Tet3 CXXC Domain and Dioxygenase Activity Cooperatively Regulate Key Genes for *Xenopus* Eye and Neural Development

Yufei Xu[1,*], **Chao Xu**[2,*], **Akiko Kato**[3,*], **Wolfram Tempel**[2], **Jose Garcia Abreu**[4], **Chuanbing Bian**[2], **Yeguang Hu**[1], **Di Hu**[1,5], **Bin Zhao**[5], **Tanja Cerovina**[2], **Jianbo Diao**[5], **Feizhen Wu**[5], **Housheng Hansen He**[6], **Qingyan Cui**[7], **Erin Clark**[1], **Chun Ma**[1,5], **Andrew Barbara**[1], **Gert Jan C. Veenstra**[8], **Guoliang Xu**[7], **Ursula B. Kaiser**[1], **X. Shirley Liu**[6,9], **Stephen P. Sugrue**[10], **Xi He**[4], **Jinrong Min**[2,#], **Yoichi Kato**[3,#], and **Yujiang Geno Shi**[1,#]

[1] Division of Endocrinology, Diabetes and Hypertension, Departments of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA [2] Structural Genomics Consortium and Department of Physiology, University of Toronto, Toronto, ON M5G 1L7, Canada [3] Department of Biomedical Sciences, Florida State University College of Medicine, Tallahassee, FL 32306, USA [4] F.M. Kirby Neurobiology Center, Children's Hospital Boston, Harvard Medical School, Boston, MA 02115, USA [5] Laboratory of Epigenetics, Institutes of Biomedical Science, Fudan University, Shanghai 200032, P.R. China [6] Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, MA 02115, USA [7] The State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, P.R. China [8] Radboud University Nijmegen, Nijmegen Center for Molecular Life Sciences, Nijmegen 6500 HB, The Netherlands [9] Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA 02215, USA [10] Department of Anatomy and Cell Biology, University of Florida, Gainesville, FL 32610, USA

SUMMARY

Ten-Eleven Translocation (Tet) family of dioxygenases offers a new mechanism for dynamic regulation of DNA methylation and has been implicated in cell lineage differentiation and oncogenesis. Yet their functional roles and mechanisms of action in gene regulation and embryonic development are largely unknown. Here, we report that *Xenopus* Tet3 plays an essential role in early eye and neural development by directly regulating a set of key developmental genes. Tet3 is an active 5mC hydroxylase regulating the 5mC/5hmC status at target gene promoters. Biochemical and structural studies further reveal a novel DNA binding mode of the Tet3 CXXC domain that is critical for specific Tet3 targeting. Finally, we show that the enzymatic activity and CXXC domain are crucial for Tet3's biological function. Together, these findings define Tet3 as a novel transcription factor and reveal a molecular mechanism by which

the 5mC hydroxylase and DNA binding activities of Tet3 cooperate to control target gene expression and embryonic development.

INTRODUCTION

The process of vertebrate development is established through the integration of several molecular pathways controlled by key regulatory genes and complex epigenetic markings. DNA methylation at the 5-position of cytosine (5mC) is a key epigenetic mark playing crucial roles in vertebrate development (Bestor and Coxon, 1993; Bird, 1986; Reik et al., 2001). Recent studies have demonstrated that the Tet family of 5mC hydroxylases can catalyze the conversion of 5mC to 5-hydroxymethylcytosine (5hmC) (Tahiliani et al., 2009) and further to 5-formylcytosine (5fC) and 5-carboxylcytosine (5CaC) (He et al., 2011; Ito et al., 2011). These studies also suggest that additional modification of 5mC modulated by Tet enzymes may regulate the dynamics of 5mC and its mediated gene regulation (Branco et al., 2011).

The mammalian Tet family has three members, Tet1, Tet2, and Tet3. It has been suggested that both Tet1 and Tet2 play important roles in ES cell lineage specification (Ito et al., 2010; Koh et al., 2011), and that Tet1 regulates DNA methylation and gene expression in mouse ES cells (Ficz et al., 2011; Williams et al., 2011; Wu et al., 2011; Xu et al., 2011b). Mutational inactivation of *TET2* has been reported to associate with decreased 5hmC levels in various myeloid leukemias (Delhommeau et al., 2009; Langemeijer et al., 2009), and Tet2 deficiency leads to increased hematopoietic stem cell self-renewal and myeloid transformation in mouse (Moran-Crusio et al., 2011; Quivoron et al., 2011). Recently, we and others also show that TET1 and TET2 play critical roles in other human cancers, such as melanoma and breast cancer (Hsu et al., 2012; Lian et al., 2012). In addition, Tet3 is the only Tet family member highly expressed in mouse oocytes and zygotes and is responsible for the hydroxylation of 5mC that occurs in the paternal pronucleus of advanced pronuclear-stage zygotes (Gu et al., 2011; Iqbal et al., 2011; Wossidlo et al., 2011). Conditional knockout of Tet3 in mouse oocytes prevents resetting of DNA methylation patterns in zygotes and impairs reprogramming of transferred somatic nuclei (Gu et al., 2011). Nevertheless, *Tet3*$^{-/-}$ knockout mice are viable through development, dying on postnatal day one (Gu et al., 2011). Taken together, while the discovery of the Tet family of 5mC hydroxylases provides a potential mechanism for the dynamic regulation of DNA methylation, it remains unclear how Tet proteins are recruited to and regulate the expression of target genes, thereby providing linkage to their specific functions in early vertebrate embryonic development.

Although all Tet family members contain a conserved C-terminal catalytic domain, only Tet1 and Tet3 contain the CXXC domain, a potential DNA binding module characterized by two CXXCXXC repeats. The CXXC domains, found in other proteins such as DNMT1, MLL, and CFP1, have been shown to specifically bind to unmethylated CpG dinucleotides and participate in gene transcription regulation (Allen et al., 2006; Pradhan et al., 2008; Xu et al., 2011a). Although our previous study has suggested an important role of the CXXC domain in targeting Tet1 enzyme to specific genomic regions in ES cells (Xu et al., 2011b), the molecular mechanism and biological importance of this domain in Tet1- and Tet3-mediated transcriptional regulation of target genes remain largely unknown.

In this report, we characterize the molecular and biochemical properties and the biological function of Tet3 using *Xenopus* as a model. Our study for the first time shows that Tet3 is essential for early eye and neural development in *Xenopus*. We also demonstrate that several master control genes essential for eye and neural development are Tet3's direct targets,

mechanistically linking Tet3 function in transcriptional regulation of these key genes to the developmental phenotypes caused by Tet3 depletion. Using structural and mutational analyses and functional rescue approaches, we show that Tet3's 5mC hydroxylase and the CXXC domain-mediated DNA binding activities cooperate to regulate target gene expression during eye and neural development.

RESULTS

Identification and characterization of *Xenopus Tet3* gene reveal that *Tet3* is essential for early eye and neural development

To understand the biological function of Tet proteins in early embryonic development, we investigated Tet family members in *Xenopus*. Database searches revealed two *Tet* orthologues in *X. tropicalis*, *Tet2* and *Tet3*, but not *Tet1* (Figure S1A). We cloned *X. laevis Tet3* (*xlTet3*) gene and subsequently discovered two *xlTet3* isoforms (HQ220207-*xlTet3a* and HQ220208-*xlTet3b*) that exhibit greater than 90% amino acid similarity to *xtTet3* (Figure S1A). Surprisingly, despite extensive searching, we were unable to identify a *Tet1* orthologue in either *X. tropicalis* or *X. laevis*. We, therefore, conclude that the *Xenopus* genome contains only two *Tet*-related genes, *Tet2* and *Tet3*, and lacks a *Tet1* gene. Sequence analysis reveals that similar to mammalian Tet3 proteins (which we cloned, validated and deposited into Genebank; HQ220209, human TET3; and HQ423151, mouse Tet3), *Xenopus* Tet3 contains a CXXC domain, a cysteine-rich domain and a double-stranded β helix (DSBH)-containing dioxygenase domain (Figure S1A).

We next examined the expression profile of *xlTet3* during embryogenesis. The temporal expression pattern of *xlTet3* by RT-qPCR reveals that, unlike the high level of *Tet3* mRNA observed in mouse oocytes (Gu et al., 2011; Iqbal et al., 2011; Wossidlo et al., 2011), the *xlTet3* mRNA level in oocytes is very low. *xlTet3* mRNA is also barely detectable at stage 2 (2-cell stage) through stage 7, but increases dramatically from stage 9 (immediately after the initiation of zygotic transcription) to stage 19 and then drops at stage 25 (Figure S1B). The spatial expression pattern of *xlTet3* by *in situ* hybridization shows that *xlTet3* mRNA is barely detectable at stage 9 (Figure 1A), but clearly detected by stage 14 in the neural plate (Figure 1B). In embryo sections, we also detect *xlTet3* mRNA in the neural plate and notochord (Figures S1C and S1D). Moreover, *xlTet3* mRNA level remains high in the neural tube at stage 19 (Figure 1C), and in the region of the developing brain, eye, branchial arches, cement gland and spinal cord at stage 25 (Figures 1D and 1E).

To address Tet3 function during embryogenesis, we performed loss-of-function studies by depleting endogenous xlTet3 protein using the Morpholino antisense oligo (MO) strategy. We first confirmed that *xlTet3* MOs efficiently deplete Tet3 protein (Figure 1F). Control MO or *xlTet3* MOs were injected into two dorsal blastomeres of 4-cell stage embryos. At stage 35, we observe striking developmental abnormalities in 96% (194/201) of *xlTet3* MOs-injected embryos, including malformation of the eye (eyeless), small head, and missing pigmentation along the lateral body, while control embryos develop normally (Figure 1G). The Tet3-depleted embryos die between stages 35 and 40. Importantly, these phenotypes are rescued by co-injecting the *xlTet3* MOs-resistant *xtTet3* mRNA (Figure 1G, right panel). Taken together, these data suggest that Tet3 plays an essential role in early embryogenesis, especially in early eye and neural development.

