# THE ROLE OF TET1 AND TET1$^{ALT}$ IN CANCER

A Dissertation
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
DOCTOR OF PHILOSOPHY

by
Charly Ryan Good
December 2017

Examining Committee Members:

Jean-Pierre Issa, MD, Advisory Chair, Fels Institute and Department of Medicine
Carmen Sapienza, PhD, Fels Institute and Department of Pathology
Nora Engel, PhD, Fels Institute and Department of Biochemistry
Alfonso Bellacosa, MD/PhD, Fox Chase Cancer Center
Xiaowei Chen, PhD, External Member, Fox Chase Cancer Center

# ABSTRACT

## THE ROLE OF TET1 AND TET1$^{ALT}$ IN CANCER

**Charly Ryan Good**

**Doctor of Philosophy**

**Temple University, 2017**

**Doctoral Advisory Committee Chair: Jean-Pierre Issa, M.D.**

DNA hypermethylation is known to contribute to the formation of cancer and this process has been widely studied. However, DNA hypomethylation has received far less attention and the factors controlling the balance between hypo and hypermethylation and its impact on tumorigenesis remains unclear. TET1 is a DNA demethylase that regulates DNA methylation, hydroxymethylation and gene expression. Full length TET1 (TET1$^{FL}$) has a CXXC domain that binds to un-methylated CG islands (CGIs). This CXXC domain allows TET1 to protect CGIs from aberrant methylation but it also limits its ability to regulate genes outside of CGIs.

This dissertation reports a novel isoform of TET1 (TET1$^{ALT}$) that has a unique transcription start site from an alternate promoter in intron 2, yielding a protein with a unique translation start site. Importantly, TET1$^{ALT}$ lacks the CXXC domain but retains the catalytic domain. TET1$^{ALT}$ is repressed in ESCs but becomes activated in embryonic and adult tissues while TET1$^{FL}$ is expressed in ESCs, but repressed in adult tissues. Overexpression of TET1$^{ALT}$ shows production of 5-hydroxymethylcytosine with distinct (and weaker) effects on DNA methylation or gene expression when compared to TET1$^{FL}$. TET1$^{ALT}$ is aberrantly activated in multiple cancer types including breast, uterine and glioblastoma and TET1 activation is associated with a worse overall survival in breast, uterine and ovarian cancers.

Indeed, we provide evidence that TET1 acts as an oncogene in triple negative breast cancer (TNBC), a subtype of breast cancer that does not overexpress the hormone receptors or the HER2/NEU oncogene. Importantly, TNBCs are one of the most hypomethylated cancers observed and are often a therapeutic challenge because of advanced presentation and lack of targeted therapies. TET1 and TET1$^{ALT}$ mRNA are upregulated specifically in TNBC tumors that display genome-wide hypomethylation and activation of genes in the PI3K-Akt-mTOR pathway. Furthermore, this hypomethylation is mutually exclusive with activating PI3K mutations, suggesting demethylation may be an alternate mechanism to activate this oncogenic pathway. In TET1 knock out (KO) cells, we observed a reduction in phospho-4E-BP1 and a downregulation of genes in the PI3K pathway, suggesting loss of PI3K/mTOR activity is concomitant with loss of TET1. Additionally, TET1 KO cells displayed a significant decrease in cellular proliferation and migration. Our work establishes TET1 as an oncogene that contributes to the aberrant hypomethylation observed in cancer and suggests TET1 could serve as a novel druggable target for therapeutic intervention.

I would like to dedicate this work to my parents,

Thomas Charles Ryan Jr and Charlotte Foretich Ryan,

who have given me confidence and encourgament for the last 30 years.

And to my husband, Austin Lewis Good, for his love and support.

# ACKNOWLEDGMENTS

I would like to start by acknowledging my mentor, Dr. Jean-Pierre Issa. I am eternally grateful for your support and guidance over the last five years. You have given me the freedom to be creative with my project and the independence to explore areas most interesting to me. You taught me by example, to be a better writer, listener, presenter and thinker. Your constant support and encouragement helped me reach my fullest potential as a scientist. Although I originally did not plan to stay in academia, with time, you convinced me that I should. You saw potential in me as a scientist and wouldn't stop until I saw that potential too. You have given me the foundation and confidence to be successful. Thank you for pushing me to write grants and apply for travel awards, and allowing me to share my research with the scientific community at conferences. Without your expertise, this project would have not been possible, and for that I am grateful.

I would also like to thank my thesis committee, Dr. Nora Engel, Dr. Carmen Sapienza and Dr. Alfonso Bellacosa who have continuously provided critical feedback throughout my graduate work. Your unique areas of expertise have been valuable as I have pushed my project forward. Most importantly, your continuous support to help in any way has been appreciated. I especially thank Dr. Bellacosa for making the drive from Fox Chase Cancer Center to attend my biannual committee meetings. I would also like to thank Dr. Xiaowei Chen for agreeing to be my external examiner. I appreciate your time and breast cancer expertise.

Being part of the Issa lab has been as great experience in part due to the wonderful lab members (past and present). Dr. Judit Garriga, you were instrumental in

my success in the lab. During the first few years you taught me almost every technique I know. You took an interest in me as a scientist and personally, and I will always be grateful for our friendship. Dr. Shoghag Panjarian and Sandra Deliard, you two have made lab time fun over the past few years, with all of our water walks and pow-wows. You two are always willing to help troubleshoot and talk about alternative approaches when things aren't working. I will miss you guys dearly. I would also like to thank Dr. Jaroslav Jelinek for teaching me everything there is to know about DNA methylation, and for being so helpful. Dr. Jozef Madzo, you were always willing to teach me bioinformatics tricks and were invaluable as I was submitting my $TET1^{ALT}$ paper. Dr. Woonbok Chung, Dr. Andrew Kelly, Dr. Takahiro Sato, Dr. Priyanka Madireddi, Dr. Somnath Pandey and Hanghang Zhang, thank you for your support and for your helpful advice/suggestions over the last five years.

I would also like to thank my parents (Chuck and Charlotte Ryan) for the emotional support they have provided not only throughout my PhD, but for the last three decades. None of this would have been possible without you. And to my loving husband Austin Good, who supported me not only as a husband, but also as a fellow scientist. You showed interest in my project, and provided great advice when I needed it most. I also appreciate the times you helped me troubleshoot pesky cloning experiments. You happily listened to me practice presentations and were a great editor for my papers/grants. I thank you for picking up the extra load around the house at my busiest times and for being the highlight of my day, everyday.

# TABLE OF CONTENTS

Page

ABSTRACT.......................................................................................................... ii

DEDICATION..................................................................................................... iv

ACKNOWLEDGMENTS ........................................................................................v

LIST OF TABLES.................................................................................................. xi

LIST OF FIGURES .............................................................................................. xii

LIST OF ABBREVIATIONS...............................................................................xv

CHAPTER 1. INTRODUCTION

1.1 Epigenetics and cell states .........................................................................2

1.2 DNA methylation in mammals ...................................................................3

1.3 DNA methylation machinery......................................................................6

1.4 DNA methylation and gene expression ......................................................8

1.5 DNA methylation in cancer ........................................................................9

1.6 TET enzymes ............................................................................................13

1.7 TET1 as a DNA demethylase ...................................................................16

1.8 TET1 interacting partners .........................................................................18

1.9 TET1 as a transcriptional co-regulator .....................................................19

1.10 TET1 in cancer.........................................................................................20

1.11 Breast cancer............................................................................................21

1.12 PI3K-Akt-mTOR signaling in cancer ......................................................22

1.13 Hypothesis and specific aims...................................................................25

CHAPTER 2. METHODS

2.1 Cell culture....28

2.2 Western blot and DNA slot blot....29

2.3 TET1 expression constructs....29

2.4 Digital restriction enzymes analysis of methylation (DREAM)....30

2.5 RNA-sequencing....31

2.6 RNA isolation and qPCR analysis (cell lines)....31

2.7 Luciferase reporter assay....32

2.8 Mouse tissues....32

2.9 CRISPR....33

2.10 Bisulfite pyrosequencing....34

2.11 PCR and clonal sequencing....35

2.12 Proliferation and cell migration....35

2.13 Bioinformatics and statistics....36

CHAPTER 3. RESULTS

3.1 Identification of a novel TET1 isoform

3.1.1 Location of TET1$^{ALT}$....42

3.1.2 Detecting mRNA of TET1$^{ALT}$....43

3.1.3 TET1$^{FL}$ and TET1$^{ALT}$ are differentially expressed during mouse development....47

3.2 TET1$^{ALT}$ promoter activity

3.2.1 Histone marks....47

3.2.2 Transcription factor binding ............................................................49

3.2.3 Confirming TET1$^{ALT}$ promoter activity..........................................57

3.3 TET1$^{ALT}$ produces a truncated protein

3.3.1 Western blot for TET1 in a panel of cell lines.................................57

3.3.2 TET1$^{ALT}$ is a catalytically active DNA demethylase..........................58

3.3.3 Knockout of TET1$^{ALT}$ reduces 5hmC...............................................59

3.4 TET1$^{ALT}$ is functional and distinct from TET1$^{FL}$

3.4.1 Gene expression targets of TET1$^{ALT}$ and TET1$^{FL}$ ..............................65

3.4.2 DNA methylation targets of TET1$^{ALT}$ and TET1$^{FL}$ ............................65

3.5 TET1$^{ALT}$ is overexpressed in cancer

3.5.1 Transcription start site reads.............................................................69

3.5.2 TET1 expression in cancer.................................................................70

3.5.3 TET1 may be playing a more oncogenic role in select cancers...........71

3.6 Expression of the TET enzymes in breast cancer

3.6.1 Expression of TET1, TET2 and TET3...............................................81

3.6.2 TET1 associates with survival in TNBC ...........................................83

3.7 TET1 expression correlates with hypomethylation in TNBC

3.7.1 TET enzyme expression vs DNA methylation changes ......................84

3.7.2 Identification of TET1 targets in TNBC and HRBC...........................89

3.7.3 TET1 targets are enriched for oncogenic signaling pathways.............96

3.8 Gene expression changes between Cluster 1 and Cluster 2 patients .....................96

3.9 Loss of TET1 leads to loss of PI3K/mTOR signaling

3.9.1 Clinical features of patients in Cluster 1 and Cluster 2 .......................97

3.9.2 Phenotype of TET1 KO ....................................................................103

3.9.3 Vulnerabilities imparted by TET1 expression in breast cancer .........103

3.10 Gene expression targets of TET1 in vitro .........................................................104

3.11 Methylation targets of TET1 in vitro ................................................................105

3.12 TET1 mediated hypomethylation in serous ovarian cancer

3.12.1 Identification of TET1 targets in ovarian cancer ............................112

3.12.2 Vulnerabilities imparted by TET1 expression in ovarian cancer .........................................................................................113

CHAPTER 4. DISCUSSION...........................................................................................119

REFERENCES CITED.....................................................................................................128

# LIST OF TABLES

Table Page

1. qRT-PCR primers……………………………………………….……………....... 39
2. Primers for luciferase reporter assay…………………………….…..…......... 39
3. Lenti CRISPR V2 gRNA sequences for TET1 KO……………………………. 39
4. gRNA sequences for CRISPR activation in pLKO.1 vector…………….…..........40
5. Pyrosequencing primers………………………………………………….…... 40
6. PCR and clonal sequencing primers……………………………………….. 40
7. Clinical characertistics of patients in Cluster 1 and Cluster 2…………………..102
8. Top correlated drugs in breast cancer……………………………………….. 107
9. Top correlated drugs in ovarian cancer…………………………………... .118

## LIST OF FIGURES

Figure Page

1. DNA methylation dynamics……………………………………….…………………… 5

2. DNMT proteins……………………..…..…………………………………… 7

3. TET enzymes……..……..........................................................................14

4. TET-mediated DNA demethylation………..……………………….........17

5. PI3K signaling…………………………..…………………………..…...24

6. The discovery of $\mathrm{TET1}^{\mathrm{ALT}}$…………………………………………………….. 44

7. TET1 and $\mathrm{TET1}^{\mathrm{ALT}}$ exon conservation ……………………………………….. 45

8. PCR of $\mathrm{TET1}^{\mathrm{ALT}}$ exons 1a and 1……………………………………………. 46

9. Differential expression of the TET1 isoforms during mouse development……. 50

10. TET1 gene model …………….……………………………………………….. 51

11. TET1 and $\mathrm{TET1}^{\mathrm{ALT}}$ promoter histone marks in human ……………………….. 52

12. Chromatin marks surrounding the $\mathrm{TET1}^{\mathrm{FL}}$ and $\mathrm{TET1}^{\mathrm{ALT}}$ promoters in human cell lines…………………………………………………………………... 53

13. TET1 and $\mathrm{TET1}^{\mathrm{ALT}}$ promoter histone marks in mouse ……………………….. 54

14. Chromatin marks surrounding the $\mathrm{TET1}^{\mathrm{FL}}$ and $\mathrm{TET1}^{\mathrm{ALT}}$ promoters in mouse tissues ……………………………………………………………………….. 55

15. Transcription factor binding at the $\mathrm{TET1}^{\mathrm{ALT}}$ promoter ……………………….. 56

16. $\mathrm{TET1}^{\mathrm{ALT}}$ promoter activity ……………………………………………………..60

17. TET1 western blot in a panel of cell lines …………………………………….. 61

18. Overexpression of $\mathrm{TET1}^{\mathrm{FL}}$ and $\mathrm{TET1}^{\mathrm{ALT}}$ increases 5hmC ……..…………….. 62

19. TET1 KO in MDA-MB-231 cells ………………………….……………….. 63

20. TET1 KO results in a loss of 5hmC …………….…………………………….. 64

21. Gene expression targets of $\mathrm{TET1}^{\mathrm{FL}}$ and $\mathrm{TET1}^{\mathrm{ALT}}$ …….……………………..67

22. Methylation targets of TET1$^{FL}$ and TET1$^{ALT}$ …………….………………….. 68

23. TET1$^{ALT}$ and TET1$^{FL}$ TSS-seq data in normal human tissues ……………….. 72

24. TET1 expression in normal tissues …………………….………………….. 73

25. TSS of TET1$^{FL}$ and TET1$^{ALT}$ in normal tissue vs cancer …………………….. 74

26. TSS-seq data from 22 lung cancer cell lines at the TET1$^{FL}$ and TET1$^{ALT}$ transcription start sites……………………………………………………..75

27. TET1$^{ALT}$ expression in breast cancer ……………………………………….. 76

28. TET1$^{ALT}$ expression in glioblastoma and uterine cancer …………………….. 77

29. TET1 exon usage in cancer ………………..…………………………….. 78

30. TET1 expression and amplification in cancer (TCGA datasets)……….…….. 79

31. Overall survival analysis based off of TET1 expression …………………….. 80

32. TET1, TET2 and TET3 expression in breast cancer ………………………….. 82

33. DNMT expression in breast cancer ………………………….…………….. 85

34. Immunohistochemical staining of TET1 in breast cancer …………………….. 86

35. Isoform specific expression for TET1 in TNBC ……………….……….. 87

36. TET1 associates with survival in TNBC …………………….…………….. 88

37. TET1 correlates with genome-wide methylation changes in TNBC ………….. 90

38. TET2 and TET3 expression vs DNA methylation changes in cancer ……….. 91

39. Spearman analysis of TET1 expression vs DNA methylation ………….…….. 93

40. TET1 hypomethylated targets……………………………. ……….……….. 94

41. TET1 hypomethylated targets with HRBCs….………………………………. 95

42. Pathways enriched in hypomethylated TET1 targets …………………………..98

43. Validation with 27K array …………………………………………………….. 99

44. Genes differentially expressed between Cluster 1 and Cluster 2….…………... 100

45. Genes upregulated and hypomethylated..…………………………………….. 101

46. Loss of TET1 leads to loss of PI3K/mTOR signaling and reduced cellular migration and proliferation …………………………………………………..106

47. Drug sensitivity vs TET1 expression in breast cancer cell lines …………..…...108

48. Gene expression targets of TET1 in vitro ……………………………………...109

49. ANGPT1 and THBS3 expression in TCGA clusters and TET1 KO cells ..…...110

50. DNA methylation targets of TET1 in vitro ………………….……………....114

51. Methylation and gene expression targets in TCGA and TET1 KO cells ……....115

52. INSR gene expression and methylation in TCGA and TET1 KO cells………...116

53. TET1 mediated hypomethylation in ovarian cancer …………………………...117

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AML | Acute Myeloid Leukemia |
| BC | Breast Cancer |
| µg/µL/µM | Microgram/microliter/micromolar |
| 5caC | 5-carboxylcytosine |
| 5fC | 5-formylcytosine |
| 5hmC | 5-hydroxymethylcytosine |
| 5mc | 5-methylcytosine |
| a-KG | Alpha Keto Glutarate |
| ATCC | American Type Culture Collection |
| bp | Base Pair |
| CAGE | Cap Analysis of Gene Expression |
| CGIs | CG Islands |
| CIMP | CG island methylator phenotype |
| DNA | Deoxyribonucleic Acid |
| DREAM | Digital Restriction Enzyme Analysis of Methylation |
| ER | Estrogen Receptor |
| GPCR | G protein Coupled Receptors |
| hESCs | Human Embryonic Stem Cells |
| HMLE | Human Mammary Epithelial Cells |
| HRBC | Hormone Receptor Positive Breast Cancer |
| KD | Knock Down |
| KDa | Kilo Daltons |
| KO | Knock-out |
| MBD | Methyl-CG-binding domain |
| mESCs | Mouse Embryonic Stem Cells |
| MSI | Microsatellite Instability |
| NBE | Normal Breast Epithelium |
| PBS | Phosphate Buffered Saline |
| PCR | Polymerase Chain Reaction |
| qRT-PCR | Quantitative Real Time Polymerase Chain Reaction |
| RNA | Ribonucleic Acid |
| RTK | Receptor Tyrosine Kinase |
| SEER | Surveillance, Epidemiology, and End Results Program |
| sgRNA | small guide RNA |
| stdev | Standard Deviation |
| TCGA | The Cancer Genome Atlas |
| TNBC | Triple Negative Breast Cancer |
| TSS | Transcription Start Site |

# CHAPTER 1

# INTRODUCTION

## 1.1 Epigenetics and cell states

Every human cell has the exact same DNA sequence, but each cell type has a diverse function thanks to an intricate network of transcriptional programs and epigenetic codes that selectively use particular parts of the DNA in a cell type-specific manner. During development, a single-celled embryo will differentiate into the 100s of cell types found in the human body. This process follows a strict step by step order of events that is led by dynamic transcriptional changes. For example, in embryonic stem cells, genes involved in pluripotency are turned on, while lineage specific genes are turned off. In addition to transcriptional programs, during differentiation epigenetic changes take place that further solidify a cell's fate. The epigenome ("epi" meaning above, above the genome) has the ability to change gene expression programs without changing the DNA sequence [1]. These epigenetics modifications, such as DNA methylation, are stable and help prevent the cell from undergoing de-differentiation [2]. Importantly, these epigenetic states are heritable through mitosis and allow for the cell to stay true to its specific cell type upon replication [2].

DNA is tightly wound around nucleosome proteins that help to compact the DNA, resulting in the dense chromatin structure found in chromosomes. Each nucleosome is made of an octamer of core histone proteins including H3-H4 tetramer and 2 H2A-H2B dimers [1]. Histones themselves are small basic proteins that contain a charged NH2 terminus (the histone tail) that sticks out of the nucleosome, making it susceptible to modifications [1]. Acetylation, phosphorylation or methylation of these histone tails make up the histone code that provides further instructions for cellular gene expression programs [1]. Importantly, histone modifications are more dynamic than DNA methylation and allow for quick activation or inactivation of gene expression programs.

Histone marks are associated with various gene states. Heterochromatin is a densely packed and closed conformation that is not amenable to active transcription [1]. Histone marks associated with heterochromatin include H3K9me3, H3K27me3 and H4K20me3 [3]. Euchromatin is more accessible and conducive to active transcription and is often associated with H3K4me2/3, H3K9me1 and hyperacetylation of histones H3 and H4 [4]. In addition, certain histone marks are associated with genic features. The methylation of histone H3 at lysine 4 (H3K4me3) is often associated with promoters, whereas H3K4me1 is often found at enhancer sites [5]. Further, acetylation of histone H3 at lysine 27 (H3K27Ac) is associated with active enhancers [6]. The histone code, along with DNA methylation, helps to determine cell states.

## 1.2 DNA methylation in mammals

DNA methylation is a highly conserved process found in most plants, fungi, bacteria, and insects and is found in all mammals. In mammals, DNA methylation mostly occurs on cytosine when followed by guanosine (CG) sites. However, in many species, including mammals, additional bases, such as adenine, have also been found to be methylated [7]. In mouse embryonic stem cells, N6-methyladenine correlates with epigenetic silencing of transposons, such as LINE1 elements [7]. However, the predominant form of methylation found in mammals is CG methylation, which is important for processes including gene regulation, early embryonic development, X chromosome inactivation, genomic imprinting and cancer [89]. In addition, it helps to repress transposable elements, where the high levels of methylation prevent the elements from moving sporadically throughout the genome [10].

The human genome contains 28 million CG sites, which are typically either methylated, hydroxymethylated or un-methylated [11]. 60-80% of all CG sites are methylated, and for the sites that are un-methylated there is a strong evolutionary mechanism protecting them from gaining methylation [12]. Most of the CG sites that are un-methylated are in CG islands (CGIs), which are segments of DNA that are highly enriched for CG dinucleotides. CGIs are on average approximately 1000 base pairs (bp) in length and the GC percentage within the CGI is greater than 50% [13]. Most promoter CG islands are un-methylated, except in cases such as X-chromosome inactivation and genomic imprinting in which case methylation is essential to maintain silencing of select alleles [14]. About 70% of genes have a CGI associated with their promoter, including all housekeeping genes [13]. Furthermore, CGI methylation is less variable between tissue types, whereas non CGI methylation is much more dynamic and cell type specific [15]. Active genes typically have un-methylated promoters and enhancers, whereas methylation varies at inactive gene promoters (Figure 1) [16]. Inactive gene promoters are usually methylated if they are in CG poor regions, but are un-methylated if they are in CG rich regions [16].

5mC is stable, as the methyl group forms a stable carbon-carbon bond with the 5-position of cytosine [12]. Although DNA methylation is stable, it can be reversed to an un-methylated state in several ways. Passive demethylation is the gradual loss of methylation following replication. A lack of the proteins that methylate DNA would

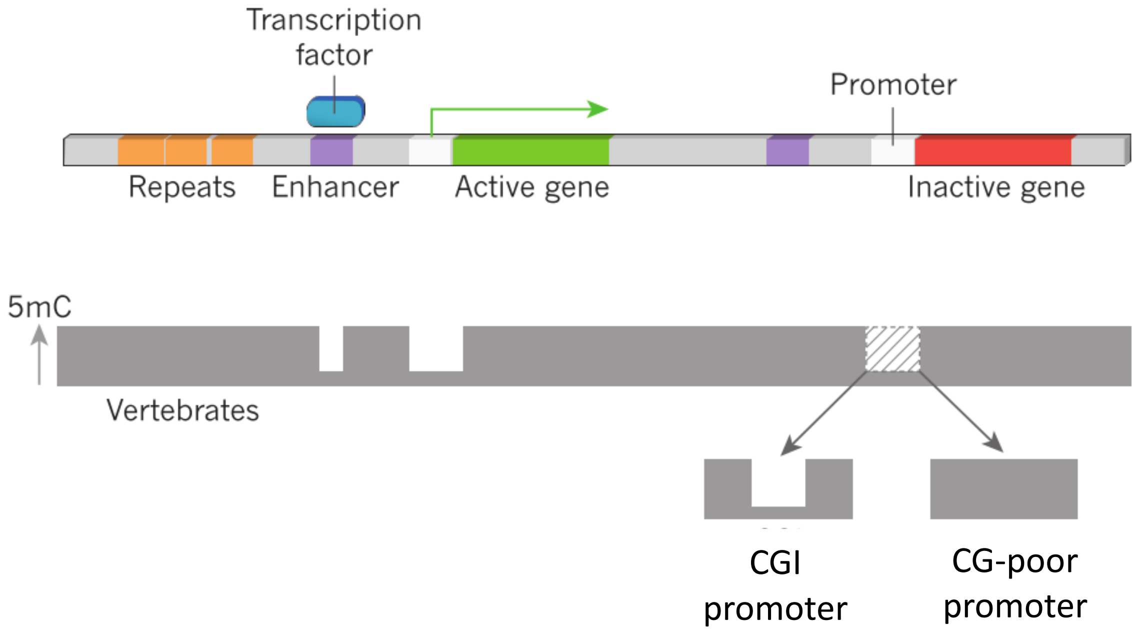


**Figure 1. DNA methylation dynamics [16].**

Distribution of DNA methylation in a typical genome for an active gene (green) and repressed gene (red). Upstream enhancers are depicted for both genes. The height of the methylation bar (in gray) indicates the levels of 5-methycytosine often observed in these genomic regions. CG islands (CGIs) often overlap with promoters and these regions often remain un-methylated, even when the gene is inactive. However, CG poor promoters are methylated when inactive.

result in passive DNA demethylation whereby following replication the newly synthesized strand is not methylated [16]. Another mechanism of demethylation is active demethylation, which is replication independent and involves the physical removal of the methyl group either through base excision repair or through the oxidation by the TET family of DNA demethylases [17]. Active demethylation has been shown to occur during development and at specific loci in somatic cells [10]. For example, activated T lymphocytes undergo robust demethylation (within 20 minutes) at the IL-2 promoter/enhancer region [18].

Methylation changes drastically during embryogenesis. Shortly after fertilization, the paternal pronucleus undergoes active and passive genome-wide DNA demethylation mediated in part by TET3 (see below) and the maternal genome undergoes a more gradual, passive demethylation through multiple rounds of replication [19]. After implantation, de-novo methylation patterns are re-established by the DNA methyltransferase enzymes (see below) [10].

## 1.3 DNA methylation machinery

The DNA methyltransferase family of enzymes (DNMT1, DNMT3A, DNMT3B) all have a methyl transferase catalytic domain that is responsible for adding methyl groups to cytosines, using S-adenosyl methionine (SAM) as a methyl donor [17] (Figure 2). DNMT3L and DNMT2 are catalytically inactive, but DNMT3L is able to modulate the activity of the other DNMT family members [20] [21]. In addition, DNMT1 also has a CXXC domain that allows the protein to bind to CG rich regions of DNA.

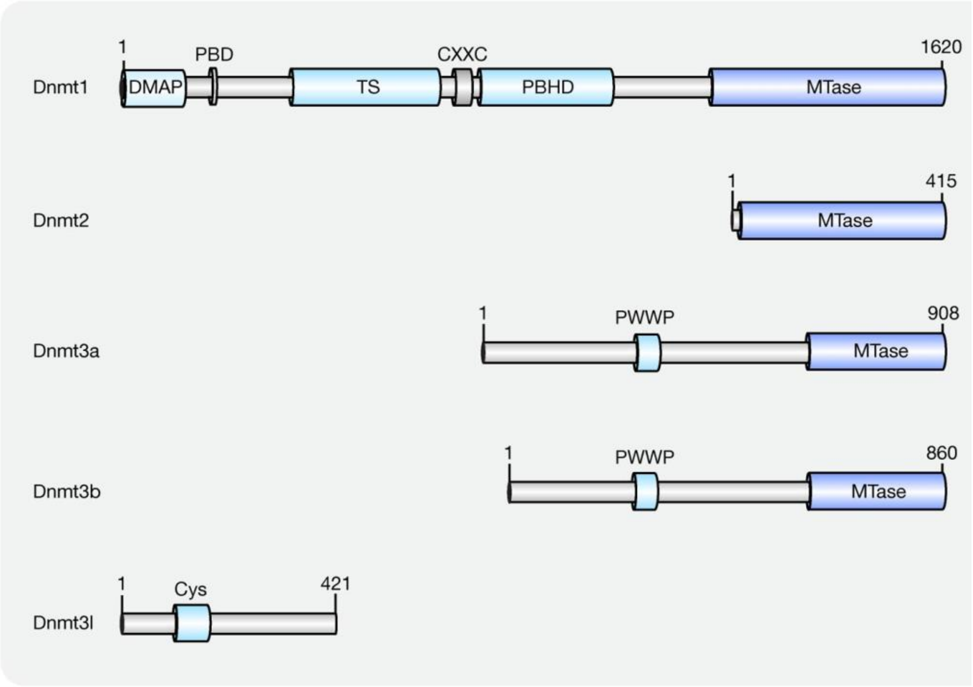


**Figure 2. DNMT proteins [22].**

Numbers represent the amino acid positions. All DNMT proteins have a methyl transferase domain (MTase), except for DNMT3L. Additional domains in DNMT1 include Dnmt1-associated protein binding domain (DMAP), proliferating cell nuclear antigen (PCNA)-binding domain (PBD), targeting sequence (TS), CXXC zinc finger domain (CXXC) and polybromo-1 protein homologous domain (PBHD). The proline-tryptophan-tryptophan-proline motif (PWWP) domain is found in DNMT3A and DNMT3B. DNMT3L has a cysteine-rich domain (Cys).

In somatic cells, methylation is maintained by the maintenance methyltransferase DNMT1. DNMT1 is found at replication foci where it serves to copy methylation patterns after DNA replication [18]. DNMT1 has a 10-40 fold preference for hemi-methylated DNA as opposed to completely un-methylated DNA [23] [24]. UHRF1 recognizes hemi-methylated DNA and recruits DNMT1 to the site for re-methylation [25]. This mechanism ensures that methylation is maintained following replication.

