

A mutation disrupting TET1 activity alters the expression of genes critical for neural tube closure in the *tuft* mouse

Keith S. K. Fong¹, Robert B. Hufnagel^{4,5}, Vedbar S. Khadka², Michael J. Corley³, Alika K. Maunakea³, Ben Fogelgren¹, Zubair M. Ahmed^{4,6}, Scott Lozanoff¹

¹Department of Anatomy, Biochemistry, and Physiology, ²Office of Biostatistics and Quantitative Health Sciences, ³Epigenomics Research Program, Department of Native Hawaiian Health, John A. Burns School of Medicine, University of Hawai'i, Honolulu, HI, 96813, ⁴Department of Pediatrics, Division of Human Genetics, Cincinnati Children's Hospital, College of Medicine, University of Cincinnati, 3333 Burnet Ave, ML 7003, Cincinnati, OH 45229, ⁵ Unit on Pediatric, Development & Genetic Ophthalmology, Ophthalmic Genetics and Visual Function Branch, National Eye Institute, National Institutes of Health, Bethesda, MD 20892, ⁶ Department of Otorhinolaryngology Head and Neck Surgery, School of Medicine, University of Maryland, BioPark Bldg1, 800 West Baltimore Street, Room 404, Baltimore, MD 21201

Corresponding Author: Keith S. K. Fong (ksfong@hawaii.edu)

Keywords: anterior cranial cephalocele, midfacial cleft, neural tube defect, encephalocele, exencephaly, anencephaly, epigenetic

Summary statement:

We propose an epigenetic mechanism establishing the regulation of genes critical for neural tube closure. This mechanism may be a novel target for resolving such birth defects and associated disorders.

Abstract

Genetic variations affecting neural tube closure along the head result in malformations of the face and brain. Neural tube defects (NTDs) are among the most common birth defects in humans. We previously reported a mouse mutant called *tuft* that arose spontaneously in our wildtype 3H1 colony. Adult *tuft* mice present midline craniofacial malformations with or without an anterior cephalocele. In addition, affected embryos presented neural tube closure defects resulting in insufficient closure of the anterior neuropore or exencephaly. Here, through whole genome sequencing, we identified a nonsense mutation in the *Tet1* gene encoding a methylcytosine dioxygenase co-segregating with the *tuft* phenotype. This mutation resulted in premature termination disrupting the catalytic domain involved in the demethylation of cytosine. We detected a significant loss of TET enzyme activity in *tuft* embryo heads homozygous for the mutation with NTDs. RNA-Seq transcriptome analysis indicated multiple gene pathways associated with neural tube closure were dysregulated in *tuft* embryo heads. Among them, the expression of *Cecr2*, *Epha7* and *Grhl2* were significantly reduced in some embryos presenting neural tube closure defects while one or more components of the non-canonical WNT signaling pathway mediating planar cell polarity and convergent extension were affected in others. We further show that the recombinant mutant TET1 protein was capable of entering the nucleus and affect the expression of endogenous *Grhl2* in IMCD-3 cells. These results indicate that TET1 is an epigenetic determinant for regulating genes critical to closure of the anterior neural tube and having implications to craniofacial development as presented by the *tuft* mouse.

Introduction

Neural tube defects (NTDs), such as anencephaly, encephalocele, and spina bifida are among the most common birth defects in humans with estimates of over 2650 annual cases in the U.S. (CDC). Despite significant reductions, largely due to prenatal supplementation of folic acid (FA) over the past 18 years in the US (mandatory fortification authorized in 1996, but not fully implemented until 1998), NTDs remain among the most common serious birth defects. Furthermore, we still do not understand how FA prevents NTDs (reviewed in Copp et al., 2013). Some studies indicate that FA may have an adverse effect, depending on the individual's genetic background (Marean et al., 2011). There are over 450 loci in mice documented in the Mouse Genome Informatics (MGI) database that are associated with neural tube closure defects. This underscores the complexity of neurulation and the different ways closure defects may arise. Thus, alternative means for treatment and preventive care are necessary to address a broader spectrum of NTDs.

Strides have been made in understanding fundamental mechanisms of neural tube closure using various animal model systems. Gene mutations affecting components mediating planar cell polarity and convergent extension in mice associate with craniorachischisis, a severe and rare type of NTD in humans (Wallingford 2012; De Marco et al., 2014; Murdoch et al., 2014). Environmental factors affecting epigenetic regulation of genes associated with neural tube closure have also been examined. Maternal diabetes and obesity, for example, are risk factors for NTDs. One of the largest studies to date indicated that the odds ratio of encephaloceles was more than 3-fold higher in infants born to diabetic mothers while anencephaly was almost 2-fold more common (Garne et al., 2012). Studies report the expression of genes associated with neural tube closure was significantly reduced in mouse models for diabetes by altering the distribution of modified histones (Salbaum and Kappen 2010, 2012; Zhang et al., 2013) or methylation patterns of gene loci in the embryos of afflicted mothers (Wei and Loeken 2014; Wang et al., 2015). Furthermore, the embryo's chromatin landscape that primes gene expression could also be affected by the maternal diet (Salbaum and Kappen 2010). Therefore, understanding the distribution of epigenetic marks in specific

disease conditions will help us identify candidate genes that are potentially affected and work towards approaches to prevent NTDs.

DNA methylation is a dynamic epigenetic mechanism for regulating gene expression that is conserved in diverse species. DNA methyl transferases (DNMTs) catalyze the methylation of 5-cytosine of CpG dinucleotides to form 5-methylcytosine (5mC) throughout the mammalian genome (reviewed in Smith and Meissner 2013). The methylated state of CpGs has pivotal roles in influencing gene transcription during embryonic development, imprinting, X-chromosome inactivation and cancer. Methylation states are dynamically reversed by TET (Ten-Eleven Translocation) enzymes. TET enzymes catalyze the conversion of 5mC to 5-hydroxymethylcytosine (5hmC), thus subsequently producing a demethylated state of cytosines and epigenetically determine which genes can be expressed in a given cell or tissue (Tahiliani et al., 2009). Recent findings indicate that 5hmC-mediated epigenetic regulation is critical to neurodevelopment, aging and human diseases (Tan and Shi 2012; Kaas et al., 2013; Pastor et al., 2013; Tsagaratou and Rao 2013; Yamaguchi et al., 2013).

There are three TET proteins encoded by separate genes (*Tet1-3*) in mammals that exhibit the same catalytic activity but are distinct in their expression levels and distribution at particular stages of development (Globisch et al., 2010; Ito et al., 2010; Szwagierczak et al., 2010; Gu et al., 2011; Yamaguchi et al., 2012). TET3 is primarily responsible for the global erasure of 5mCs in the paternal genome upon fertilization of the oocyte (Iqbal et al., 2011; Wossidlo et al., 2011) and in later stages of murine primordial germ cell (PGC) development to re-establish a pluripotent state (reviewed in Wu and Zhang 2014). Without TET3, mouse embryos died during early embryogenesis (Gu et al., 2011). TET1 and TET2, however, appear to have more specific roles in establishing gene transcriptional programs by defining the genomic landscape in developing cell populations.

Gene knockouts for either *Tet1* or *Tet2* suggest they can partially compensate the function of each other. Mice deficient in TET1 protein alone appeared normal with no lack of predicted progeny despite reductions in the level of 5hmC (Dawlaty et al., 2011). However, about a third of mice homozygous for the *Tet1*^{-/-} knock out were

smaller than normal. Some of them eventually grew to normal size and weight following one month in age. Mice deficient in TET2 alone were similarly viable, but about a third developed malignancies resembling myeloid leukemia, indicating its significance in hematopoiesis (Ko et al., 2011; Li et al., 2011). Double knock out mice deficient in both TET1 and TET2 developed to term, but many died and had severe gross abnormalities such as exencephaly, cranial hemorrhaging or growth retardation, indicating their role during embryonic development (Dawlaty et al., 2013). How TET1 or TET2 affects the expression of genes associated with neural tube closure or cranial development is not yet known.

The *tuft* mouse presents traits resembling *Tet1/Tet2* double knockout mice. Affected newborn mice presented anterior facial malformations, resulting in a midfacial cleft, cranial cephalocele, or both to varying severity (Fong et al., 2012). However, the most severe defects were observed early as E8.5-9.0 during closure of the anterior part of the neural tube (Fong et al., 2014). Tips of the anterior folds failed to adhere and curled. These likely resulted in embryos with exencephaly, which we also observed in E10.5 and later stages, or to a lesser extent, incomplete closure of the anterior neuropore. Through whole genome sequencing of the *tuft* mouse, we have identified a nonsense mutation within the *Tet1* gene that disrupts its catalytic function. RNA-Seq analysis and quantitative RTPCR indicated that the mutation affected the expression of genes in multiple pathways associated with neural tube closure. We hypothesize these defective TET1 proteins were able to bind their designated targets needed for neural tube closure and morphogenesis of the frontonasal region, but altered their expression. We propose an epigenetic mechanism for regulating closure of the anterior neural tube in the *tuft* mouse. *Tet1* is thus a candidate gene locus for predicting defects to neural tube closure and craniofacial development in humans.

Results

1. Affected *tuft* mice were homozygous for a point mutation within the *Tet1* gene

The genomes of five mice that were family members exhibiting the *tuft* phenotype or carrying the affected allele, and one wildtype mouse of the same background strain were sequenced. The family of mice consisted of an affected male that exhibited craniofacial malformations with a cephalocele we previously described as a *tuft* trait (Fong et al., 2012), a normal appearing female predicted to be a carrier for the *tuft* allele and three of their pups. Two pups were affected, one with a cephalocele and the other with a midfacial cleft. The third pup did not appear affected. Therefore, the predicted genotype for mice with visible *tuft* traits was homozygous for the mutant allele (*tu/tu*) and the predicted genotype for normal-appearing carriers was heterozygous (*tu/+*).

Sequences were assembled and analyzed against the NCBI37/mm9 reference sequence from the UCSC Genome Bioinformatics database. A list of nucleotide variations based on the predicted genotypes of the mice was first obtained for chromosome 10, since it was initially linked to the *tuft* phenotype (Fong et al., 2012). From a list of 72 variants on Chr 10, 43 were known strain-specific polymorphisms. Of the remaining variations, only one affected the coding sequence of a gene, *Tet1*. A single cytosine was substituted by thymine in the first nucleotide position of a codon in exon 11 (c.5167C>T) encoding for arginine, resulting in a termination codon (p.R1723*) compared to the reference sequence (MGI) and background strain mouse (Fig. 1). This disrupted a potential nuclear localization signal sequence and excluded the last 318 amino acids of the catalytic domain (CD), thus likely rendering it non-functional. The HxD motif that coordinates with Fe(II) that is required for catalyzing 5mC to 5hmC (Tahiliani et al., 2009) remained intact.