Tet3 directly regulates a set of genes critical for eye and neural development

To gain insight into the underlying molecular mechanisms linking Tet3 function to early eye and neural development, we examined the effect of Tet3 depletion on the expression of a set of key developmental genes, including *pax6* (eye and neural marker), *rx* and *six3* (eye

markers), *sox2* (pan-neural marker), *otx2* (anterior neural marker), *sox9* and *snail* (neural crest markers), neurogenin related 1 (*ngn2*) and n-tubulin (*tubb2b*) (primary neuron markers), and *shh* and *ptc-1* (sonic hedgehog signaling) in stage 14 embryos. MOs were injected into one dorsal blastomere of 4-cell stage embryos so that the uninjected side could be used as an internal control. For *shh* and *ptc-1* measurement, we injected MOs into two dorsal blastomeres of 4-cell stage embryos since the expression of *shh* and *ptc-1* is in the midline. As shown by *in situ* hybridization, the expression of master eye developmental genes, *pax6*, *rx* and *six3*, is greatly reduced on the *xlTet3* MOs-injected side compared to the uninjected side at stage 14 (Figure 2A). Reduced expression of these eye genes is also seen in the potential eye field on the Tet3-depleted side at stage 19 (Figure S2A). Depletion of Tet3 also inhibits the expression of two primary neuron markers, *ngn2* and *tubb2b*, and two neural crest markers, *sox9* and *snail*, supporting a critical role of Tet3 in neural and neural crest development (Figure 2A). The expression of two major shh signaling components, *shh* and *ptc-1*, is also abolished in Tet3-depleted embryos at stage 14 (Figure 2A), while the expression of *otx2* and *sox2* shows no significant changes (Figure S2B). Importantly, the diminished expression of all affected genes is rescued by *xtTet3* mRNA coinjection (Figure 2A), demonstrating the specific regulation of these genes by Tet3. Results by *in situ* hybridization are further independently confirmed by RT-qPCR assays (Figures 2B and S2C).

To address whether Tet3 directly regulates these key genes, we performed chromatin immunoprecipitation (ChIP) assays. We employed an epitope-tagged ChIP assay using Flag antibody and *flag-xtTet3* mRNA-injected embryos to explore Tet3 occupancy at the promoters of these genes. We detect strong Tet3 binding at the promoters of *pax6*, *rx*, *six3*, *ptc-1, ptc-2, sox9* and *ngn2*, but not at the promoters of control genes, myosin light chain 2 (*myl2*) and cardiac actin (*actc*) (Figure 2C).

Finally, we tested if ectopic expression of any one of these downstream Tet3 target genes can rescue the phenotypes caused by Tet3 depletion. We examined the rescue effects of *pax6*, *rx* or *shh* overexpression. As shown in Figure S2D, none of these can rescue the phenotypes, suggesting that it may be the cumulative effect of the altered expression of a group of target genes, rather than a single gene, that causes the observed developmental defects. Nevertheless, our data strongly suggest that Tet3 is an upstream transcriptional regulator specifically and directly controlling a set of key genes important for early eye and neural development.

Tet3 is an active 5mC hydroxylase that regulates the 5mC/5hmC status at target gene promoters

As DNA methylation at gene promoter has regulatory effects on the expression of the associated gene, we asked whether *Xenopus* Tet3 modulates the 5mC/5hmC status at its target gene promoters. To address this question, we first demonstrate that *Xenopus* Tet3 has conserved 5mC hydroxylase activity as mammalian Tet proteins. The xlTet3 catalytic domain (CD) can convert 5mC to 5hmC in cells and *in vitro* (Figures 3A and 3B). Utilizing two independent methods, anti-5hmC antibody-based dot-blot and T4 Phage β-glucosyltransferase-mediated 5hmC glucosylation assays, we demonstrate that 5hmC exists in the genome of *Xenopus* embryos, and is globally reduced after Tet3 depletion (Figure 3C). However, consistent with undetectable or extremely low levels of 5fC and 5CaC in ES cells and various tissues (Globisch et al., 2010; Ito et al., 2011), we failed to detect 5fC and 5CaC in *Xenopus* embryos (data not shown). Collectively, these data suggest that Tet3 possesses intrinsic 5mC hydroxylase activity and is, at least in part, responsible for modulating 5hmC levels in *Xenopus* embryos.

We next asked whether Tet3 regulates 5hmC levels at target gene promoters. We first examined the dynamic 5hmC level changes by hydroxymethylated DNA immunoprecipitation (hMeDIP)-qPCR (Xu et al., 2011b). Consistent with the increasing *Tet3* expression from stage 10 to 19 (Figure S1B), we observe increasing 5hmC levels at a Tet3 target gene promoter, *rx* promoter, but not at the non-targeted *actc* promoter (Figures 3D and 3E). Importantly, Tet3 depletion results in a significant reduction in 5hmC at *rx* promoter at stages 14 and 19, thereby significantly abolishing the increase in 5hmC from stage 10 to 14 to 19 (P<0.05), while no 5hmC level changes are observed at *actc* promoter (Figure 3E). However, we failed to detect 5hmC level changes at other Tet3 target gene promoters including *pax6, six3, sox9* and *ptc-2* by hMeDIP-qPCR perhaps due to the low 5hmC levels at those promoters and the detection limits of hMeDIP-qPCR. Thus, we applied a more sensitive CCGG site-dependent strategy, *EpiMark* method (Ficz et al., 2011). As exemplified by *pax6* and *ptc-2*, depletion of Tet3 results in a significant reduction in 5hmC at CCGG sites of both promoters (P<0.05), while no 5hmC changes are observed at the non-targeted *myl2* and *actc* genes (Figure 3F). To next determine the 5hmC status at multiple CpG sites, we employed the newly developed Tet-assisted bisulfite sequencing (TAB-Seq) approach (Yu et al., 2012), the most sensitive method to specifically detect 5hmC at base-pair resolution. We first confirm that more than 90% of 5mCs in the internal control DNA oligo have been converted by TET1 CD (Figures S3A–S3C). Using *pax6* as an example, we clearly observe a significant decrease in 5hmC at specific CpG sites within the *pax6* promoter after Tet3 depletion (P<0.05 by one-way ANOVA), while the 5hmC levels at *actc* promoter are not changed (Figure 3G), consistent with that observed via the *EpiMark* method. Taken together, by employing multiple approaches to analyze changes in 5hmC at promoters after Tet3 depletion, we demonstrate that Tet3 regulates the 5hmC status at its target gene promoters.

Finally, we addressed the extent to which 5hmC level decrease by Tet3 depletion results in a subsequent 5mC level increase. Notably, examining the alteration of 5mC directly resulting from the change of 5hmC at Tet3 target gene promoters is extremely technically challenging. Like most *Xenopus* genes, *Tet3* exhibits specific spatial-temporal expression in embryos (Figures 1B–1E). Thus, changes in 5mC caused by Tet3 depletion are expected to occur only in the small fraction of the embryo expressing Tet3, while 5mC remains unchanged in the rest of embryo, effectively diluting the signal from Tet3-expressing cells. Indeed, we attempted and failed to detect 5mC changes after Tet3 depletion using the whole embryos by multiple approaches, such as methylated DNA immunoprecipitation (MeDIP) and methylated CpG island recovery assay. We therefore developed a targeting strategy, employing hMeDIP-MeDIP qPCR to separate the signal from the noise and examine the 5mC level changes directly resulting from 5hmC level changes by Tet3 depletion (Figure 3H). We first performed hMeDIP to enrich 5hmC-containing genomic DNA fragments which presumably are from Tet-expressing cells. Given that Tet3 depletion results in a 5hmC level decrease but not complete removal, Tet3-targeted DNAs in *xlTet3* MOs-treated embryos still contain 5hmC and can be enriched, albeit with a lower amount than the control embryos. We then carried out MeDIP using the hMeDIPed DNA as input to examine relative 5mC levels in those 5hmC-containing DNAs given the dynamic conversion of 5mC to 5hmC and the coexistence of 5mC and 5hmC in the same region (Ficz et al., 2011; Yu et al., 2012). We first confirm that the different amount of starting DNA (hMeDIPed DNA) will not introduce a MeDIP efficiency bias, as validated by the equal MeDIP enrichment of the human *NBR2* promoter from control DNA spiked in the hMeDIPed DNA from control MO- or Tet3 MOs-injected embryos (Figure S3D). For proof-of-principle, we examined the *rx* gene promoter since it has the most dramatic changes in 5hmC after Tet3 depletion and can be easily detected (Figure 3E). Indeed, we clearly observe a significant 5mC level increase at *rx* gene promoter after Tet3 depletion by this hMeDIP-MeDIP qPCR method in stage 14 and 19 embryos (Figure 3I), while the 5mC level at *actc* promoter is not changed

(Figure 3I), validating the reliability of hMeDIP-MeDIP qPCR procedure. Thus, we conclude that Tet3 regulates target gene expression, at least partially, through control of 5mC/5hmC status at the promoter of target genes.