DNMT3A and DNMT3B are the de-novo methyltransferases, which establish new methylation patterns at un-methylated DNA [18]. Importantly, study of the DNMTs led scientists to conclude that DNA methylation is essential for mammals. Germline deletion of DNMT1 or DNMT3B is embryonically lethal and DNMT3A deletion results in death by four weeks, and thus methylation is essential to mammalian development [26] [27].

### 1.4 DNA methylation and gene expression

DNA methylation is often associated with blocking transcription, especially if found at promoters or enhancers. This is because DNA methylation physically blocks transcription factors from binding and methyl binding proteins (MBDs) that read the methylation signal, recruit chromatin modifying enzymes to establish a closed chromatin configuration at the methylated loci [13]. However, how DNA methylation affects transcription is largely dependent on the location of the CG site. For instance, methylation in gene bodies is often associated with active transcription, as methylation in the body prevents cryptic transcription from occurring within the gene and has been shown to stimulate transcriptional elongation [15] [28].

CG sites are often discussed by their location relative to CGIs.  A CG site can be in a CGI, CGI shore (which is +/- 2000 bp from the CGI), CGI shelf (+/- 2000 bp from the CGI shore) or noted as being away from a CGI [29].  It was reported a decade ago that CGIs are protected from gaining methylation [14] but it was not until recently that one of the proteins involved in protecting CGIs from methylation was identified [30] (see below).  It is well known that methylation of a CGI promoter can result in silenced gene expression, especially if the gene is active.  However, in embryonic stem cells (ESCs), one-fifth of CGI promoters are bivalent (contain both H3K4me3 and H3K27me3) and are inactive but poised for transcription [13].  Importantly, even though these bivalent CGI promoters are often un-methylated, they are not associated with active transcription.

Methylation at shores and shelves is less clear, although methylation at these sites is known to be much more dynamic and tends to be more cell-type specific [31].  For example, comparing DNA methylation in the brain, liver, and spleen revealed CGI methylation is not drastically different between the tissue types; however, methylation of CGI shores was very different between the three tissue types [31].  Further, methylation in shores is inversely correlated to gene expression levels, as long as the shores are within 2000 base pairs (bp) from the transcription start site [31].

## 1.5 DNA methylation in cancer

DNA methylation is a frequent mechanism utilized by cancer cells to promote tumorigenesis.  Cancer cells typically display a global loss of methylation, while at the same time obtain gains of methylation at specific CGI promoters [14].  Frequently, the gain of methylation at CGI promoters is found in tumor suppressor genes, which can promote tumorigenesis [14]. It is not uncommon to observe promoter methylation of one allele, with a truncating mutation on the other allele, suggesting methylation can serve as another hit in the Knudson two hit hypothesis [3].

The promoters of *RB1*, *MLH1*, *CDKN2A* and *BRCA1* are frequently methylated in cancer, including in retinoblastoma, colon, lung and ovarian cancers [15]. Further, promoter methylation of these genes has been shown to disrupt gene function [32]. For example, in breast cancer, the *BRCA1* promoter was found to be hypermethylated in patients who had decreased BRCA1 protein expression and the methylation was associated with lymph node metastasis and histological grade [33].

In select cancers, CGI promoter methylation occurs so frequently that patients can be classified as having the CG island methylator phenotype or CIMP. CIMP was originally described in a subset of colon cancers where *MINT1, MINT2, MINT31*, *CDKN2A* and *MLH1* promoters were found to be hypermethylated compared to normal tissue [14]. Further, a link between microsatellite instability (MSI) and CIMP has also been described, where tumors with MSI can arise either through mutations in mismatch repair genes or through promoter hypermethylation of *MLH1* [14]. Over the last decade, CIMP has been found in multiple cancer types, owing to the widespread use of this mechanism to silence genes in cancer [34] [35] [36]. Another negative implication of increased DNA methylation is that it greatly increases the rate of C→T transition mutations, which could be a major factor behind disease causing mutations [15].

Some of the factors that might lead to aberrant hypermethylation in cancer includes changes in the DNA methylation machinery. *DNMT3A* and *IDH1*/*IDH2* mutations have been linked to methylation changes in leukemia, providing mechanistic insight into hypermethylation in cancer [37] [38]. However, *DNMT3A* and *IDH1/IDH2* alterations do not explain all promoter hypermethylation and therefore more is needed to fully understand the complex regulation of methylation in cancer.

Although hypermethylation is widely studied, the first reported epigenetic change in human cancer was hypomethylation [39]. This includes global hypomethylation that can result in loss of methylation at repetitive elements, leading to genome instability [40]. In addition, hypomethylation of specific regulatory elements can result in aberrant gene activation [41]. For example, the gene encoding the protease urokinase (PLAU/uPA) is hypomethylated and overexpressed in breast and prostate cancer and correlates with tumor progression [41]. Additional genes that are specifically hypomethylated and activated in cancer include *S100A4*, *CLDN4*, *TFF2*, *SERPINB5*, and *UCHL1* [41].

In breast cancer, methylation levels strongly associate with breast cancer molecular subtypes. Hormone receptor positive breast cancers (HRBCs) are often hypermethylated compared to other breast cancer subtypes and normal breast controls and triple negative (TN) tumors have widespread genome-wide hypomethylation [42] [43] [44] [45]. Furthermore, this hypomethylation is independently associated with a worse overall survival, independent of stage, age, nodal status and hormone receptor status [34]. How tumors become hypomethylated remains unknown.

Methylation states of particular genes have been found to be associated with clinical parameters in patient samples, including disease risk, prognosis, and survival. One example is in breast cancer where methylation of *CDH13* is associated with tumor size, and methylation of *RASSF1A* is associated with a worse overall survival [46]. Since methylation levels associate with clinical outcomes, it is important to delineate the cause of aberrant methylation and investigate whether altering methylation levels in cancer may be a therapeutic option to combat disease progression. This has worked well in hematologic cancers, where DNMT1 inhibitors (Azacitidine and Decitabine) have been approved by the FDA for treatment of myelodysplastic syndromes and for select patients with acute myelogenous leukemia (AML) [35]. These hypomethylating drugs work, in part, by reactivating silenced tumor suppressor genes [47].

Importantly, both hyper and hypomethylation in cancer are defined by comparing the cancer tissue to normal control tissue. Several confounding variables complicate the use of normal tissue controls. For example, tissues are often composed of heterogeneous cell populations. In the breast, in addition to mammary epithelial cells, fat, connective tissue, lymph nodes and blood vessels also fill the region. Therefore, the location of the biopsy sample could greatly impact the observed methylome. A gene in fat tissue may be completely methylated, whereas in mammary cells it may be completely un-methylated. In this case, the normal tissue sample will appear to have 50% methylation, which might mask actual hypermethylation in cancer tissue. One way around this problem would be to separate mammary epithelial cells from the population after microdissection.

Another confounding variable is the cell of origin of the tumor tissue, which can be difficult to determine. Therefore, changes in methylation observed between normal and cancer could be due to differences in the cell of origin. Lastly, DNA methylation changes with age and differences between normal and cancer could be due to the normal aging process and not be cancer specific [48]. However, if aging is the predominant cause for the methylation change, filtering for sites that do not differ in normal should help reduce the aging effect on methylation.

## 1.6 TET enzymes

The Ten-Eleven Translocation protein (*TET1*) was originally identified due to a rare translocation with the histone methyl transferase gene MLL in AML, that resulted in a 5' MLL-TET1 3' chimera [49]. All three TET enzymes (*TET1*, *TET2*, *TET3*) have an oxidase domain that convert 5-methylcytosine (5mC) into 5-hydroxymethylcytosine (5hmC), which can then be further oxidized or converted to un-methylated cytosine [50] [51] [52] [53] (Figure 3). 5hmC can be further converted by the TET enzymes to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). These oxidized forms 5hmC, 5fC and 5caC can be converted to un-methylated cytosine either by replication-dependent dilution or by removal of the group by thymine DNA glycosylase (TDG) mediated excision followed by base excision repair [12] (Figure 4). TET1 and TET3 proteins also have CXXC domains, which recognize and bind to un-methylated CGIs [54]. TET2 originally had a CXXC domain, but during evolution, the CXXC domain of TET2 separated from the rest of the protein [55]. The former CXXC domain of TET2 is known as IDAX. Interestingly, IDAX has been shown to recruit TET2 to CGIs [56].

All TET proteins require oxygen and alpha ketoglutarate (a-KG) as substrates and Fe(II) as a co-factor to generate $CO_2$ and succinate [12]. The double-stranded beta helix (DSBH) domain at the C terminus of the protein, brings together iron, a-KG, and 5mC together for oxidation [12]. A-KG is an intermediate of the citric acid cycle and is the

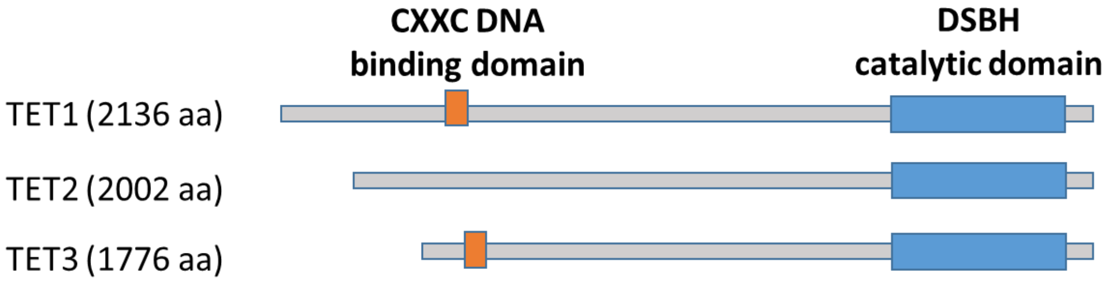


**Figure 3. TET enzymes.**

All three TET proteins have the DSBH catalytic domain that allows the proteins to oxidize 5-methyl cytosine. TET1 and TET3 also have CXXC domains which allow them to recognize and bind to CG rich regions of DNA.

product of the isocitrate dehydrogenase reaction, which is catalyzed by the isocitrate dehydrogenase enzymes (IDH) [57]. Interestingly, IDH1/2 is mutated in several cancers and mutant IDH can convert a-KG to 2-hydroxyglutarate which has been shown to be a potent inhibitor of the TET enzymes [12]. As such, patients with mutant IDH display abnormal methylation levels, likely due to limited activity of the TET enzymes [37]. Further, since IDH is responsible for producing a major substrate for the TET enzymes, overexpression of IDH1 or IDH2 has been shown to increase TET activity (increase 5hmC) [12]. Another factor that has been shown to modulate TET activity is Vitamin C. Vitamin C interacts with the catalytic domain of the TET enzymes, and increases their activity [58].

The TET enzymes display tissue specific expression, where TET1 is primarily expressed in embryonic stem cells (ESCs) but has been shown to be expressed at low levels in select adult tissues [59] [60] [61]. Although it was reported that loss of Tet1 in ESCs leads to defects in differentiation and self-renewal, these findings were shown to be an artifact of the shRNA used in the study [62] [63]. Indeed, loss of Tet1 in ESCs does not affect pluripotency or embryonic development, and Tet1 KO mice remain viable and fertile [59]. Even Tet1 and Tet2 double knockout mice produce some viable and fertile pups [64]. However, Tet3 KO is lethal as embryos die either at embryonic day 11.5 or after birth, highlighting the importance of Tet3 in development [65].

The TET proteins can be post transcriptionally regulated. For example, several miRNAs have been implicated to regulate TET1, including miR-22, miR-26 and miR-29 [12] [66]. Additionally, TET proteins have been reported to be post-translationally modified by phosphorylation, GlcNAcylation and PARylation [12].

## 1.7 TET1 as a DNA demethylase

Much of what we know about TET1's role as a demethylase has come from studies in mice, with a particular focus on embryonic stem cells and methylation changes that occur during development. Tet1 knockout in ESCs skews differentiation towards specific lineages; however, Tet1 knockout mice are viable, suggesting loss of Tet1 does not abrogate pluripotency and development [59]. Tet1 and Tet2 double KO mice display increased DNA methylation and reduced 5hmC and have aberrant methylation at various imprinted loci [64] [67]. In addition, low levels of Tet1 in male mice results in abnormal methylation specifically at imprinted loci in primordial germ cells and sperm cells [67].

In HEK293T cells, TET1 binds to CGIs and protects them from gaining aberrant methylation, as overexpression of TET1 leads to the accumulation of 5hmC at the borders of CGIs while loss of TET1 leads to increased DNA methylation specifically at CG sites within and around CGIs [30]. However, other studies have implicated TET1 in the dynamic regulation of DNA methylation outside of CGIs, such as at CTCF sites where TET1 and 5hmC are involved in nucleosome repositioning [68].

In mESCs, 5hmC is most highly associated with low to intermediate CG dense sites and is found less frequently at high CG dense sites (even though this is where TET1 is mostly bound, see below) [12]. 5hmC is often absent from high CG dense regions because these regions are typically un-methylated and 5mC is required to generate 5hmC. In addition, 5hmC is found in gene bodies, enhancers, DNase sensitive sites and transcription factor binding sites, as these sites typically have low to intermediate CG density and usually have moderate levels of methylation [69].

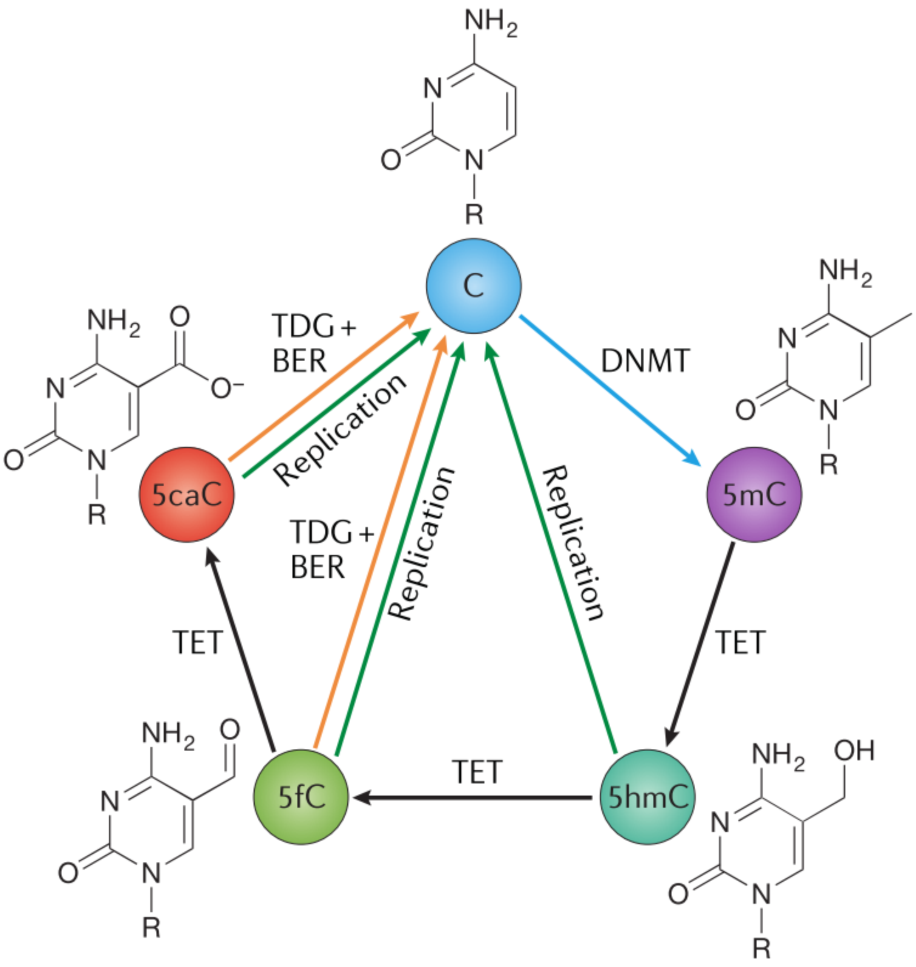


**Figure 4. TET-mediated DNA demethylation** [12]**.**

DNA methyltransferases (DNMTs) add a methyl group to the 5’ position of cytosine. The TET family of enzymes can oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). These oxidative modifications can be converted back to un-methylated cytosine either through replication (passive demethylation) or through a TDG/base excision repair mechanism (active demethylation).

## 1.8 TET1 interacting partners

The TET proteins can be recruited to DNA by interacting partners, which gives the proteins target specificity. Several TET1 interacting partners have been described in the literature, mostly in mESCs. In mESCs, Tet1 and Tet2 physically interact with Nanog, where Tet1 synergizes with Nanog to enhance the efficiency of reprogramming [70]. Endogenous Tet1 was also found to interact with several factors including Sin3A, Hdac1/2, Mta3 and Chd4 which are all part of major chromatin remodeling complexes, suggesting Tet1 may be involved in chromatin remodeling [71]. In addition, Tet1 interacts with the O-GlcNAc transferase enzyme Ogt, where TET proteins recruit Ogt to chromatin [71] [72] [73] [74]. Ogt has been implicated to act as a transcriptional activator and co-repressor, but it also modifies proteins by transferring O-GlcNAc from UDP-GlcNAc to the hydroxyl group of threonine or serine [72]. TET proteins themselves can be modified by Ogt, which stabilizes the protein [71]. Other proteins shown to recruit Tet1 to DNA includes Prc2, Prdm14 and Lin28A [75] [76] [77].

In addition to mESCs, TET1 interacting partners have also been identified in a few cancer cell lines. In the glioblastoma cell line (U251 cells), TET1 interacts with SIN3A, EZH2, LSD1, HDAC1/6/7, MBD1, UHRF1 and MECP2 which are also part of chromatin repressor complexes [78]. In prostate cancer cells, TET1 can be targeted to DNA via FOXA1, where these proteins work together to modify the epigenetic signature at linage-specific enhancers [79].

Lastly, a report in HEK293T cells found that all three TET proteins interact with PARP1, LIG3, XRCC1, TDG and MBD4 [80]. Interestingly, several of these proteins are involved in the TDG/base excision repair process, which removes 5fC and 5caC from cytosine following TET oxidation [17].

It remains unknown what determines which partner TET1 interacts with in any given scenario. Further experiments should be performed in additional cell types and under different biological conditions to obtain a better understanding of TET1 function in differentiated adult cells.

## 1.9 TET1 as a transcriptional co-regulator

ChIP-seq experiments in mouse ESCs revealed Tet1 is enriched at CG islands, active promoters and bivalent promoters (marked by H3K27me3 and H3K4me3) [12]. In addition, as CG density increases, Tet1 occupancy increases, which is likely due to the Tet1 CXXC domain [12].

TET1 binds to un-methylated promoters, where it can recruit co-factors to help regulate transcription [12]. Several reports have shown that upon knock down (KD) of TET1, close to an equal number of genes are up and down regulated, indicating TET1 may have positive and negative effects on transcription [30] [63].

Tet1 has been linked to epigenetic repression complexes including Sin3a and Prc2 [63] [81]. Tet1 recruits Sin3a to a subset of promoters [63]. These target genes showed increased expression upon Tet1 KD or Sin3a KD and ChIP-seq data found Tet1 and Sin3a are co-bound at these loci [63]. Also, at bivalent promoters, Prc2 recruits Tet1 to chromatin, facilitating a repressed chromatin state [75].

TET1 can also act as a transcriptional co-activator. For example, TET1 acts as a transcriptional co-activator with HIF1a independent of its catalytic activity [82]. In a lung cancer cell line (H1299) an enzymatically dead mutant TET1 showed a very similar phenotype to the wild type (WT) TET1 [82]. Also, in HEK293T cells, overexpression of WT TET1 and a catalytically dead mutant TET1 resulted in nearly identical gene expression profiles [30].

## 1.10 TET1 in cancer

Several reports have found TET1 to be downregulated in cancer, where it has been described as a tumor suppressor [60 83 84]. Reported causes for this downregulation include promoter hypermethylation, inhibition by HMGA2 and miR-22 [83 66 61]. One study found TET1 binds to TIMP2 and TIMP3 gene promoters and protects them from aberrant methylation [60]. Upon loss of TET1 in breast cells, TIMP2 and TIMP3 gained methylation, and ultimately led to increased invasion [60]. Furthermore, overexpression of TET1 in a mouse xenograft model resulted in decreased tumor volume [60]. Importantly, in the supplemental section of this paper, they showed TET1 KD led to decreased cellular proliferation and migration, implicating a more complex role for TET1 in cancer [60]. Another breast cancer study found TET1 overexpression decreased tumor volume in a breast cancer xenograft mouse model and decreased bone metastasis through the HMGA2/TET1/HOXA9 axis [83]. Also, low TET1 levels have been linked to a worse overall survival in breast cancer [60 84]. However, additional reports contradict these findings and show no correlation between TET1 expression and survival [83].

Although *TET1* is not frequently mutated in hematological malignancies, a report in 2013 revealed TET1 as playing an essential oncogenic role in MLL-rearranged leukemias [85]. The authors showed that TET1 is a transcriptional target of MLL-fusion proteins and that TET1 is overexpressed in MLL rearranged leukemias, resulting in increased 5hmC [85]. Importantly, this is the only study implicating TET1 as an oncogene.

### 1.11 Breast cancer

In the United States, breast cancer is the most commonly diagnosed cancer, with an estimated 255,000 new cases estimated in 2017 alone (seer.cancer.gov). Breast cancer is the fourth leading cause of cancer death in the United States, with approximately 40,000 deaths per year. However, breast cancer has become a more treatable disease with the overall five year survival rate of 89.7%, thanks to early detection and targeted therapies.

Tests are often performed at diagnosis to help physicians identify the best treatment options for patients. The most frequently observed parameters are whether the tumor overexpresses the estrogen receptor, progesterone receptor or HER2/NEU oncogene. This is an important clinical feature that dictates treatment options. For example, a patient that overexpresses the estrogen receptor (ER) is treated with Tamoxifen, an ER antagonist that blocks the receptor from driving the proliferative signal. Targeted therapies have been successful in the clinic, although resistance to therapy remains an issue for some patients [86].

Most breast cancers are ER and progesterone receptor (PR) positive, with a smaller portion of patients that are HER2 positive. However, between 10-20% of breast cancer patients do not overexpress any of these proteins (ER, PR or HER2) and this molecular subtype is defined as triple negative [87]. In addition to molecular subtyping, there are also distinct groups of breast cancer patients defined by their gene expression profiles, including Luminal A, Luminal B, HER2-enriched and Basal-like breast cancers. The luminal subtypes are the ER positive patients, whereas the HER2 enriched group have HER2 activation but lack expression of ER or PR [88]. Although not all basal breast cancers are triple negative and not all TNBCs are basal, a report found more than 80% of TNBCs are classified as basal-like [88].

TNBCs are often a therapeutic challenge because of advanced presentation and because they lack options for targeted therapy. Further, TNBCs have a worse overall survival compared to non-TNBCs and are more likely to relapse within the first three years [89]. Currently, TNBCs are treated with surgery, chemotherapy and/or radiation therapy. *TP53* is the only gene found to be mutated frequently in TNBC; however, this pathway has been difficult to target therapeutically [42]. A better understanding of the basic biology and molecular underpinnings of TNBCs is crucial to identifying better treatment options. Another feature of TNBC tumors is widespread genome-wide hypomethylation [42] [43] [44]. Furthermore, this hypomethylation is independently associated with a worse overall survival [34]. How tumors become hypomethylated and why it is associated with a worse prognosis remains unknown.

## 1.12 PI3K-Akt-mTOR signaling in cancer

The phosphoinositide 3-kinase (PI3K) pathway is involved in multiple cellular functions, including cell proliferation, migration and survival [88]. There are three classes of PI3Ks, with Class 1 PI3K being the most frequently mutated in cancer [88]. PI3K functions as a heterodimer, consisting of the catalytic (p110) subunit and the regulatory (p85) subunit [90]. Once a receptor tyrosine kinase (RTK) is activated by a ligand, the regulatory subunit is recruited, causing a change in conformation that allows the catalytic subunit to catalyze the conversion of 4,5-phosphoinositide (PIP2) phosphorylation to 3,4,5-phosphoinositide (PIP3) [88]. PIP3 results in the activation of AKT which itself has many downstream targets, including TSC1/2 which are negative regulators of mTOR [88]. Upon Akt activation, repressors of mTOR are inhibited, allowing for the activation of mTOR. Once mTOR is activated, it can lead to the activation of its major downstream targets (S6K and 4EBP1) [91]. When mTOR phosphorylates 4EBP1, it leads to the global translation of CAP dependent mRNAs [92]. However, under nutrient deprivation, mTOR is not activated and 4EBP1 is not phosphorylated, resulting in reduced translation of mRNAs in a CAP independent manner, resulting in reduced cell growth. One control mechanism for the cell is PTEN, a protein phosphatase that converts PIP3 to PIP2, which prevents the activation of the PI3K pathway [93]. Importantly, there are several types of receptors that can lead to the activation of PI3K, including receptor tyrosine kinases (EGFR, INSR, IGF1R, PDGFR, FGFR, KIT, etc) [91]. Additional activators include the G protein coupled receptors (GPCR), as well as integrin, cytokine, toll-like, and B-cell receptors [91].

The PI3K pathway is frequently altered in cancer, including in all subtypes of breast cancer. In fact, PI3K is the second most frequently mutated gene in breast cancer (behind TP53) with 36% of breast cancers having a mutation [42]. Mutations are mostly found in the helical and kinase domains, which results in the constitutive activation of the

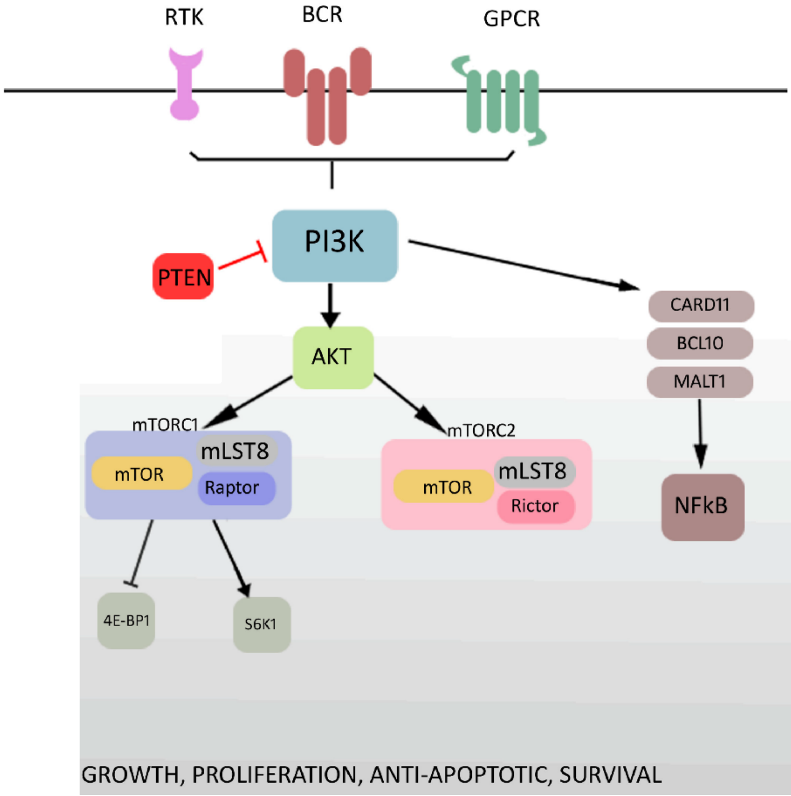


**Figure 5. PI3K signaling [94].**

PI3K is activated by receptor tyrosine kinases (RTK), B cell receptor (BCR) and G protein coupled receptors (GPCR). PI3K activates AKT, which in turn activates mTOR. Activation of mTORC1 leads to the phosphorylation and inhibition of 4E-BP1. PTEN is a negative regulator of PI3K signaling.

kinase [88]. PI3K is most frequently mutated in the hormone receptor positive breast cancers (Luminal A (45%), Luminal B (29%), HER2 enriched (39%)) [42]. However, in TNBCs or basal like breast cancers, PIK3CA is mutated in only 9% of cases [42]. This is true for PTEN as well where only 1% of basal like breast cancers have alterations in PTEN [42]. Interestingly, gene expression and proteomic studies have revealed hyperactivity of the PI3K pathway in TNBCs, even though they are not enriched for genetics alterations in the pathway [42]. This raises an interesting question as to how the PI3K pathway is activated in TNBCs and whether TNBCs patients may benefit from treatment with PI3K inhibitors.

A multitude of PI3K inhibitors are on the market, with numerous on-going clinical trials testing them in a variety of tumor types. Most of the trials in breast cancer have been in hormone receptor positive patients that have activating mutations in PIK3CA [95]. However, early results revealed PI3K inhibitors have shown limited efficacy in patients with activating PI3K mutations [95]. This could be due to lack of specificity of the drug itself or the cells overcoming the deficiency by genetically or epigenetically activating compensatory pathways [95]. Every year more specific/selective PI3K inhibitors are being developed, and the efficacy of these drugs should be tested. In addition, more research should be done to identify select groups of patients who may benefit from the therapy.

## 1.13 Hypothesis and specific aims

Although DNA hypermethylation in cancer has been well established, much less is known about hypomethylation and its effects on tumorigenesis. In addition, mechanisms regulating the hyper and hypomethylation observed in cancer remains unclear. We discovered a novel truncated isoform of TET1 that retains its demethylase domain but lacks the CXXC DNA binding domain. We hypothesize that TET1 and TET1$^{ALT}$ have distinct functions in normal and cancer cells. In addition, we hypothesize that TET1 and TET1$^{ALT}$ are dysregulated in cancer, leading to aberrant hypomethylation and activation of genes involved in tumorigenesis. The following aims were devised to test our hypotheses.