We confirmed our genome analysis by Sanger sequencing the region containing the putative mutation in over 50 mice and embryos. These included samples that were used for whole genome sequencing, some of the affected animals used for the

initial linkage analysis (Fong et al., 2012) and others spanning generations including normal appearing mice predicted to be carriers, and with other mice with phenotypes unrelated to *tuft* that were housed in the same facility serving as external negative controls. Nearly all affected mice (48/50) exhibiting one of the craniofacial traits from adults, newborns (ocular hypertelorism, cranial cephalocele, bifid nose, midfacial cleft, or anencephaly) and embryos with neural tube closure defects we previously described (Fong et al., 2014) were homozygous for the c.5167C>T mutation in the *Tet1* gene (*Tet1^{tuft/tuft}*). Each of the other two mice had a very small cephalocele that we predicted to be homozygous for the mutation were heterozygous, which may reflect a dominant effect. All normal appearing predicted carriers we sampled (6/6) were heterozygous for cytosine and thymine at the same c.5167 position (*Tet1^{tuft/+}*). Wildtype strains (4/4) including 3H1 (*tuft* background strain), BALB/c, C57BL/6J and *Brachyrrhine* mouse housed in the same room with *tuft*, were homozygous for cytosine. The relative location of this mutation was consistent with our initial candidate region between 27-45 cM on Chr 10 (32.48 cM) predicted by linkage analysis (Fong et al., 2012). No other significant variations affecting the coding sequences (exons) of the mouse genomes were detected. Therefore, the mutation in *Tet1*, we will refer to as *Tet1^{tuft}*, is likely to be the primary genetic defect responsible for the *tuft* phenotypes.

2. Reduced body size and TET activity in *Tet1^{tuft}* mice

In light of the *Tet1^{tuft}* mutation and phenotype of *Tet1* knockout mice, we noted a significant number of pups from *tuft* matings were smaller than typical littermates or normal mice of the same age and sex (Fig. 2A). About 30% (28/96) of 30 day-old mice from *Tet1^{tuft/tuft}* x *Tet1^{tuft/+}* matings were reduced in body weight and length when 50% would be expected based on Mendelian inheritance for a recessive trait. This percentage may be an overestimate, since mice that died *in utero* or stillborn were not accounted. Male runts were almost 40% less than the average body weight of sex and age-matched wildtype background mice or normal appearing siblings (12.38g ± 1.39 SD, vs 20.11g ± 0.93 SD, $p = 0.0001$). Their nose to rump length differed by less than 10% (69.67 mm ± 3.08 SD, vs. 79.67 mm ± 3.28 SD, $p = 0.00013$). Female runts, on the other hand, were about 20% less than the normal

body weight ($13.25\text{g} \pm 1.20\text{ SD}$, vs $16.70\text{g} \pm 0.6\text{ SD}$, $p = 0.0011$) and just about 6% less in length ($71.33\text{mm} \pm 1.63\text{ SD}$ vs $75.71\text{mm} \pm 1.25\text{ SD}$, $p = 0.0031$). There were few extreme cases where mice weighed less than 10 grams ($n = 5$), or about 50% of the normal weight at one month in age regardless of sex (shown in Fig. 2A but not included in the data set). Almost a third of runt mice (12/42) also exhibited one or more of the craniofacial traits (cranial cephalocele, hypertelorism, bifid nose) characteristic of *tuft* traits we previously described (Fong et al., 2012). But, like the craniofacial traits, the occurrence of the runt phenotype was lower than expected for complete penetrance. This observation is consistent to the phenotype in mice deficient in *Tet1*, with the exception of the developmental defects and lower penetrance. Genetic knockout mice for the *Tet1* gene were reported to be viable and fertile, but about 75% of homozygous mutant pups (13/17) had a smaller body size at birth (Dawlaty et al., 2011). Homozygous *Tet1* knockout pups averaged 8g compared with 11g or about 27% less than normal. Some grew to normal body weight following one month in age as in the case for *tuft* runts.

We then sequenced the *Tet1* gene in mice with the runt phenotype with ($n = 6$) or without ($n = 6$) the *tuft* craniofacial traits. Runt mice that did not exhibit a craniofacial trait, hence reflecting the phenotype observed in *Tet1* knockout mice, were either homozygous ($n = 2/6$) or heterozygous ($n = 4/6$) for the same nonsense mutation. Whereas runts that also presented one or more of the *tuft* craniofacial traits were always homozygous ($n = 6/6$). This indicates a tight association between the craniofacial traits and homozygosity of the *Tet1^{tuft}* mutation. Furthermore, the occurrence of a runt phenotype in mice heterozygous for the mutation suggests it can have a dominant negative effect in *tuft* mice and partially prevent compensation by normal protein.

Since premature termination within the C-terminal domain would conceivably render the catalytic function of TET1 nonfunctional, we compared the amount of genomic 5hmC in *tuft* embryos homozygous for the mutation with corresponding tissues from the wildtype background strain. We found the amount of genomic 5hmC in the anterior or rostral part of E9 embryos (18-22 somites) homozygous for the *Tet1^{tuft}* allele was significantly less ($0.034\% \pm 0.004\text{ SD}$, $p = 0.002$) than wildtype

embryos ($0.046\% \pm 0.003$ SD), or 74% of normal (Fig. 2B). This difference was not as large than what was observed in *Tet1* knockout mice (60% of normal), whereas mice deficient of both *Tet1* and *Tet2* were approximately 75% of normal (Dawlaty et al., 2011; 2013). This decline signified a loss of TET1 function in *tuft* embryos expressing a truncated protein. Furthermore, this loss could not be fully compensated by TET2 activity in *tuft* embryos.

We also examined the relative amount of genomic 5mC since we hypothesized that a reduction in TET1 activity would result in similar or higher amount of methylated DNA than normal as seen in ES cells undergoing differentiation or depleted of TET1 (Ficz et al., 2011). We found the amount of 5mC was not significantly different in E9 *Tet1^{tuft}* embryo heads ($3.62\% \pm 0.93$ SD) than wildtype ($4.07\% \pm 0.75$ SD). This difference, however, was not statistically significant ($p = 0.379$). These values were comparable to what was observed in *Tet1/Tet2* double knockout mice where 4% of cytosines were methylated compared with 3.5% in wildtype (Dawlaty et al., 2013).

3. Expression of *Tet1* RNA and protein in *tuft* embryos

Since the *Tet1^{tuft}* point mutation resulted in premature termination, we wanted to determine whether *Tet1* mRNA was still present and producing protein in *tuft* embryos. We compared the expression of *Tet1* in *tuft* embryos with its wildtype 3H1 Balb background strain by whole mount *in situ* hybridization (WMISH) and quantitative real-time polymerase chain reaction (qPCR). The expression of *Tet1* appeared most prominent in cells along the apical side of the neuroectoderm facing the ventricular space and along the dorsal midline in normal E9.0-9.5 (14-24 somites) embryonic heads by WMISH (Fig. 3A-D). This was consistent with what was observed in E9 embryos containing a *Tet1*-LacZ genetrapp construct (Yamaguchi et al., 2012). Staining in *tuft* embryos with neural tube closure defects was similar in localization, though somewhat less intense (Fig. 3B). We then analyzed the expression of *Tet1* RNA in the rostral part of the head in embryos from different stages during and following closure of the anterior neural tube (14-24 somites) by qPCR. The amount of *Tet1* RNA from E9 *Tet1^{tuft}* embryos homozygous for the mutation was less than half of what was measured in wildtype embryos of similar

stage (Fig. 3E). Levels of *Tet2* and *Tet3* RNA were not significantly different from wildtype embryos ($p = 0.57, 0.78$ respectively).

We assessed whether mutant protein was still present or degraded in *tuft* mice, despite the reduced levels of RNA. Western blots using an antibody specific to the N-terminal region of TET1 detected bands approximating the predicted molecular weights in tissue samples from both 3H1 wildtype and *tuft* mice homozygous for the *Tet1* mutation (Fig. 3F). Two bands appeared in the nuclear fraction flanking the 250 kD molecular weight marker. The band corresponding to the lower molecular weight was only present in the cytoplasmic fraction. These bands were larger than the calculated molecular weights of the two, alternatively spliced TET1 variants (220 and 200 kD). The band larger than the 250 kD marker in our Western blots was similarly observed by others (Jin et al., 2014) whereas the lower band was similar seen in mESCs (Dawlaty et al., 2011). Since none of these reported bands corresponded exactly to the predicted molecular weights, it's likely they are post-translationally modified forms. The predicted molecular weight of the truncated TET1^{tuft} protein, however, is about 190 kD. We could not clearly resolve this difference from wildtype no matter how far samples were run through the gel. To verify these two bands corresponded to TET1, we overexpressed recombinant TET1 and mutant TET1^{tuft} protein fused to a poly histidine epitope tag in IMCD-3 cells. Two bands of similar molecular weights seen in our tissue samples were also detected using antibodies against either the histidine epitope tag (lower blot in Fig. 3F) or TET1. These bands were also visible in Coomassie stained gels, but not in cells transfected with the control plasmid expressing just poly histidine (not shown). These results indicate that TET1 protein is expressed and present to some degree in *tuft* mice homozygous for the mutation.

4. Gene expression associated with anterior neural tube closure was altered in *Tet1*^{tuft} E9 heads.

The neural tube closure defects we observed in *tuft* embryos homozygous for the mutation suggests they were affected by reduced TET1 protein or function. Neural tube closure defects were not reported in mice deficient in either *Tet1*, *Tet2* or *Tet3* individual knockout mice (Dawlaty et al., 2011; Gu et al., 2011; Ko et al., 2011;

Li et al., 2011). But, exencephaly and cranial defects were reported in *Tet1/Tet2* double knockout mice (Dawlaty et al., 2013). We wanted to determine how the truncated TET1^{tuft} protein resulted in neural tube closure defects we observed in *tuft* embryos.