The Tet3 CXXC domain possesses novel DNA binding properties

In addition to the dioxygenase domain that confers 5mC hydroxylase activity, Tet3 also contains a potential DNA binding domain, the CXXC domain. In general, the selective DNA binding activity of a transcription factor serves as a key mechanism for action of the transcription factor in gene transcriptional regulation. Sequence alignment indicates that Tet CXXC domains exhibit a conserved overall structure with other CXXC domains (Figure S4A). However, Tet CXXC domains lack a short sequence motif (KFGG) (Figure S4A), which has been shown to be important for the DNA binding activity of the MLL and CFP1 CXXC domains (Allen et al., 2006; Xu et al., 2011a). To address whether the Tet3 CXXC domain may possess unique DNA binding features, we examined the DNA binding ability and specificity of the *Xenopus* Tet3 CXXC domain (aa 58–111) by isothermal titration calorimetry (ITC) assays. The Tet3 CXXC domain strongly binds to various C/G-rich DNA oligos, but exhibits virtually no binding activity to the A/T-only DNA oligo (Figures 4A–4E and Table S1 #1–8). Significantly, these ITC results also reveal previously uncharacterized DNA binding properties of the CXXC domain. Firstly and most importantly, the Tet3 CXXC domain strongly binds to both non-CpG (Table S1 #5–7) and CpG DNA oligos (Table S1 #1–4 and #9) with a slight preference for CpG DNA oligos. Secondly, the Tet3 CXXC domain strongly binds to CmCGG DNA (Figure 4F and Table S1 #10). These binding properties are in stark contrast to the well-characterized DNA binding property of the CFP1 and MLL CXXC domains, which absolutely require unmethylated CpG dinucleotides (Allen et al., 2006; Xu et al., 2011a). In contrast to the strong binding to CmCGG, CTAG and ACGT DNA oligos, the Tet3 CXXC domain does not bind to mCmCGG, mCTAG or AmCGT DNA oligos (Figure 4G and Table S1 #11–13). These data suggest that although strict CpG content is not required for the Tet3 CXXC domain to interact with DNA, an unmodified cytosine is essential for the Tet3 CXXC binding to DNA. Using the same ITC approach, we also demonstrate that the human TET3 CXXC domain has similar DNA binding properties to the *Xenopus* Tet3 CXXC domain (Table S1 #17–22). Thus, the CXXC domain of the Tet3 family has unique newly identified DNA binding characteristics conserved among vertebrates.

Given the similar binding properties between human and *Xenopus* Tet3 CXXC domains, we next determined the DNA binding specificity of the TET3 CXXC domain across the whole genome of HEK293T cells. The specific TET3 CXXC-bound genomic DNA fragments were enriched by GST pull-downs and analyzed by deep DNA sequencing, as we previously described (Xu et al., 2011b). The TET3 CXXC domain selectively binds to restricted genomic regions (Figure S4B). Importantly, consistent with Tet3 occupancy in *Xenopus* embryos (Figure 2C), the TET3 CXXC domain strongly and selectively binds to the promoters of *PAX6*, *PTCH1*, *NGN2*, *TUBB2B* and *SHH*, but not the promoter of *MYL2* (Figures 4I and S4C). Furthermore, bioinformatics analyses identify 17,953 TET3 CXXC-bound peaks, more than half of which are located at gene promoters (Figure 4J). Interestingly, we also observe a significant enrichment of CpG (and a less degree of CpC) dinucleotides in TET3 CXXC-bound regions (Figure 4K). Further *de novo* motif discovery analyses identify three C-rich sequences among the top 15 ranking motifs within TET3 CXXC-bound regions (Figure 4L), whose consensus sequences are SSGCSGCGCG (*P*=1e-30), CSSCGCSCRC (*P*=3.4711e-26) and SCWGCWGCBS (*P*=4.6563e-25), respectively. Indeed, we validate that these motifs are present at the promoters of several TET3 target genes (Figure S4D). Together with the ITC binding data, these genome-wide analyses suggest that the TET3 CXXC domain is able to bind to the unmodified C followed

by A, T, C or G with a slight preference for CpG dinucleotides. These data also indicate that the specific DNA binding activity of the Tet3 CXXC domain may contribute to Tet3 targeting, thereby serving as another important mechanism for Tet3-mediated gene transcriptional regulation.

Crystal structures of the Tet3 CXXC domain in complex with DNA oligos

To gain further mechanistic insight into the unique DNA binding properties of the Tet3 CXXC domain, we next determined the crystal structures of the *Xenopus* Tet3 CXXC domain (aa 58–111) in complex with 5mC-containing DNA (CmCGG) and CpG-containing DNA (ACGT), respectively (Table 1). Like other CXXC domains, the Tet3 CXXC domain contains eight conserved cysteine residues coordinating two zinc ions. These two zinc ions play a structural role by holding the mainly unstructured CXXC domain together, and forming a crescent-shaped architecture to bind DNA (Figures 5A and 5C). The DNA-binding surface is predominantly positively charged, and wedged into the major groove of DNA to extensively interact with DNA bases by means of hydrogen bonds and electrostatic interactions (Figures 5A-5D).

Surprisingly, structural comparison of the Tet3 CXXC domain in complex with CmCGG or ACGT DNA reveals that the Tet3 CXXC domain binds to the 12mer target DNAs with one nucleotide shift. In the crystal structure of the Tet3 CXXC domain in complex with the unmethylated CpG DNA (ACGT), the Tet3 CXXC domain binds to the DNA centering on the cytosine in the CpG dinucleotide (underlined). Methylation of the cytosine in the CpG dinucleotides (CmCGG) shifts the binding to the cytosine (underlined) preceding the mCpG dinucleotides. Therefore, consistent with the ITC binding results, the Tet3 CXXC domain binds to an unmodified cytosine, which is not restricted by CpG-content, distinct from the CpG-dependent binding of the MLL, CFP1 and DNMT1 CXXC domains (Allen et al., 2006; Pradhan et al., 2008; Xu et al., 2011a).

In both Tet3 CXXC-ACGT and Tet3 CXXC-CmCGG complex structures, the main chain carbonyl oxygen of residue His90 forms a hydrogen bond with the target cytosine, while the side chain of His90 forms another hydrogen bond with the complementary guanine (Figures 5E, 5G, S5A and S5C). In the Tet3 CXXC-ACGT complex structure, residues Gln91 and Ser89 form hydrogen bonds with the guanine-cytosine base pair following the target cytosine-guanine base pair (Figures 5F and S5B). In contrast, in the Tet3 CXXC-CmCGG complex structure, methylation of C6 introduces steric clashes with the side chain of Gln91 and pushes the side chain of Gln91 towards Ser89 which causes the side chain of Gln91 to become partially disordered. As a result, Ser89 flips peptide plane and loses its hydrogen bond interaction with DNA (Figures 5H and S5D). These data provide structural explanations for the binding to CmCGG DNA and the slight binding preference for CpG compared to CpH DNAs (H= A, T, C or 5mC). Substituting the target cytosine with any other nucleotide or 5mC would introduce steric clashes with His90 (Figure S5E), consistent with the undetectable binding of the Tet3 CXXC domain to ATAT DNA (Figure 4E). To determine the role of residue His90 of Tet3, which is conserved among all Tet CXXC domains (Figure S4A), we generated the Tet3 CXXC H90A mutant and tested its DNA binding activities by ITC assay. As expected, the Tet3 CXXC H90A mutant loses its DNA binding abilities (Table S1 #14–15), supporting an essential role of His90 in DNA binding.

To further understand the novel DNA binding property of the Tet3 CXXC domain, we superimposed the complex structures of the Tet3 and CFP1 CXXC domains (Xu et al., 2011a) (Figure S6A). Although the Tet3 CXXC domain adopts a similar fold to that of the CFP1 CXXC domain, the loop region preceding His90 of Tet3 is dramatically different from that of CFP1 (Figure S6A). In the CFP1 CXXC domain, Asp189 forms three hydrogen bonds with the backbone of Lys198, Ile199 and Arg200, making the loop very rigid and

only allowing CpG binding (Xu et al., 2011a). Notably, this Asp residue is highly conserved in the CXXC domains of CFP1, MLL, DNMT1 and KDM2A (Figure S4A). However, the corresponding loop between the α3 and α4 helices of the Tet3 CXXC domain is much shorter and more flexible due to lack of hydrogen bonds seen in the rigid loop of the CFP1 CXXC domain (Figure S6A). Therefore, the shorter and less rigid loop of the Tet3 CXXC domain can accommodate other nucleotides besides G, such as T, C, A, or 5mC, following the target C, which confers cytosine-specific instead of CpG-specific binding property. In addition, we also observe that the Tet3 CXXC domain binds to TCGA and mCCGG DNA oligos (Table S1 #2 and #9) with slight weaker binding affinities compared to other CpG DNA oligos (Table S1 #1, #3 and #4). Although the methyl group of thymine or 5mC does not cause any steric clash with the Tet3 CXXC domain, the hydrophobic methyl group points to the solvent and no residues in the Tet3 CXXC domain can recognize and stabilize it, which is not energetically favorable (Figures S6B–S6E). Taken together, the Tet3 CXXC domain binds to cytosine-containing DNA with a slight preference for a G following the target cytosine and modestly disfavoring a T or mC preceding the target cytosine.

Both 5mC hydroxylase activity and CXXC domain-mediated specific DNA binding are required for Tet3 function in target gene regulation and embryonic development

Having defined the regulation of 5mC/5hmC status at Tet3 target gene promoters by the 5mC hydroxylase activity and the unique DNA binding properties of the Tet3 CXXC domain, we next determined the role of these two functional domains in Tet3 function. We generated wild type (*xtTet3*), CXXC-deleted (*xtTet3ΔCXXC*), His90-to-Ala (*xtTet3Δ90A*) and iron binding site-disrupted catalytically inactive (*xtTet3 ΔHD*) *xtTet3* expression constructs (Figure 6A). We confirm that both the wild type and mutant xtTet3 proteins are properly expressed and localized to the nucleus (Figures S7A and S7B), and that the 5mC hydroxylase activity is retained in xtTet3ΔCXXC and xtTet3H90A, but not in xtTet3ΔHD (Figure S7B). We also performed ChIP-qPCR assays to compare the occupancy of wild type and mutant Tet3 proteins at target gene promoters. Both the CXXC domain deletion and H90A point mutation abolish the occupancy of mutant proteins at target gene promoters (Figures 6B and S7C), suggesting an essential role of the CXXC domain-mediated DNA binding in Tet3 targeting to specific genes. Importantly, although both xtTet3 and xtTet3ΔHD exhibit similar occupancy, co-injection of *xtTet3*, but not *xtTet3 ΔHD*, significantly rescues the decreased 5hmC levels caused by Tet3 depletion at target gene promoters including the *rx* gene promoter (Figure S7D), suggesting that the enzymatic activity of Tet3 is primarily responsible for the dynamic regulation of 5hmC at its target genes.

