534. doi: 10.1038/nrg.2017.33
29. Pfeifer GP, Kadam S, Jin SG. 5-hydroxymethylcytosine and its potential roles in development and cancer. Epigenetics Chromatin. 2013;6(1):10. doi: 10.1186/1756-8935-6-10
30. Dawlaty MM, Breiling A, Le T, et al. Loss of tet enzymes compromises proper differentiation of embryonic stem cells. Dev Cell. 2014;29:102–111. doi: 10.1016/j.devcel.2014.03.003
31. Dawlaty MM, Breiling A, Le T, et al. Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. Dev Cell. 2013;24:310–323. doi: 10.1016/j.devcel.2012.12.015
32. Lopez-Moyado IF, Tsagaratou A, Yuita H, et al. Paradoxical association of TET loss of function with genome-wide DNA hypomethylation. Proc Natl Acad Sci U S A. 2019;116:16933–16942. doi: 10.1073/pnas.1903059116
33. Lu F, Liu Y, Jiang L, et al. Role of tet proteins in enhancer activity and telomere elongation. Genes Dev. 2014;28:2103–2119. doi: 10.1101/gad.248005.114
34. Verma N, Pan H, Dore LC, et al. TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. Nat Genet. 2018;50:83–95. doi: 10.1038/s41588-017-0002-y
35. Wu H, D'Alessio AC, Ito S, et al. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature. 2011;473:389–393. doi: 10.1038/nature09934
36. Abou-Jaoude A, Huang CY, Flores JC, et al. Idax and Rinf facilitate expression of tet enzymes to promote neural and suppress trophectodermal programs during differentiation of embryonic stem cells. Stem Cell Res. 2022;61:102770. doi: 10.1016/j.scr.2022.102770

37. Jin SG, Kadam S, Pfeifer GP. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res. 2010;38(11):e125. doi: 10.1093/nar/gkq223
38. Nestor CE, Ottaviano R, Reddington J, et al. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. Genome Res. 2012;22:467–477. doi: 10.1101/gr.126417.111
39. Chen Q, Chen Y, Bian C, et al. TET2 promotes histone O-GlcNAcylation during gene transcription. Nature. 2013;493:561–564. doi: 10.1038/nature11742
40. Sepulveda H, Li X, Yue X, et al. OGT prevents DNA demethylation and suppresses the expression of transposable elements in heterochromatin by restraining TET activity genome-wide. bioRxiv. 2024. doi: 10.1101/2024.01.31.578097
• **Characterization of OGT as a global regulator of TET activity.**
41. Sturm D, Orr BA, Toprak UH, et al. New brain tumor entities emerge from molecular classification of CNS-PNETs. Cell. 2016;164:1060–1072. doi: 10.1016/j.cell.2016.01.015
42. Ehrlich M. DNA methylation in cancer: too much, but also too little. Oncogene. 2002;21:5400–5413. doi: 10.1038/sj.onc.1205651