Since TET1 was shown to affect gene transcription in mouse embryonic stem cells (Williams et al., 2011, Wu et al., 2011), we wanted to determine whether genes associated with neural tube closure were affected in *tuft* embryos. We compared the levels of RNA from the anterior part of E9 (16-22 somites) *Tet1^{tuft}* heads that presented curled neural folds with corresponding 3H1 Balb wildtype background strain embryos by RNA-Seq. Following statistical analysis, RNA levels from 2,957 of the 24,487 genes identified (12.08%) were significantly different ($q < 0.05$) compared to wildtype. 1,530 of the 2,957 genes were downregulated (51.7%) while 1,427 genes were elevated (48.3%). The expression of 104/465 gene loci (22%) associated with neural tube closure defects listed in the MGI database for mouse phenotypes (MP:0002151) was significantly altered ($q < 0.05$). About 45% of those (47/104) were associated with TET1 binding in mESCs based on the data from Williams et al., 2011. 30 of those genes were downregulated (64%) while 17 remained elevated (36%). Among those were genes associated with incomplete rostral or anterior neuropore closure (MP:0000928), where expression was significantly changed in 14/65 of those genes (21.5%). We conducted similar surveys with genes associated with cellular adhesion, since that process appeared to be affected in *Tet1^{tuft}* embryos. From these surveys, all together, we noticed a number of genes associated with the noncanonical *Wingless* (WNT) signaling pathway were significantly dysregulated ($P_{NOI} > 0.95$) in *Tet1^{tuft}* embryo heads (Table 1).

Table 1. Expression of genes associated with non-canonical WNT signaling pathway significantly affected in *Tet1^{tuft}* embryos. Fold change in rostrum of E9 *Tet1^{tuft}* embryos compared to its background strain determined from RNA-Seq data. Significance indicated as P_{NOI} . TET1 binding sites in mESCs determined by ChIP analysis using antibodies specific for either the N or C-terminus (Williams et al., 2011) is indicated.

Gene	Fold Change	P_{NOI}	TET1 ChIP
<i>Setdb1</i>	0.099	0.999	
<i>Rhoa</i>	8.845	0.998	
<i>Chd7</i>	0.142	0.996	
<i>Fzd1</i>	5.006	0.988	
<i>Lrp6</i>	0.209	0.986	
<i>Prickle2</i>	5.193	0.982	N and C term
<i>Wnt5a</i>	3.755	0.971	
<i>Dkk1</i>	4.004	0.968	N term
<i>Dvl2</i>	3.442	0.965	N term
<i>Gsk3a</i>	0.311	0.961	N term
<i>Celsr2</i>	3.071	0.960	N and C term
<i>Daam1</i>	0.028	0.959	C term
<i>Vangl2</i>	0.327	0.951	N term

Considering that the initial RNA template used for analysis was pooled from three embryos bearing the same genotype, similar phenotype and age, we then analyzed the expression of several genes in individual samples by qPCR, particularly those with putative TET1 binding sites. We found *Dkk1* expression was significantly elevated ($p = 0.002$) in the anterior rostrum from a number of *tuft* E9 embryos (16-22 somites) with neural tube closure defects (4/6 affected embryos) compared with wildtype embryos of similar stage (Fig. 4A) as indicated by our initial RNA-Seq analysis. Whereas, no significant change was detected in other affected *tuft* embryos (2/6). We also found that expression of either *Celsr1* (2/6 embryos) or *Celsr2* (2/6) was elevated in some of those same embryos with elevated *Dkk1*, even though we did not detect a statistically significant change in the expression of *Celsr1* from our initial RNA-Seq samples ($P_{NOI} = 0.818, 0.592$ fold change) based on our criteria. We were not able to detect consistent changes in the expression of *Dvl2*, *Daam1*, *Prickle2* and *Vangl2* from individual embryos by qPCR (data not shown) as indicated by our RNA-Seq results. Therefore, we could not firmly conclude WNT-signaling was significantly affected.

Some embryos, however, did not indicate changes in the levels of *Celsr1* and *Celsr2*, but had reductions in the expression of *Cecr2* (Cat Eye Chromosome Region candidate 2) or *Grhl2* (*Grainyhead like 2*). *Cecr2* and *Grhl2* are two genes associated with neural tube closure with defects similar to *tuft* mice (Banting et al, 2005; Pyrgaki et al., 2011) that were also associated with TET1 binding in mESCs (Williams et al., 2011). The expression of *Cecr2* and *Grhl2* declined following anterior closure of the neural tube in wildtype embryos (about 16-18 somites) but was lower in 3/6 *tuft* embryos of comparable stage (Fig. 4B). We further found that the expressions of both Ephrin receptor alpha 7 (*Epha7*) transcript variants were nearly depleted in these embryos. *Epha7* is a putative downstream target of CECR2 (Fairbridge et al., 2010) critical for closure of the anterior part of the neural tube (Holmberg et al., 2000).

5. *Grhl2* expression was reduced by TET1^{tuft} protein in cell culture

To determine whether the mutant TET1^{tuft} protein had an effect on gene expression, we overexpressed wildtype TET1 and mutant TET1^{tuft} recombinant protein in murine IMCD-3 cells that were known to express appreciable levels of *Grhl2* (Werth et al., 2010). We first determined whether the mutant TET1^{tuft} recombinant protein was capable of translocating into the nucleus, since the mutation disrupted a putative nuclear bipartite localization signal sequence in the linker region of the catalytic domain. IMCD-3 cells expressing either TET1 or TET1^{tuft} recombinant protein fused with a poly histidine epitope tag (HIS) at the N-terminus were primarily detected in the nucleus within 48 hrs from transfection (Fig. 5A-H). Cells transfected with the control plasmid expressing just the poly HIS tag were positive for histidine immunostaining generally throughout the cell but negative for the TET1 antibody (Fig. 5I). Transfected cells incubated with just the secondary antibodies were also negative (Fig. 5J).

We monitored the expression levels of endogenous *Grhl2* in cells transfected with poly HIS-tagged TET1 or TET1^{tuft} mutant protein up to 72 hrs of similar confluence with or without the addition of ascorbic acid. Ascorbic acid is a cosubstrate for TET enzyme activity (Blaschke et al., 2013; Yin et al., 2013). Levels of *Grhl2* RNA did not significantly change following 48 hrs of overexpressing TET1

compared with cultures transfected with the empty pcDNA-HIS vector as a negative control with or without ascorbic acid (Fig. 5K). There was a slight elevation of *Grhl2* transcription when comparing cultures supplemented with ascorbic acid. Cultures transfected with the mutant TET1^{tuft} recombinant protein, however, showed reductions in *Grhl2* expression, with ($p = 0.03$) or without ($p = 0.15$) the addition of ascorbic acid. After 72 hrs, the expression of *Grhl2* was elevated in cultures expressing wildtype TET1, especially in those supplemented with ascorbic acid ($p = 0.14$, Fig. 5L). However, little or no change in expression was detected in the presence of mutant protein following 72 hrs ($p = 0.66$). Levels of *Grhl2* in control cultures following 72 hrs were much lower than that detected for 48 hrs with or without ascorbate (data not shown). So, levels of *Grhl2* seemed to remain low in 72 hrs cultures with mutant TET1^{tuft} protein, but could be elevated in the presence of recombinant TET1 protein and ascorbate. These results are consistent with the observation of *tuft* embryos presenting neural tube closure defects having reductions in *Grhl2* RNA. This further supports that mutant TET1^{tuft} protein can exert a dominant negative effect on gene transcription by affecting their normal target sites.

Discussion

Animal model studies have unraveled a large number of genes associated with neural tube closure and underscore the complexity of this process (Harris and Juriloff 2007, 2010). Although gene mutations associated with neural tube closure in humans are being identified, there is a need for understanding how environmental conditions influence their expression in order to address a broader spectrum of cases. It has been shown, for example, that maternal conditions can dictate cellular programs for gene expression by defining the chromatin landscape. Altered distributions of modified histones (Pavlinkova et al., 2009; Salbaum and Kappen 2010; 2012; Kappen et al., 2011) or methylated CpGs (Ichi et al., 2010; Zhang et al., 2013; Wei and Loeken 2014; Wang et al., 2015) have been shown to directly affect genes associated with neural tube closure. Since these marks target multiple genes involved with neural tube closure, the genes affected may differ in a particular environment. Furthermore, their occurrence may not be carried over generations if

these epigenetic marks are not maintained. This may account for the sporadic occurrence of specific NTDs and other birth defects within a family history. Variance will also depend on the maternal condition prior and during pregnancy. Thus, NTDs may not only arise from mutations in genes directly involved with the mechanics of closure. We will need to know how epigenetic factors govern the organization of chromatin in order for us to understand how gene expression can be manipulated and alter the functionality of cells.

We found a heritable mutation disrupting the catalytic function of TET1 in the *tuft* mouse that presents defects to neural tube closure or craniofacial development. These defects were generally restricted to the anterior midline or rostral part of the neural tube but varied in severity (Fong et al., 2012, 2014). Mice homozygous for the mutation could also exhibit one or more traits such as a midfacial cleft, cranial cephalocele, and smaller body mass. Despite the reduced level of enzymatic activity, the mutation did not necessarily have a null effect as in *Tet1* knockout mice (Dawlaty et al., 2011), but had gross defects resembling mice that were deficient in both TET1 and TET2 (Dawlaty et al., 2013). The traits in *tuft* mice, however, cannot be completely attributed to a double null of TET1 and TET2. TET1^{*tuft*} RNA and protein was present in *tuft* mice homozygous for the mutation. We did not detect a significant difference in the expression of *Tet2* or *Tet3* from our RNA-Seq and qPCRs, or mutation affecting their coding region from genomic sequence analysis (data not shown). Thus, normal amounts of TET2 and TET3 were likely present and active in *Tet1^{tuft}* mice, though unable to fully compensate for the loss of TET1 activity. Therefore, mutant TET1^{*tuft*} protein was present and interfered with compensatory activity, thus having a dominant negative effect. However, it's also possible that mutant TET1^{*tuft*} protein formed ectopic interactions with other molecules to generate a neomorphic trait, as in cases when a lipomatous cephalocele was formed. The presence of a lipomatous cephalocele in *Tet1/Tet2* double knockouts was not reported.