**Aim 1: Identify and characterize the function of TET1$^{ALT}$.**

**Aim 2: Identify the proteins involved in regulating the balance between hypo and hypermethylation in cancer.**

**Aim 3: Characterize the function and role of TET1 and TET1$^{ALT}$ in breast cancer progression.**

# CHAPTER 2

# MATERIALS AND METHODS

## 2.1 Cell culture

Breast cancer cell lines (BT549, HCC2218, HCC1599, MCF-7, MDA-MB-231), immortalized breast, but non-malignant (MCF10A), human embryonic kidney (HEK293T), chronic myelogenous leukemia (K562), and the PC-3 prostate cancer cell line were all obtained from American Type Culture Collection (ATCC). Normal breast epithelium (NBE) was a kind gift from Dr. Xiaowei Chen at the Fox Chase Cancer Center. The immortalized human mammary epithelial cells HMLE cells were a generous gift from Dr. Sendurai A. Mani at the University of Texas MD Anderson Cancer Center. The lymphoblastoid cell line (GM12878) was a kind gift from Dr. Italo Tempera at the Fels Institute for Cancer Research and Molecular Biology. GM12878 were cultured in RPMI 1640, 2mM L-glutamine with 15% fetal bovine serum (FBS). K562 cells were cultured in Iscove's Modified Dulbecco's Medium, supplemented with 10% FBS. PC-3 cells were cultured in ATCC-formulated F-12K Medium with 10% FBS. To culture MCF-10A cells, we used DMEM/F12 with 5% horse serum (treated to remove divalent cations), 20 ng/ml EGF, 100 ng/ml Cholera toxin, 10 µg/ml Insulin, 500 ng/ml Hydrocortisone, 1.39 mM Calcium and 1% Penicillin-Streptomycin antibiotics. HMLE cells were grown as previously described [96]. MCF-7 cells were cultured in Eagle's Minimum Essential Medium with 10% FBS. HEK293T and MDA-MB-231 cells were cultured in Dulbecco's Modified Eagle's Medium supplemented with 10% FBS. HCC1599, BT549 and HCC2218 were cultured in RPMI-1640 Medium with 10% FBS. All cell lines routinely tested negative for mycoplasma contamination.

## 2.2 Western blot and DNA slot blot

Protein was extracted using a lysis buffer consisting of (50mM Tris-HCl pH 7.4, 5mM EDTA, 250mM NaCl, 50mM NaF, 0.1% Triton X-100, 0.1 mM $Na_3VO_4$) supplemented with 1x protease inhibitor cocktail solution (Roche). Extracts were quantified using Qubit protein assay (ThermoFisher) and were run on polyacrylamide gels. Gels were transferred to PVDF membranes using a wet transfer method in CAPS buffer. Primary antibodies were incubated overnight at 4°C. The following primary antibodies were used in this study: anti-FLAG (A8592, Sigma), anti-TET1 (GTX124207, GeneTex), anti-TET1 (GT1462, Sigma), anti-5hmC (catalog # 39769, Active Motif), anti-phospho 4EBP1-Thr37/46 (ab32130, Abcam) and anti-β-actin (A5316, Sigma). For the DNA slot blot analysis, we followed the protocol established previously with the exception of using a slot blot apparatus instead of a dot blot [30]. Blots were imaged using FluorChem Q and unsaturated bands were quantified using the multiplex band analysis tool followed by normalizing to local background and β-Actin.

## 2.3 TET1 expression constructs

The $TET1^{FL}$ pIRES hrGFP II expression plasmid along with the $TET1^{CD}$ plasmid was generated in our lab previously [30]. To generate the $TET1^{ALT}$ plasmid, $TET1^{FL}$ plasmid was digested with restriction enzymes BamHI and BglII which removed the first 1,850 bp of the $TET1^{FL}$ cDNA sequence thus excluding the $TET1^{FL}$ start codon. 163 bp of the TET1$^{FL}$ cDNA is upstream of the TET1$^{ALT}$ ATG, however no in-frame ATG is contained within this region. The plasmids were transfected into HEK293T cells using Lipofectamine 2000 following the manufacturer's instructions.

### 2.4 Digital restriction enzyme analysis of methylation (DREAM)

DREAM is a quantitative, deep sequencing based method for DNA methylation analysis and it was performed as previously described [30] [97]. Briefly, 2 µg of genomic DNA from HEK293T cells expressing empty vector, TET1$^{ALT}$ or TET1$^{FL}$ were digested with 20 units of SmaI (8 h at 25°C, NEB) and 20 units of XmaI (~16 h at 37°C, NEB), resulting in a distinct DNA methylation signature at CCCGGG sites. 3' ends of the DNA fragments were repaired using Klenow fragment (3'→5' exo-) DNA polymerase and dCTP, dGTP, and dATP nucleotides. Illumina sequencing adapters were ligated to the DNA fragments and the libraries were sequenced by paired-end 40 nt sequencing on Illumina HiSeq2500. The sequencing reads were mapped to the hg19 genome and methylation values were calculated as the ratio of the number of the reads with the methylated XmaI signature over the total number of tags mapped to a given SmaI/XmaI site. The coverage threshold was set to greater than 50 reads per sample. TET1$^{ALT}$ and TET1$^{FL}$ were compared to empty vector control and data were filtered for sites that change methylation by greater than 5%. High-throughput DNA methylation data generated in this study have been deposited in the GEO database under accession number GSE93617 and GSE100640.

## 2.5 RNA-sequencing

RNA was isolated using Qiagen's RNeasy Plus Mini kit following manufacturer's instructions from experiments done in biological triplicates. RNA was quantified by Nanodrop and purity was checked using the Agilent Bioanalyzer. Strand-specific RNA libraries were generated from 1 µg of RNA using TruSeq stranded total RNA with Ribo-Zero Gold (Illumina). Sequencing was performed using single end reads (50 bp, average 30 million reads per sample) on the HiSeq2500 platform (Illumina). Sequenced reads were aligned to the hg19 genome assembly using TopHat2 software suite [98]. The expression level and fold change of each treatment group was evaluated using Cuffdiff [99]. Genes that had 0 reads across all samples were excluded. High-throughput RNA-seq data generated in this study have been deposited in the GEO database under accession number GSE100483 and GSE93619.

## 2.6 RNA isolation and qPCR analysis (cell lines)

RNA isolation, reverse transcription and RT-PCR analysis were performed as described previously [30]. Total RNA (in biological triplicates) was extracted using TRIzol following manufacturer's protocol. RNA was DNase-treated using TURBO DNA-free kit following manufacturer's protocol (Ambion). cDNA was synthesized using 1 µg of RNA using the High-Capacity cDNA reverse transcription kit following manufacturer's protocol (Applied Biosystems). qPCR was performed on Applied Biosystems 7500 machine using iTaq Universal SYBR Green Super Mix following manufacturer's instructions (BioRad). qPCR was performed in technical and biological triplicates and the average Ct values were determined for each gene. Samples were normalized to HPRT. The primers used are listed in Table 1.

### 2.7 Luciferase reporter assay

Luciferase reporter assays were conducted using the Dual-Luciferase Reporter Assay System (Promega) following the manufacturer's protocol. HEK293T cells were seeded in 6 well plates and co-transfected with Renilla expression plasmid and the pGL4.10 [luc2] reporter constructs containing either empty vector, intron control or the TET1$^{ALT}$ promoter region. Cells were transfected using Lipofectamine 2000. Luciferase activities were measured 48 hours post transfection and normalized to Renilla and to empty vector control. The primers used are listed in Table 2.

### 2.8 Mouse tissues

Adult tissues were harvested from C57BL/6 male mice 6.2 months after birth. RNA was extracted and DNase treated as described above under RNA isolation.

To obtain staged embryonic tissues from C57BL/6 mice, matings were set up and plugs checked the following morning. Noon of the day of the plug was designated E0.5. Pregnant dams were killed at the appropriate day of gestation (10.5, 12.5, 14.5, 16.5 dpc) and fetal tissues were dissected and frozen for further analysis. Neonatal tissues were obtained from day 3 pups. C57BL/6 mouse embryonic stem (ES) cells were maintained in Leukemia Inhibitory Factor (LIF) on inactivated mouse embryonic fibroblasts (MEFs), on gelatin-coated tissue culture plates in a 5% CO2 humid incubator at 37°C. RNA extraction was performed following two consecutive 60 minute MEF-depletion steps. RNA from ES cells, fetal and neonatal tissues was extracted using the Roche High Pure RNA Isolation Kit (#11828665001), following the manufacturer's protocol. All RNA samples were subjected to DNAse treatment using Turbo DNA-free (Ambion #AM1907). Three to five biological samples were collected for each tissue.

## 2.9 CRISPR

To knockout TET1, we used the Lenti CRISPR V2 plasmid (Addgene) [100]. CRISPR gRNAs were designed using http://crispr.mit.edu/ to target three different TET1 exons. gRNAs can be found in Table 3. Oligonucleotides were annealed and ligated into the Lenti CRISPR V2 plasmid that was previously digested with BsmBI. The cloning protocol associated with the plasmid was followed: http://genome-engineering.org/gecko/wp-content/uploads/2013/12/lentiCRISPRv2-and-lentiGuide-oligo-cloning-protocol.pdf [100]. Ligated plasmids were propagated and verified by restriction enzyme digest and by sequencing. Lentiviruses were generated using HEK293T cells by transfecting with packaging plasmid (psPAX2, Addgene 12260), envelope plasmid (pMD2.G, Addgene 12259) and lenti-CRISPR V2 plasmid (Addgene 52961). Cells were transfected using Lipofectamine 2000. Viral supernatant was collected at 48 and 72 hours, filtered with 0.45µm membrane and incubated on MDA-MB-231 cells for 7 hours in the presence of polybrene (6µg/mL, Millipore). We puromycin selected (1µg/mL) the cells for three days, followed by single cell cloning using serial dilution in 96 well plates.  After selection, cells were maintained in normal media supplemented with 0.25 µg/mL puromycin.

To activate the TET1$^{ALT}$ and TET1$^{FL}$ promoter, we used plasmids previously generated following published protocols [101].  Plasmids used included pLKO.1-puro U6 sgRNA CAG (Addgene 50927), pLKO.1-puro U6 sgRNA BfuAI stuffer (Addgene 50920), and pHAGE EF1α dCas9-VP64 (Addgene 50918).  sgRNA's were designed as described above, but target regions were limited to the promoter regions of TET1$^{ALT}$ and TET1$^{FL}$.  Promoters were identified using UCSC genome browser by locating the transcription start site, RNA polymerase II binding, and H3K4me$^{3}$ enrichment.  Two gRNAs were simultaneously used to target the TET1$^{FL}$ promoter and 3 gRNAs to target the TET1$^{ALT}$ promoter.  gRNAs used can be found in Table 4.  A sgRNA targeting the CAG (CMV-IE, chicken actin, rabbit beta globin) promoter was used as an off-target control.  First, stable cell lines were generated that overexpress the dCas9-VP64 fusion protein.  Lentiviruses were made in HEK293T cells, and then MCF10A cells were transduced and puromycin selected for one week.  Next, lentiviruses were generated in HEK293T cells using the pLKO.1 gRNA plasmid mentioned above.  After viral collection, MCF10A-dCas9-VP64 expressing cell lines were transduced with the viral containing gRNAs.  Cells were selected for three days using puromycin (1µg/mL).

## 2.10 Bisulfite pyrosequencing

Genomic DNA from HEK293T cells expressing either empty vector, TET1$^{CD}$, TET1$^{ALT}$ or TET1$^{FL}$ was bisulfite converted using the EpiTect bisulfite kit (Qiagen) following the manufacturer's protocol. For bisulfite pyrosequencing of LINE1, a PCR for amplification was used as previously described [102]. Pyro Q-CG Software (Qiagen) was used to analyze the data. PCR and pyrosequencing primers can be found in Table 5.

### 2.11 PCR and clonal sequencing

Genomic DNA from MDA-MB-231 TET1 CRISPR knockout cells were amplified using primers surrounding the gRNA target sequencing and Phusion High-Fidelity DNA polymerase (Neb). Amplified PCR products were cloned into the Zero Blunt TOPO pCR4 system and OneShot Top10 chemically competent E. Coli were transformed following the manufacturers protocol. (ThermoFisher). Twelve clones were sequenced and analyzed using Serial Cloner. PCR and sequencing primers can be found in Table 6.

### 2.12 Proliferation and cell migration

20,000 cells were seeded per well in 24 well plates for proliferation assays. Biological triplicates were counted twice for a total of six counts per sample. Cells were counted every 24 hours up to 120 hours and the data were plotted as the total number of cells vs the number of hours. For the wound healing assay, cells were seeded in 24-well plates and incubated to confluency. Wounds were scratched in the cell monolayer using a

200 µl sterile pipette tip. After scratching, detached cells were removed by washing three times with PBS and replenished with DMEM 10% FBS. The migration of the cells at the edge of the scratch was photographed at 0, 24, 48 hours. The gap distance was quantitatively evaluated using ImageJ Wound Healing Tool (http://rsb.info.nih.gov/ij/download.html). To reduce variability, multiple wells of each cell line were evaluated.

### 2.13 Bioinformatics and statistics

DNA methylation data (Illumina HumanMethylation 27K and 450K array platform beta-values) for normal breast tissue (N=41), breast cancer (N=500) and ovarian cancer (N=304) was downloaded from The Cancer Genome Atlas public data portal in 2015. CG sites with NA values were excluded from the analysis. Unsupervised hierarchical clustering analysis was performed using ArrayTrack. DNA methylation data (450K array) from 63 human cell lines were downloaded from the UCSC genome browser track HAIB Methyl450 (wgEncodeHaibMethyl450) as part of the Encode project [103]. The methylation values for each CG site associated with TET1 (N=30) was averaged for all normal breast samples and for all cell lines and plotted versus the distance to the TET1 transcription start site.

RNA-sequencing BAM files were downloaded from the Genomic Data Commons Portal for breast, ovarian and uterine cancer, AML and glioblastoma patient samples. Access was granted by the NCI Center for Biomedical Informatics and Information Technology to obtain TCGA controlled access datasets. Reads mapping to exon 1 or the

TET1$^{ALT}$ exons were extracted and normalized to exon length, and then to all reads on chromosome 10, and multiplied by 1 million to obtain normalized exon reads.

Level 3 RNA-sequencing data (RSEM) were downloaded from TCGA in 2015 for normal breast (N=105) and breast cancer patient samples (N=866). For our differential gene expression analysis, genes with RSEM<1 in more than half of patients in Cluster 1 or Cluster 2 were excluded. Genes were called as significantly upregulated with FC>2 and downregulated with FC<0.5, p<0.05. TET1 expression in ovarian cancer and pancreatic cancer were downloaded using CBioPortal (RNA-seq, RSEM).

To illustrate the exon usage of TET1 in selected cancers, we plotted median length normalized exon quantification based on RNA-seq level 3 TCGA data in breast and uterine cancer, glioblastoma and AML. To exclude subjects that do not express TET1 at sufficient levels, we filtered out the samples with average usages of exon 3-12 below a 0.5 cut off.

DNA mutation data were downloaded using CBio Portal for the TCGA Provisional cohort of breast cancer patient samples.

Drug sensitivity data in breast cancer cell lines and ovarian cancer cell lines were downloaded from the Genomics of Drug Sensitivity in Cancer database (cancerrxgene.org). Gene expression data (Affymetrix U133 array) for cancer cell lines were downloaded from the Cancer Cell Line Encyclopedia. IC50 values for 265 drugs were correlated to TET1 expression values in all breast and ovarian cancer cell lines available using Spearman analysis.

Survival curves were generated using GraphPad Prism 4.0. Survival data were downloaded from CBioPortal using the following studies: Breast cancer (METABRIC, Nature 2012 & Nature Communications 2016)[104] [105], Glioblastoma (TCGA, Provisional), Uterine cancer (TCGA, Nature 2013)[106], Ovarian Serous Cystadenocarcinoma (TCGA, Provisional), Acute Myeloid Leukemia (TCGA, NEJM 2013)[107]. TET1 high was considered >1 standard deviation above the mean and all other patients were classified as TET1 low for uterine cancer (RNA-seq), AML (RNA-seq), glioblastoma (Microarray) and ovarian cancer (RNA-seq). TET1 high was considered stdev>2 above the mean for breast cancer (Microarray U133). TET1, TET2, and TET3 expression in TNBC specific analyses was considered high if >1 standard deviation above the mean and all other patients were classified as low. Significance of survival curves were calculated using the Gehan-Breslow-Wilcoxon test and/or the log-rank test, as indicated.

Calculations were done in GraphPad. The Student's t-test was used to calculate significant p values unless otherwise stated. All p-values are two-sided. * $p<0.05$, ** $p<0.01$, *** $p<0.001$ denotes significance. Mann-Whitney test was used to test significance of TET1$^{ALT}$ exons or TET1 exon 1 reads in vivo. Error bars are standard error of the mean (SEM). Significance of clinical characteristics, including mutation levels, for each cluster was calculated using Fisher's exact test. Significance of overlap for the datasets was calculated using the Chi-squared test. The following graphs/statistical tests were generated using R software: principal component analysis, permutations, Spearman correlations, histogram distribution of R values, hierarchical clustering, differential TCGA gene expression plot and venn digrams. All pathway analyses were performed using Consensus Path Database, and pathway analyses of methylation data were background corrected dependent on the assay performed [108].

**Table 1. qRT-PCR primers**

| | |
|---|---|
| Mouse TET1-ALT (Forward) | CCGTGAAGAATGCAGAAGCTAA |
| (Reverse) | CTCTGGGGCCTCTTGTTTTCT |
| Mouse TET1-FL (Forward) | ATCGAAAGAACAGCCACCAGA |
| (Reverse) | GGGGCCTCTTGTTTTCTTTTG |
| Mouse HPRT (Forward) | TGCTCGAGATGTCATGAAGGA |
| (Reverse) | CCAGCAGGTCAGCAAAGAACT |
| Human TET1-ALT DB11 (Forward) | CAAGCAAGATGGCTACCTCGT |
| (Reverse) | GGGGCCTCTTGTTTTCCTTTA |
| Human TET1-ALT DB15 (Forward) | TTGAAGCCTCCTGTGATTTCG |
| (Reverse) | GGGGCCTCTTGTTTTCCTTTA |
| Human TET1-FL (Forward) | GCGCGAGTTGGAAAGTTTG |
| (Reverse) | GCTCAGTCACACAAGGTTTTGG |
| Human HPRT (Forward) | TGAGGATTTGGAAAGGGTGTT |
| (Reverse) | GAGCACACAGAGGGCTACAATG |
| Human b-Actin (Forward) | GAAGAGCTACGAGCTGCCTGA |
| (Reverse) | GTTTCGTGGATGCCACAGGA |

**Table 2. Primers for luciferase reporter assay**

| | |
|---|---|
| Intron control (Forward) | atatGGTACCctcactctgttcctgatttctggttg |
| Intron control (Reverse) | atatAAGCTTgggcatttctgatgaccttcatt |
| TET1-ALT promoter (Forward) | atatGGTACCatgagacacgcagcccaacag |
| TET1-ALT promoter (Reverse) | atatAAGCTTttacCTTTAAAACTTTGGGCTTCTTTTCC |

**Table 3. Lenti CRISPR V2 gRNA sequences for TET1 KO**

| | |
|---|---|
| **gRNAs for TET1 KO-1** | |
| Targets Exon 6 (Sense) | CACCGATAGAAATAGTAGTGTACAC |
| (Antisense) | AAACGTGTACACTACTATTTCTATC |
| **gRNAs for TET1 KO-2** | |
| Targets Exon 3 (Sense) | CACCGCTGATTACCTTTAAAACTT |
| (Antisense) | AAACAAGTTTTAAAGGTAATCAGC |
| **gRNAs for TET1 KO-3** | |
| Targets Exon 11 (Sense) | CACCGTTCCGCTTGATTCGGGGAAT |
| (Antisense) | AAACATTCCCCGAATCAAGCGGAAC |

**Table 4. gRNA sequences for CRISPR activation in pLKO.1 vector**

| **gRNAs for dCas9-VP64 fusion experiment** | |
|---|---|
| TET1-ALT promoter gRNA #1 (Sense) | ACCGTCTGGTCGCGCCGAAATCAC |
| (Antisense) | AAACGTGATTTCGGCGCGACCAGA |
| TET1-ALT promoter gRNA #2 (Sense) | ACCGGCAGAACTGTTCCACTGTAG |
| (Antisense) | AAACCTACAGTGGAACAGTTCTGC |
| TET1-ALT promoter gRNA #3 (Sense) | ACCGGCCTAGCCCTTCCTAGACAA |
| (Antisense) | AAACTTGTCTAGGAAGGGCTAGGC |
| TET1-FL promoter gRNA #1 (Sense) | ACCGGCGAGAGACAAAACGCGAGC |
| (Antisense) | AAACGCTCGCGTTTTGTCTCTCGC |
| TET1-FL promoter gRNA #2 (Sense) | ACCGAACTGTGCAGGGTCCAGCGA |
| (Antisense) | AAACTCGCTGGACCCTGCACAGTT |

**Table 5. Pyrosequencing primers**

| **LINE1 pyrosequencing assay** | |
|---|---|
| Forward | TTTTGAGTTAGGTGTGGGATATA |
| Biotinylated Reverse | Biotin-AAAATCAAAAAATTCCCTTTC |
| Sequencing primer | AGTTAGGTGTGGGATATAGT |

**Table 6. PCR and clonal sequencing primers**

| | |
|---|---|
| Forward (PCR) | gctctttaggttctgcctagc |
| Reverse (PCR) | ctccaaatatacccaagtgcag |
| Sequencing primer | gctctttaggttctgcctagc |

# CHAPTER 3

# RESULTS

## 3.1 Identification of a novel TET1 isoform

### 3.1.1 Location of TET1$^{ALT}$

NCBI lists one TET1 gene (NM_030625.2), with no alternate isoforms. It has been reported that the canonical TET1 promoter can be hypermethylated in cancer [61]. In examining TET1 methylation using publicly available data from The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov) and ENCODE, we found that the TET1 transcription start site (TSS), which is in a CG island (CGI), is occasionally hypermethylated in cancer. However, we noted a CG site (that is not in a CGI) in intron 2 of TET1 that was unmethylated in 62/63 cell lines and in all 41 normal breast tissue samples surveyed (Figure 6A). Given that most CG sites outside of CGIs are highly methylated unless they are in a regulatory region, we examined this genomic area in greater detail.

This region containing the unmethylated CG site aligned with the start site of 2 expressed sequence tags (ESTs), which we called TET1$^{ALT}$ exon 1a and 1b and which spliced into TET1 exon 3 (Figure 6B). This suggested the presence of an alternate transcription start site. We next queried Poly A (+) CAGE data (which identify TSSs) and found peaks in multiple cell lines, including MCF-7, GM12878 and IMR90 in the TET1$^{ALT}$ putative promoter region (Figure 6B) [109].

The 1 kb region upstream of TET1 exon 3 is highly conserved at the DNA level in primates and placental mammals but not among more distant vertebrates such as chicken and zebrafish (Figure 6B, black represents conserved, white represents non-conserved regions). Investigation of conservation in additional species and for additional TET1 exons show TET1 exon 7 is highly conserved in 44/45 species, however, TET1$^{ALT}$ exons and exon 3 is conserved in only 35/45, suggesting more recent evolution for this region (Figure 7). The 10 species lacking conservation for the TET1$^{ALT}$ region are boxed in red.

### 3.1.2 Detecting mRNA of TET1$^{ALT}$

We put the predicted TET1$^{ALT}$ sequence into open reading frame finder and found two possible start codons in exon 4, which would encode proteins with molecular weights of 147 and 162 kilodaltons (kDa) [110]. Both alternate ATGs have moderate to strong Kozak sequences that are amenable to translation initiation (Figure 8A).

To confirm that TET1$^{ALT}$ is indeed expressed, we designed forward PCR primers to target either TET1$^{ALT}$ exon 1a or 1b (the noncoding exons of TET1$^{ALT}$) and placed the reverse primer in TET1 exon 3. In MDA-MB-231 cells, the primer set for exon1a amplified a PCR product of 104bp in the cDNA samples, and 315bp in the DNA control samples, indicating that the splicing machinery splices out 211bp between TET1$^{ALT}$ exon1a and TET1 exon 3 (Figure 8B). To confirm our findings, we gel extracted and sequenced the PCR products (boxed in red) and aligned them to the genome using BLAT (Figure 8C). The spliced PCR product was also found in additional cell lines including MCF10A, HCC1599, BT549, HCC2218 and HMLE (data not shown). Furthermore, PCR of TET1$^{ALT}$ exon 1b in MCF10A cDNA also led to the amplification of the spliced product of 107bp, while no bands were detected at the unspliced size of 637bp (Figure 8D).

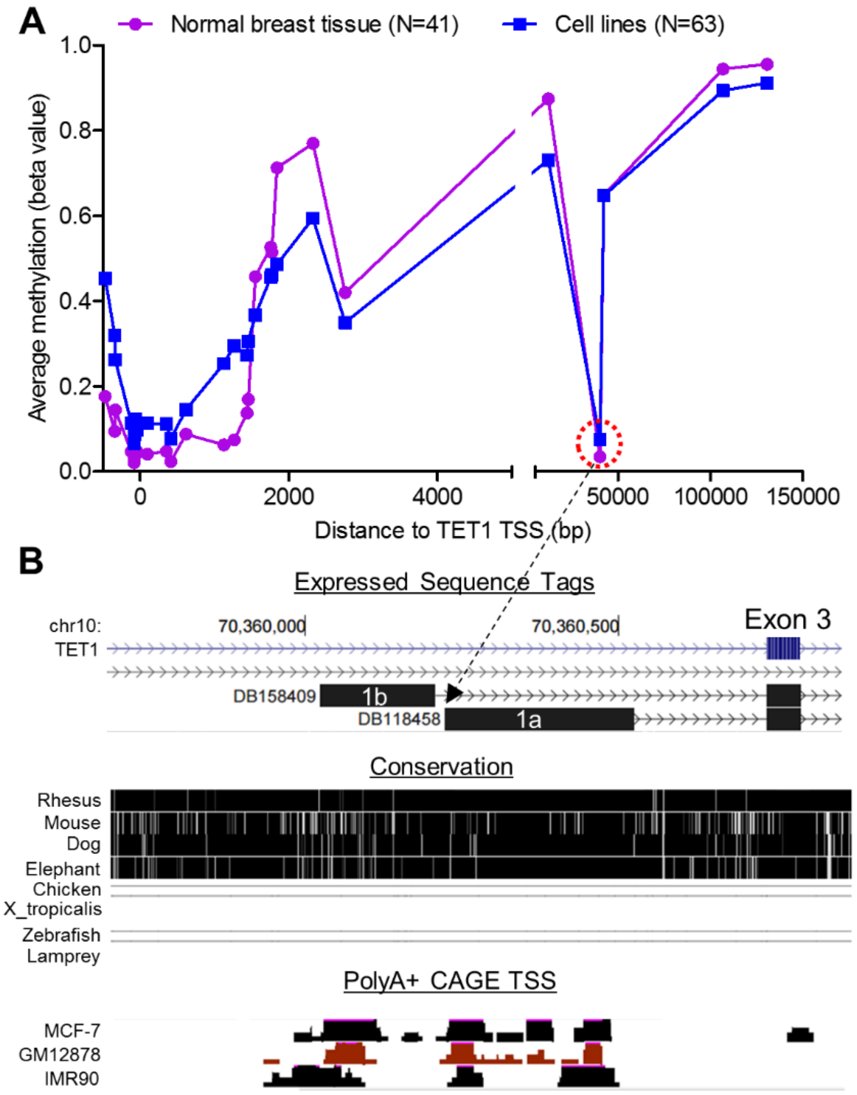


**Figure 6. The discovery of TET1$^{ALT}$.**

(**A**) Average DNA methylation values from 450K array data of 41 normal breast tissues (TCGA) and 63 human cell lines (ENCODE) across the TET1 gene. An unmethylated CG site is indicated by a red circle, located ~40,000bp downstream from the TET1 TSS. **(B)** Mapping of the CG site to intron 2 of TET1 using the UCSC genome browser. Characteristics of this region include the start sites of two ESTs, high conservation, and PolyA+ CAGE plus TSS peaks (ENCODE).