We next used these constructs to determine which domain(s) are critical for Tet3 functions *in vivo*. We first employed a functional domain rescue approach to interrogate the role of Tet3 catalytic domain using *pax6* gene expression as a readout. The expression of *pax6* is completely rescued in 79% of *xtTet3* co-injected embryos, while the complete rescue effect of *xtTet3 ΔHD* co-injection only reaches 37%, significantly lower than that of *xtTet3* co-injection (*P*<0.01) (Figure 6C). These results are corroborated by phenotypic rescues. Only 25% of embryos are completely rescued by *xtTet3 ΔHD* co-injection, again significantly lower than that of *xtTet3* co-injection, in which 62% are completely rescued (*P*<0.01) (Figure 6D). Thus, these data suggest that the 5mC hydroxylase activity of Tet3 not only controls the dynamics of 5mC/5hmC at target gene promoters but is also required for target gene regulation and biological function in early embryonic development. Next, we employed the same strategy to examine the importance of the CXXC domain/His90 mediated specific DNA binding in Tet3 function. Strikingly, *xtTet3 H90A* or *xtTet3 ΔCXXC* co-injection shows no rescue effects on the inhibited *pax6* expression or developmental phenotypes caused by Tet3 depletion (Figure 6C and 6D), highlighting an

essential role of the CXXC domain-mediated DNA binding in Tet3 function in target gene regulation and embryonic development.

DISCUSSION

In this study, we report that *Xenopus* Tet3 is a new class of transcription factor playing an essential role in early eye and neural development by directly regulating a set of key genes critical for these developmental processes. We also uncover that the dynamic regulation of 5mC/5hmC status at target gene promoters is an important mechanism underlying Tet3-mediated target gene regulation during embryonic development. The Tet3 CXXC domain-mediated specific DNA binding is essential for targeting Tet3 to its target genes, thus, providing another layer of regulation on the transcription of Tet3 target genes. Our findings support a molecular model of Tet3 action in target gene regulation that involves Tet3 binding to unmethylated cytosines (with a slight preference for CpG content) at target gene promoters through its CXXC domain and its intrinsic 5mC hydroxylase activity converting adjacent 5mC to 5hmC, which is an intermediate for further DNA demethylation (Figure 6E). Recruitment to a specific gene promoter and subsequent conversion of 5mC to 5hmC cooperatively activate the expression of Tet3 target genes, including those identified key developmental genes, to ensure normal and precisely regulated embryogenesis. Inactivation of either one of these functional domains therefore will have adverse impact on Tet3 function. Noteworthy, our data also show that the catalytically inactive Tet3 mutant retains partial rescue effects *in vivo*, suggesting potential enzymatic activity-independent mechanisms, such as recruiting or interacting with other transcription factors, may also contribute to the full function of Tet3 in gene regulation (Figure 6E). Indeed, TET3 forms a stable complex with several critical transcription factors and histone modifiers (data not shown). It has also been shown that an enzymatic activity-independent mechanism is involved in gene regulation by Tet1 in mouse ES cells (Williams et al., 2011). Moreover, it has been suggested that 5hmC is a stable epigenetic mark that participates in regulating gene expression through unidentified mechanisms such as recruiting unidentified 5hmC specific "readers" (Branco et al., 2011; Matarese et al., 2011). Thus, this model does not exclude the possibility that Tet3 and 5hmC may participate in target gene regulation through this yet to be identified regulatory circuit.

Unlike well-characterized DNA methylation (5mC) and DNMTs, the mechanism by which 5hmC and the enzymatic activity of TET proteins contribute to gene regulation has been elusive. Our findings here offer the first biological model demonstrating that the dynamic regulation of 5mC/5hmC by the Tet family of 5mC hydroxylases has an important role in gene regulation during early eye and neural development. We show that Tet3 is an active 5mC hydroxylase highly expressed in the region of the developing brain, eye and spinal cord in *Xenopus* embryos. As exemplified by *rx* and *pax6*, our data clearly show that Tet3 target genes have significant alterations in 5hmC status at their promoters after Tet3 depletion. Importantly, 5hmC level alterations and dysregulated gene expression can be completely rescued by wild type Tet3. Yet the rescue effect by the catalytically inactive Tet3 mutant is significantly impaired, highlighting the importance of 5mC hydroxylase activity in Tet3 function. Furthermore, consistent with the eye developmental phenotypes in *Xenopus*, we find that homozygous deletion of the Tet3 catalytic domain in pure B6 genetic background mice (Gu et al., 2011) results in eye-related phenotypes including the Eyelid Open at Birth (EOB) phenotype (data not shown), suggesting an evolutionarily conserved function in eye development from vertebrate to mammals and the critical role of the 5mC hydroxylase activity in Tet3 function. Taken together, our study clearly indicates that precise regulation of the dynamics of DNA modification status at specific gene loci by Tet3-mediated conversion of 5mC to 5hmC is an important and conserved epigenetic mechanism for target gene regulation.

How epigenetic enzymes are targeted and/or confined to their functional sites is a fundamental question in understanding the mechanism underlying epigenetic transcription regulation. In this study, our biochemical characterizations reveal that the Tet3 CXXC domain binds to DNA in a cytosine-dependent manner with a slight preference for CpG dinucleotides, distinct from the CpG-dependent binding of other well-characterized CXXC domains. We also notice that the DNA binding property of the Tet3 CXXC domain is different from that of the TET1 CXXC domain (Xu et al., 2011b), suggesting that Tet1 and Tet3 may have different functions in gene regulation besides their distinct expression patterns during early embryonic development (Tan and Shi, 2012). Moreover, our genome-wide mapping and *de novo* motif analyses of the TET3 CXXC domain binding sites strongly suggest that the TET3 CXXC domain selectively binds to several consensus sequences, therefore offering a mechanism for targeting Tet3 to its target genes. Finally, the crystal structure analysis of the Tet3 CXXC domain in complex with DNA provides the following structural mechanisms underlying its novel DNA binding ability and specificity. a) It reveals that the major binding force of the Tet3 CXXC domain to CmCGG DNA is through the specific interaction between residue His90 and the unmethylated cytosine, which explains why the Tet3 CXXC domain can still bind to partially-methylated DNA, as long as there is an unmethylated cytosine available in the target sequence; b) it explains the slight binding preference of the Tet3 CXXC domain for CpG compared to CpH (H= A, T, C or 5mC) and the modest disfavor for a T or 5mC preceding the target C, suggesting that sequences flanking the target C contribute to the overall binding affinity and specificity of the Tet3 CXXC domain.

The importance of this CXXC domain-mediated specific DNA binding activity of Tet3 is further demonstrated in our functional studies using the Tet3 mutants with the CXXC domain deletion (ΔCXXC) or H90A point mutation. These mutants lose the specific association with Tet3 target gene promoters and functional rescue abilities *in vivo*. Taken together, our study reports a novel DNA binding property and functionality of the CXXC domain which is essential for Tet3 function both in gene regulation and embryonic development. It not only provides a key mechanistic layer of Tet3-mediated target gene regulation but also significantly advances our current understanding of the molecular and biological function of the Tet family CXXC domains. Noteworthy, even though the CXXC domain is essential for targeting Tet3 to specific genomic regions, we do not exclude the involvement of other potential cellular mechanisms for targeting or recruiting Tet3 to its functional sites. We favor the hypothesis that Tet3 is likely in complex with many other cellular factors including sequence-specific DNA binding transcription factors and cofactors, to execute its molecular, cellular and biological functions. Therefore, the associated transcription factors may in part coordinate with the Tet3 CXXC domain and enzymatic activity and contribute to the overall mechanism of action of Tet3 in gene regulation and embryonic development.

Experimental Procedures

Embryo manipulation and microinjection

X. laevis eggs were artificially fertilized using testis homogenate and cultivated in 0.1×MMR as previously described (Sakano et al., 2010). Capped synthetic mRNAs were generated by *in vitro* transcription with sp6 polymerase. Embryos were transferred to 3% Ficoll 400 in 0.1×MMR and injected embryos were cultured in 0.1×MMR until the desired stage. For phenotype experiments, 80 ng of *xlTet3* MOs (40ng of each *xlTet3* MO) or control MO were injected into two dorsal blastomeres of 4-cell stage embryos. For phenotype rescue experiments, 1ng of mRNA was co-injected with 80ng of *xlTet3* MOs into two dorsal blastomeres of 4-cell stage embryos. For *pax6*, *rx*, *six3*, *sox2*, *otx2*, *sox9*, *snail*, *tubb2b* and *ngn2 in situ* hybridization experiments, 160 ng of *xlTet3* MOs or control MO

were injected into one dorsal blastomere of each 4-cell stage embryo. Since the expression of *shh* and *ptc-1* is on the midline and thin, for *shh* and *ptc-1 in situ* hybridization experiments, two dorsal blastomeres of each 4-cell stage embryo were injected with 160 ng of *xlTet3* MOs or control MO. In all injection studies, 200 pg of *nucβ-gal* RNA were co-injected as the injection tracer.