The truncated TET1^{*tuft*} protein interfered with the expression of genes associated with neural tube closure in *tuft* mice. This is consistent with the role of its C-terminal catalytic domain in regulating gene transcription. Since the methylation of gene promoters is typically associated with transcriptional repression, TET1 can promote

transcription by catalytically removing 5mCs and preventing remethylation (Fig. 6A). Without this function, we would expect genes regulated this way may remain repressed in *tuft* mice. Indeed we found significant reductions in the amount of 5hmC and the expression of genes associated with the *tuft* phenotype, thus TET1 activity. Many of the genes associated with neural tube closure were associated with TET1 binding based on ChIPs from mESCs (Williams et al., 2011). However, the expression in all the genes with putative TET1 binding sites were not necessarily affected in each *tuft* embryo homozygous for the mutation presenting an NTD. Some embryos, for example, were deficient in *Cecr2* but others were not and instead affected by dysregulated components of the noncanonical WNT signaling pathway. We suspect that mutant TET1^{tuft} protein was present, but insufficient to bind every site, possibly due to degradation and partial compensation by TET2 or TET3.

Mice deficient in *Cecr2* closely resembled the traits of newborn *tuft* mice. Mice deficient in *Cecr2* primarily exhibited exencephaly (Banting et al., 2005) but also presented a midline facial cleft with exencephaly or forebrain cephalocele in a different genetic background (Fairbridge et al., 2010). CECR2 is a chromatin modifier that heterodimerizes with SNF2L (mouse SMARCA1) to form CERF (Banting et al., 2005). CERF (CECR2-containing Remodeling Factor) modifies the position of nucleosomes in an ATP-dependent manner, thus altering the accessibility of transcriptional regulatory sites. *Cecr2* appears to be expressed throughout the mouse embryo, but prominently at the margins of neural folds during closure in E9 and in neural tissue through E13.5 (Banting et al., 2005). Disruption of *Cecr2* led to a significant reduction in levels of *Alx1* (*Cart1*), *Dlx5*, and *Epha7* RNA (Fairbridge et al., 2010). Mice deficient in any one of these genes resulted in neural tube closure defects (Zhao et al., 1996; Acampora et al., 1999; Depew et al., 1999; Holmberg et al., 2000). Consistent with this finding, the expression of both *Epha7* isoforms were nearly depleted in the same *tuft* embryos with reduced levels of *Cecr2*. Therefore, closure was likely prohibited by the lack of *Epha7* through deregulation of *Cecr2* by mutant TET1^{tuft} protein in these cases. We hypothesize TET1^{tuft} protein was unable to catalytically reverse the repressive methylated state of the *Cecr2* promoter and its presence prohibited compensatory efforts by TET2 or TET3 (Fig. 6B). We demonstrated that mutant TET1^{tuft} protein could still enter the nucleus and deregulate the expression of *Grhl2* in IMCD-3 cells. However, not all genes

associated with neural tube closure were suppressed in *Tet1^{tuft}* embryos. This indicates a mechanism for transcriptional repression could also be disrupted.

Tet1^{tuft} embryos that did not have significant changes in *Cecr2* usually had abnormal expression levels in one or more components of the planar cell polarity proteins that mediate convergent extension. Deficiencies to one of these components in mice typically led to craniorachischisis, the most severe type of NTD involving caudal closure, but also a range of neural tube closure defects (Murdoch et al., 2014). Mutations have also been linked to human cases with NTDs, including lipomas (DeMarco et al., 2014). Thus, we were somewhat surprised to find that the expression in one or more of these components remained elevated in *Tet1^{tuft}* embryos with neural tube closure defects. The anterior neural folds of these embryos curled inward but did not appear to have problems with closure along the caudal portion of the neural tube as in craniorachischisis (Fong et al., 2014). CELSR1 and CELSR2 are protocadherin transmembrane adhesion proteins that are both expressed in the neural ectoderm during neural tube closure and brain development (Formstone and Little 2001). CELSR1 recruits PDZ-RhoGEF, RhoA and Rho kinase (ROCK1) through interactions with DVL and DAAM1 to promote medio-lateral contraction along the floor plate to bend the neural tube (Nishimura et al., 2012). Abnormally high levels of *Celsr1* may have prompted elevated levels of *RhoA* we detected in our initial RNA-Seq analysis, thus apical actin-myosin contractile activity exaggerated curling of the anterior neural folds. Whether abnormal expression levels of *Celsr2* affects neural tube closure is unclear. Mice deficient in *Celsr2* impaired ciliogenesis in ependymal cells and their function resulting in hydrocephalus (Tissir et al., 2010). *Celsr2* is necessary for correct positioning of cilia at the apical surface of ependymal cells. Elevated levels of *Celsr2* may result in an imbalance of CELSR-CELSR interactions thus altering cell polarity or cellular organization. Since TET1 associated with the promoters of *Celsr1* and *Celsr2* (Williams et al., 2011) and other chromatin modifying factors (Wu et al., 2011; Cartron et al., 2013), elevated expression indicates that the truncated TET1^{tuft} protein lost its ability to repress transcription of these genes (Fig. 6C).

The significance of elevated *Dkk1* (*Dickkopf1*) in *tuft* embryos with neural tube closure defects is not clear. Especially since the expression of *Cecr2*, *Grhl2*, or PCP

genes could also have been affected in the same embryos. DKK1 is the major factor for anterior identity and head formation (Mukhopadhyay et al., 2001). Normal expression of *Dkk1* in the anterior region is thought to prevent the formation of neural crest cells by antagonizing canonical WNT signaling (Fossat et al., 2011). Thereby, anterior structures are contributed by cranial neural crest cells (CNCC) emigrating from the midbrain region. It's possible that extended expression of *Dkk1* in *tuft* embryos limited adequate amount of CNCCs migrating toward the frontonasal region for normal craniofacial development. TET1 has been associated with *Dkk1* in mESCs (Williams et al., 2011). As in the case with elevated *Celsr1/2*, mutant TET1^{*tuft*} protein may have prevented timely down regulation of *Dkk1* during craniofacial development. Disruptions to WNT signaling by elevated *Dkk1* may account for craniofacial anomalies in *tuft* mice.

NTDs or craniofacial anomalies associated with mutations to *Tet1* have not been reported thus far. We focused our investigation on genes critical to neural tube closure that are putative targets for TET1 based on the normal state of mESCs. Since TET2 or TET3 could not fully compensate the loss of activity in *tuft* embryos, it is likely that these genetic loci involved with neural tube closure were targets for TET1 activity. If so, we may be able to reduce or eliminate the negative effect of dysfunctional TET1^{*tuft*} protein and allow TET2 or TET3 to compensate its function. Alternatively, we may overcome the effects of dysfunctional TET1 protein by overexpressing normal enzyme or possibly augment TET activity by modifying the parental diet. We demonstrated that TET1 could promote the transcription of *Grhl2* in IMCD-3 cells supplemented with ascorbic acid, but was deregulated in the presence of mutant TET1^{*tuft*} protein. Variable activity due to genetic background or dietary intake may dictate particular traits, physical or even mental as TET1 is thought to play significant roles in neural development and behavior. Mice deficient in TET1 alone have deficits in memory extinction, a behavioral mechanism allowing adaptation that may be impaired in cases of posttraumatic stress disorder (Rudenko et al., 2013). Heritable variants altering the function of TET1 in humans might result in such defects. It would be of value to know when TET1 is essential to elicit critical steps. The *tuft* mouse, can thus serve as a model system to assess therapeutic strategies addressing NTDs and TET1-associated disorders. We must consider, however, that epigenetic marks regulating the same genes in humans may differ

from mice or even between individuals. This might explain the variance in phenotype between species and susceptibility of disease.

MATERIALS AND METHODS

Animals

All procedures were carried out in accordance with IACUC specifications and approved by the University of Hawai'i Laboratory of Animal Services. Mice strains were housed under standard conditions and bred as previously described (Fong et al., 2012). Timed matings were determined by noting the presence of a vaginal plug as day 0.5 for staging embryo collections and estimating date of births. Developmental stages of embryos were determined based on the number of somites. Genotyping of DNA isolated from tail clips or embryonic tissue done by PCR (Primers: Forward 5'-GGTAGAACAGGCCTTATTCCTC-3', Reverse 5'-GGTGAGAAGTAGATGAGGCTG-3') followed by Sanger sequencing (Sequencing primer: 5'-GGATGAACAACCTCCACGTCCTG-3').

Library Preparation and DNA Sequencing

1 μ g of dsDNA determined by Invitrogen Qbit (Life Technologies, Carlsbad, CA) high sensitivity spectrofluorometric measurement was sheared by sonication to an average size of 200 bp on a Covaris S2 instrument (Covaris, Woburn, MA). Library construction was performed in an automated fashion on an IntegenX Apollo324 (IntegenX, Pleasanton, CA) which size-selects fragments by double-SPRI binding with different concentrations of PEG for a high cut and a low cut. Each library was fitted with an adapter containing a different 6 base molecular barcode for high level multiplexing. After 9 cycles of PCR amplification using the Clontech Advantage II kit (Clontech Laboratories, Mountain View, CA), 1 μ g of genomic library was recovered for genome enrichment using the PE125 kit. Libraries were enriched according to the manufacturer's recommendations and sequenced on an Illumina HiSeq2500

(Illumina, San Diego, CA), generating around 30 million 100 bp paired end reads each equivalent to 6GB of usable high quality sequence per sample.

Data Alignment and Analysis Methods

Analysis methods utilize the Broad Institute's Genome Analysis Toolkit (GATK) and follow the pipeline described in (DePristo et al., 2011), along with the modifications listed in the "Best Practices" document on their website (http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit). Briefly, reads that pass Illumina's Chastity Filter are aligned with the Burrows Wheeler Aligner (BWA) (Li & Durbin, 2009). GATK's UnifiedGenotyper module is used to call variant sites (both single nucleotide and small indel) in all samples simultaneously. Finally, SNV calls are filtered using the variant quality score recalibration method described in (DePristo et al., 2011). Indel calls will be filtered with a set of hard filters, as there are not enough indels in an exome to use the Gaussian method.

Variants were filtered using Golden Helix SNP & Variation Suite (SVS; www.GoldenHelix.com). Variants were filtered for quality >20. Variants from chromosome 10 and matching segregation through expected genotypes were isolated and analyzed for pathogenicity and known frequency in healthy populations.