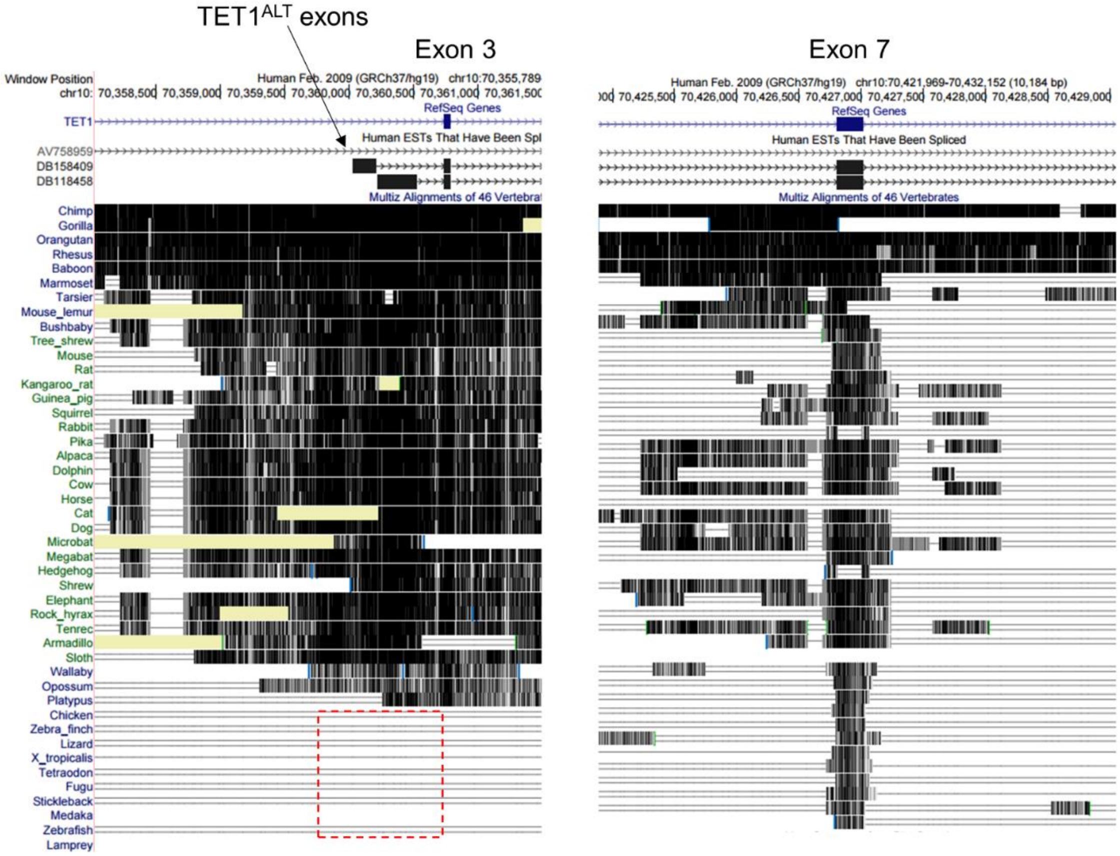


**Figure 7. TET1 and TET1$^{ALT}$ exon conservation.**

TET1$^{ALT}$ exons and TET1 exon 7 conservation as shown by the UCSC Multiz Alignments of 46 Vertebrates (black is conserved, white/gray is not conserved). Exon 7 is generally conserved in vertebrates, however TET1$^{ALT}$ exons are not conserved across lower vertebrates such as Zebrafish, Lizard and Chicken.

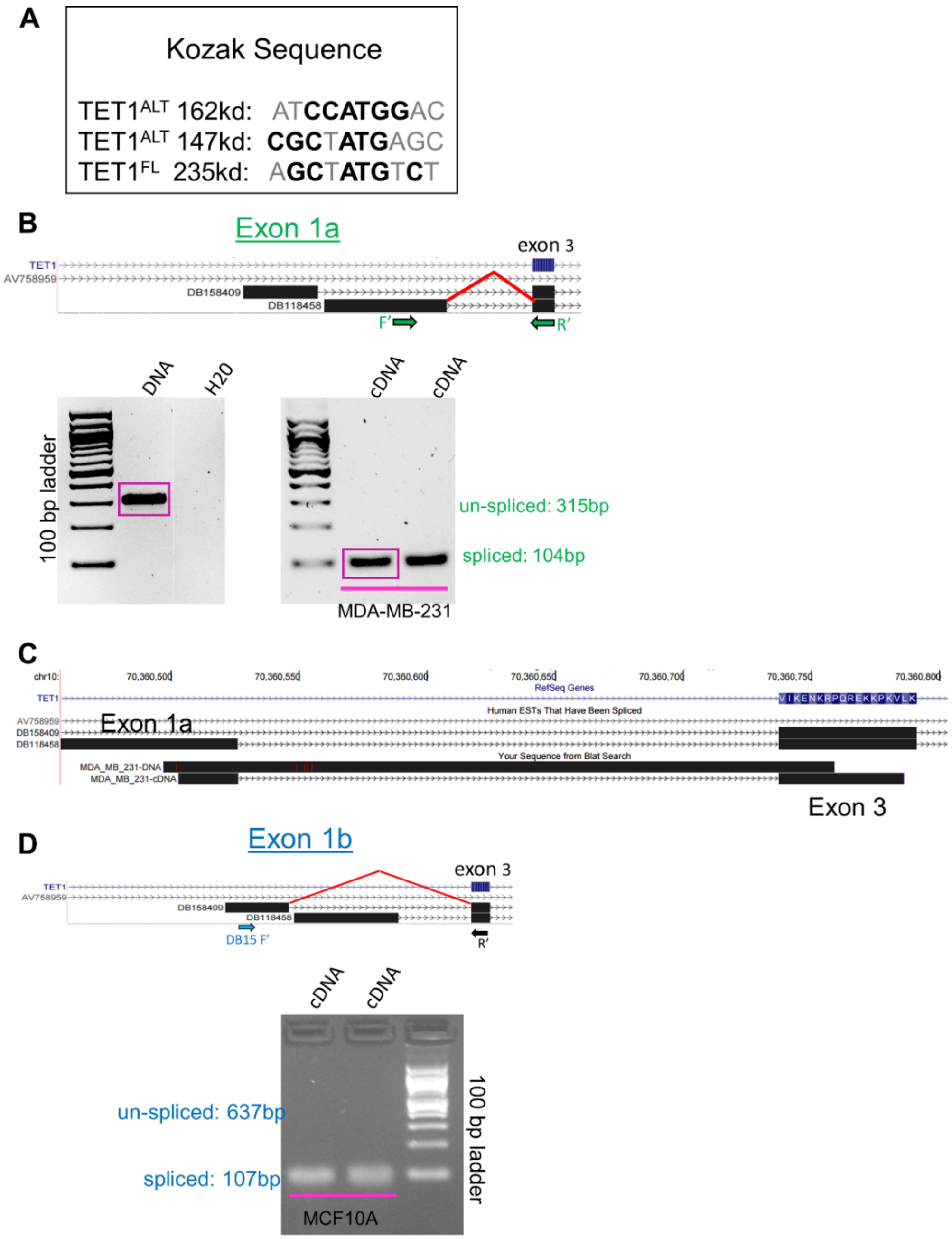


**Figure 8. PCR of TET1$^{ALT}$ exons 1a and 1b.**

**(A)** Kozak sequence of TET1$^{ALT}$ isoforms and TET1$^{FL}$, bolded are the most highly conserved Kozak sequences. **(B)** PCR amplification of TET1$^{ALT}$ exon 1a in MDA-MB-231 DNA and cDNA. PCR products were run on an agarose gel and bands boxed in pink were gel extracted and sent for sequencing. **(C)** Sequenced PCR products from (B) aligned to the hg19 genome using the UCSC blat tool. **(D)** PCR amplification of TET1$^{ALT}$ exon 1b in MCF10A cDNA. PCR products were run on an agarose gel, spliced PCR products are observed at 107bp.

### 3.1.3 TET1$^{FL}$ and TET1$^{ALT}$ are differentially expressed during mouse development

To measure relative abundance of the TET1 isoforms, we used isoform specific qRT-PCR. We found that TET1$^{FL}$ is highly expressed in mESCs, whereas TET1$^{ALT}$ is repressed (Figure 9). During embryonic development, we observed an isoform switch, where the TET1$^{FL}$ isoform slowly decreases in expression, as TET1$^{ALT}$ appears and progressively increases. This is most evident in brain development (see Figure 9 inset), where TET1$^{ALT}$ becomes the dominant isoform expressed during development and in the adult olfactory bulb. Note that neither TET1$^{ALT}$ nor TET1$^{FL}$ are expressed in adult liver.

Aligning the predicted amino acid sequences of the alternate proteins to the canonical protein, we discovered that the alternate isoforms are in frame with TET1 and contain the catalytic domain, but lack the CXXC DNA binding domain (Figure 10).

## 3.2 TET1$^{ALT}$ promoter activity

### 3.2.1 Histone marks

To identify the chromatin architecture surrounding the TET1$^{ALT}$ promoter, we analyzed publicly available ChIP-seq datasets for H3K4me$^{3}$, H3K27Ac, and H3K27me$^{3}$ [111]. H3K4me$^{3}$ is a histone mark permissive for transcription and generally marks active or poised promoters, while H3K27me$^{3}$ is generally a repressive mark. 17/19 cell lines were marked by H3K4me$^{3}$ and 0/19 had H3K27me$^{3}$, indicating that TET1$^{ALT}$ is active or permissive for transcription in these cell lines (data not shown). In figure 11, histone marks from 3 representative cell lines H1-hESCs, GM12878 and HeLa cells are displayed for both the canonical and $TET1^{ALT}$ promoter. GM12878, a lymphoblastoid cell line, is enriched for $H3K4me^3$ at the $TET1^{ALT}$ promoter but not at the $TET1^{FL}$ promoter. In contrast, $H3K4me^3$ is enriched at the $TET1^{FL}$ promoter but not at the $TET1^{ALT}$ promoter. HeLa cells are used as a negative control, where no enrichment for either promoter is found. Additional cell lines are shown in figure 12, where HMECs, HepG2, Dnd41, NHEK and K562 show enrichment for H3K4me3 at the $TET1^{ALT}$ promoter.

We next analyzed ChIP-seq datasets for mouse tissues, including both embryonic and adult (Figure 13) [111]. In agreement with with the human data, mouse ESCs have active marks for $TET1^{FL}$ but not for $TET1^{ALT}$, indicating that $TET1^{ALT}$ is likely inactive in both mouse and human ESCs. However, during embryonic development we see that active promoter marks are gained at the $TET1^{ALT}$ promoter in several tissues with an isoform switch in many tissues. For example, in embryonic day 14.5 brain, the canonical promoter becomes poised (marked by $H3K27me^3$ and $H3K4me^3$) and the $TET1^{ALT}$ promoter becomes active ($H3K4me^3$, H3K27Ac). Histone marks in additional tissues can be found in Figure 14, which show heart, spleen, kidney, cerebellum, and thymus are enriched for $H3K4me^3$ at the $TET1^{ALT}$ promoter. Importantly, the histone marks at the $TET1^{ALT}$ promoter agree with our qRT-PCR RNA expression data. Active histone marks are found at the $TET1^{ALT}$ promoter in tissues where the alternate isoform is expressed (embryonic brain), but are absent in tissues where the isoform is not expressed (mESCs).

Another good example is the liver, which lacks active histone marks at the TET1$^{ALT}$ promoter and transcript levels are undetectable by qPCR (Figure 9 and 14).

### 3.2.2 Transcription factor binding

We also analyzed publicly available ChIP-seq datasets for transcription factors and found that the TET1$^{ALT}$ promoter is bound by a multitude of factors that are likely regulating its activity, including CMYC (MCF-7, NB4), YY1 (GM12878, GM12892, K562), and NFKB (GM12891). (Figure 15). The diversity of transcription factors binding to the TET1$^{ALT}$ promoter highlight its complex regulation. Of note is a region 2.5 kb downstream of the TET1$^{ALT}$ promoter that shows p300, HDAC2, FOXA1 and FOXA2 binding in HepG2 cells. The HepG2 cell line is a hepatocellular carcinoma cell line that shows active promoter marks at the TET1$^{ALT}$ promoter (Figure 12). We downloaded TET1 expression levels from the cancer cell line encyclopedia (CCLE) for 1,036 cell lines and found HepG2 ranks 29/1036 in TET1 expression, and thus is one of the top TET1 expressing cell lines (data not shown). Since TET1$^{ALT}$ transcript expression is undetectable in normal liver, we speculate TET1$^{ALT}$ could be an interesting target to study in liver cancer, where the transcript is highly expressed, the promoter histone marks are conducive to active transcription, and the FOXA1/FOXA2 proteins (which are extremely important in liver development and in liver cancer) show transcription factor binding by ChIP-seq near the TET1$^{ALT}$ promoter.

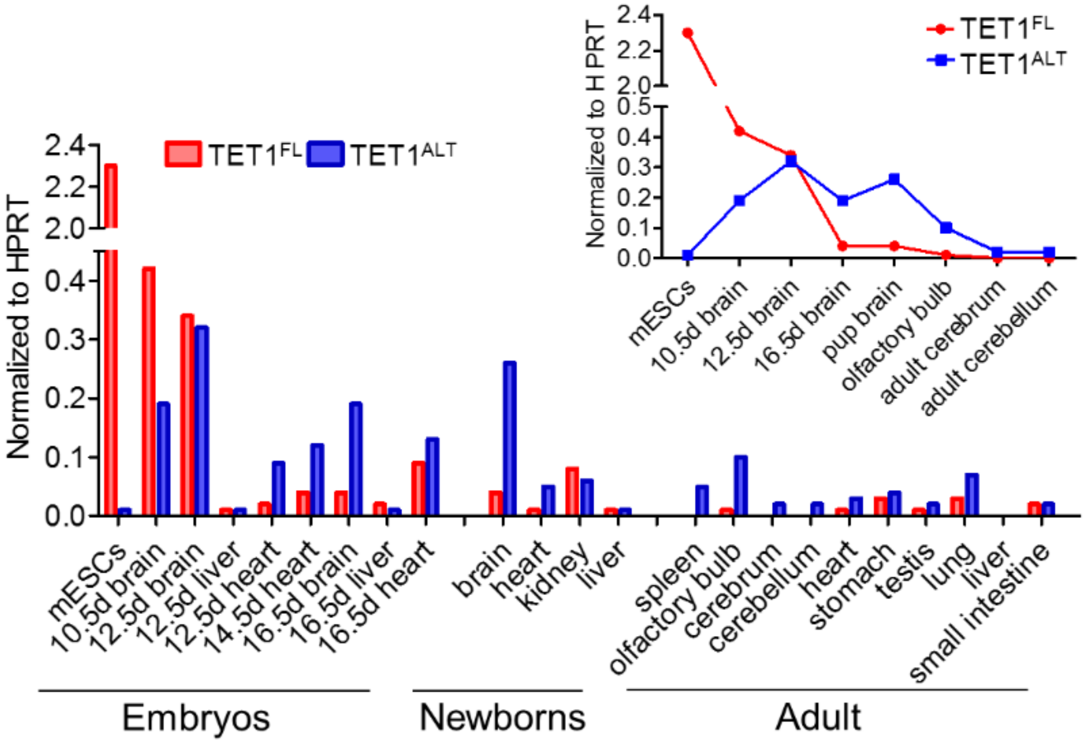


**Figure 9. Differential expression of the Tet1 isoforms during mouse development.**

qRT-PCR for Tet1 and Tet1$^{ALT}$ in mouse embryonic and adult tissues, normalized to HPRT ($2^{-dCt}$). Inset depicts expression of the isoforms in the developing brain. (N=1, two independent experiments performed).

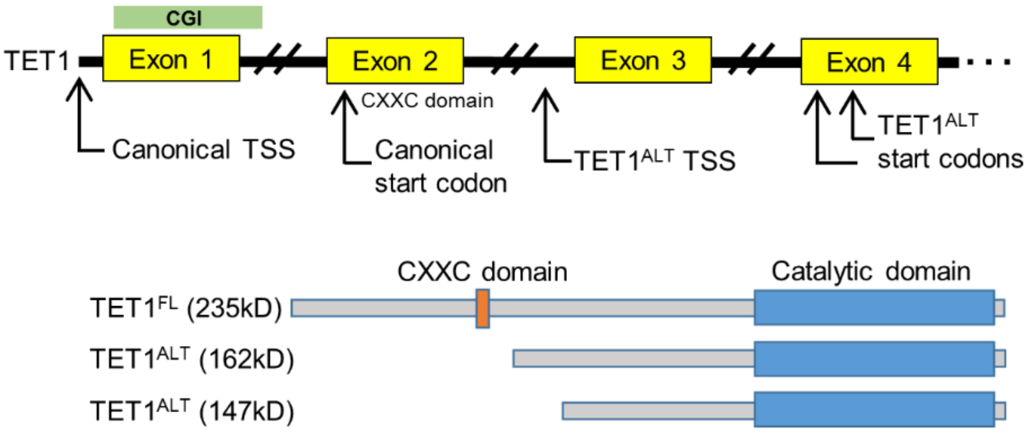


**Figure 10. TET1 gene model.**

Schematic illustrating the gene models for TET1 and TET1$^{ALT}$.

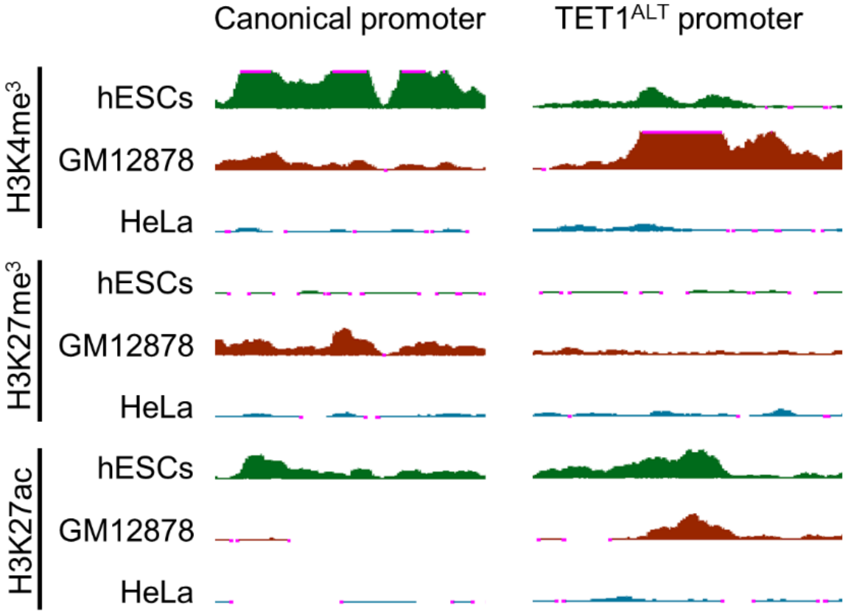


**Figure 11. TET1 and TET1$^{ALT}$ promoter histone marks in human.**

ChIP-seq histone marks (ENCODE) for H3K4me$^{3}$, H3K27me$^{3}$ and H3K27Ac in H1-hESCs, GM12878 and HeLa cells for the TET1 canonical and TET1$^{ALT}$ promoters.

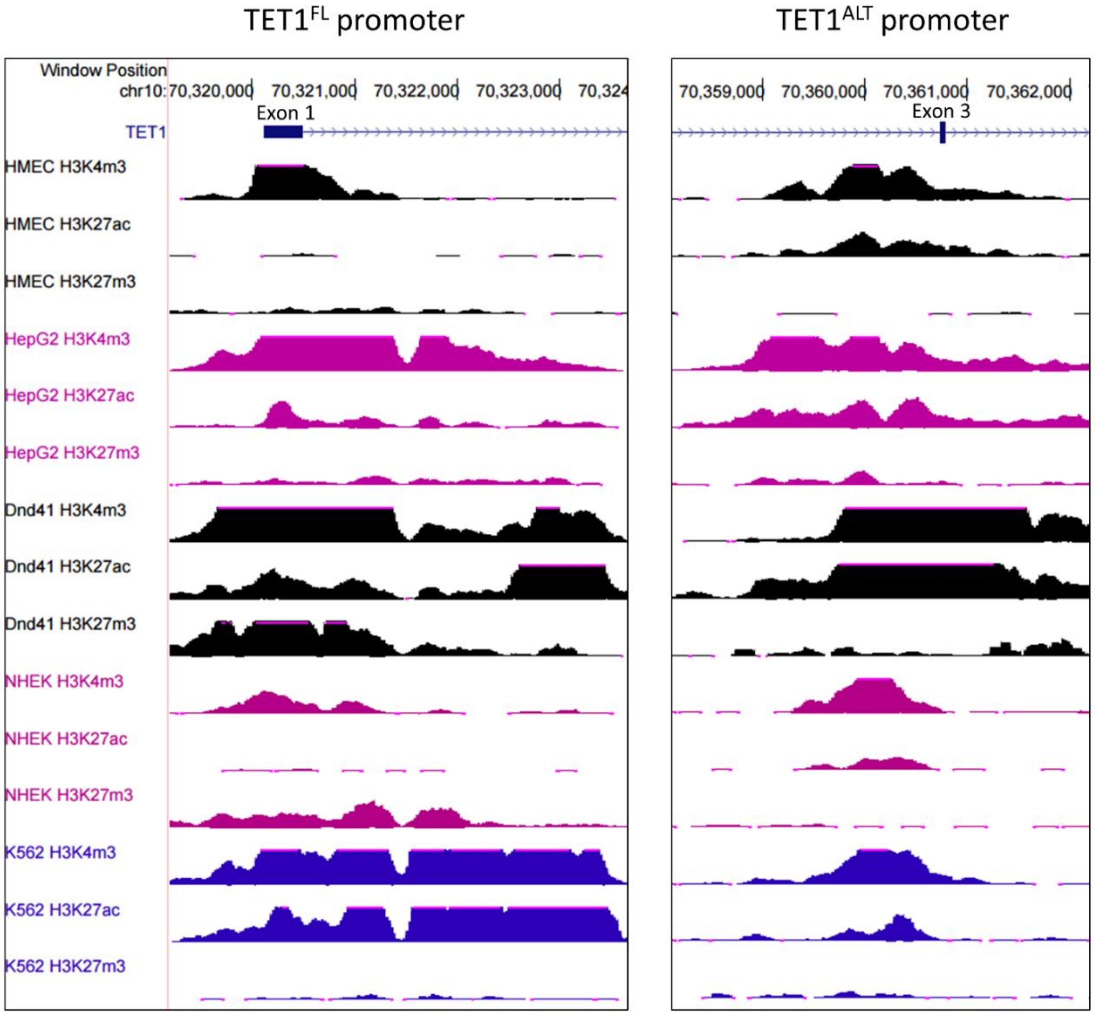


**Figure 12. Chromatin marks surrounding the TET1$^{FL}$ and TET1$^{ALT}$ promoters in human cell lines.**

Histone marks occupying TET1$^{ALT}$ and TET1$^{FL}$ promoters in HMEC, HepG2, Dnd41, NHEK and K562 cells. Histone marks include H3K4me3, H3K27Ac, and H3K27me3.

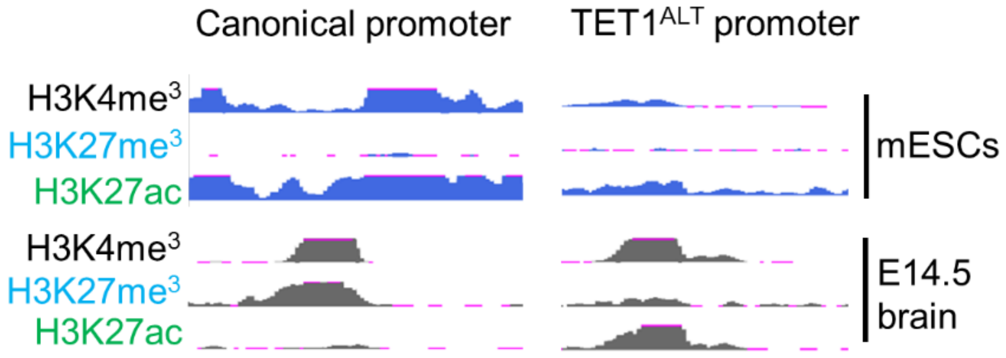


**Figure 13. TET1 and TET1$^{ALT}$ promoter histone marks in mouse.**

ChIP-seq histone marks H3K4me$^3$, H3K27me$^3$ and H3K27Ac in mESCs and embryonic day 14.5 brain (ENCODE).

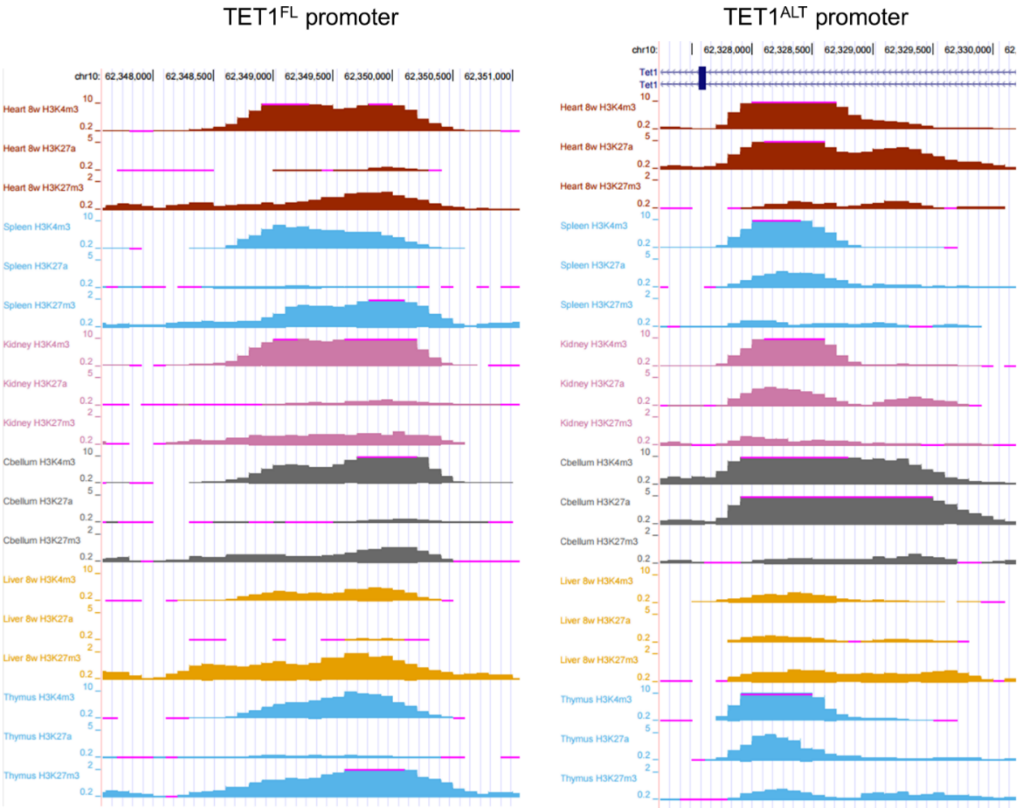


**Figure 14. Chromatin marks surrounding the TET1$^{FL}$ and TET1$^{ALT}$ promoters in mouse tissues.**

Histone marks surrounding the TET1$^{ALT}$ and TET1$^{FL}$ promoters in heart (red), spleen (blue), kidney (pink), cerebellum (gray), liver (yellow), and thymus (light blue) from adult mouse tissues. Histone marks include H3K4me3, H3K27Ac, and H3K27me3.

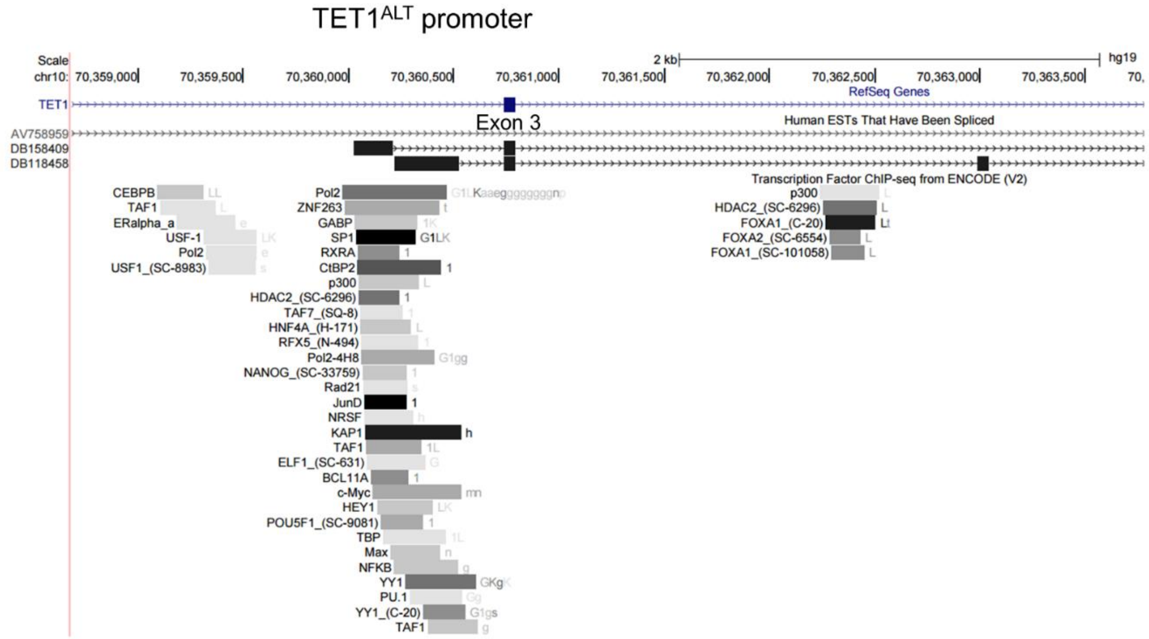


**Figure 15. Transcription factor binding at the TET1$^{ALT}$ promoter.**

Transcription factors enriched at the TET1$^{ALT}$ promoter based off ChIP-seq datasets from ENCODE (V2).

### 3.2.3 Confirming TET1$^{ALT}$ promoter activity

To confirm the promoter activity of TET1$^{ALT}$ in vitro, we cloned the promoter into a luciferase reporter assay and transfected it into HEK293T cells. A 996bp region corresponding to the alternate promoter sequence drove 40-fold higher levels of luciferase compared to empty vector and intron control (Figure 16A). This indicates the promoter region has DNA elements amenable to activate transcription, however, this assay does not establish if the region is a promoter or an enhancer.