Crystallization

Each pair of single-stranded DNAs was mixed with a molar ratio of 1:1 and annealed to form double-stranded DNAs. Before co-crystallization, purified xtTet3 CXXC protein was mixed with different DNAs in a molar ratio of 1:1.2. Crystals of xtTet3 CXXC in complex with ACGT DNA (GCCAACGTTGGC) were obtained using sitting drop vapour diffusion. 1.0 μl of complex was mixed with 1.0 μl of well solution containing 0.1M HEPES (pH 7.5), 0.2M NaCl, 30% PEG 1500, against 800 μl of reservoir buffer at 18 °C. Crystals grow to a mountable size in three days. Crystals of xtTet3 CXXC in complex with CmCGG DNA (GCCACmCGGTGGC) were obtained in a similar way in the buffer containing 0.1 M HEPES, pH 7.5, 0.2 M NaCl, 30% PEG 1500, 5% MPD. Both crystals were flash-frozen in liquid nitrogen directly without cryoprotectant.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES

Allen MD, Grummitt CG, Hilcenko C, Min SY, Tonkin LM, Johnson CM, Freund SM, Bycroft M, Warren AJ. Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. The EMBO journal. 2006; 25:4503–4512. [PubMed: 16990798]

Bestor TH, Coxon A. Cytosine methylation: the pros and cons of DNA methylation. Curr Biol. 1993; 3:384–386. [PubMed: 15335739]

Bird AP. CpG-rich islands and the function of DNA methylation. Nature. 1986; 321:209–213. [PubMed: 2423876]

Branco MR, Ficz G, Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat Rev Genet. 2011; 13:7–13. [PubMed: 22083101]

Delhommeau F, Dupont S, Della Valle V, James C, Trannoy S, Masse A, Kosmider O, Le Couedic JP, Robert F, Alberdi A, et al. Mutation in TET2 in myeloid cancers. N Engl J Med. 2009; 360:2289–2301. [PubMed: 19474426]

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature. 2011; 473:398–402. [PubMed: 21460836]

Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, Koch S, Bruckl T, Biel M, Carell T. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS One. 2010; 5:e15367. [PubMed: 21203455]

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, et al. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature. 2011; 477:606–610. [PubMed: 21892189]

He Y-F, Li B-Z, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, et al. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science. 2011; 333:1303–1307. [PubMed: 21817016]

Hsu C-H, Peng K-L, Kang M-L, Chen Y-R, Yang Y-C, Tsai C-H, Chu C-S, Jeng Y-M, Chen Y-T, Lin F-M, et al. TET1 Suppresses Cancer Invasion by Activating the Tissue Inhibitors of Metalloproteinases. Cell Reports. 2012; 2:568–579. [PubMed: 22999938]

Iqbal K, Jin S-G, Pfeifer GP, Szabo PE. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proceedings of the National Academy of Sciences. 2011

Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 2010; 466:1129–1133. [PubMed: 20639862]

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science. 2011; 333:1300–1303. [PubMed: 21778364]

Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G, et al. Tet1 and Tet2 Regulate 5-Hydroxymethylcytosine Production and Cell Lineage Specification in Mouse Embryonic Stem Cells. Cell Stem Cell. 2011; 8:200–213. [PubMed: 21295276]

Langemeijer SM, Kuiper RP, Berends M, Knops R, Aslanyan MG, Massop M, Stevens-Linders E, van Hoogen P, van Kessel AG, Raymakers RA, et al. Acquired mutations in TET2 are common in myelodysplastic syndromes. Nat Genet. 2009; 41:838–842. [PubMed: 19483684]

Lian CG, Xu Y, Ceol C, Wu F, Larson A, Dresser K, Xu W, Tan L, Hu Y, Zhan Q, et al. Loss of 5-Hydroxymethylcytosine Is an Epigenetic Hallmark of Melanoma. Cell. 2012; 150:1135–1146. [PubMed: 22980977]

Matarese F, Carrillo-de Santa Pau E, Stunnenberg HG. 5-Hydroxymethylcytosine: a new kid on the epigenetic block? Mol Syst Biol. 2011; 7

Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, Figueroa Maria E, Vasanthakumar A, Patel J, Zhao X, et al. Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. Cancer Cell. 2011; 20:11–24. [PubMed: 21723200]

Pradhan M, Estève P-O, Chin HG, Samaranayke M, Kim G-D, Pradhan S. CXXC Domain of Human DNMT1 Is Essential for Enzymatic Activity. Biochemistry. 2008; 47:10000–10009. [PubMed: 18754681]

Quivoron C, Couronné L, Della Valle V, Lopez Cécile K, Plo I, Wagner-Ballon O, Do Cruzeiro M, Delhommeau F, Arnulf B, Stern M-H, et al. TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. Cancer Cell. 2011; 20:25–38. [PubMed: 21723201]

Reik W, Dean W, Walter J. Epigenetic Reprogramming in Mammalian Development. Science. 2001; 293:1089–1093. [PubMed: 11498579]

Sakano D, Kato A, Parikh N, McKnight K, Terry D, Stefanovic B, Kato Y. BCL6 canalizes Notch-dependent transcription, excluding Mastermind-like1 from selected target genes during left-right patterning. Dev Cell. 2010; 18:450–462. [PubMed: 20230751]

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009; 324:930–935. [PubMed: 19372391]

Tan L, Shi YG. Tet family proteins and 5-hydroxymethylcytosine in development and disease. Development. 2012; 139:1895–1902. [PubMed: 22569552]

Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PAC, Rappsilber J, Helin K. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature. 2011; 473:343–348. [PubMed: 21490601]

Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M, Arand J, Nakano T, Reik W, Walter J. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun. 2011; 2:241. [PubMed: 21407207]

Wu H, D'Alessio AC, Ito S, Xia K, Wang Z, Cui K, Zhao K, Eve Sun Y, Zhang Y. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature. 2011; 473:389–393. [PubMed: 21451524]

Xu C, Bian C, Lam R, Dong A, Min J. The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain. Nat Commun. 2011a; 2:227. [PubMed: 21407193]

Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, Barbera AJ, Zheng L, Zhang H, Huang S, et al. Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. Mol Cell. 2011b; 42:451–464. [PubMed: 21514197]

Yu M, Hon Gary C, Szulwach Keith E, Song C-X, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B, et al. Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. Cell. 2012; 149:1368–1380. [PubMed: 22608086]

Highlights

Tet3 regulates key genes essential for eye and neural development in *Xenopus*

Gene regulation by Tet3 partly through modulating 5mC/5hmC status at the promoter

Tet3 CXXC domain processes novel DNA binding properties critical for Tet3 targeting

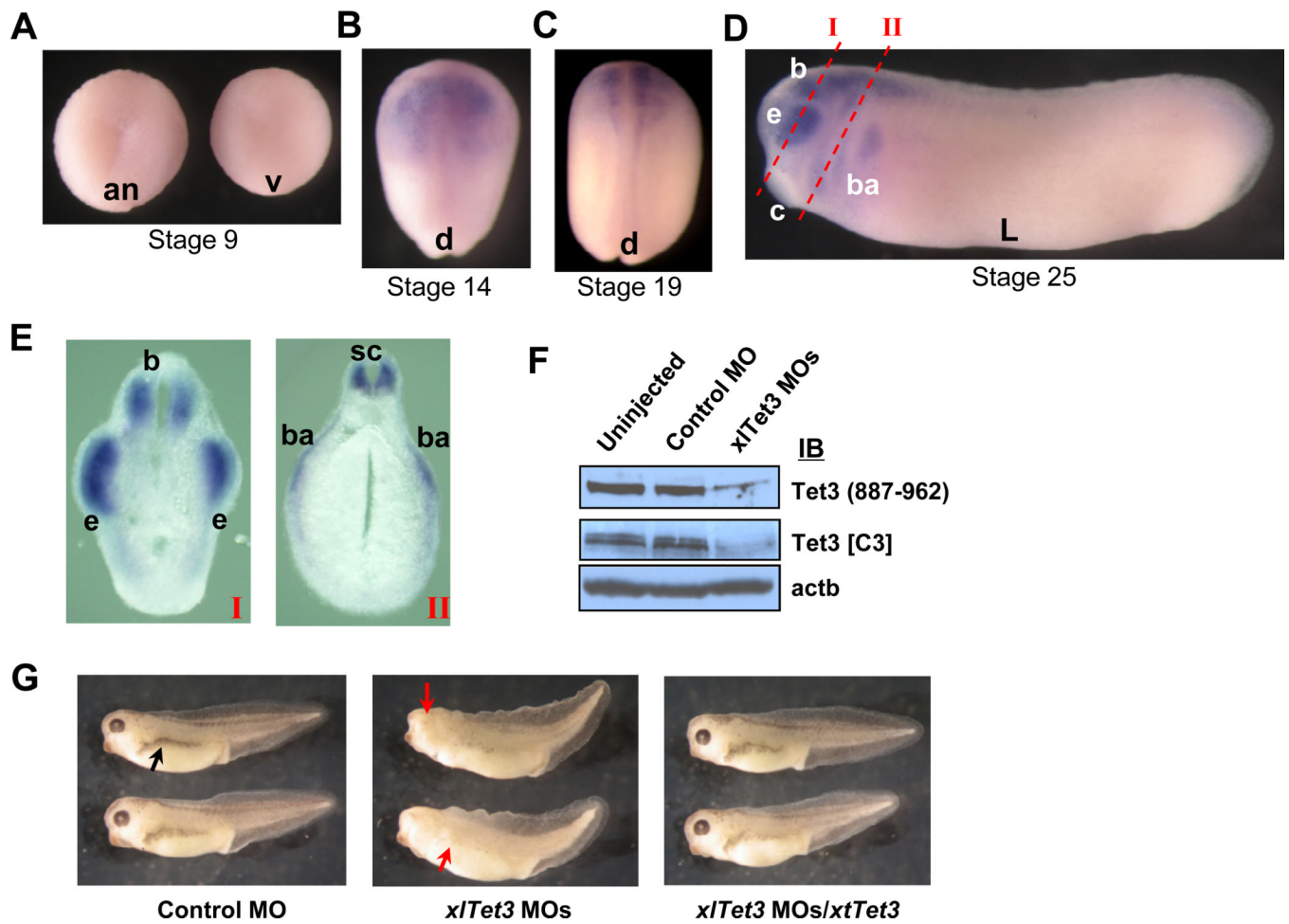Both the CXXC domain and hydroxylase activity are required for Tet3 function

Figure 1. Tet3 is important for early eye and neural development

(A–E) Spatial expression profile of *xlTet3* by *in situ* hybridization at stage 9 (A), 14 (B), 19 (C) and 25 (D and E). The sites of sections I and II in panel E are noted by red dashed lines in panel D. Animal view (an), vegetal view (v), dorsal view (d), lateral view (L), brain (b), eye (e), cement gland (c), branchial arches (ba) and spinal cord (sc).