TET Activity Assay

TET1 activity was indirectly estimated using the Quest 5-hmC and 5-mC DNA ELISA Kits (Zymo Research, Irvine, CA). 200 ng of genomic DNA purified from the anterior part of E9 heads (14-22 somites) from *tuft* litters with NTDs and 3H1 Balb wildtype background embryos was measured for 5-mC or 5-hmC content according to the manufacturer's protocol. Assays were done in triplicate.

Whole Mount *In Situ* Hybridization

Embryos collected and preserved in 100% methanol were processed for whole mount *in situ* hybridization (WMISH) as previously described (Fong et al., 2014). The template for generating mouse *Tet1* riboprobes was a 1.8 kb cDNA consisting part of the 3' coding region (c.4333-6146) cloned into pBluescript (Stratagene, La Jolla, CA).

RNA-Seq and Statistical Analysis

RNA was purified from the anterior part of E9 heads (18-22 somites) using the RNeasy Plus Universal mini kit (QIAGEN Inc, Valencia, CA). Quality and concentration was determined using a 2100 Bioanalyzer (Agilent Technologies) and NanoDrop (NanoDrop, Wilmington, DE). RNA with a RIN of at least 9 was pooled for a total of 300 ng from 3 biological replicates for each condition (wildtype and mutant). RNA was poly-A selected using Dynabeads (Thermo Fisher Scientific, Waltham, MA). cDNA libraries were constructed using the Ion Total RNA-Seq Kit v2 (Thermo Fisher Scientific) according to the manufacturer's protocol for poly-A selected RNA. RNA-Seq libraries were templated using an Ion OneTouch 2 System (ThermoFisher Scientific) and sequenced using a 200 bp kit on an Ion Proton sequencing system (Life Technologies) according to the manufacturer's instructions.

Ion Torrent Suite was used to obtain FASTQ sequencing data. Sequenced single-end reads (66,187,511 for wildtype and 53,803,778 for *tuft*) were trimmed and filtered using PRINSEQ (Schmieder and Edwards, 2011). Low quality sequences were trimmed from the ends until a base pair of Phred quality score ≥ 20 (at least 99% accurate) not found, and filtered out sequences having below 20 nucleotides.

The *Mus musculus* UCSC mm10 reference genome was indexed by Bowtie2 v2.2.5. Processed reads were aligned to the reference genome using Tophat v2.0.14 (Kim et al. 2013). Tophat2 incorporates the Bowtie2 (Langmead and Salzberg 2012) algorithm to perform the alignment. Resulting alignment (.BAM) files were analyzed with Cufflinks v2.1.1 (Trapnell et al 2010) which quantified transcript

abundance in terms of Reads Per Kilobase of exon model per Million mapped reads (RPKM). SAMtools v0.1.18 (Li et al., 2009) was used for sorting and BAM conversion, and htseq-count script on HTSeq package was used to count reads mapped to mouse gene models.

Differential gene expression from the count data was identified using the non-parametric NOISeq-sim program (Tarazona et al. 2011) with default parameters, a trimmed mean of M-values normalization and estimated probability of differential expression $P_{NOI} > 0.95$ as a threshold. The probability ($1-P_{NOI}$) reported in NOISeq can be considered equivalent to q-value (FDR adjusted p-value) (Zheng and Moriyama 2013). Gene set enrichment analysis on the expressed genes was conducted using GSEA (<http://www.broadinstitute.org/gsea/>) with recommended default parameters of 1000 permutations and FDR < 0.25 as a threshold for enrichment in phenotype. Data deposited into the Gene Expression Omnibus (GEO), accession number GSE75001.

Quantitative RT-PCR

RNA was purified using the Total RNA Plus Universal Kit (QIAGEN). 0.1-0.3 μ g of RNA was used as a template for first-strand DNA synthesis (Bio-Rad Laboratories, Hercules, CA). Quantitative real-time polymerase chain reactions (qPCR) performed using SYBR Green Universal Master Mix Dyes (Bio-Rad Laboratories), amplified and detected with the CFX96 Real-Time System C1000 Thermocycler (Bio-Rad Laboratories). Annealing temperatures and data capture for analysis was determined from melt curves and amplification efficiencies for each oligonucleotide primer pair (Integrated DNA Technologies, Coralville, IA) (Table 2). One microliter of first-strand DNA was used per reaction. Reactions done in triplicate for each biological replicate (n) and normalized using the $2^{-\Delta\Delta C(t)}$ method (Livak and Schmittgen, 2001). Mean values of biological replicates were charted with calculated standard deviation (SD) indicated by error bars. Samples with differences in expression levels were charted. Statistical significance (p) was determined by two-tailed, unpaired Student's t -test (R version 3.2.2) using ΔC_t values to account for technical variance and sample size.

Expression constructs and Cell culture

Full length clones of mouse *Tet1* mRNA (NM001253857.1, longer isoform 1) and one containing the c.5167C>T transition were assembled with gBlocks Gene Fragments (Integrated DNA Technologies) using the Gibson Assembly Method (New England BioLabs, Ipswich, MA). Sequences were confirmed followed by Maxi Prep purification (Sigma-Aldrich, St. Louis, MO) prior to cell transfections. Mouse mIMCD-3 cells (CRL-2123) (ATCC, Manassas, VA) were transfected using the Gene Pulser Xcell electroporator (Bio-Rad Laboratories). Cells were seeded in 6-well plates and cultured for 24-72 hours in DMEM:F12 with 10% fetal bovine serum and antibiotics (Mediatech Inc., Manassas, VA). Media was replenished daily with or without 100 $\mu\text{g ml}^{-1}$ of sodium L-ascorbate (Santa Cruz Biotechnology, Dallas, TX) following transfections. RNA or protein was harvested 24, 48 and 72 hrs following transfection. Transfected cells for immunostainings were seeded on sterile glass coverslips coated with PureCol EZ Gel Bovine Type I Collagen Solution (Advanced BioMatrix, San Diego, CA) diluted 1:6 following 20 min at room temperature. Cells were washed and fixed in 4% paraformaldehyde, permeabilized in 0.1% Triton-X100 (Sigma-Aldrich) prior to incubations with antibodies.

Protein Extraction and Detection

Protein was extracted from tissue or cell cultures using the EpiSeeker Nuclear Protein Extraction Kit (Abcam, Cambridge, MA) according to the manufacturer's protocol. Protein concentrations were estimated against BSA standards (Thermo Fisher Scientific) using Bradford reagent (Bio-Rad Laboratories). 40 μg of protein were loaded per well of 4-15% gradient polyacrylamide gels (Bio-Rad Laboratories). Proteins were transferred onto Immobilon-FL nylon membranes (EMD Millipore, Billerica, MA) using Dunn's bicarbonate buffer with 0.1% sodium dodecyl sulfate. Proteins were detected using anti-TET1 specific for the amino terminus (Sigma-Aldrich, SAB2700188) or anti-polyhistidine epitope tag (Cell Signaling Technology, Danvers, MA, mAb 27E8) diluted 1:1000 in phosphate buffered saline (PBS) with 5% BSA overnight at 4°C. Bands were detected with secondary antibodies conjugated

with IRDye (LI-COR Biosciences, Lincoln, NE) diluted 1:10,000 and scanned with the LI-COR Digital Imager (LI-COR Biosciences). Immunolocalizations were detected using goat anti-rabbit or mouse secondary antibodies conjugated with DyLight diluted 1:800 in PBS (Rockland Immunochemicals Inc., Limerick, PA) and counterstained with DAPI (Life Technologies) for nuclei. Images taken at exposures < 0.5 second based on negative controls using a DP73 microscope digital camera and cellSens Standard 1.12 software (Olympus Corporation, Tokyo, Japan).

Acknowledgments

We would like to thank Dr. Yusuke Marikawa for kindly providing oligonucleotide primers for mouse *Celsr1*. Special thanks go to Vanessa Lui, Dr. Noemi Polgar and Geoffrey Yee for their technical contributions.

Competing interests

The authors declare no competing financial interests.

Author contributions

K.F. conceived, designed the experiments, conducted gene expression experiments, data analysis, and wrote the paper. R.H. and Z.M.A. performed, analyzed genomic sequence data and contributed to writing. V.K. conducted statistical and RNA-Seq analysis and interpretation. M.J.C. and A.K.M. conducted and provided input to the RNA-Seq experiment. B.F. designed and generated plasmid constructs and contributed to the cell culture experiments. S.L. provided the *tuft* mouse and resources. All authors were involved in editing the manuscript and approving the submitted version.

Funding

This work was supported by grants from the Hawai'i Community Foundation 13ADVC-60316 (K.F.) and in part by the National Institutes of Health U54MD007584 and P20GM103466 (V.K.), 1K01DK087852, R03DK100738 (B.F.), R01DC012564 (Z.M.A.), R01DK064752 (S.L.), the Cincinnati Children's Hospital Medical Center (R.B.H.), the March of Dimes #5-FY14-56 (B.F.), and the Queen's Health Systems Native Hawaiian Health Initiative (A.K.M.). Sequencing service at the University of Hawai'i was enabled in part by the National Institutes of Health (P20GM103457). Use of the Histopathology Core was supported by grants from the National Institute on Minority Health and Health Disparities (G12MD007601) and the National Institute of General Medical Sciences (P30GM103341).

Data availability

RNA-Seq expression data deposited into the Gene Expression Omnibus (GEO), accession number GSE75001.

References

Acampora D., Merlo G. R., Paeleari L., Zerega B., Postiglione M. P., Mantero S., Bober E., Barbieri O., Simeone A., Levi G. (1999). Craniofacial, vestibular and bone defects in mice lacking the Distal-less-related gene *Dlx5*. *Development* **126**, 3795-3809.

Banting G. S., Barak O., Ames T. M., Burnham A. C., Kardel M. D., Cooch N. S., Davidson C.E., Godbout R., McDermid H. E., Shiekhattar R. (2005). CECR2, a protein involved in neurulation, forms a novel chromatin remodeling complex with SNF2L. *Hum. Mol. Genet.* **14**, 513-524.

Blaschke K., Ebata K. T., Karimi M. M., Zepeda-Martinez J. A., Goyal P., Mahapatra S., Tam A., Laird D. J., Hirst M., Rao A. et al. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222-226.