To confirm that the TET1$^{ALT}$ promoter is driving expression of the TET1$^{ALT}$ transcripts, we used CRISPR dCas9 fused to the VP64 activator domain to activate transcription of each isoform independently. We designed gRNAs to target either the canonical or the alternate TET1 promoter and transfected them into an immortalized mammary cell line (MCF10A) that overexpressed the dCas9-VP64 fusion protein. By qRT-PCR, we found that only gRNAs tethered to the alternate promoter lead to an increase in TET1$^{ALT}$, while gRNAs tethered to the canonical promoter did not affect TET1$^{ALT}$ transcription (Figure 16B). Taken together, our data show that a conserved intragenic alternate promoter is used to activate transcription of an alternate isoform of TET1 that potentially retains the catalytic domain but lacks the CXXC DNA binding domain.

## 3.3 TET1$^{ALT}$ produces a truncated protein

### 3.3.1 Western blot for TET1 in a panel of cell lines

To test if TET1$^{ALT}$ produces a detectable protein, we performed a Western blot analysis using protein from a variety of cell lines and found that in addition to the TET1$^{FL}$ band at 235 kDa, there is a strong band at ~162 kDa and a smaller more variable band at 150 kDa (Figure 17). The smaller bands (marked by asterisks) may be non-specific as they persist after CRISPR knockout (see below). This was observed with multiple antibodies and is consistent with Western blots from published studies of the TET1 protein [79]. We evaluated expression levels in human ESCs because they express TET1$^{FL}$, but not TET1$^{ALT}$. As expected, a band is observed for TET1$^{FL}$ in the hESCs (Figure 17). Additionally, the smaller band at ~150 kDa is also observed, which is further evidence that this band is indeed non-specific. A closer look at the TET1$^{ALT}$ band at ~162 kDa reveals overexpression in several breast cancer cell lines (HCC2218, HCC1599, MCF7, MDA-MB-231) compared to the untransformed, immortalized breast lines (HMLE, MCF10A).

### 3.3.2 TET1$^{ALT}$ is a catalytically active DNA demethylase

To verify that the bands are indeed TET1$^{ALT}$, we used a pIRES expression construct to overexpress either empty vector, TET1$^{ALT}$ or TET1$^{FL}$ in HEK293T cells. Western blot analysis of these lysates using a TET1 antibody (Figure 18A) and a FLAG tagged antibody (Figure 18B) revealed that TET1$^{ALT}$ overexpression resulted in 1 major band around 162 kDa, which is not observed upon overexpression of the TET1$^{FL}$ isoform. Protein produced by the TET1$^{FL}$ isoform is found at ~235 kDa, as expected. We quantified the non-saturated bands (normalized to β-actin) and found TET1$^{ALT}$ is overexpressed 9.4 fold and TET1$^{FL}$ is overexpressed 5.4 fold compared to empty vector control. To test TET1$^{ALT}$ catalytic activity, we blotted DNA from the overexpressing HEK293T cells onto a membrane and probed with 5hmC (Figure 18C). As predicted, both TET1$^{ALT}$ and TET1$^{FL}$ display increased 5hmC, indicating TET1$^{ALT}$ can oxidize methylated DNA.

### 3.3.3 Knockout of TET1$^{ALT}$ reduces 5hmC

To complement the overexpression experiments and to verify that the band is specific to TET1, we designed CRISPR gRNAs to target a common exon shared between TET1 and TET1$^{ALT}$ (targeting the noncoding TET1$^{ALT}$ exons 1a or 1b alone was not successful). We generated knockouts in MDA-MB-231 cells, as this cell line expresses high levels of TET1$^{ALT}$ and low levels of TET1$^{FL}$. Upon knockout, we see a loss of the TET1$^{ALT}$ band at 162 kDa, and see minimal to no change in the lower band, further suggesting the lower band is likely non-specific (Figure 19A). Cloning and sequencing of the knockout cells confirmed that CRISPR induced frameshift mutations resulting in an early stop codon for both TET1 alleles (Figure 19B). These data confirm that TET1$^{ALT}$ expression results in a detectable protein that is truncated compared to TET1$^{FL}$.

To estimate the extent to which 5hmC is dependent on TET1$^{FL}$ or TET1$^{ALT}$, we performed a 5hmC slot blot in our TET1 knock out MDA-MB-231 cells. Upon loss of TET1$^{ALT}$, we see a 30% reduction compared to empty vector control (Figure 20). This is consistent with previous reports suggesting that TET2 and/or TET3 can compensate for 5hmC production in the absence of TET1 [59].

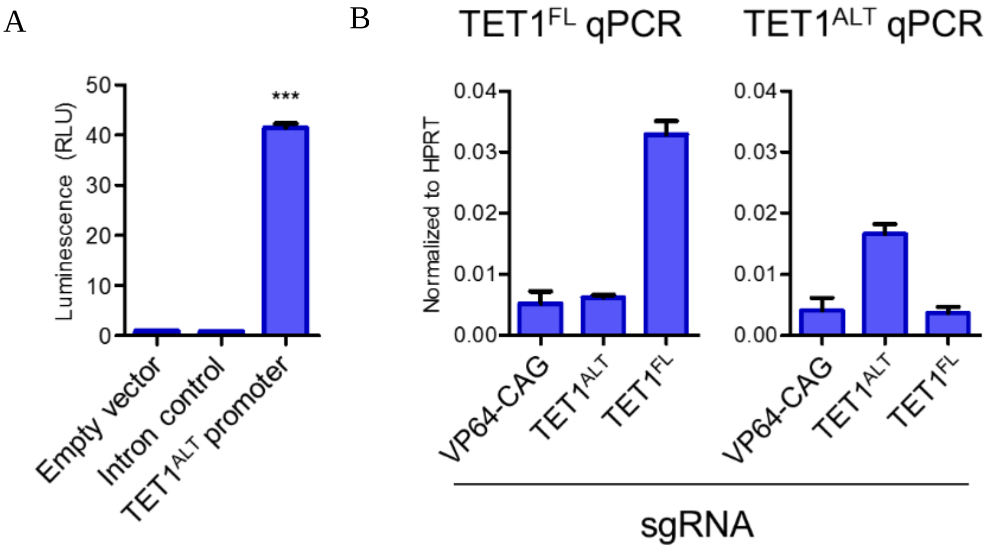


**Figure 16. TET1$^{ALT}$ promoter activity.**

**(A)** DNA fragments from a control intron region and the TET1$^{ALT}$ promoter region were each cloned into pGL4 luciferase reporter construct and transfected into HEK293T cells, along with empty vector control. Luminescence (RLU) is normalized to Renilla and empty vector control (technical and biological triplicates). **(B)** MCF10A cells were infected with lentiviruses expressing dCas9-VP64 with scrambled gRNAs (VP64-CAG) or gRNAs targeting the TET1$^{ALT}$ promoter or the TET1$^{FL}$ promoter. TET1$^{FL}$ and TET1$^{ALT}$ transcript expression was assayed by qRT-PCR (data shown as the average of biological duplicates, technical triplicates).

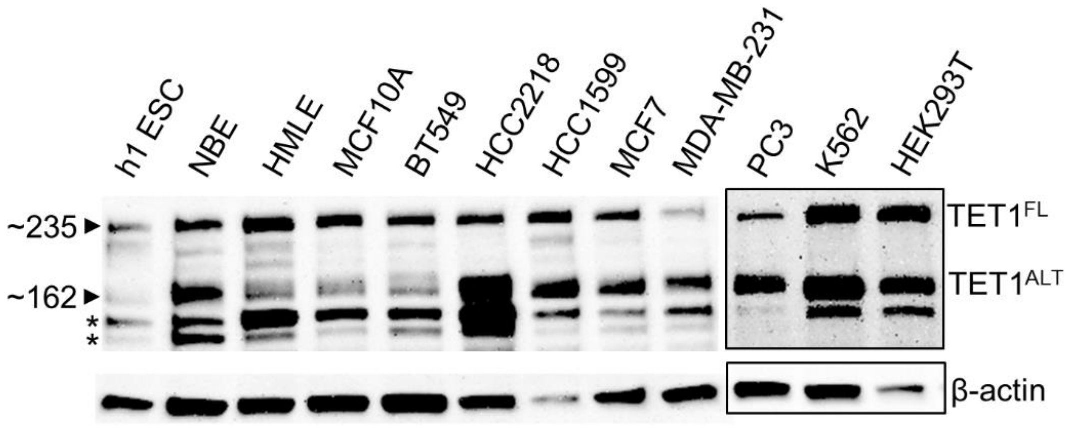


**Figure 17. TET1 western blot in a panel of cell lines.**

Western blot of the TET1 isoforms in human cell lines (two independent experiments performed, asterisk denotes non-specific bands).

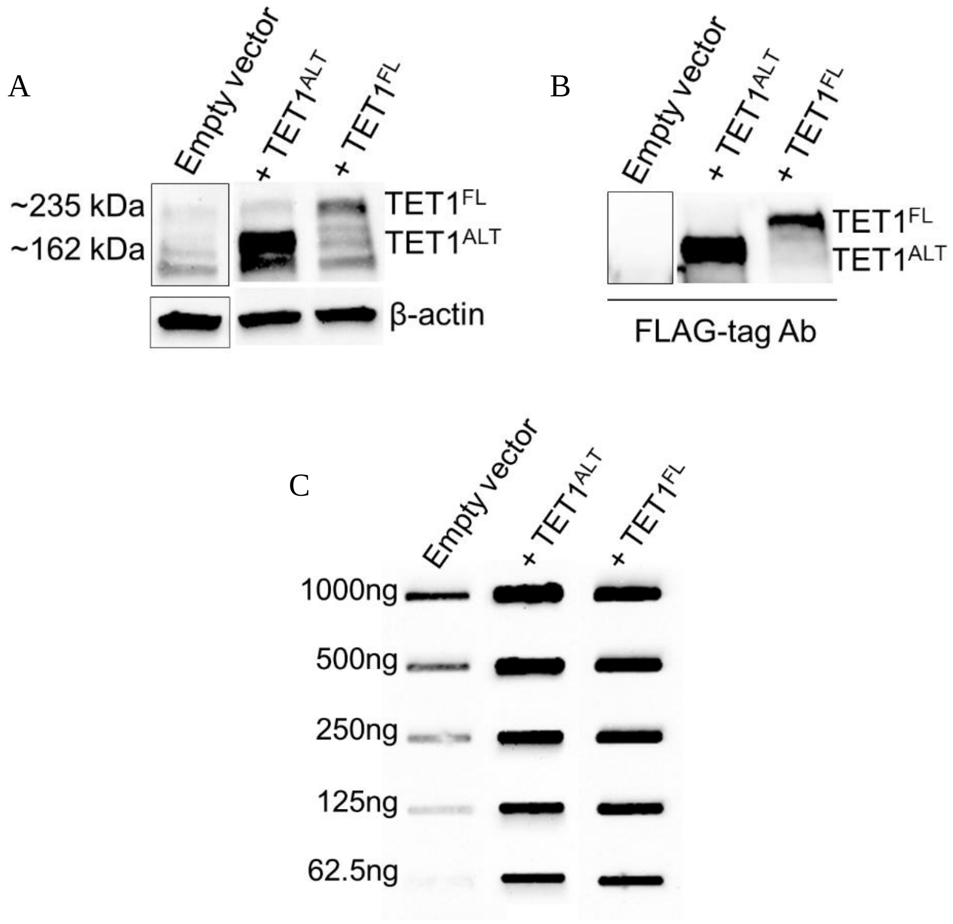
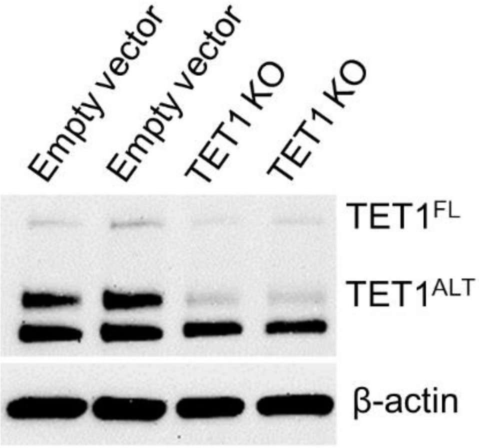


**Figure 18. Overexpression of TET1$^{FL}$ and TET1$^{ALT}$ increases 5hmC.**

**(A)** Western blot analysis of empty vector, TET1$^{ALT}$ or TET1$^{FL}$ overexpression in HEK293T cells with anti-TET1 antibody (two independent experiments performed). **(B)** Western blot analysis of empty vector, TET1$^{ALT}$ or TET1$^{FL}$ overexpression in HEK293T cells with anti-FLAG antibody. **(C)** 5hmC DNA slot blot of HEK293T cells expressing either empty vector, TET1$^{ALT}$ or TET1$^{FL}$.

A



B **Allele #1**

```
Wildtype      AGGTTCTTGCACATAAGATAAGGGCAGTGGAAAAGAAACCTATTCCCCGAATCAAGCGGA
              |||||||||||||||||||||||||||||||||||||##############|||||||||
TET1 KO       AGGTTCTTGCACATAAGATAAGGGCAGTGGAAAAGAA--------------TCAAGCGGA

Wildtype      AGAATAACTCAACAACAACAAACAACAGTAAGCCTTCGTCACTGCCAACCTTAGgtgagc
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TET1_KO       AGAATAACTCAACAACAACAAACAACAGTAAGCCTTCGTCACTGCCAACCTTAGGTGAGC
```

**Allele #2**

```
Wildtype      AGGTTCTTGCACATAAGATAAGGGCAGTGGAAAAGAAACCTATTCCCCGAATCAAGCGGA
              |||||||||||||||||||||||||||||||||||||#######################
TET1_KO       AGGTTCTTGCACATAAGATAAGGGCAGTGGAAAAGAA-----------------------

Wildtype      AGAATAACTCAACAACAACAAACAACAGTAAGCCTTCGTCACTGCCAACCTTAGgtgagc
              ############################################################
TET1_KO       ----------------------------------------------------------T-

Wildtype      cctatggaa
              |||||#|||
TET1_KO       CCTATTGAA
```

**Figure 19. TET1 KO in MDA-MB-231 cells.**

**(A)** Western blot of CRISPR TET1 knockout in MDA-MB-231 cells using an anti-TET1 antibody (technical duplicates, three independent experiments performed). **(B)** The TET1 knockout MDA-MB-231 single clone cells (TET1_KO) is aligned to the TET1 sequence (Wildtype). The gRNA target sequence is highlighted in purple. 12 clones/alleles were analyzed and sequenced. Allele #1 (top) has a 14bp deletion in the gRNA target sequence that results in a frameshift and introduces a premature stop codon. Allele #2 (bottom) has an 82bp deletion in the gRNA target sequence that results in a frameshift and introduces a premature stop codon.

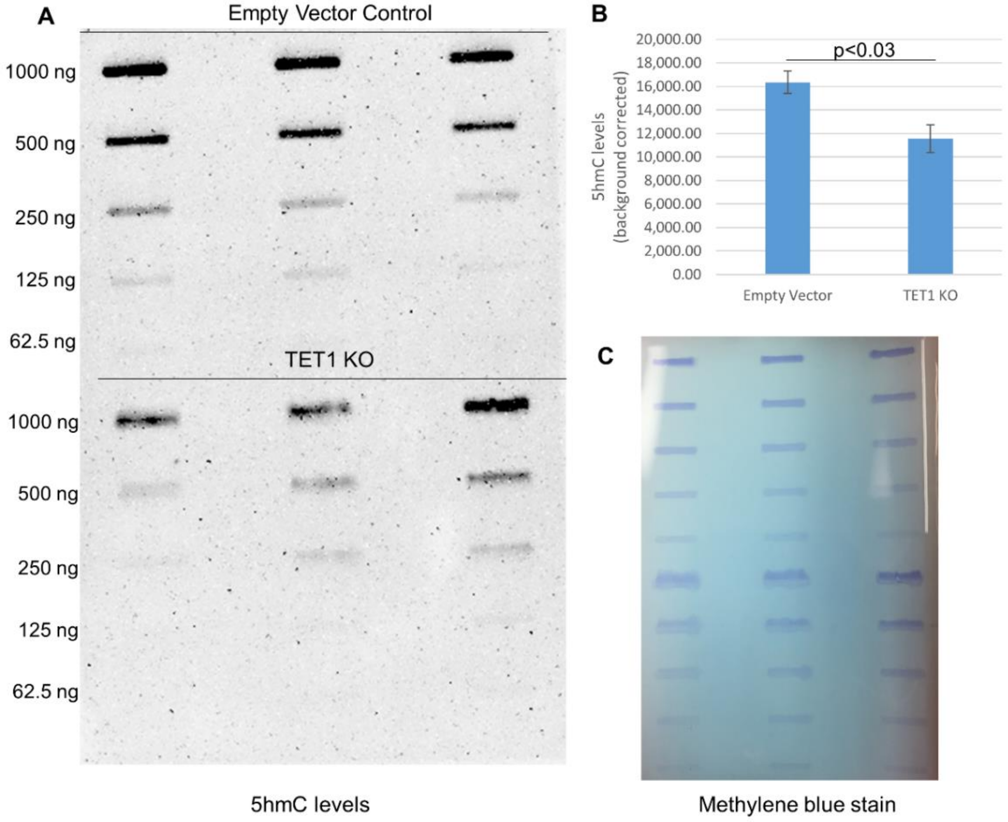


**Figure 20. TET1 KO results in a loss of 5hmC.**

**(A)** 5hmC slot blot in MDA-MB-231 cells in empty vector control (top) and TET1 KO (bottom) with varying concentrations of DNA all done in triplicates. **(B)** Quantification of 5hmC levels in empty vector control and TET1 KO cells. Error bars represent SEM. **(C)** Methylene blue stain of membrane to confirm even loading.

## 3.4 TET1$^{ALT}$ is functional and distinct from TET1$^{FL}$

### 3.4.1 Gene expression targets of TET1$^{ALT}$ and TET1$^{FL}$

TET1 has previously been reported to affect gene transcription, as loss of TET1 leads to both the upregulation and downregulation of genes [81 30]. To determine if TET1$^{ALT}$ also affects gene expression, we overexpressed empty vector, TET1$^{ALT}$ and TET1$^{FL}$ in HEK293T cells and performed RNA-seq. An unsupervised hierarchal cluster analysis and a principal component analysis (PCA) of the data showed that TET1$^{FL}$ clusters separately from empty vector and TET1$^{ALT}$ (Figure 21A, B). Although both isoforms have only modest effects on transcription (changes in <1.5% of the transcriptome), TET1$^{FL}$ overexpression induced substantially more gene expression changes than TET1$^{ALT}$: TET1$^{FL}$ led to the upregulation of 7.5-fold more genes than TET1$^{ALT}$ and repressed ~3 fold more genes. There was only a moderate overlap between the gene expression targets of TET1$^{FL}$ and TET1$^{ALT}$ with a large number of unique targets (Figure 21C).

### 3.4.2 DNA methylation targets of TET1$^{ALT}$ and TET1$^{FL}$

Previous reports found that the TET1 catalytic domain (TET1$^{CD}$, which lacks the CXXC domain) induced widespread hypomethylation, but TET1$^{FL}$ produced small changes in methylation at sites with low basal methylation [30]. Since TET1$^{ALT}$ lacks the CXXC domain, we wondered if it would induce widespread or targeted DNA demethylation. We first examined LINE1 methylation using bisulfite pyrosequencing. As expected, the expression of TET1$^{CD}$ led to demethylation of the LINE1 repetitive elements. However, neither TET1$^{FL}$ nor TET1$^{ALT}$ had major effects on LINE1 methylation (change of methylation <1%), indicating preserved target specificity (Figure 22A). We next analyzed genome-wide DNA methylation data using the DREAM method (Digital Restriction Enzyme Analysis of Methylation) [97]. Hierarchal cluster analysis of HEK293T cells overexpressing the TET1 isoforms showed that TET1$^{ALT}$ clusters more closely with empty vector than TET1$^{FL}$ (Figure 22B), indicating weaker effects on DNA methylation. To visualize the sites that changed methylation, we used volcano plots and found that both TET1$^{FL}$ and TET1$^{ALT}$ expression led to changes in DNA methylation compared to empty vector control (Figure 22C, D). TET1$^{FL}$ demethylated 3-fold more target genes than TET1$^{ALT}$ (770 CG sites), but TET1$^{ALT}$ still decreased the methylation of 225 CG sites by at least 5%. Furthermore, TET1$^{FL}$ and TET1$^{ALT}$ mostly have their own individual target CG sites (Figure 22E). A small number of genes gained methylation with overexpression of either isoform, which may represent background/false positives.

Taken together, our data show that TET1$^{ALT}$ is functional but relatively weak when overexpressed alone, likely because it needs to be recruited to DNA by specific co-factors, as previously shown [79 82]. Our study does not address whether TET1$^{ALT}$ is physiologically different from TET1$^{FL}$, but instead illustrates that the two proteins have different gene expression and methylation targets and are thus functionally distinct.

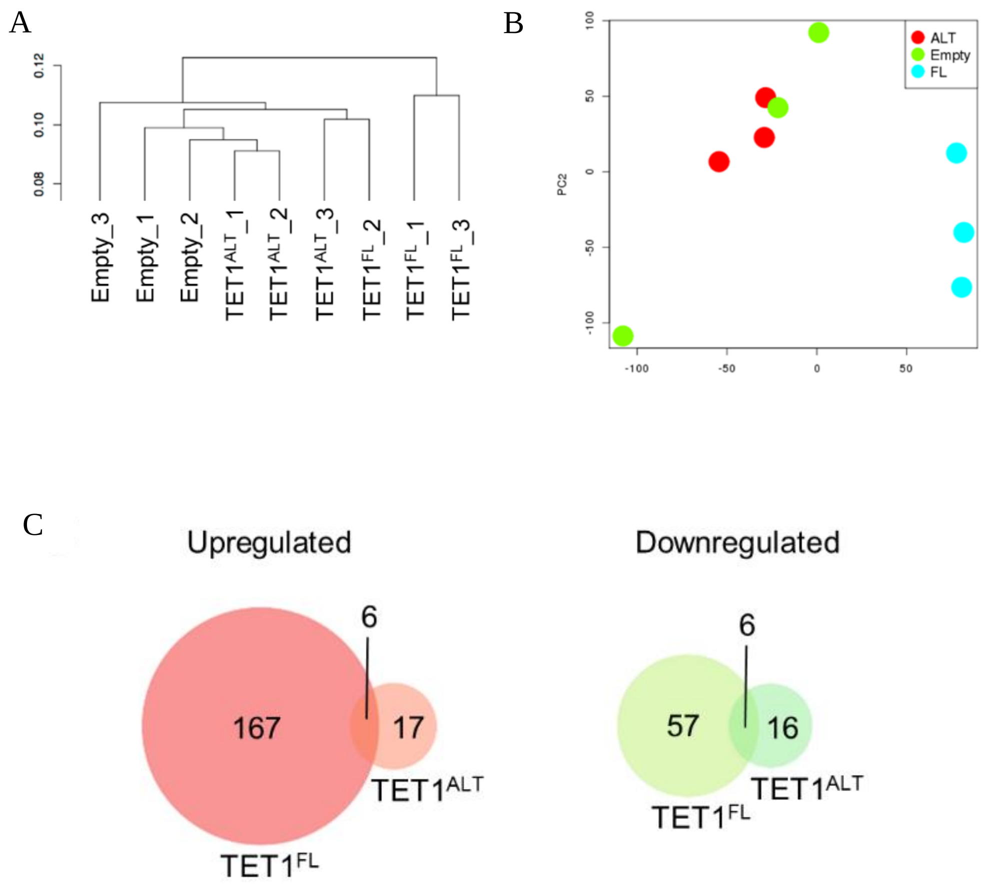


**Figure 21. Gene expression targets of TET1$^{FL}$ and TET1$^{ALT}$.**

**(A)** Dendrogram of RNA-seq hierarchical clustering of HEK293T cells overexpressing empty vector, TET1$^{ALT}$ or TET1$^{FL}$ (N=3, biological triplicates). **(B)** Principal component analysis of RNA-seq data. **(C)** Overlap of genes upregulated and downregulated in TET1$^{ALT}$ vs TET1$^{FL}$ HEK293T cells, FC>1.5 or FC<0.66, p<0.05).

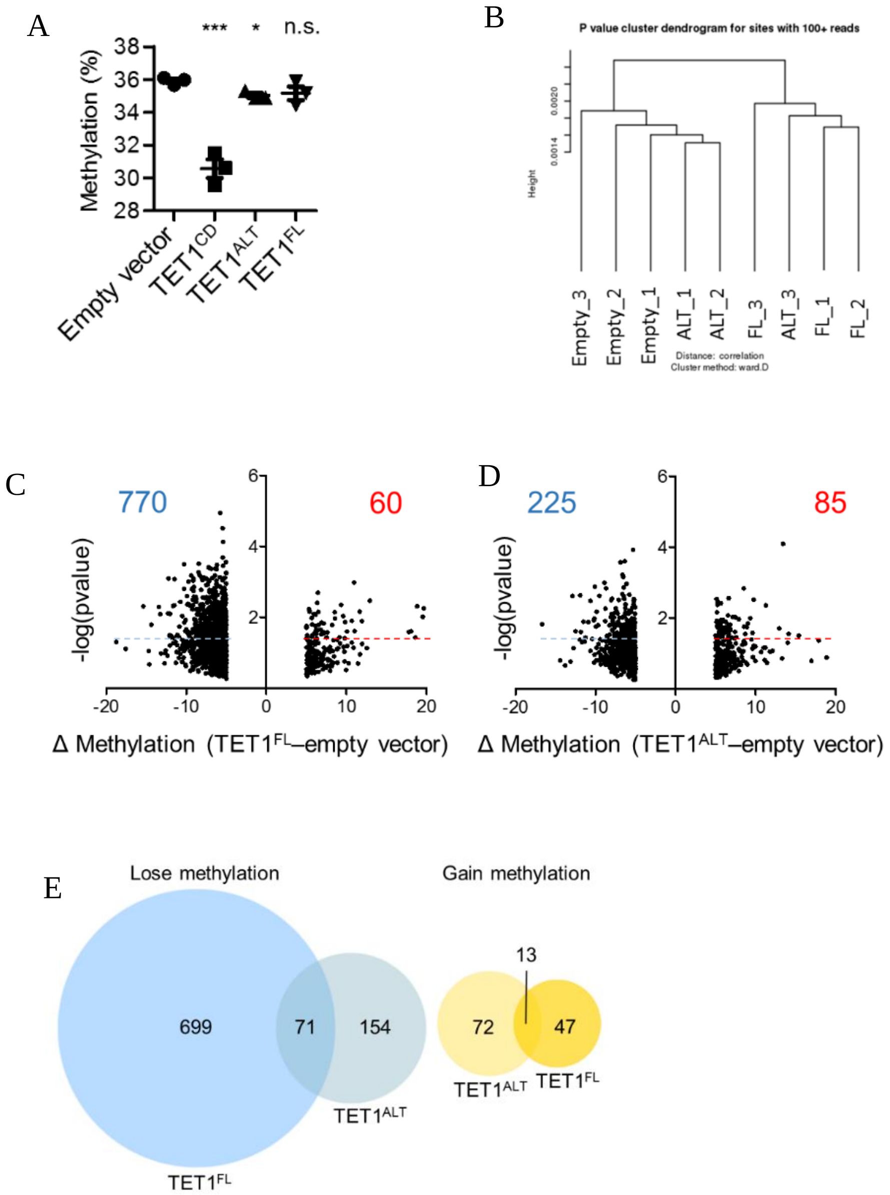


**Figure 22. Methylation targets of TET1$^{FL}$ and TET1$^{ALT}$.**

**(A)** LINE-1 pyrosequencing assay of methylation in HEK293T cells overexpressing empty vector, TET1$^{CD}$, TET1$^{ALT}$ or TET1$^{FL}$ (technical and biological triplicates). **(B)** Dendrogram of DNA methylation data filtered for CG sites with greater than 50 tags in HEK293T cells overexpressing empty vector, TET1$^{ALT}$ or TET1$^{FL}$. Volcano plot analysis from genome-wide DNA methylation data (DREAM) comparing empty vector to TET1$^{FL}$ **(C)** or TET1$^{ALT}$ **(D)**. CG sites were filtered for minimum sequencing depth of 50 reads and methylation changes >5%. Each point is the average methylation value for biological triplicates. **(E)** Overlap of genes that lose and gain methylation in HEK293T cells that overexpress either TET1$^{ALT}$ or TET1$^{FL}$ compared to empty vector control.

## 3.5 TET1$^{ALT}$ is overexpressed in cancer

### 3.5.1 Transcription start site reads

The role of TET1 in cancer remains under debate and previous reports of “loss” of TET1 may relate to the downregulation of TET1$^{FL}$ post-development. We therefore sought to examine isoform specific TET1 expression in cancer. We first examined publicly available TSS-seq datasets generated by a method that combines oligo capping with Illumina GA technology to map the exact positions of transcriptional start sites [112]. We found that very few reads map to the TET1$^{FL}$ or TET1$^{ALT}$ TSS in normal tissues (Figure 23A, B), which corroborates RNA-seq data showing that TET1 is expressed at low levels in normal adult tissues (GTEx Portal) (Figure 24) [112 113]. However, breast, lung cancer and Burkitt's lymphoma have substantially more TET1$^{ALT}$ TSS reads than their normal tissue counterparts (Figure 25A). Interestingly, there is little to no increased activity at the canonical TET1$^{FL}$ promoter in these cell lines (Figure 25B) indicating that TET1$^{ALT}$ is specifically activated in cancer according to TSS data. For example, 17/22 lung cancer cell lines have multiple reads mapping to the TET1$^{ALT}$ alternate exons (Figure 26B) but only 2 lung cancer cell lines show activation at the TET1$^{FL}$ promoter (Figure 26A). Taken together, these data show that the TET1$^{ALT}$ TSS is aberrantly activated in multiple cancers, suggesting that TET1$^{ALT}$ may be a cancer specific alternate isoform involved in cancer progression.