(F) Western blot showing depletion of endogenous Tet3 protein by *xlTet3* MOs in stage 14 embryos.

(G) Developmental defects in stage 35 embryos caused by Tet3 depletion. Small head, eyeless and missing pigmentation phenotypes in *xlTet3* MOs-injected embryos are noted by red arrows, and the normal pigmentation in control embryos is noted by a black arrow.
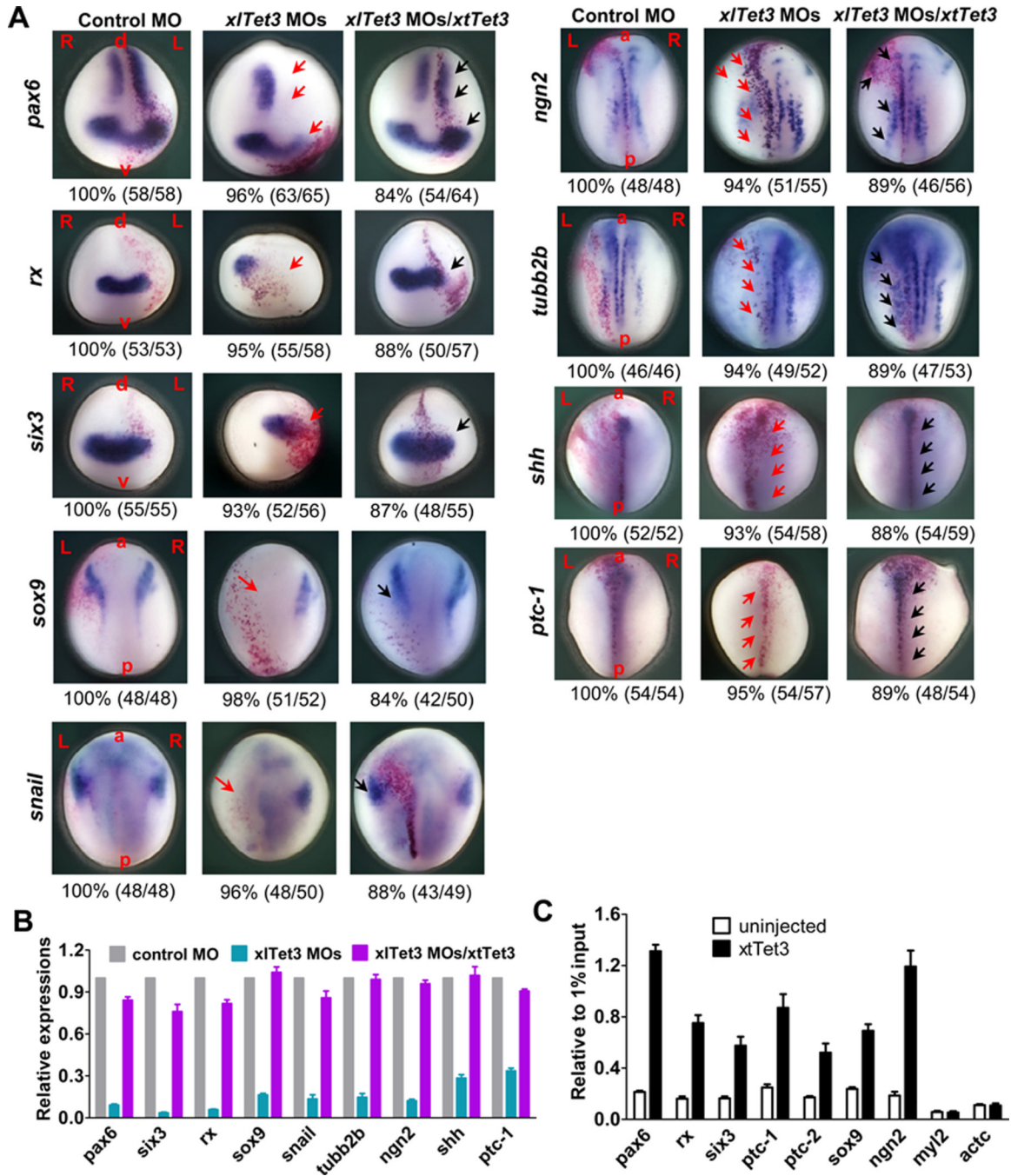
See also Figure S1.

Figure 2. Tet3 directly regulates key developmental genes

(A) Expression level changes of developmental genes resulting from Tet3 depletion as shown by *in situ* hybridization at stage 14. Red arrows show the inhibited expression by *xlTet3* MOs injection, while black arrows show rescued expression by *xtTet3* co-injection. L: left; R: right; a: anterior; p: posterior; d: dorsal; v: ventral. The red dots are injection tracer by *β-gal* staining.

(B) RT-qPCR confirms the differentially expressed genes after Tet3 depletion at stage 14. Control MO, *xlTet3* MOs or *xlTet3* MOs/*xtTet3* was injected into two dorsal blastomeres at 4-cell stage. Relative gene expression was normalized to *odc*. Results are shown as mean ± SD (n =3).

(C) Tet3 occupancy at target gene promoters by ChIP-qPCR assay. Data are presented as mean ± SEM (n=3).
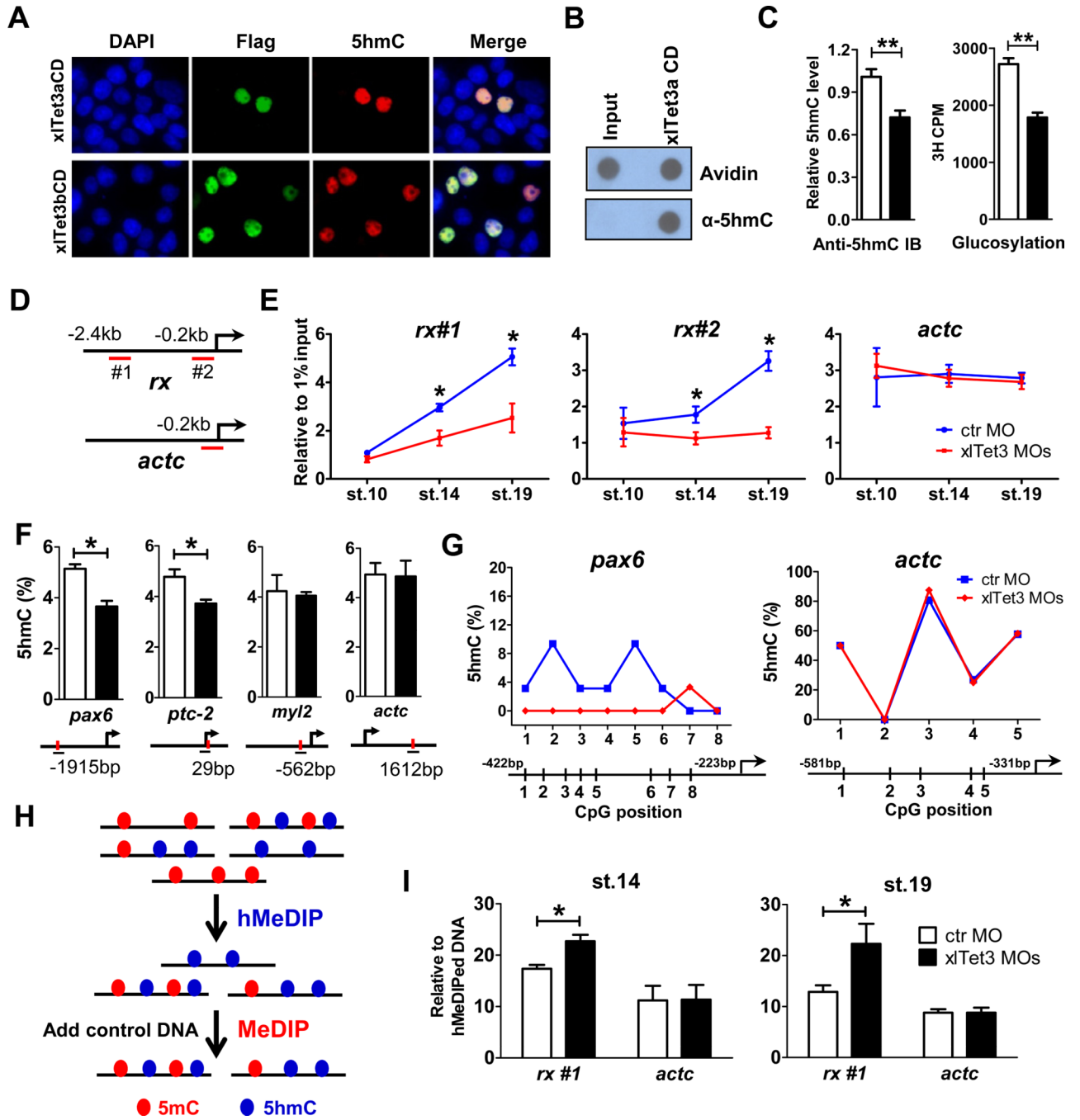See also Figure S2.