Cartron P.-F., Nadaradjane A., LePape F., Lalier L., Gardie B., Vallette F.M. (2013). Identification of TET1 partners that control its DNA-demethylating function. *Genes & Cancer* **4**, 235-241.

Copp A. J., Stanier P., Greene N. D. E. (2013). Neural tube defects: recent advances, unresolved questions, and controversies. *Lancet Neurol.* **12**, 799-810.

Dawlaty M. M., Ganz K., Powell B. E., Hu Y.-C., Markoulaki S., Cheng A. W., Gao Q., Kim J., Choi S.-W., Page D. C., Jaenisch R. (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* **9**, 166-175.

Dawlaty M. M., Breiling A., Le T., Raddatz G., Barrasa M. I., Cheng A. W., Gao Q., Powell B. E., Li Z., Xu M. et al. (2013). Combined deficiency of Tet1 and Tet2

causes epigenetic abnormalities but is compatible with postnatal development. *Dev. Cell* **24**, 310-323.

Depew M. J., Liu J. K., Long J. E., Presley R., Meneses J. J., Pedersen R. A., Rubenstein J. L. R. (1999). Dlx5 regulates regional development of the branchial arches and sensory capsules. *Development* **126**, 3831-3846

De Marco P., Merello E., Piatelli G., Cama A., Kibar Z., Capra V. (2014). Planar cell polarity gene mutations contribute to the etiology of human neural tube defects in our population. *Birth Defects Res. A Clin. Mol. Teratol.* **100**, 633-641.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel G., Rivas M. A., Hanna M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**, 491–498.

Fairbridge N. A., Dawe C. E., Niri F. H., Kooistra M. K., King-Jones K., McDermid H. E. (2010). Cecr2 mutations causing exencephaly trigger misregulation of mesenchymal/ectodermal transcription factors. *Birth Defects Res. Part A, Clin. Mol. Teratol.* **88**, 619-625.

Ficz G., Branco M. R., Seisenberger S., Santos F., Krueger F., Hore T. A., Marques C. J., Andrews S., Reik W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398-402.

Fong K. S. K., Cooper T. B., Drumhiller W. C., Somponpun S. J., Yang S., Ernst T., Chang L., Lozanoff S. (2012). Craniofacial features resembling frontonasal dysplasia with a tubulonodular interhemispheric lipoma in the adult 3H1 tuft mouse. *Birth Defects Res. A Clin. Mol. Teratol.* **94**, 102-113.

Fong K. S. K., Adachi D. A. T., Chang S. B., Lozanoff S. (2014). Midline craniofacial malformations with a lipomatous cephalocele are associated with

insufficient closure of the neural tube in the tuft mouse. *Birth Defects Res A Clin. Mol. Teratol.* **100**, 598-607.

Formstone C. J., Little P. F. R. (2001). The flamingo-related mouse *Celsr* family (*Celsr1-3*) genes exhibit patterns of expression during embryonic development. *Mech. Dev.* **109**, 91-94.

Fossat N., Jones V., Khoo P.-L., Bogani D., Hardy A., Steinder K., Mukhopadhyay M., Westphal H., Nolan P. M., Arkell R., et al. (2011). Stringent requirement of a proper level of canonical WNT signalling activity for head formation in mouse embryo. *Development* **138**, 667-676.

Garne E., Loane M., Dolk H., Barisic I., Addor M.-C., Arriola L., Bakker M., Calzolari E., Dias C. M., Doray B., et al. (2012). Spectrum of congenital anomalies in pregnancies with pregestational diabetes. *Birth Defects Res. A Clin. Mol. Teratol.* **94**, 134-140.

Globisch D., Münzel M., Müller M., Michalakis S., Wagner M., Koch S., Bruckl T., Biel M., Carell T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367. doi:10.1371/journal.pone.0015367.

Gu T.-P., Guo F., Yang H., Wu H.-P., Xu G.-F., Xie Z.-G., Shi L., He X., Jin S.-G., Iqbal K., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606-610.

Harris M. J., Juriloff D. M. (2007). Mouse mutants with neural tube closure defects and their role in understanding human neural tube defects. *Birth Defects Res A Clin. Mol. Teratol.* **79**, 187-210.

Harris M. J., Juriloff D. M. (2010). An update to the list of mouse mutants with neural tube closure defects and advances toward a complete genetic perspective of neural tube closure. *Birth Defects Res. A Clin. Mol. Teratol.* **88**, 653-669.

Holmberg J., Clarke D. L., Frisen J. (2000). Regulation of repulsion versus adhesion by different splice forms of an Eph receptor. *Nature* **408**, 203-206.

Ichi S., Costa F. F., Bischof J. M., Nakazaki H., Shen Y.-W., Boshnjaku V., Sharma S., Maria-Farnell B., McLone D. G., Tomita T., et al. (2010). Folic acid remodels chromatin on Hes1 and Neurog2 promoters during caudal neural tube development. *J. Biol. Chem.* **285**, 36922-36932.

Ito S., D'Alessio A. C., Taranova O. V., Hong K., Sowers L. C., Zhang Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129-1133.

Iqbal K., Jin S.-G., Pfeifer G. P., Szabó P. E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl. Acad. Sci.* **108**, 3642-3647.

Jin C., Lu Y., Jelinek J., Liang S., Estecio M. R., Barton M. C., Issa J. P. (2014). TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. *Nucleic Acids Res.* **42**, 6956-6971.

Kaas G. A., Zhong X., Eason D. E., Ross D. L., Vachhani R. V., Ming G. L., King J. R., Song H., Sweatt J. D. (2013). TET1 controls CNS 5-methylcytosine hydroxylation, active DNA demethylation, gene transcription, and memory formation. *Neuron* **79**, 1086-1093.

Kappen C., Kruger C., MacGowan J., Salbaum J. M. (2011). Maternal diet modulates the risk for neural tube defects in a mouse model of diabetic pregnancy. *Reprod. Toxicol.* **31**, 41-49.

Kim D., Pertea G., Trapnell C., Pimentel H., Kelley R., Salzberg S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36.

Ko M., Bandukwala H. S., An J., Lamperti E. D., Thompson E. C., Hastie R., Tsangarartou A., Rajewsky K., Koralov S. B., Rao A. (2011). Ten-eleven-translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc. Natl. Acad. Sci.* **108**, 14566-14571.

Langmead B., Salzberg S. L. (2012). Fast gapped-read alignment with Bowtie2. *Nature Methods* **9**, 357-359.

Li, H., Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. (2009). Genome Project Data Processing Subgroup: The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

Li Z., Cai X., Cai C.-L., Wang J., Zhang W., Petersen B. E., Yang F.-C., Xu M. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood* **118**, 4509-4518.

Livak K. J., Schmittgen T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402-408.

Marean A., Graf A., Zhang Y., Niswander L. (2011). Folic acid supplementation can adversely affect murine neural tube closure and embryonic survival. *Hum. Mol. Genet.* **20**, 3678-3683.

Mukhopadhyay M., Shtrom S., Rodriguez-Esteban C., Chen L., Tsukui T., Gomer L., Dorward D. W., Glinka A., Grinberg A., Huang S.-P., et al. (2001). Dickkopf1 is required for embryonic head induction and limb morphogenesis. *Dev. Cell* **1**, 423-434.

Murdoch J. N., Damrau C., Paudyal A., Bogani D., Wells S., Greene N. D. E., Stanier P., Copp A. J. (2014). Genetic interactions between planar cell polarity genes cause diverse neural tube defects in mice. *Dis. Model. Mech.* **7**, 1153-1163.

Nishimura T., Honda H., Takeichi M. (2012). Planar cell polarity links axes of spatial dynamics in neural-tube closure. *Cell* **149**, 1084-1097.

Pastor W. A., Aravind L., Rao A. (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* **14**, 341-356.

Pavlinkova G., Salbaum J. M., Kappen C. (2009). Maternal diabetes alters transcriptional programs in the developing embryo. *BMC Genomics* **10**, 274.

Pyrgaki C., Liu A., Niswander L. (2011). Grainyhead-like 2 regulates neural tube closure and adhesion molecule expression during neural fold fusion. *Dev. Biol.* **353**, 38-49.

Rudenko A., Dawlaty M. M., Seo J., Cheng A. W., Meng J., Le T., Faull K. F., Jaenisch R., Tsai L.-H. (2013). Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* **79**, 1109-1122.

Salbaum J. M., Kappen C. (2010). Neural tube defect genes and maternal diabetes during pregnancy. *Birth Defects Res. A Clin. Mol. Teratol.* **88**, 601-611.

Salbaum J. M., Kappen C. (2012). Response of the embryonic epigenome to maternal diabetes. *Birth Defects Res. A Clin. Mol. Teratol.* **94**, 770-781.

Schmieder R., Edwards R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863-864.

Smith Z. D., Meissner A. (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204-220.

Szwagierczak A., Bultmann S., Spada F., Leonhardt H. (2010). Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **38**, e181

Tahiliani M., Koh K.-P., Shen Y., Pastor W. A., Bandukwala H., Brudno Y., Agarwal S., Iyer L. M., Liu D. R., Aravind L., Rao A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935.

Tan L., Shi Y. G. (2012). Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development* **139**, 1895-1902.

Tarazona S., Alcalde F.G., Dopazo J., Ferrer A., Conesa A. (2011). Differential expression in RNA-seq: A matter of depth. *Genome Res.* **21**, 2213-2223.

Tissir F., Qu Y., Montcouquiol M., Zhou L., Komatsu K., Shi D., Fujimori T., Labeau J., Tyteca D., Courtoy P, et al. (2010). Lack of cadherins Celsr2 and Celsr3 impairs ependymal ciliogenesis, leading to fatal hydrocephalus. *Nat. Neurosci.* **13**, 700-707.

Trapnell C., Williams B. A., Pertea G., Mortazavi A., Kwan G., Baren M. J. V., Salzberg S. L., Wold B. J., Pachter L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnology* **28**, 511-515.

Tsagaratou A, Rao A. (2013). TET proteins and 5-methylcytosine oxidation in the immune system. *Cold Spring Harb. Symp. Quant. Biol.* **78**, 1-10.

Wang X., Guan Z., Chen Y., Dong Y., Niu Y., Wang J., Zhang T., Niu B. (2015). Genomic DNA hypomethylation is associated with neural tube defects induced by methotrexate inhibition of folate metabolism. *PLoS ONE* **10**, e0121869. doi:10.1371/journal.pone.0121869.