### 3.5.2 TET1 expression in cancer

We performed isoform specific qRT-PCR to identify TET1$^{ALT}$ expression levels in breast cells. Figure 27A shows TET1$^{ALT}$ expression in two untransformed, immortalized breast cell lines (MCF10A and HMLE) and four breast cancer cell lines (MCF7, HCC2218, HCC1599, and BT549). TET1$^{ALT}$ was overexpressed in several breast cancer cell lines, with the highest expression being in HCC1599.

To determine if TET1$^{ALT}$ is overexpressed in primary human samples, we used RNA-sequencing files from TCGA to quantify TET1$^{ALT}$ by counting the reads aligning to the TET1$^{ALT}$ exons (exon 1a and exon 1b). TET1$^{ALT}$ is expressed at low levels in normal breast (N=107) but is substantially overexpressed in breast cancer patient samples (N=807), Mann-Whitney p=0.03 (Figure 27B). Compared to the average TET1$^{ALT}$ expression in normal breast, most breast cancers overexpress TET1$^{ALT}$, however, a subset of cases show dramatic overexpression. Furthermore, reads aligning to TET1$^{ALT}$ exon 1a showed splicing to exon 3 in breast cancer patient samples (Figure 27C), similar to our in vitro splicing analyses.

We also analyzed additional cancer types, including uterine cancer and glioblastoma. Compared to their normal tissue counterparts, uterine cancer and glioblastoma significantly overexpress the TET1$^{ALT}$ isoform (Figure 28).

Next, we investigated differential TET1 exon usage in cancer. Since TET1$^{FL}$ and TET1$^{ALT}$ both use exons 3-12, and TET1$^{ALT}$ does not use exons 1-2, a discrepancy in usage of the first 2 exons indicates differential isoform expression. In breast cancer, glioblastoma, uterine cancer and AML exons towards the 3' end of the gene are used much more frequently than exons 1-2, evidence of an alternate truncated isoform (Figure 29). Finally, because these datasets indicated that TET1$^{ALT}$ is the predominant isoform overexpressed in cancer, we next looked at TET1 expression across multiple cancers using TCGA data. We find TET1 has a wide range of expression in several cancer types including AML, ovarian, breast and lung, but has very low and tight expression in colorectal, renal, pancreatic and prostate cancer (Figure 30A).

### 3.5.3 TET1 may be playing a more oncogenic role in select cancers

TET1 appears to be frequently amplified in some cancers, indicating that TET1$^{ALT}$ may play an oncogenic role in cancer (Figure 30) [104 114 115 116 117 118]. Next, we determined whether TET1 expression associates with overall survival outcomes in patient samples using data downloaded from TCGA and METABRIC [105 106 107]. Patients were considered TET1 high if expression levels were greater than 1 standard deviation (stdev) above the mean for all cancers except breast (breast stdev>1 p=0.05). For breast cancer, a more stringent cutoff of stdev>2 was used as there were many patients that fit this criteria (N=83), whereas in glioblastoma only 1 patient has stdev>2. Interestingly, TET1 expression is associated with a worse overall survival in cancers that are predominantly found in women (uterine p=0.001, breast p=0.01, ovarian p=0.007), is associated with a better survival in glioblastoma (p=0.04) and is not associated with survival in AML (Figure 31).

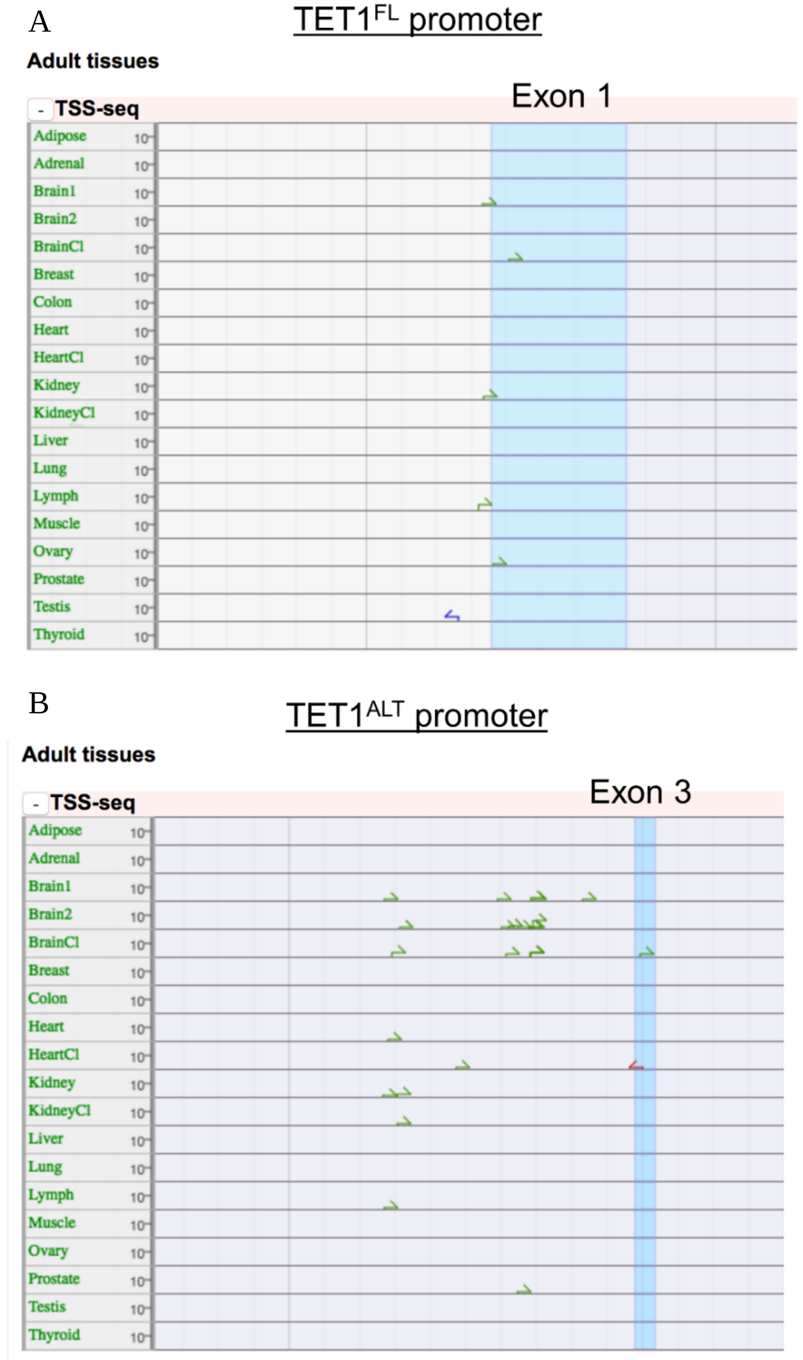


**Figure 23. TET1$^{ALT}$ and TET1$^{FL}$ TSS-seq data in normal human tissues [112].**

**(A)** TSS-seq reads in normal tissues at the TET1$^{FL}$ promoter. **(B)** TSS-seq reads in normal tissues at the TET1$^{ALT}$ promoter.

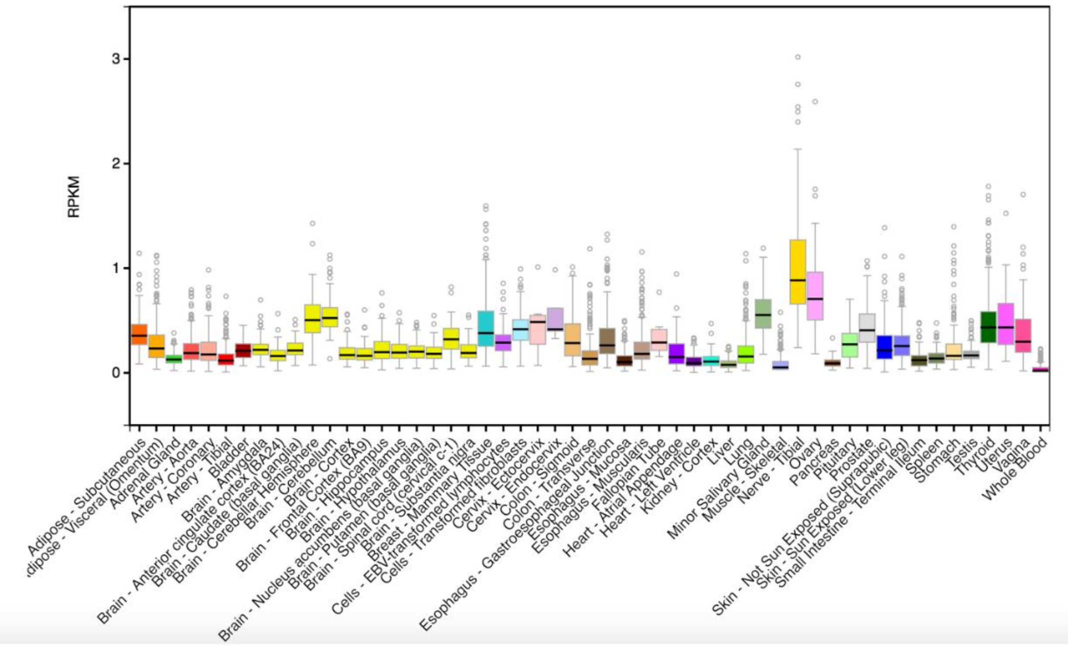


**Figure 24. TET1 expression in normal tissues** [113]**.**

RNA-seq TET1 expression data (RPKM) across normal human tissues using the Broad Institutes GTEx Portal.

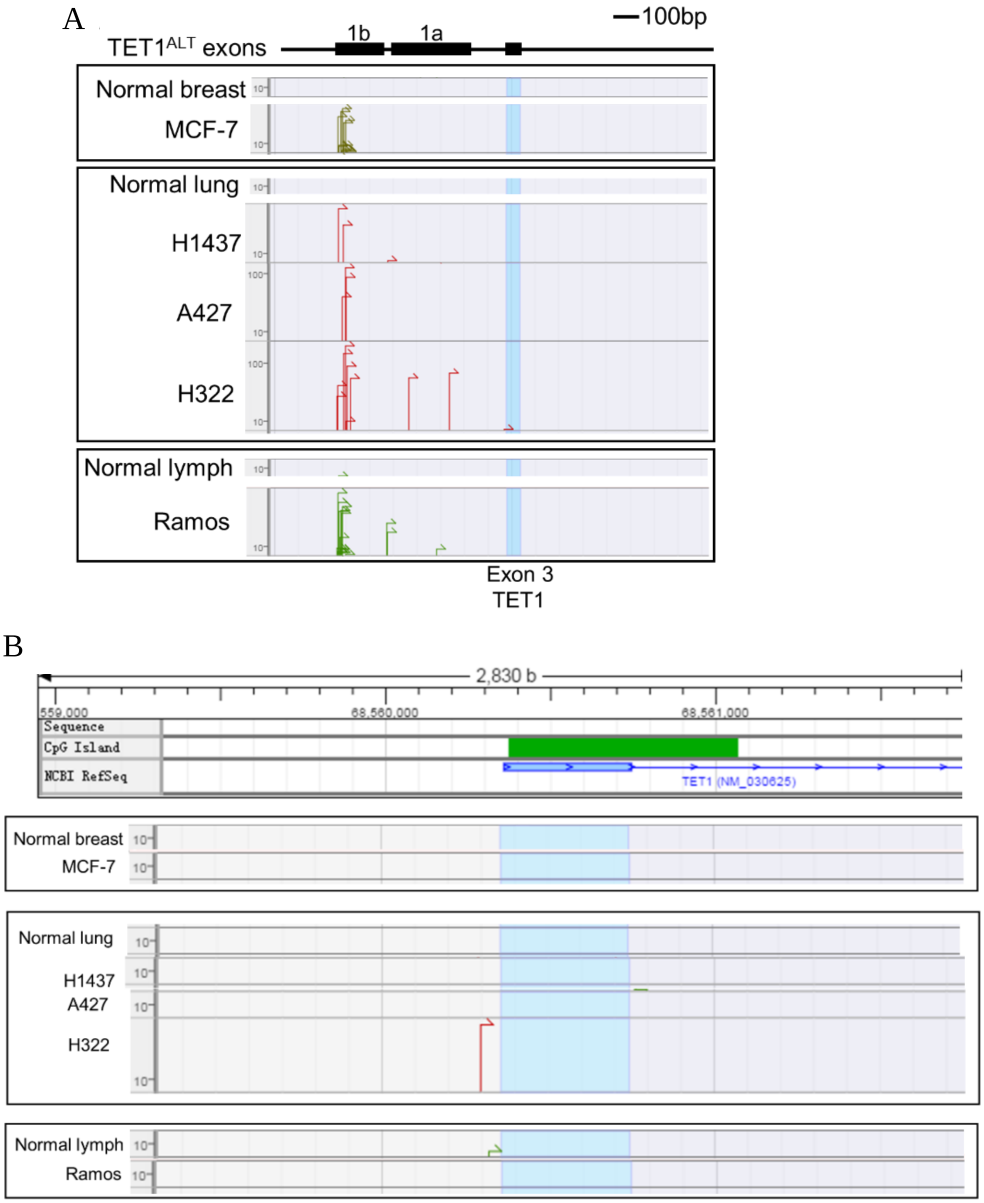


**Figure 25. TSS of TET1$^{FL}$ and TET1$^{ALT}$ in normal tissue vs cancer [112].**

**(A)** TSS-seq datasets for normal breast vs a breast cancer cell line, normal lung vs three lung cancer cell lines and normal lymph vs lymphoma cell line showing TSS peaks at the TET1$^{ALT}$ promoter [112]. **(B)** TSS-seq data at the TET1$^{FL}$ transcription start site in normal breast, lung and lymph compared to the breast cancer cell line (MCF-7), lung cancer cell lines (H1437, A427, H322) and lymphoma cell line (Ramos).

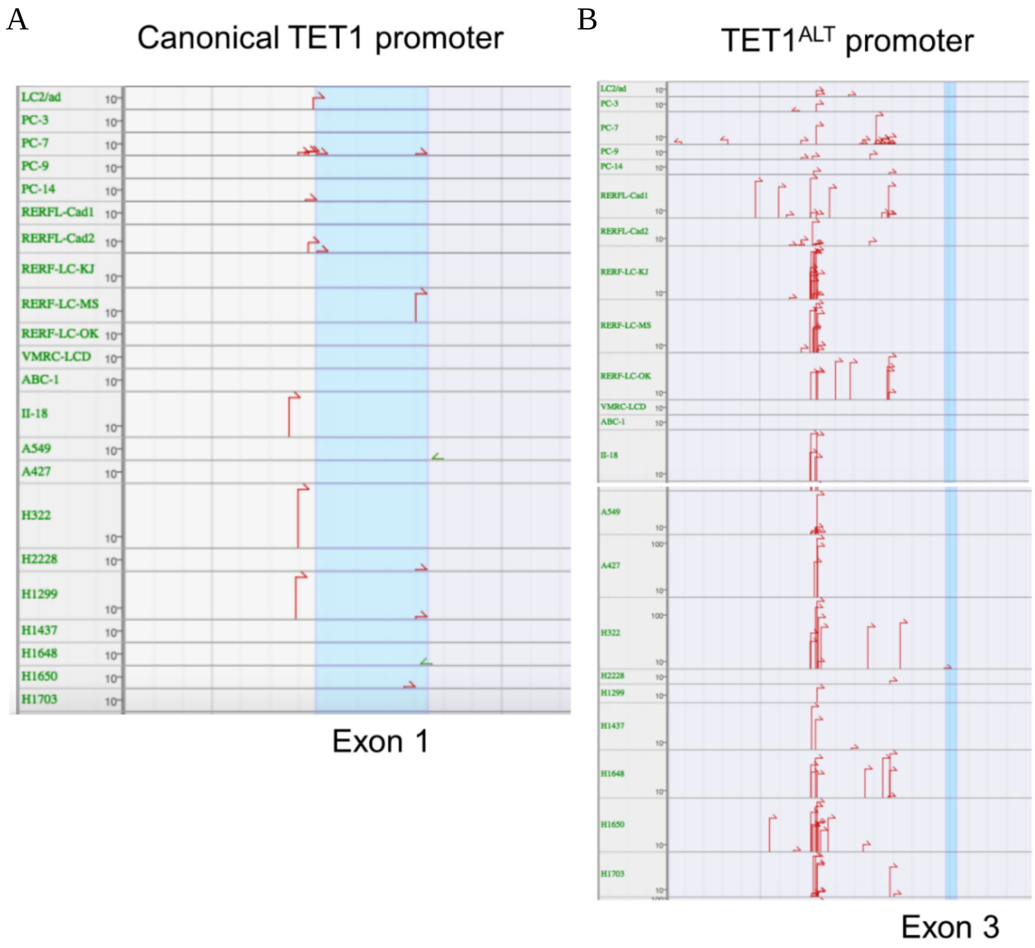


**Figure 26. TSS-seq data from 22 lung cancer cell lines at the TET1$^{FL}$ and TET1$^{ALT}$ transcription start sites [112].**

**(A)** TSS-seq reads in 17 lung cancer cell lines at the TET1$^{FL}$ promoter [112]. **(B)** TSS-seq reads in 17 lung cancer cell lines at the TET1$^{ALT}$ promoter.

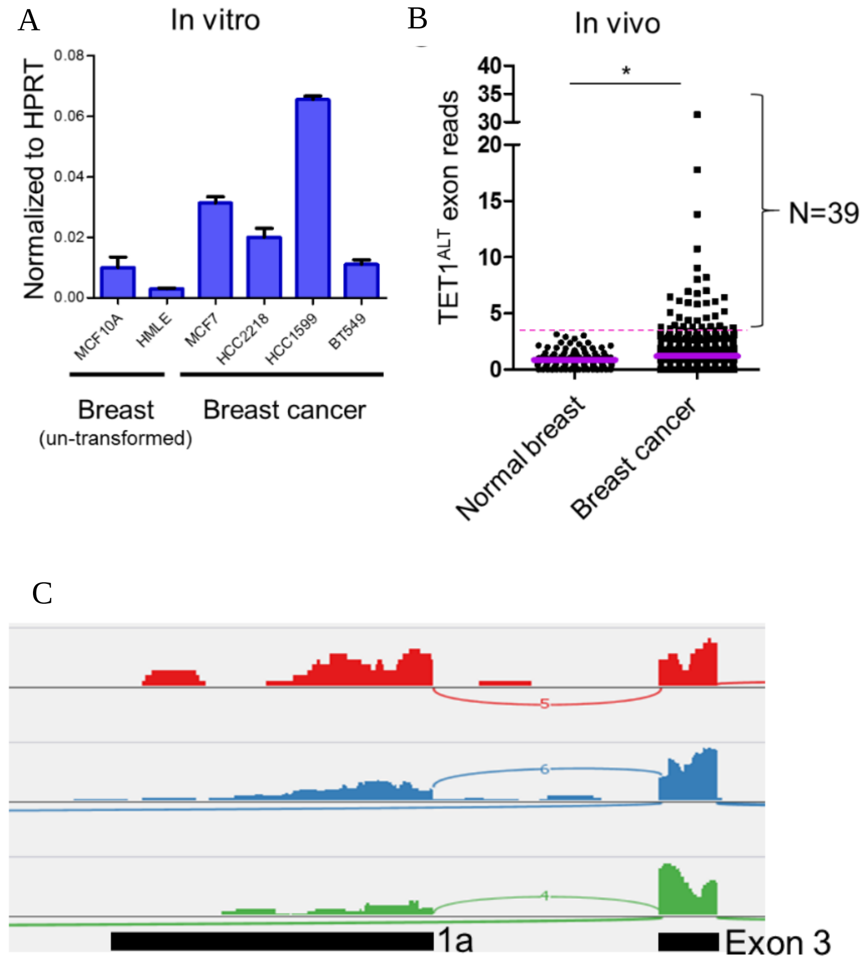


**Figure 27. TET1$^{ALT}$ expression in breast cancer.**

**(A)** In vitro analysis of TET1$^{ALT}$ RNA expression across a variety of human cell lines by qRT-PCR (technical and biological triplicates). **(B)** In vivo analysis of TCGA data of TET1$^{ALT}$ RNA-sequencing reads from normal breast tissue (N=107) and tissue from breast cancer patients (N=807). Y axis is TET1$^{ALT}$ reads (TET1$^{ALT}$ exon 1a + exon 1b)/chromosome 10 reads x 1e$^{6}$. Mann-Whitney test, p=0.03. **(C)** Sashimi plot of RNA-sequencing reads from 3 representative breast cancer patient samples aligned to the TET1$^{ALT}$ exons.

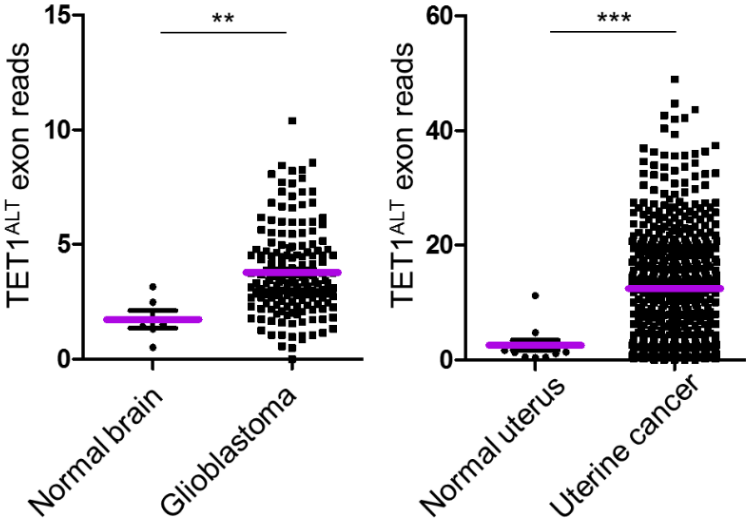


**Figure 28. TET1$^{ALT}$ expression in glioblastoma and uterine cancer.**

In vivo analysis of TCGA data of TET1$^{ALT}$ RNA-sequencing reads from normal brain (N=6) and glioblastoma tumor samples (N=167, p=0.004) (left) and normal uterine (N=12) and uterine tumor samples (N=574, p<0.0001) (right).

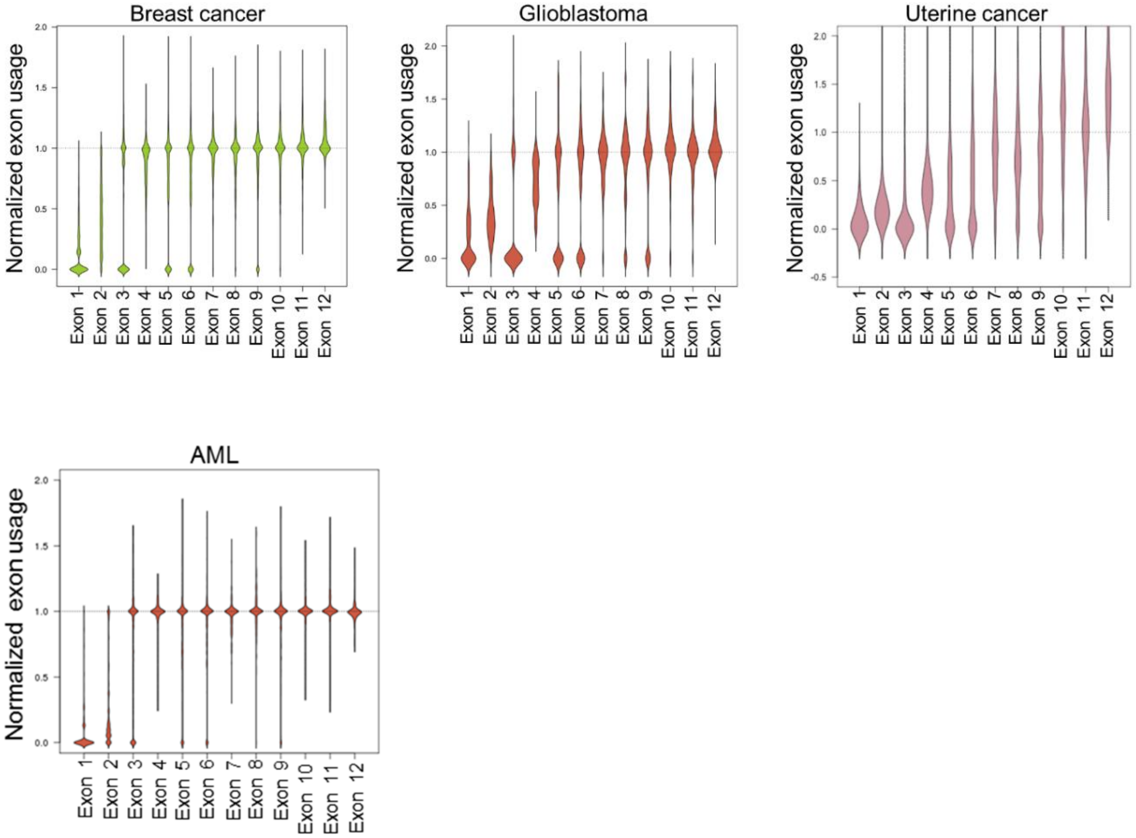


**Figure 29. TET1 exon usage in cancer.**

Normalized exon usage for TET1 in breast cancer, glioblastoma, uterine cancer and AML using TCGA datasets.

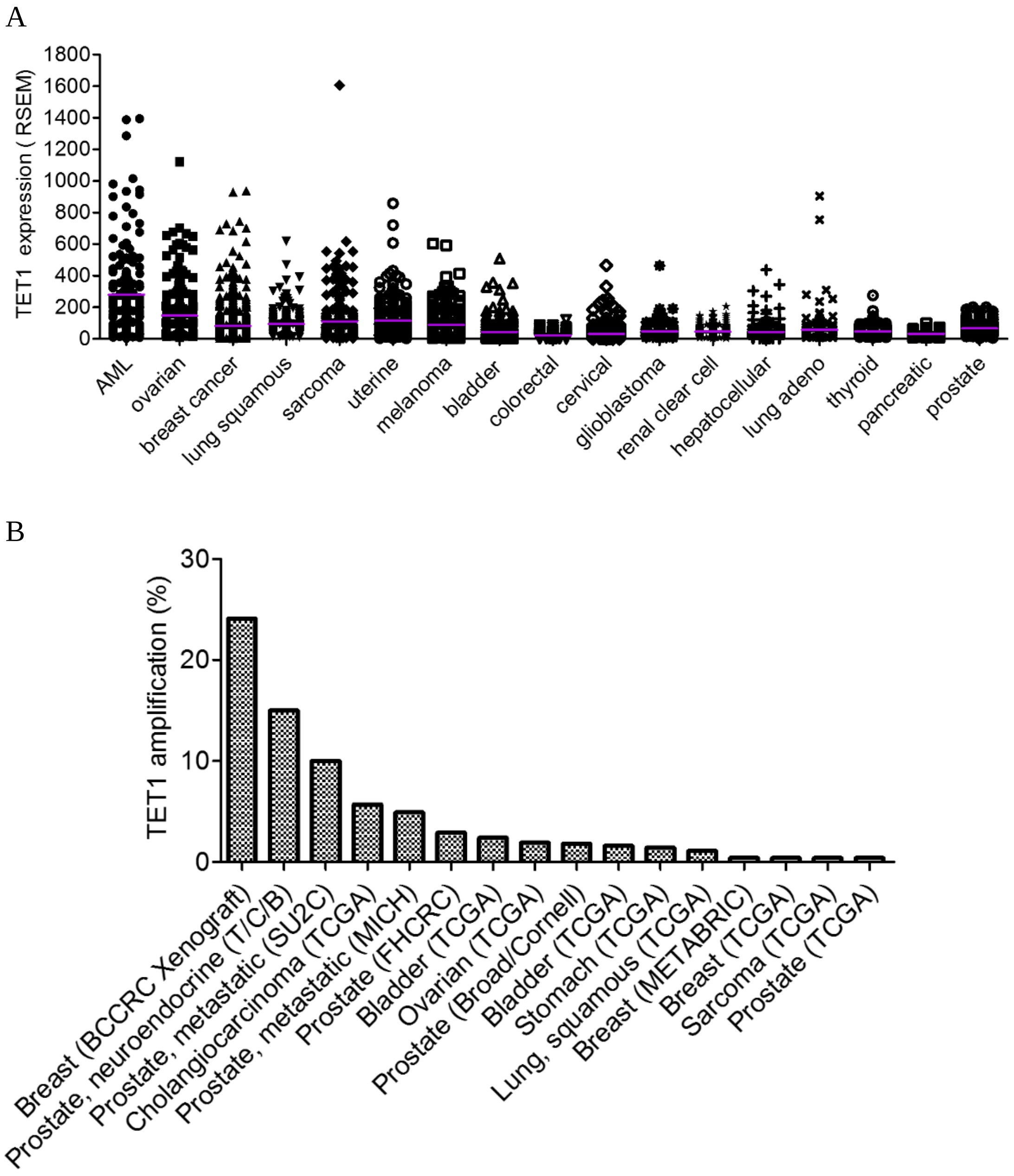


**Figure 30. TET1 expression and amplification in cancer (TCGA datasets).**

**(A)** TET1 expression (RNA-seq, RSEM) in 17 cancer types. Data downloaded from TCGA, CBioportal. **(B)** Amplification frequency in TCGA datasets.