Figure 3. Tet3 is an active 5mC hydroxylase regulating the 5mC/5hmC status at target gene promoters

(A) xlTet3 CD is sufficient to convert 5mC to 5hmC in HEK293T cells by immunofluorescence analysis. Flag-tagged xlTet3 CD protein was detected using Flag antibody.

(B) xlTet3a CD protein converts 5mC to 5hmC *in vitro* by dot-blot assay. Avidin-HRP is used to detect total biotin-labeled DNA, showing equal loading.

(C) Tet3 depletion results in globally decreased 5hmC in stage 14 embryos by dot-blot (left) and 5hmC glucosylation (right) assays. Open bar: control MO, filled bar: *xlTet3* MOs. Data are presented as mean ± SD (n=3). **P<0.01.

(D–E) hMeDIP-qPCR to detect dynamic 5hmC level changes in stage 10, 14 and 19 embryos. The targeting region for each primer set is underlined in panel D. Results are shown as mean ± SD (n=3) in panel E. *$P<0.05$.

(F) Site-specific 5hmC level changes by Tet3 depletion in stage 14 embryos using the *EpiMark* 5mC/5hmC analysis kit. Open bar: control MO, filled bar: *xlTet3* MOs. Red dot indicates MspI/HapII recognition site and each PCR amplified region is underlined. Arrow denotes promoter orientation. Data are shown as mean ± SD (n=3). *$P<0.05$.

(G) TAB-seq analyses of 5hmC status at the promoter of *pax6* (left) and *actc* (right) in stage 14 embryos. The average percent at each CpG site is derived from sequencing of 30–32 clones for *pax6* promoter and 24–26 clones for *actc* promoter.

(H) Schematic diagram of hMeDIP-MeDIP strategy.

(I) hMeDIP-MeDIP qPCR to detect 5mC level changes after Tet3 depletion in stage 14 and 19 embryos. Data are shown as mean ± SD (n=3). *$P<0.05$. The targeting region for each primer set is shown in panel D.
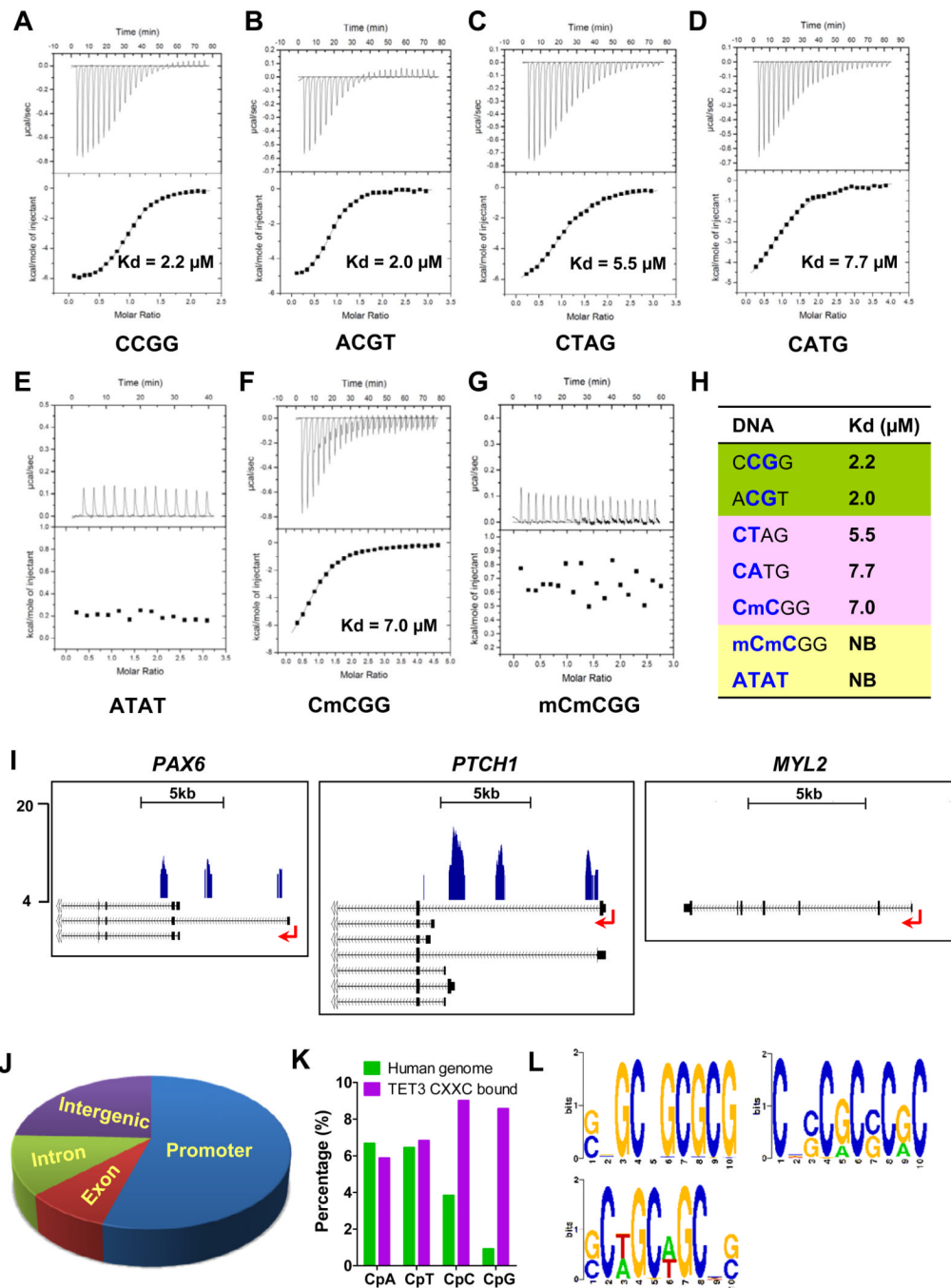
See also Figure S3.

Figure 4. The unique DNA binding properties of the Tet3 CXXC domain

(A–H) Binding affinities of the Tet3 CXXC domain to various DNA oligos by ITC assays. The sequence of the central four nucleotides of each double-stranded DNA probe is shown under the corresponding panel. Detailed sequence information for all DNA oligos used in this study is listed in Table S1. NB: no binding.

(I) Representative TET3 CXXC GST pull down-sequencing results. Arrow denotes promoter orientation.

(J) Genomic distribution of the TET3 CXXC-bound loci. Promoter is defined as −2 kb to +2 kb relative to transcription start site (TSS).

(K) The percentage of CpA, CpT, CpC and CpG in human genome and the TET3 CXXC-bound loci.

(L) DNA motifs that are enriched in the TET3 CXXC-bound loci.
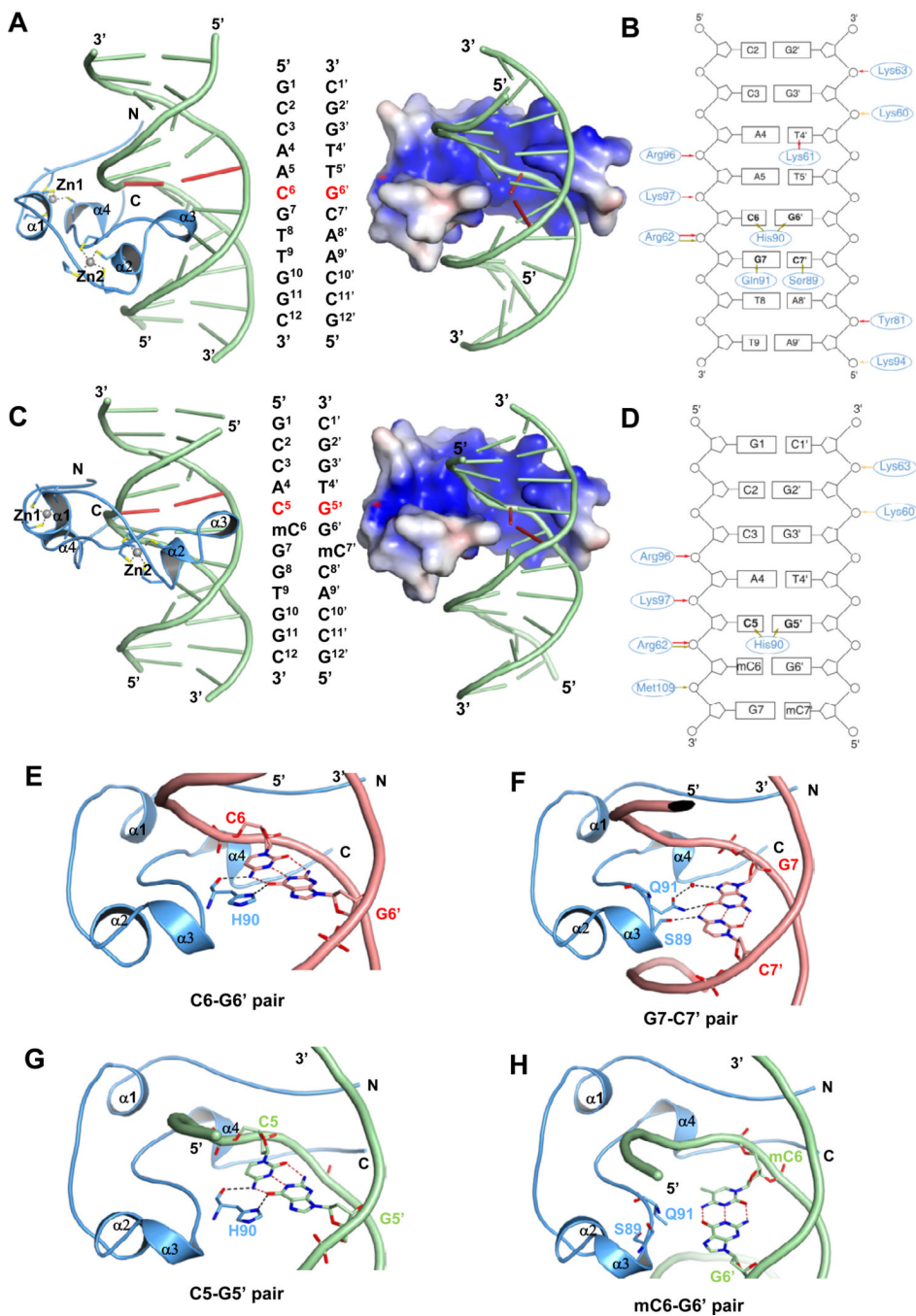
See also Figure S4 and Table S1.