Wallingford J. B. (2012). Planar cell polarity and the developmental control of cell behavior in vertebrate embryos. *Annu. Rev. Cell Dev. Biol.* **28**, 627-53.

Wei D., Loeken M. R. (2014). Increased DNA methyltransferase 3b (Dnmt3b)-mediated CpG island methylation stimulated by oxidative stress inhibits expression of a gene required for neural tube and neural crest development in diabetic pregnancy. *Diabetes* **63**, 3512-3522.

Werth M., Walentin K., Aue A., Schonheit J., Wuebken A., Pode-Shakked N., Vilianovitch L., Erdmann B., Dekel B., Bader M., et al. (2010). The transcription factor grainyhead-like 2 regulates the molecular composition of the epithelial apical junctional complex. *Development* **137**, 3835-3845.

Williams K., Christensen J., Pedersen M. T., Johansen J. V., Cloos P. A. C., Rappsilber J., Helin K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343-348.

Wossidlo M., Nakamura T., Lepikhov K, Marques C. J., Zakhartchenko V., Boiani M., Arand J., Nakano T., Reik W., Walter J. (2011). 5-hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* **2**, 241.

Wu H., D'Alessio A. C., Ito S., Xia K., Wang Z., Cui K., Zhao K., Sun Y. E., Zhang Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**, 389-393.

Wu H., Zhang Y. (2014). Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **156**, 45-68.

Yamaguchi S., Hong K., Liu R., Shen L., Inoue A., Diep D., Zhang K., Zhang Y. (2012). Tet1 controls meiosis by regulating meiotic gene expression. *Nature* **492**, 443-447.

Yamaguchi S., Shen L., Liu Y., Sandler D., Zhang Y. (2013). Role of Tet1 in erasure of genomic imprinting. *Nature* **504**, 460-464.

Yin R., Mao S.-Q., Zhao B., Chong Z., Yang Y., Zhao C., Zhang D., Huang H., Gao J., Li Z., et al. (2013). Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *J. Am. Chem. Soc.* **135**, 10396-10403.

Zhang Q., Xue P., Li H., Bao Y., Wu L., Chang S., Niu B., Yang F., Zhang T. (2013). Histone modification mapping in human brain reveals aberrant expression of histone H3 lysine 79 dimethylation in neural tube defects. *Neurobiol. Dis.* **54**, 404-413.

Zhao Q., Behringer R. R., de Crombrughe B. (1996). Prenatal folic acid treatment suppresses acrania and meroanencephaly in mice mutant for the *Cart1* homeobox gene. *Nat. Genet.* **13**, 275-283.

Zheng X., Moriyama E. N. (2013). Comparative studies of differential gene calling using RNA-Seq data. *BMC Bioinformatics* **14**, S7.

Figures

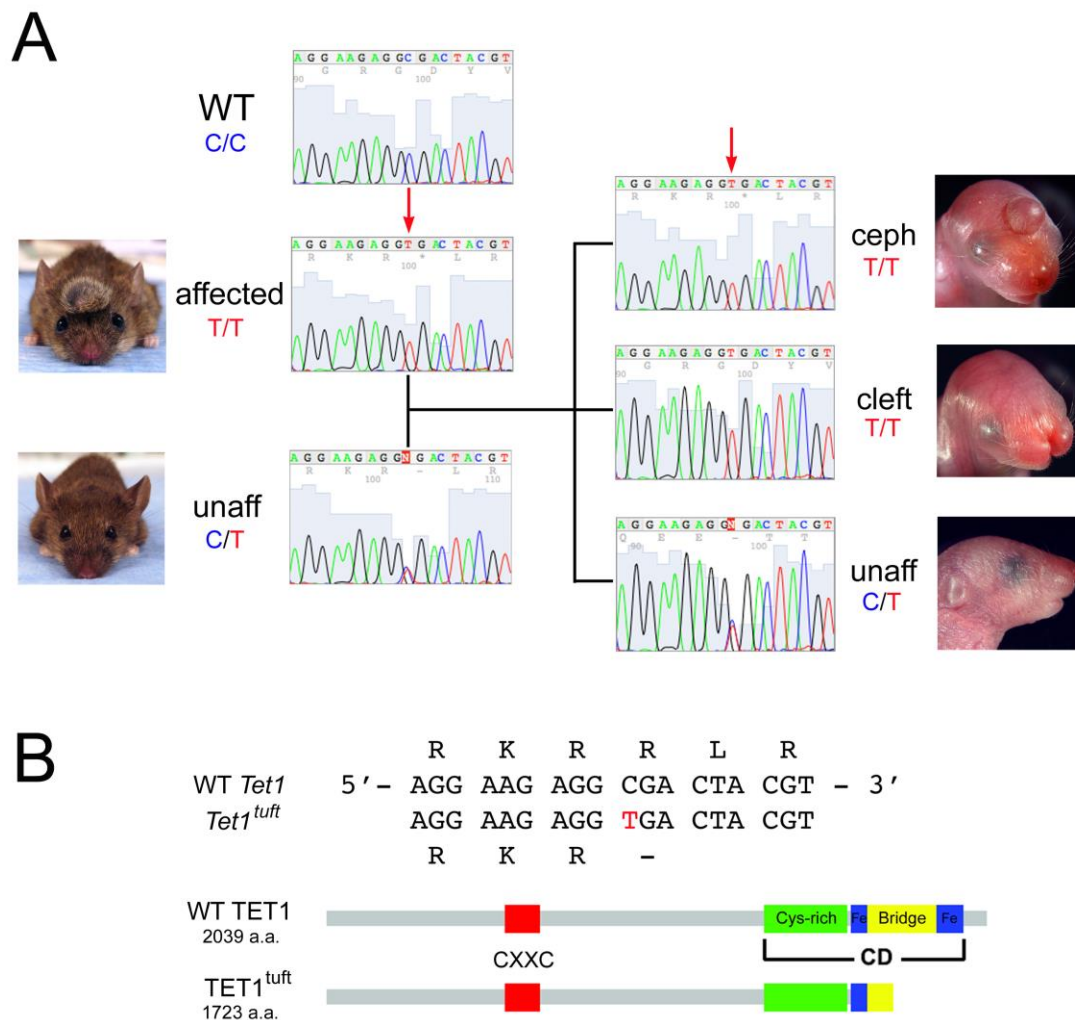
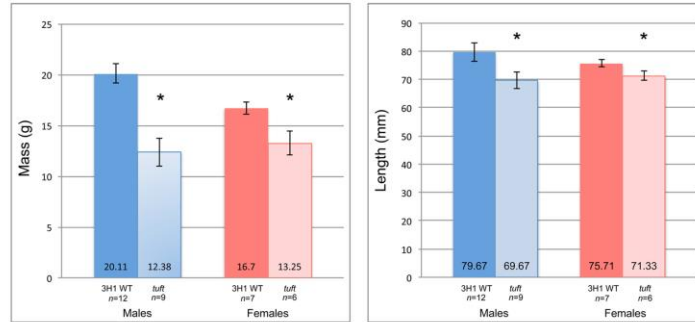
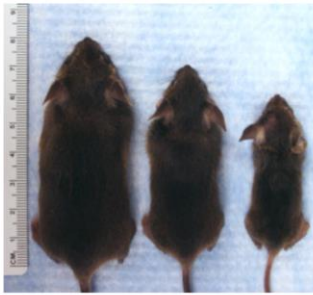


Fig. 1. Nonsense mutation in the *Tet1* gene of *tuft* mice. (A) Example Sanger sequence showing mutation (arrow) from C in wildtype (WT) or 3H1 background strain to T in affected *tuft* mice which is carried by normal appearing mice (unaff). (B) Sequence indicating the C>T mutation resulted in a truncated TET1^{tuft} protein lacking the C-terminal iron-binding site (blue boxes) within the catalytic domain (CD). Domain structure adapted from Tan and Shi 2012.

A



B

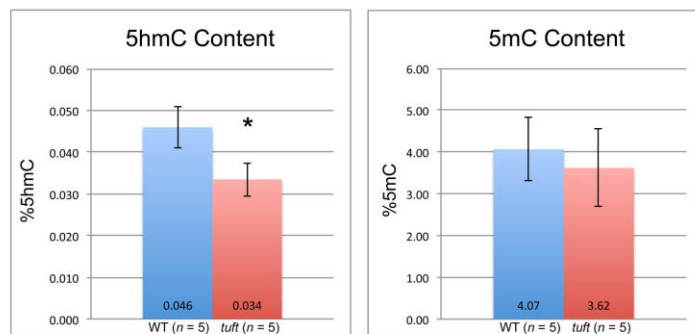


Fig. 2. Reductions in size and TET1 activity in *tuft* mice. (A) 30 day-old mice heterozygous for *Tet1^{tuft}* appearing normal (left), homozygous *Tet1^{tuft}* runt (middle), and extreme case (right). Differences in body mass (grams) and length (millimeters) between 30 day runts and 3H1 Balb wildtype (WT) male and female mice. Mean values measured from *n* mice indicated at the base. Error bars indicate SD. Significance determined by two-tailed unpaired Student's *t* test indicated $p < 0.005$ (*). (B) DNA purified from dissected rostrums of E9 embryos demarcated by the red line. Amount of 5hmC and 5mC in rostrums of E9 embryos homozygous for the *Tet1^{tuft}* mutation compared with 3H1 Balb wildtype background strain (WT). $p = 0.002$ (*).