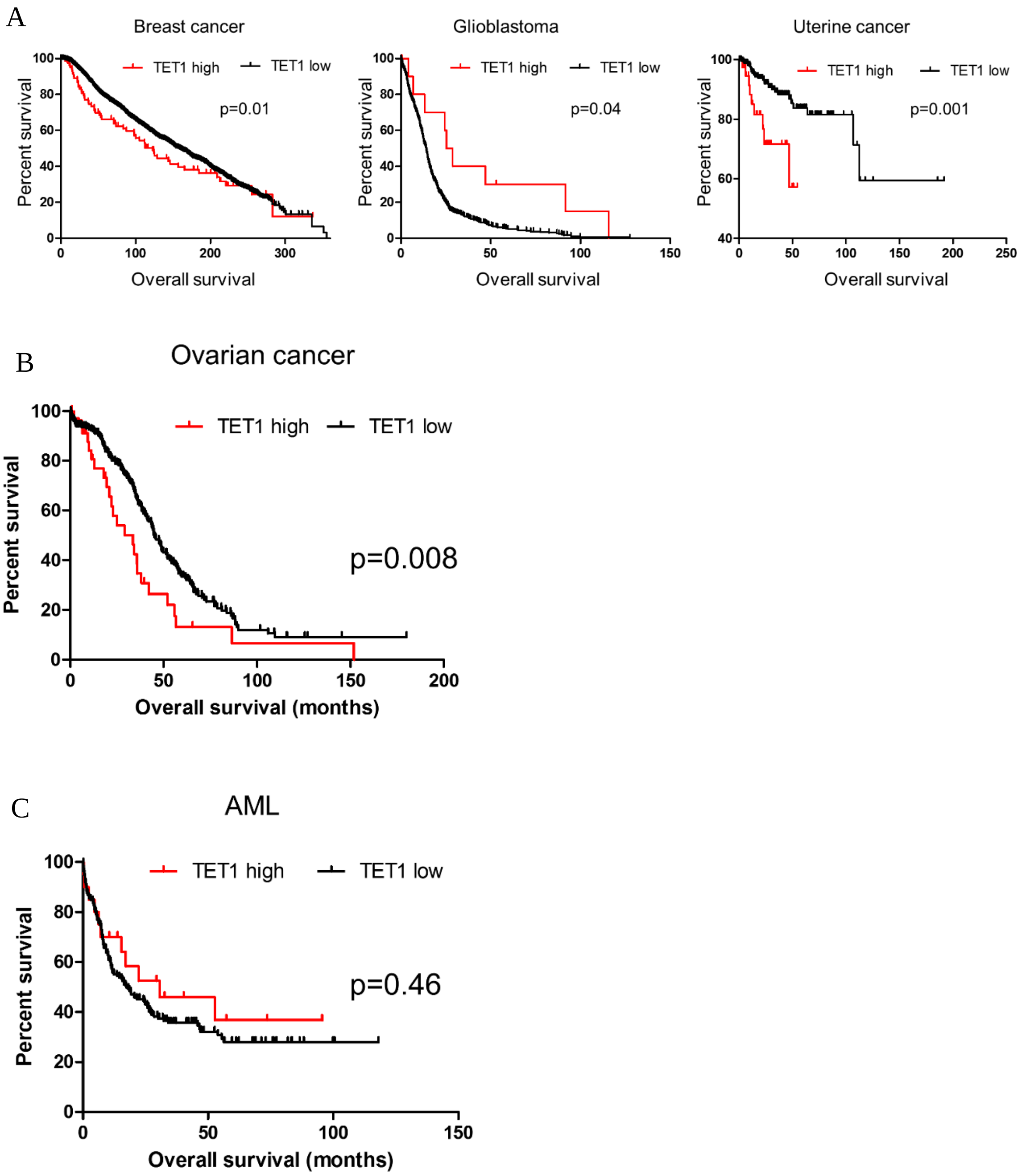


**Figure 31. Overall survival analysis based off of TET1 expression.**

**(A)** Survival curves based on TET1 expression in breast cancer (TET1 high is stdev>2 above the mean, stdev>1 p=0.05), glioblastoma and uterine cancer (stdev>1 above the mean). **(B)** TET1 high expression is associated with a worse overall survival in ovarian cancer (TCGA, Provisional). **(C)** TET1 is not associated with survival in Acute Myeloid Leukemia (AML) (TCGA, NEJM 2013). TET1 high is classified as expression >1 standard deviation above the mean.

## 3.6 Expression of the TET enzymes in breast cancer

### 3.6.1 Expression of TET1, TET2 and TET3

Because TET1 is overexpressed in breast cancer, and overexpression associates with a worse overall survival, we set out to explore the role of the TET enzymes and DNMTs in breast cancer. DNA methylation can divide breast cancers into multiple groups, with TNBCs characterized as having the lowest levels of methylation compared to the other breast cancer subtypes [42] [43] [44]. We used RNA-Seq data downloaded from The Cancer Genome Atlas (TCGA) to investigate whether expression of DNMTs or TETs could explain the methylation differences between TNBCs (N=100 patients) and hormone receptor positive breast cancers (HRBCs) (N=732 patients). Compared to normal breast (N=105), TET1 showed dramatic differences; it was significantly repressed in HRBCs ($p<0.0001$) and substantially overexpressed in TNBCs ($p<0.0001$), while unchanged in HER2 cases ($p=0.65$) (Figure 32A). TET2 was downregulated in all subtypes while TET3 was overexpressed in both TNBCs and HRBCs (Figure 32B, C). The fact that TET3 is high in all patients is interesting but does not explain the differences in methylation between the subtypes. Overexpression of TET1 in TNBC was confirmed in two independent datasets (METABRIC and GSE27447) (Figure 32D, E). DNMTs were overexpressed in cancer but equally so in all subtypes (Figure 33). We focused on TET1 because it showed this dichotomy where it was high in TNBC but low in HRBC and we hypothesized TET1 could be a candidate to explain the methylation

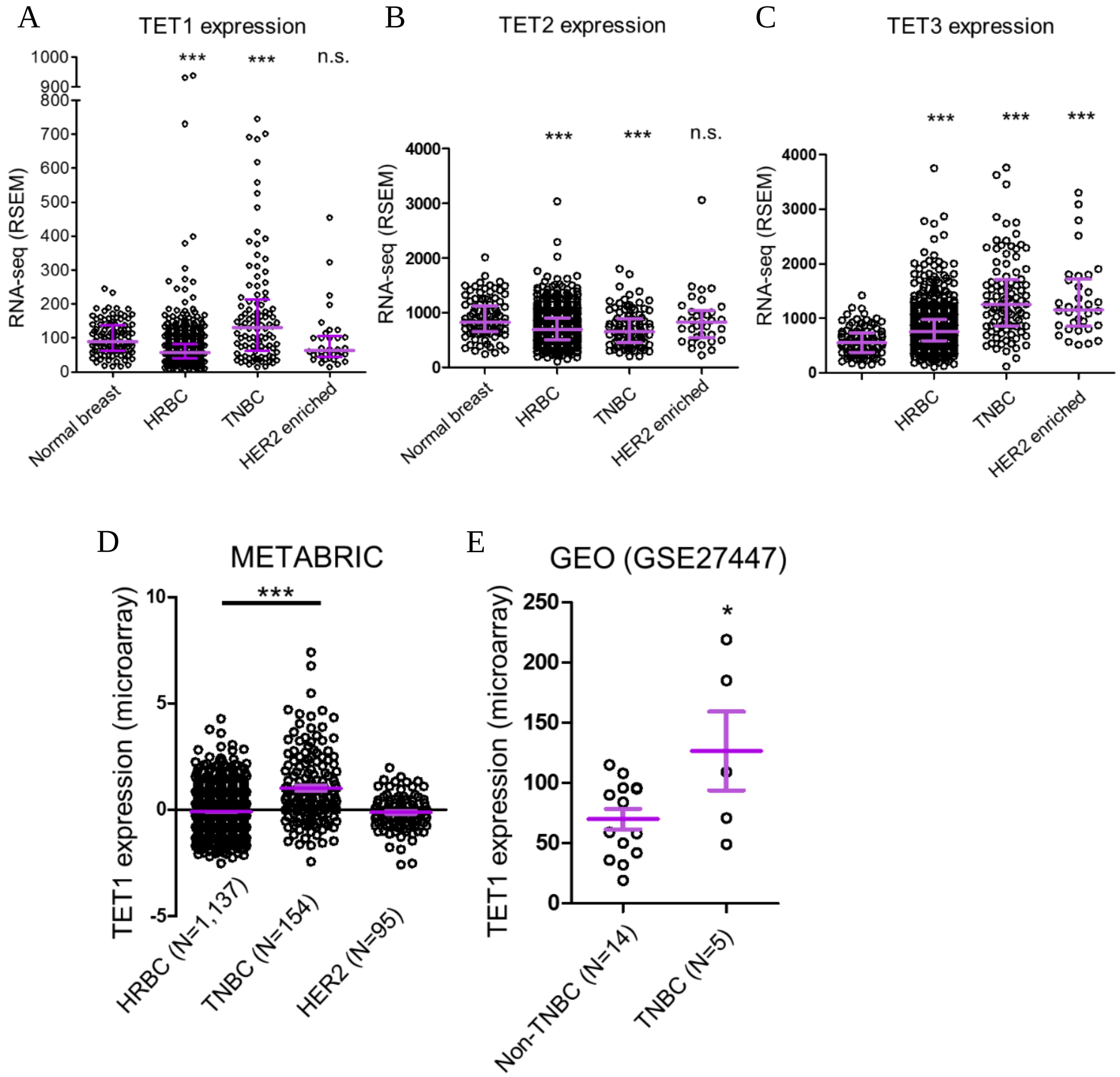


**Figure 32. TET1, TET2 and TET3 expression in breast cancer.**

(**A**) TET1 expression (RNA-seq, RSEM values) for normal breast (N=105), hormone receptor + (N=732), triple negative (N=100) and HER2 enriched (N=34). Error bars are median with interquartile range. **(B)** TET2 expression (RNA-seq, RSEM values). **(C)** TET3 expression (RNA-seq, RSEM values). **(D)** TET1 expression in METABRIC cohort (microarray) in HRBC (N=1,137), TNBC (N=154) and HER2 enriched (N=95). **(E)** TET1 expression in GEO dataset (GSE27447, mircroarray) for TNBC (N=5) and non-TNBC (N=14).

differences between TNBCs and the other breast cancer subtypes; TET1 repression could contribute to the hypermethylation characteristic of HRBCs, while TET1 overexpression could potentially account for the TNBC-specific hypomethylation. Further, TET1 immunohistochemical staining from three representative breast cancer patient samples showed TET1 protein is expressed in tumor cells (Figure 34, www.proteinatlas.org) [119].

To determine whether the TNBCs are expressing the full length TET1 (TET1$^{FL}$) or the truncated TET1 (TET1$^{ALT}$), we analyzed RNA-seq read counts for normal breast (N=107) and for TNBCs (N=91) for exon 1 (only expressed in TET1$^{FL}$, left) and TET1$^{ALT}$ exons (only expressed in TET1$^{ALT}$, right) and found both isoforms are overexpressed in TNBC (Figure 35A, Mann-Whitney test p=0.007 and p=0.02 respectively). Next, we plotted TET1$^{ALT}$ reads against TET1$^{FL}$ reads (Figure 35B). We identified four groups of TNBC patients: Group 1 overexpress only TET1$^{FL}$ (N=5), Group 2 overexpress only TET1$^{ALT}$ (N=12), Group 3 overexpress both TET1$^{FL}$ and TET1$^{ALT}$ (N=10, note one patient with very high levels of both isoforms is not depicted) and Group 4 overexpress neither isoform (N=64). Thus, TET1$^{ALT}$ is the predominant isoform overexpressed in TNBC.

### 3.6.2 TET1 associates with survival in TNBC

Because TET1 expression is variable within TNBC, we asked whether TET1 expression associates with survival. We analyzed survival data in the METABRIC cohort and found TNBC patients with high TET1 (> 1 standard deviation above the mean) had a significantly worse overall survival compared to all other TNBC patients (p=0.04) (Figure 36A). Importantly, this was not observed with TET2 or TET3 expression (p=0.38 and 0.96, respectively) nor was TET1 associated with survival in HRBCs (p=0.4), suggesting this is a TET1-TNBC specific event (Figure 36B-D).

## 3.7 TET1 expression correlates with hypomethylation in TNBC

### 3.7.1 TET enzyme expression vs DNA methylation changes

To examine if TET1 expression is associated with DNA methylation, we analyzed RNA-seq and 27K methylation array data from the TCGA. For this analysis, we chose the 27K arrays because a larger number of TNBC cases were studied on this platform. We calculated the number of sites that gain or lose methylation (change > 20% compared to normal breast) for each patient and plotted it against TET1 expression values. TNBCs (N=95) displayed a significant correlation, with high TET1 patients having the most hypomethylation (top, p=0.001) and the least hypermethylation (bottom, p=0.01) (Figure 37A). On the contrary, HRBCs (N=368) displayed no correlation between TET1 expression and DNA methylation (Figure 37B), likely due to the relatively low TET1 expression in HRBCs. Next, we determined whether TET2 or TET3 expression was correlated with DNA methylation. TET2 displayed no correlation in TNBCs or HRBCs (Figure 38A, B), while TET3 was weakly correlated in TNBC, and not correlated in HRBC (Figure 38C, D).

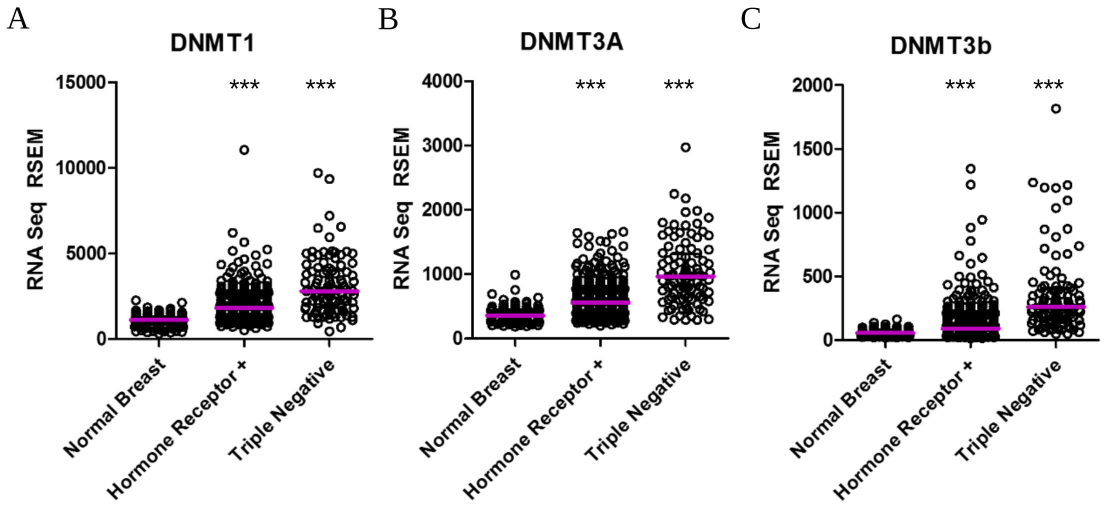


**Figure 33. DNMT expression in breast cancer.**

(**A**) DNMT1 expression (RNA-seq, RSEM values) for normal breast (N=105), hormone receptor + (N=732) and triple negative (N=100). **(B)** DNMT3A expression (RNA-seq, RSEM values). **(C)** DNMT3b expression (RNA-seq, RSEM values).

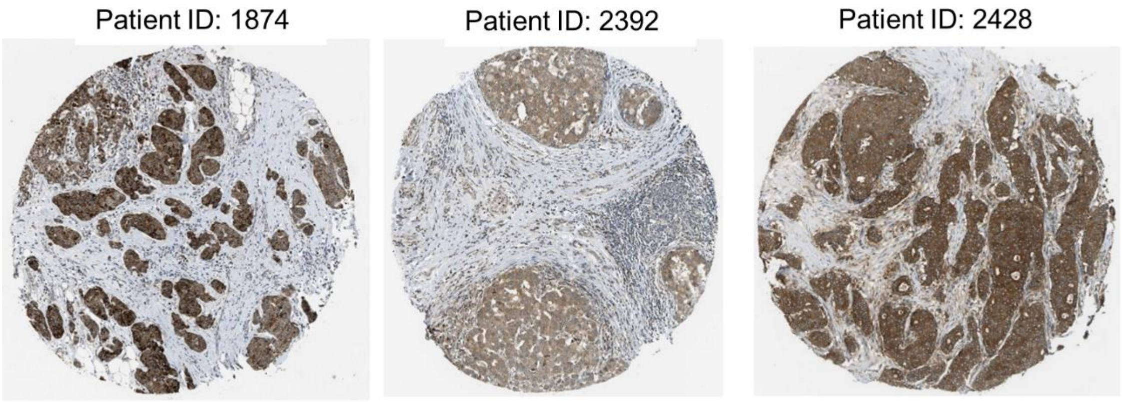


**Figure 34. Immunohistochemical staining of TET1 in breast cancer.**

TET1 immuno staining images for three representative patient samples were downloaded from www.proteinatlas.org[119].

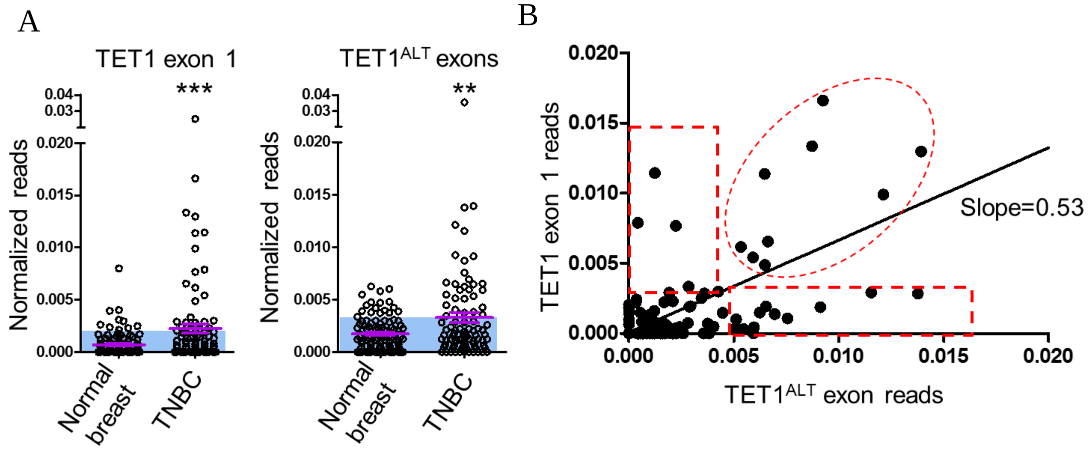


**Figure 35. Isoform specific expression for TET1 in TNBC.**

**(A)** In vivo analysis of TCGA RNA-seq reads from normal breast (N=107) and TNBC patients (N=91). Normalized exon reads for TET1$^{FL}$ exon 1 (left) and TET1$^{ALT}$ exons (right). Y axis is reads from TET1$^{ALT}$ exon 1a + exon 1b or Exon 1/exon length/chromosome 10 reads x 1e6. Mann-Whitney test, p=0.007 and p=0.02 respectively. Shaded blue boxes represent 1 standard deviation above the mean of normal. **(B)** TET1 exon 1 reads vs TET1$^{ALT}$ exon reads for TNBC patients. Expression was considered high if it was >2 standard deviations above the mean of normal for each isoform. Boxes encapsulate patients that overexpress either TET1$^{FL}$ or TET1$^{ALT}$ and the cicle denotes patients that overexpress both isoforms.

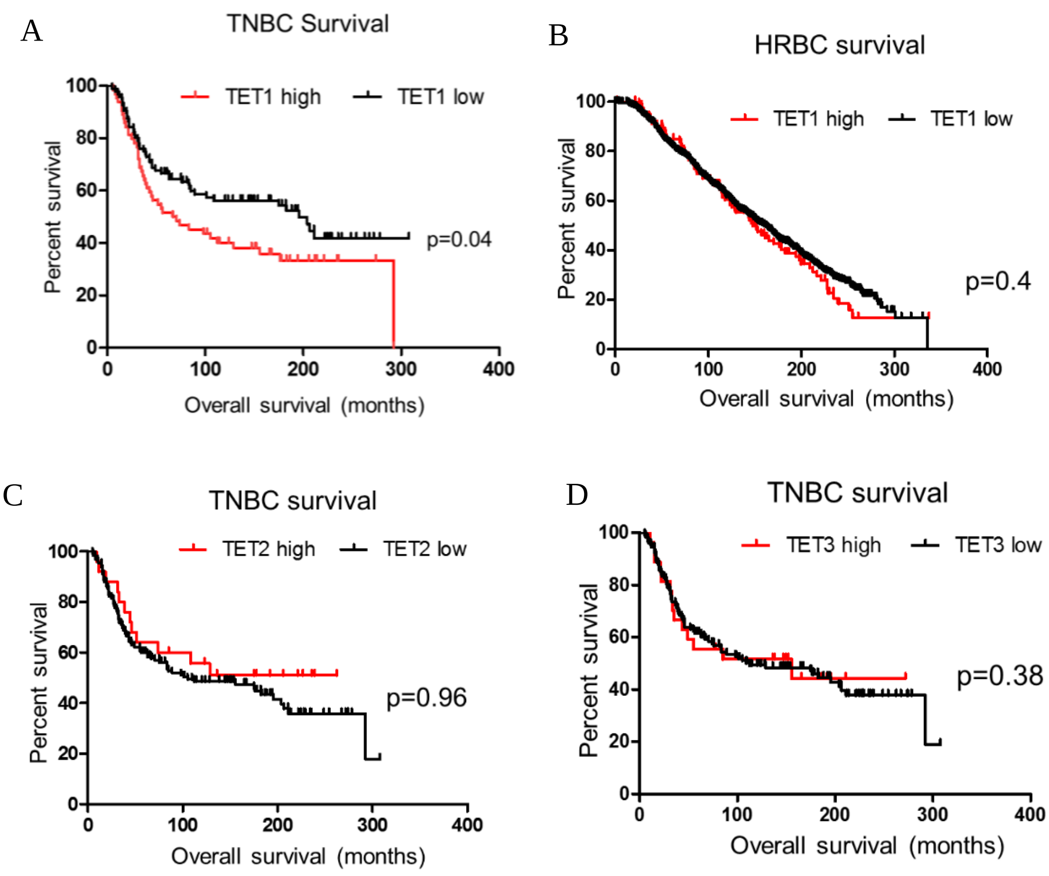


**Figure 36. TET1 associates with survival in TNBC.**

**(A)** Survival curve based on TET1 expression in triple negative breast cancer. TET1 high is classified as patients with stdev>1 above the mean, log rank test p=0.04. **(B)** Survival curve based on TET1 expression in HRBC (N=1,171). TET1 high is classified as patients with stdev>1 above the mean, log rank test p=0.4. **(C)** Survival curve based on TET2 expression in triple negative breast cancer. TET2 high is classified as patients with stdev>1 above the mean, log rank test p=0.96. **(D)** Survival curve based on TET3 expression in triple negative breast cancer. TET3 high is classified as patients with stdev>1 above the mean, log rank test p=0.96.

### 3.7.2 Identification of TET1 targets in TNBC and HRBC

To identify potential TET1 target genes in breast cancer, we used the more comprehensive 450K arrays and computed Spearman correlations between TET1 expression and methylation values for 450,000 CG sites across 469 patients, including 67 TNBC patients. In this analysis, negative r values indicate sites that hypomethylate when TET1 is high, while positive r values correspond to sites that hypermethylate. We plotted histograms of the r values and found a predominance of negative values in TNBC, consistent with the hypothesis of TET1 mediated hypomethylation (Figure 39A). To statistically analyze this observation, we performed another Spearman analysis following 1000 permutations of the data (green distribution), which is expected if no correlation exists between TET1 expression and methylation. The distribution of the actual correlations (orange) showed a marked excess of negative r values compared to the random data. We identified 42,559 probes with r<-0.3, a striking number suggesting that TET1 could potentially regulate up to 10% of the methylome. In contrast, we find a narrow range of r values in HRBC, with none exceeding the low correlation of -0.2 (Figure 39B). This lack of correlation between TET1 expression and methylation in HRBC is consistent with the uniform repression of the gene in that subtype.

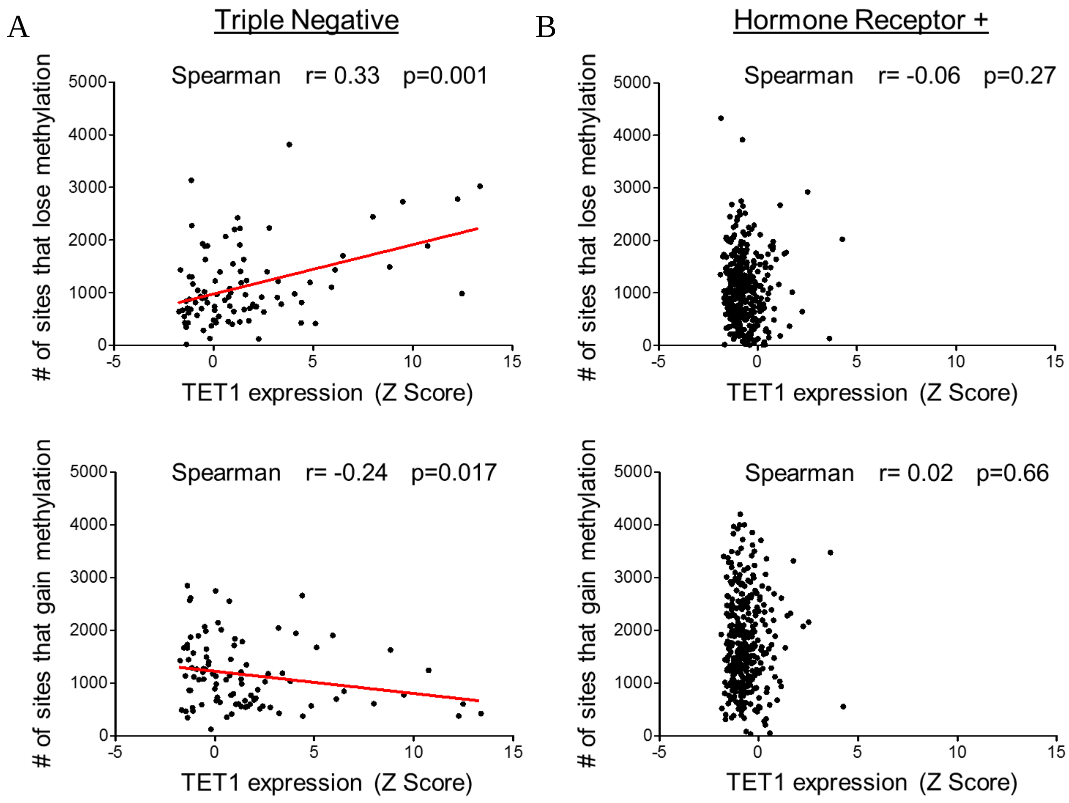


**Figure 37. TET1 correlates with genome-wide methylation changes in TNBC.**

**(A)** Triple negative breast cancer (N=95) analysis of # of sites that lose methylation (top) and # of sites that gain methylation (bottom) vs TET1 expression z score. Change in methylation >20% compared to the average of normal breast. **(B)** Hormone receptor + breast cancer (N=368) analysis of # of sites that lose methylation (top) and # of sites that gain methylation (bottom) vs TET1 expression z score. Change in methylation >20% compared to the average of normal breast.