Figure 5. The Tet3 CXXC domain specifically recognizes cytosine through a conserved residue His90

(A, C) Crystal structures of the Tet3 CXXC domain in complex with ACGT DNA (5'-GCCA**ACGT**TGGC-3') (A) or CmCGG DNA (5'-GCCA**CmCGG**TGGC-3') (C) in cartoon (left) and electrostatic representations (right), respectively. The double-stranded DNA sequence is shown in the middle of each corresponding panel.

(B, D) Detailed interactions between the Tet3 CXXC domain and ACGT DNA (B) or between the Tet3 CXXC domain and CmCGG DNA (D). Red arrow: salt bridge interaction; olive arrow: hydrogen bond interaction; yellow arrow: electrostatic interaction.

(E–F) Detailed interactions between the Tet3 CXXC domain (blue cartoon) and the target CG pair (E) or the following GC pair (F) in the ACGT DNA (red cartoon).

(G–H) Detailed interactions between the Tet3 CXXC domain (blue cartoon) and the target CG pair (G) or the following mCG pairs (H) in the CmCGG DNA (green cartoon).
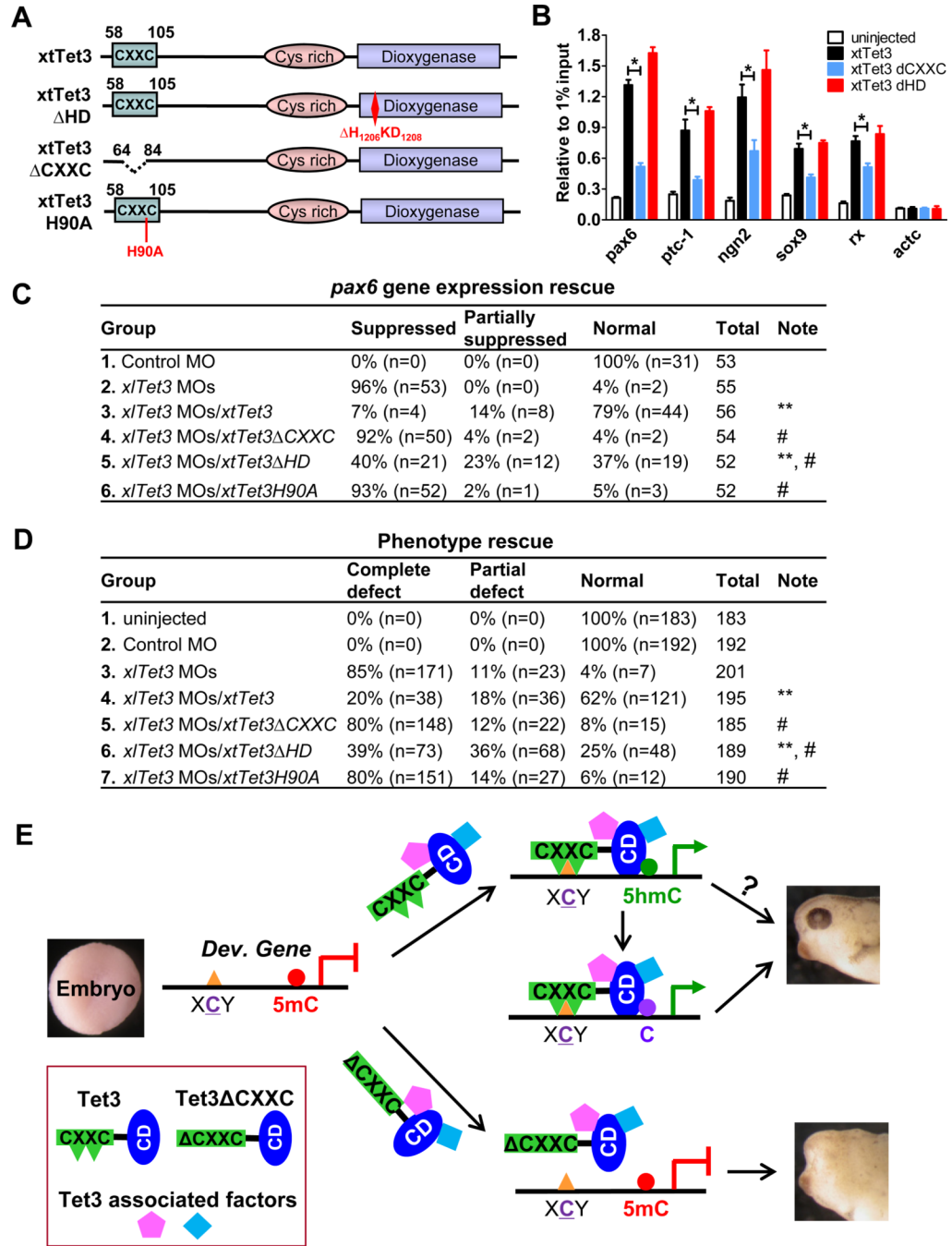
See also Figures S5 and S6.

Figure 6. Both 5mC hydroxylase activity and the CXXC domain are important for Tet3 function

(A) Schematic representation of xtTet3 mutants.

(B) The CXXC domain deletion disrupts Tet3 occupancy at specific gene promoters by ChIPqPCR assay. Data are presented as mean ± SEM (n=3). *P<0.05.

(C) Summarized results of five independent *pax6* expression rescue experiments. "Suppressed" means significantly suppressed *pax6* expression in posterior and anterior neural plates; "Partially suppressed" means *pax6* expression is detected but not intact in posterior and anterior neural plates; "Normal" means intact *pax6* expression in posterior and anterior neural plates. **P<0.01 compared to *xlTet3* MOs, # P<0.01 compared to *xlTet3* MOs*/xtTet3*.

(D) Summarized results of five independent phenotypic rescue experiments. "Partial defect" means mild abnormal head structure, small eyes or one eye; "Complete defect" means abnormal head structure and no eye. **P<0.01 compared to *xlTet3* MOs, # *P*<0.01 compared to *xlTet3* MOs*/xtTet3*.

(E) A model of Tet3 action in gene transcription regulation. The Tet3 CXXC domain specifically binds to unmodified cytosine (underlined)-containing sequence motifs with a slight preference for G at "Y" position and a mild disfavor for T or 5mC at "X" position, targeting Tet3 to the promoter of target developmental genes. Then, the 5mC hydroxylase activity of Tet3 converts adjacent 5mC to 5hmC, an intermediate for further DNA demethylation, thus activates the gene expression. CD: catalytic domain. Please refer to the related text for more details.

See also Figure S7.

Table 1

Data collection and refinement statistics

	xtTet3 CXXC-GCCAACGTTGGC	xtTet3 CXXC-GCCACmCGGTGGC
PDB code	4HP3	4HP1
Space Group	C2	C2
Cell dimensions		
a, b, c (Å)	69.9, 39.5, 54.1	71.0, 39.5, 57.6
α, β, γ (°)	90.0, 99.9, 90.0	90.0, 91.0, 90.0
Resolution range (Å)	34.42-2.05 (2.16-2.05)	35.51-2.25 (2.37-2.25)
Number of unique HKLs	9281 (1345)	7738 (1110)
Completeness (%)	99.8 (99.8)	99.9 (100.0)
Friedel Redundancy	3.7 (3.7)	4.1 (4.2)
Rsym (%)	5.3 (66.9)	4.6 (48.7)
$<I/\sigma(I)>$	12.6 (2.1)	14.7 (2.8)
Resolution limits (Å)	30.00-2.05	35.00-2.25
Number of unique HKLs work / free	8829 / 452	7385 / 351
R_{work} / R_{free} (%)	21.6 / 24.3	22.0 / 25.0
Number of atoms / $<B>$ (Å^2)	898 / 54.8	886 / 73.9
DNA	486 / 58.2	487 / 78.8
Protein	384 / 51.7	387 / 68.5
Zn^{2+}	2 / 33.2	2 / 46.8
RMSD bonds (Å) / angles (°)	0.013 / 1.4	0.012 / 1.4
Ramachandran plot favored residues, no outliers (Lovell et al., 2003)	47 of 49	48 of 49

RMSD: Root Mean Squared Deviation.

Average B-factors calculated with MOLEMAN (G.J. Kleywegt, Uppsala Univ.) The highest resolution shell is shown in parentheses.

Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. (2003). Structure validation by Cα geometry: φ,ψ and Cβ deviation. Proteins: Structure, Function, and Bioinformatics *50*, 437-450.