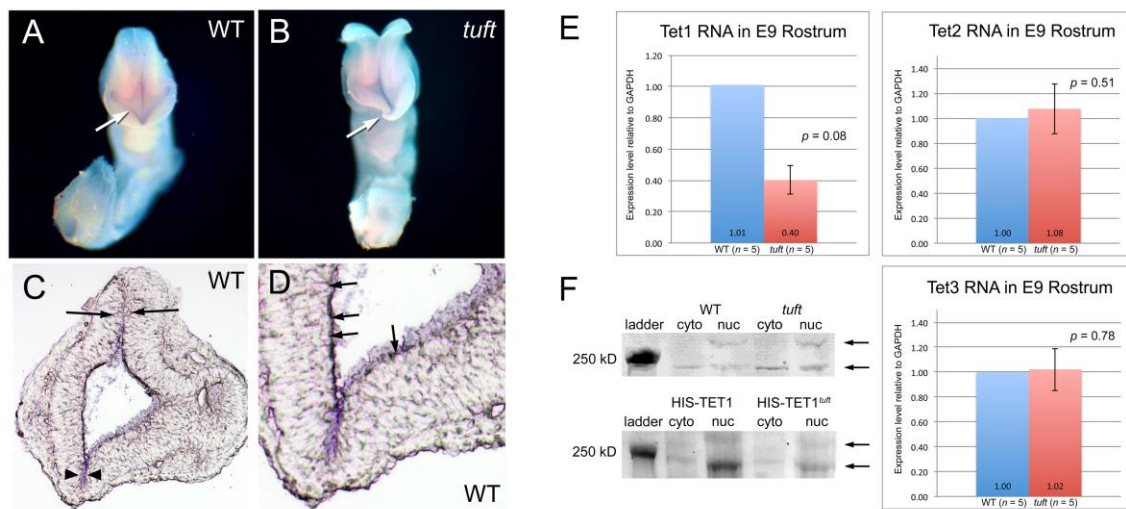


Fig. 3. Tet1 expressed in neuroectoderm of E9 heads. (A) Whole mount staining revealed *Tet1* riboprobes (purple) along the neuroectoderm (arrow) of E9 wildtype (WT) and (B) *Tet1^{tuft}* embryos. Curled rostrum (arrow), open midbrain and less intense purple staining along the basal neuroectoderm is noted in the *Tet1^{tuft}* mutant. (C) Coronal section of WT embryo head stained for *Tet1* riboprobe as shown in (A), indicating expression along ventral (arrowheads) and dorsal midline (arrows) where the neural folds adhere, enclosing the neuroectoderm. (D) Magnified area of the ventral region shown in (C) indicating stained basal cells along neuroectoderm (arrows). (E) Fold change differences of *Tet1*, *Tet2*, *Tet3* RNA in *Tet1^{tuft}* E9 rostrums compared with 3H1 Balb (WT) normalized to GAPDH levels. Mean values of biological replicates (*n*) at the base with error bars marking SD. Statistical significance indicated as *p* values. (F) Western blots indicating TET1 protein (arrows) in cytoplasmic (cyto) and nuclear fractions (nuc) in wildtype (WT) and *Tet1^{tuft}* mice near the 250 kD marker. Lower blot indicates recombinant poly histidine-tagged TET1 and TET1^{tuft} protein expressed in IMCD-3 cells (arrows).

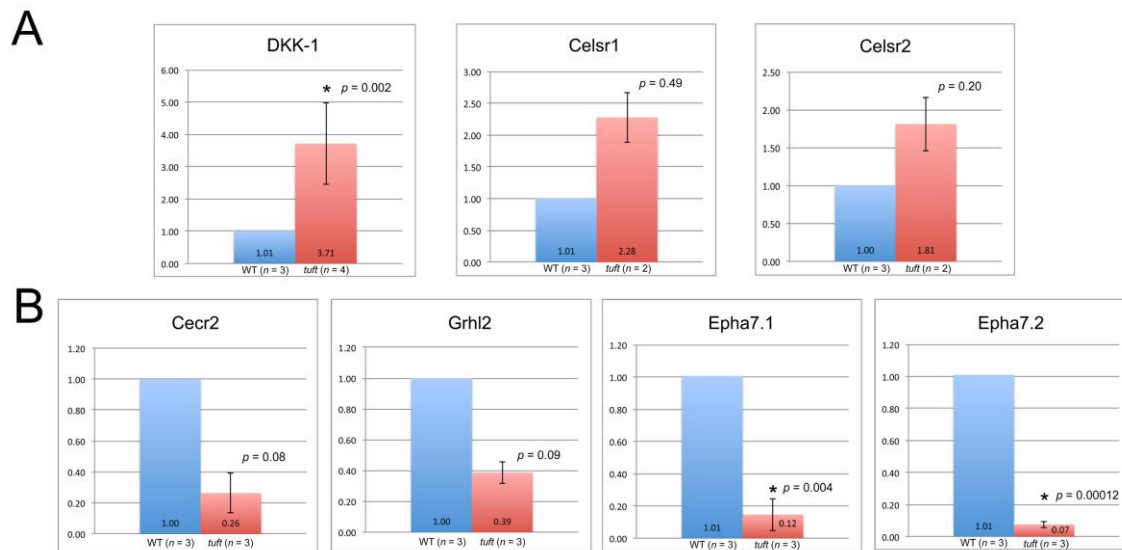


Fig. 4. Levels of RNA transcribed from genes associated with planar cell polarity and neural tube closure altered in E9 *Tet1^{tuft}* embryos. (A) Fold changes in RNA levels of *Dkk1*, *Celsr1*, *Celsr2* from *Tet1^{tuft}* embryo rostrums normalized against corresponding wildtype (WT) relative to GAPDH by qPCR. Mean values of biological replicates (*n*) indicated at the base with error bars marking SD. (B) Fold changes in RNA levels of *Cecr2*, *Grhl2*, *Epha7.1*, *Epha7.2* as in (A). Statistically significant difference when $p < 0.05$ indicated (*).

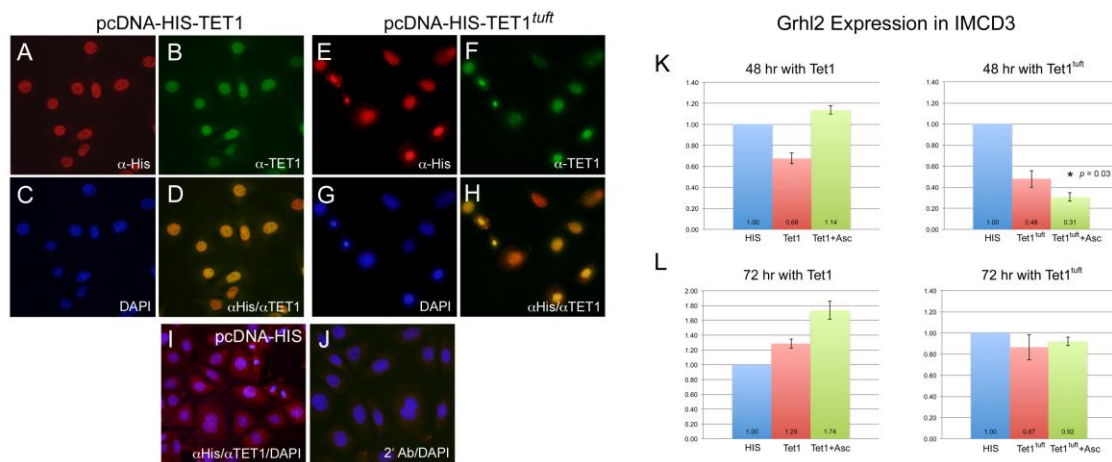


Fig. 5. Recombinant TET1^{tuft} protein reduced *Grh12* RNA in IMCD-3 cells. (A-D) Immunocytochemical localization of HIS-Tet1 and (E-H) HIS-Tet1^{tuft} recombinant protein overexpressed in IMCD-3 cells. (A, E) Cells stained for anti-poly His epitope tag (red), (B, F) the amino terminal part of TET1 (green), and (C, G) DAPI to reveal nuclei (blue). (D, H) Merged image colocalizing anti-His and anti-TET1 to nuclei (yellow). (I) IMCD-3 cells transfected with pcDNA-HIS vector stained for anti-His, anti-TET1 and DAPI. (J) HIS-Tet1^{tuft} transfected cells incubated with just secondary antibodies stained for all three colors. (K, L) Expression levels of *Grh12* in IMCD-3 cells transfected with pcDNA-HIS vector (HIS), pcDNA-HIS-Tet1 (Tet1) or pcDNA-HIS-Tet1^{tuft} (Tet1^{tuft}) with and without ascorbate (Asc) cultured for (K) 48 and (L) 72 hours, relative to GAPDH. Mean values of replicates indicated at the base with error bars marking s.d. $p = 0.03$ (*).

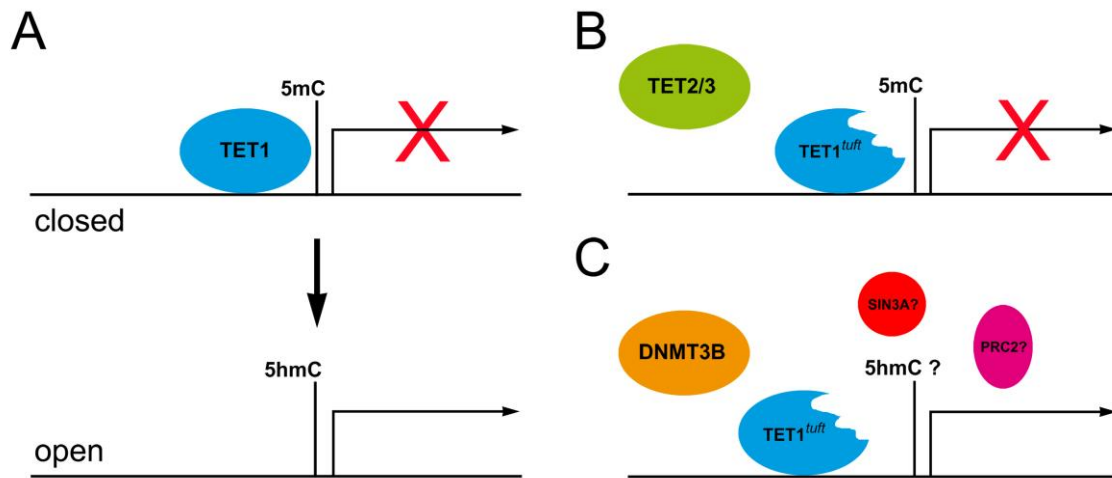


Fig. 6. Modeling TET1 function and dominant negative effect of TET1^{tuft} protein. (A) TET1 normally associates with 5mC, where gene transcription is generally silent or closed. Transcriptional complexes open following catalysis of 5mC to 5hmC by TET1, allowing for regulation. (B) In the case of *tuft* mice, TET1^{tuft} binds to target sites but unable to catalyze conversion of 5mC to 5hmC, thus permitting transcriptional repression. Presence also interferes with compensation from TET2 or TET3 activity. (C) TET1^{tuft} may also be unable to interact with repressor complexes (SIN3A, PRC2) or allow methyltransferase activity (DNMT3B) to suppress gene transcription.