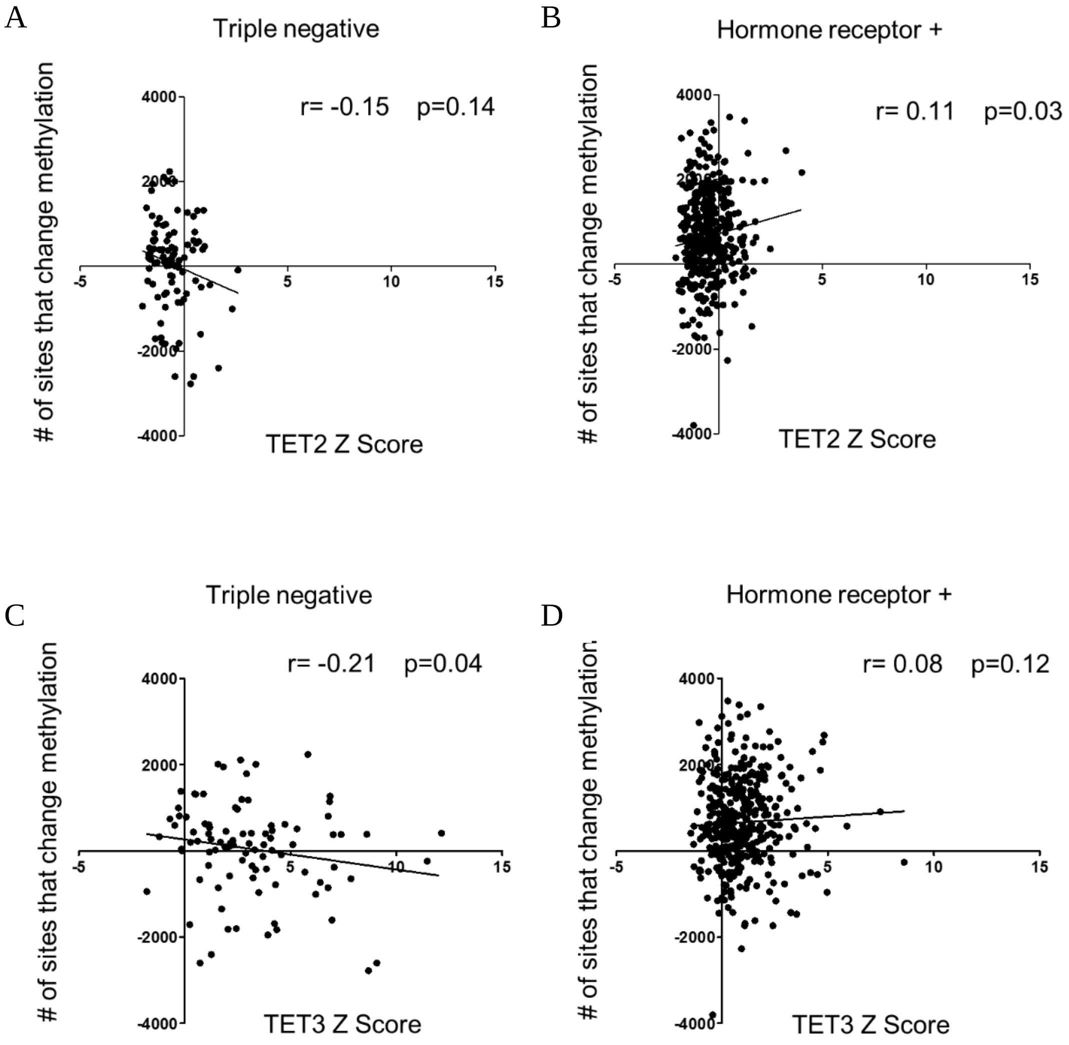


**Figure 38. TET2 and TET3 expression vs DNA methylation changes in cancer.**

**(A)** Triple negative breast cancer (N=95) analysis of # of sites that change methylation vs TET2 expression. Change in methylation >20% compared to the average of normal breast. **(B)** Hormone receptor + breast cancer (N=368) analysis of # of sites that change methylation vs TET2 expression. **(C)** Triple negative breast cancer (N=95) analysis of # of sites that change methylation vs TET3 expression. **(D)** Hormone receptor + breast cancer (N=368) analysis of # of sites that change methylation vs TET3 expression.

To identify potential breast cancer driver genes affected by TET1-mediated hypomethylation, we filtered the negatively correlated probes (r<-0.3) for sites that have an average methylation >40% in normal breast. This strategy identified 17,251 CG sites corresponding to 6,962 genes. Interestingly, compared to non-TET1 targets (r values between -0.05 to 0.05, and methylated >40% in normal breast) TET1 targets are significantly enriched for CG island shores (p<0.001) (Figure 40A), consistent with previous data linking TET1 to methylation of CG island borders [30]. Cluster analysis of the TNBC and normal breast cases using these probes (Fig. 40B) revealed two main clusters, with Cluster 1 (N=27) having the highest levels of TET1 and the most hypomethylation (blue). As expected, normal breast clustered separately from TNBCs and patients in Cluster 2 (N=38) are slightly more methylated than normal breast.

Next we asked if this pattern of methylation is specific to TNBC or if HRBC with high TET1 expression are also hypomethylated at these sites. We used TCGA methylation data for HRBC that overexpress TET1 (N=8) and 16 randomly selected cases that do not overexpress TET1 and added them to the TNBC cluster of TET1 targets in Figure 40B. As can be seen in Figure 41A, 5/8 TET1 high HRBCs clustered in the hypomethylated Cluster 1. In contrast, all 16 patients that do not overexpress TET1 were found in Cluster 2, with the exception of 2 patients that were found in the normal breast cluster. The HRBC patients found in Cluster 1 have even higher TET1 expression than the TNBCs found in Cluster 1, and in Cluster 2 the TNBCs and HRBCs have nearly identical TET1 expression levels (Figure 41B).

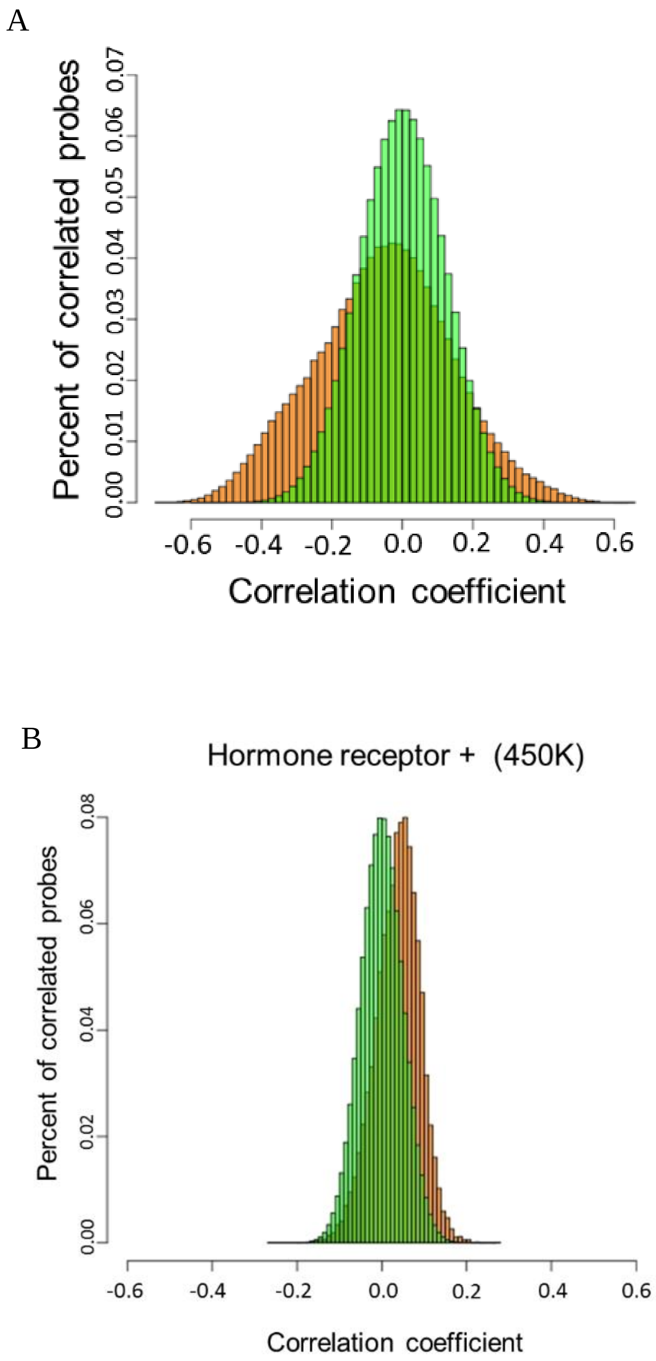


**Figure 39. Spearman analysis of TET1 expression vs DNA methylation.**

**(A)** Frequency distribution of Spearman correlation r values (TET1 expression vs 450K DNA methylation values) in TNBC patients. X axis (correlation coefficient) and Y axis (percent of correlated probes). 1000 permutations of the data (green) and real dataset (orange). **(B)** Frequency distribution of Spearman correlation r values in HRBC patients.

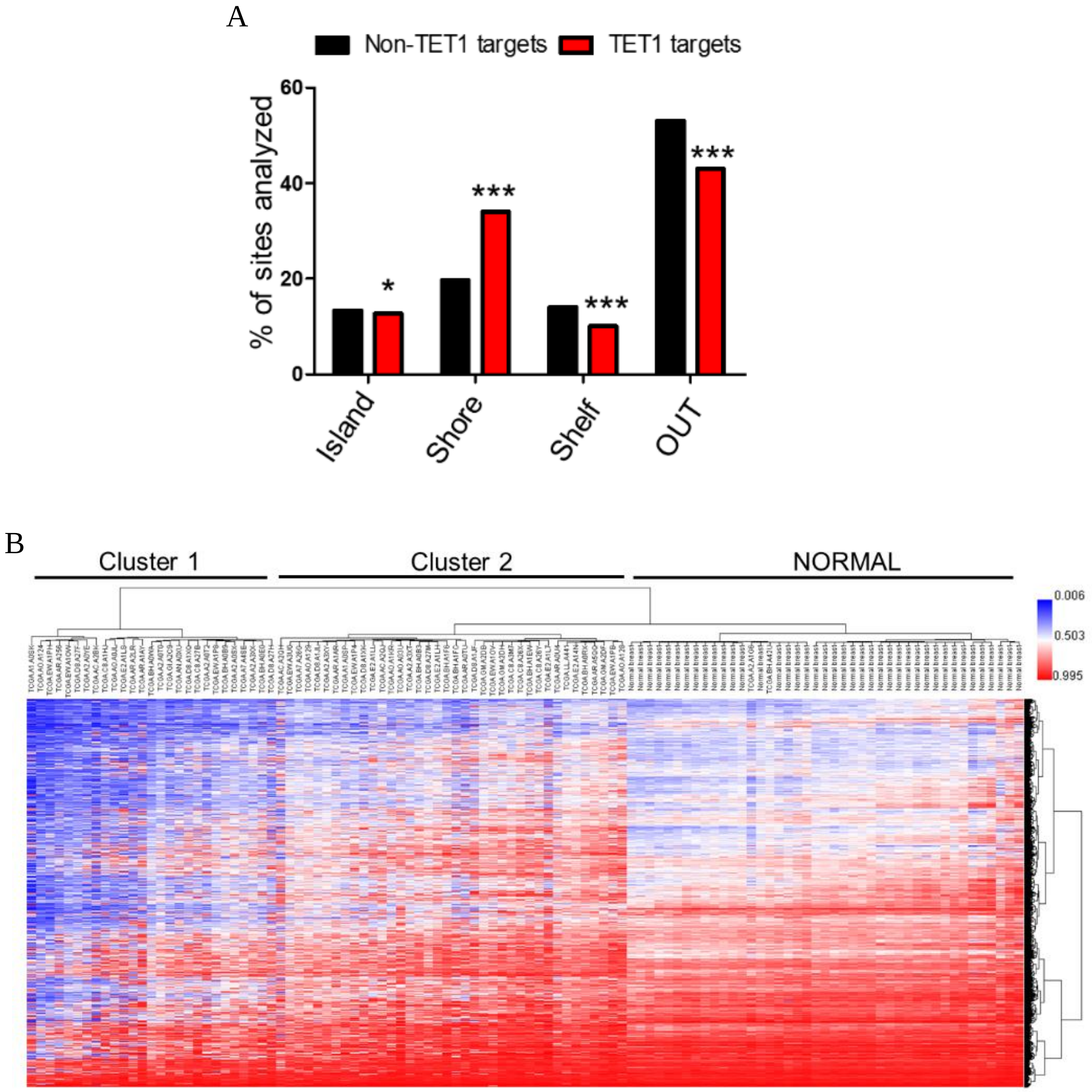


**Figure 40. TET1 hypomethylated targets.**

**(A)** Location of TET1 targets and TET1 non-targets relative to CG islands. TET1 targets identified as probes with Spearman r<-0.3 and normal methylation >40%. **(B)** Unsupervised cluster analysis of methylation in TNBC for the TET1 targets (N=17,251 sites), including 41 normal breast controls.

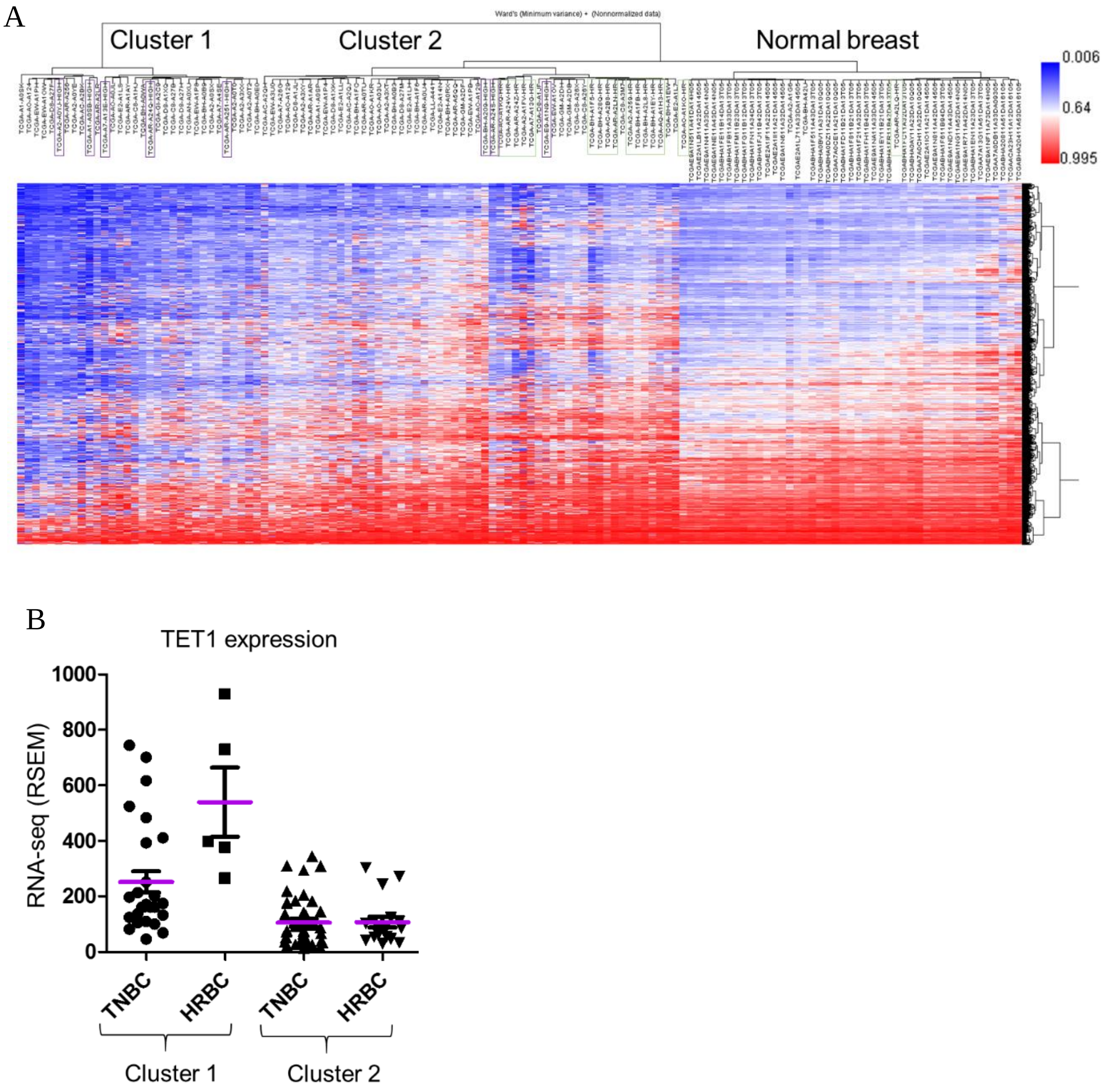


**Figure 41. TET1 hypomethylated targets with HRBCs.**

**(A)** Unsupervised cluster analysis of TET1 methylation targets in TNBC, including 41 normal breast samples, 8 HRBC patients with high TET1 (>1 stdev above normal mean, boxed in purple) and 16 HRBC patients that do not overexpress TET1 (boxed in purple). TET1 targets identified as probes with Spearman r<-0.3 (in TNBC) and normal methylation >40%. **(B)** TET1 expression (RNA-seq, RSEM) for TNBC and HRBC patients for each methylation cluster.

These data suggest that the TET1 associated demethylator phenotype is likely not breast cancer subtype specific but rather is dependent on TET1 expression levels.

### 3.7.3 TET1 targets are enriched for oncogenic signaling pathways

Pathway analysis of the putative TET1 targets from above showed enrichment for many cancer relevant pathways, including a striking number of genes in the PI3K/mTOR pathway (115 genes enriched out of 299 total genes in the pathway) as well as Hippo signaling, pathways in cancer, and extracellular matrix organization (Figure 42). We decided to focus on the PI3K pathway because PI3K is targetable by drugs [120] and many of the top candidates also feed into this pathway (indicated with *), including EGFR, PDGF, KIT, G protein coupled receptors (GPCR), etc. For validation, we analyzed a larger cohort of patients (N=95) studied on the 27K array platform (27,000 CG sites total) and obtained nearly identical results despite the lower number of sites (Figure 43).

## 3.8 Gene expression changes between Cluster 1 and Cluster 2 patients

To address if these changes in methylation are leading to changes in gene expression, we performed a differential gene expression analysis between Cluster 1 and Cluster 2. We identified 240 genes that are upregulated and 680 genes downregulated in Cluster 1 (Figure 44A). Pathway analysis of the upregulated genes revealed enrichment for the PI3K pathway, and the downregulated genes were enriched for immune system pathways (Figure 44B, C). Overlap of the genes upregulated in Cluster 1 with the genes identified as TET1 methylation targets above, revealed 119/240 (50%) of the genes upregulated in Cluster 1 are also hypomethylated in Cluster 1, (Chi-square p<0.0001, compared to overlap with genes downregulated in Cluster 1). Interestingly, the overlapping genes are enriched for the PI3K-Akt-mTOR pathway (Figure 45A). Upon further analysis of these genes, we identified that 8/9 are upstream of the PI3K pathway, including tyrosine kinase receptors, G protein coupled receptors and integrin receptors which all activate PI3K. Gene expression and methylation levels for some of these target genes (e.g. KIT, ITGA10) can be found in Figure 45B, C.

## 3.9 Loss of TET1 leads to loss of PI3K/mTOR signaling

### 3.9.1 Clinical features of patients in Cluster 1 and Cluster 2

Phosphoinositide 3-kinase (PI3K) regulates cell proliferation, survival and migration [120]. Activating PI3K mutations are found in 41% of HRBCs, but only 7% of TNBCs [42]. However, gene expression and proteomic data has revealed the PI3K pathway is more active in TN tumors compared to non-TN tumors, but it is unknown why [42]. We were interested in the fact that TNBCs can be divided into TET1 high and low based on their methylation levels (Clusters 1 and 2 in Figure 40B). When we compared these two groups, they had similar clinical characteristics but differed significantly in the rate of mutations affecting PI3K-AKT signaling (Table 7). 0% of Cluster 1 patients have

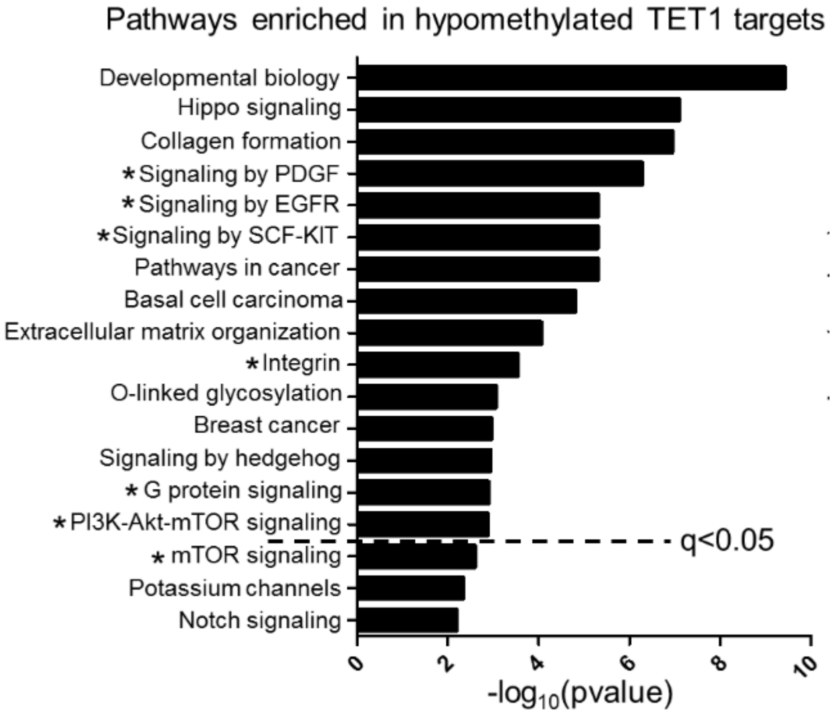


**Figure 42. Pathways enriched in hypomethylated TET1 targets.**

X axis is significance. * denotes pathways that are directly related to the PI3K pathway.

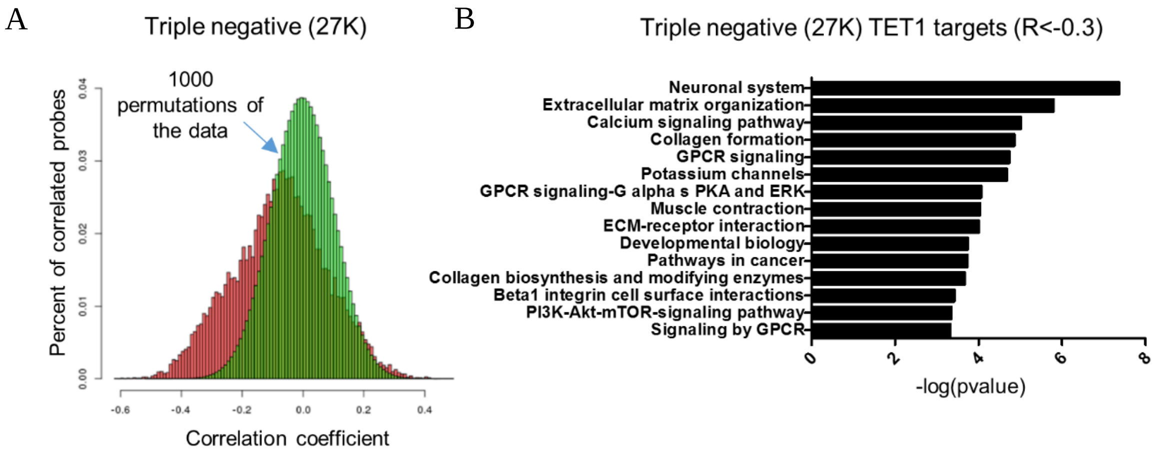


**Figure 43. Validation with 27K array.**

**(A)** Frequency distribution of Spearman correlation r values (TET1 expression vs DNA methylation 27K array values) in TNBC patients (N=95). X axis (correlation coefficient) and Y axis (percent of correlated probes). **(B)** Pathway analysis (Consensus Path DB) of TET1 targets in 27K array. X axis is –$\log_{10}$(pvalue).

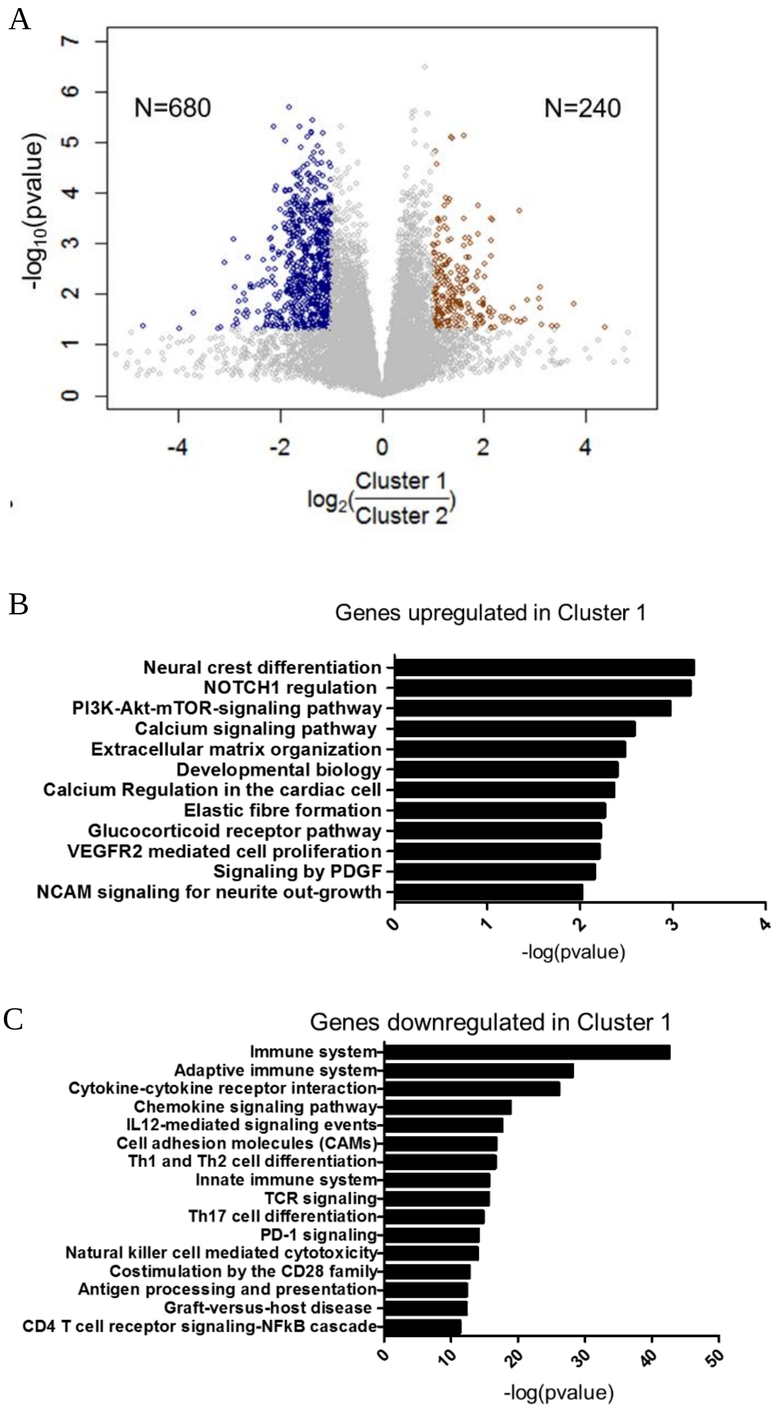


**Figure 44. Genes differentially expressed between Cluster 1 and Cluster 2.**

**(A)** Differential gene expression analysis between Cluster 1 and Cluster 2. Y axis is –$\log_{10}$(pvalue) and x axis is log2(Cluster 1/Cluster 2). Genes were considered upregulated if FC>2 or FC<0.5 and $p<0.05$. **(B)** Pathway analysis of genes upregulated in Cluster 1 (compared to Cluster 2). **(C)** Pathway analysis of genes downregulated in Cluster 1.

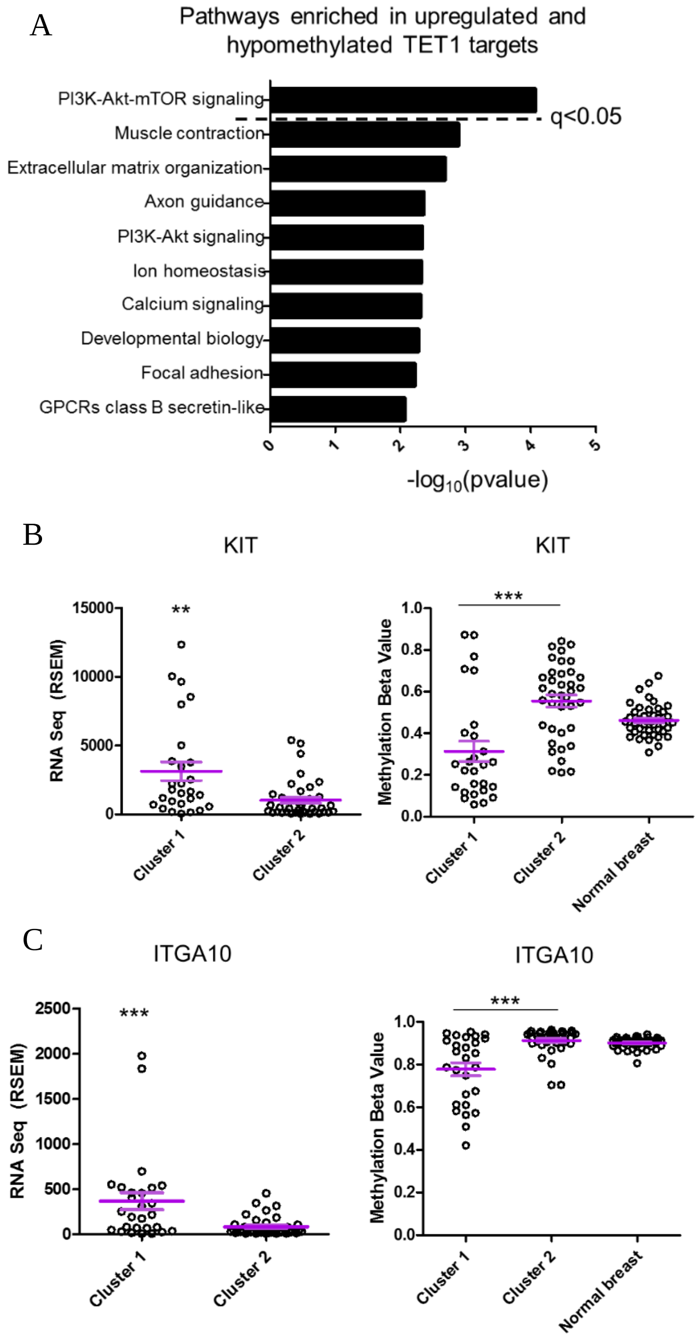


**Figure 45. Genes upregulated and hypomethylated.**

**(A)** Pathway analysis of genes that are hypomethylated and upregulated in Cluster 1. **(B)** RNA-seq (RSEM) for KIT expression in Cluster 1 and Cluster 2 (left) and DNA methylation beta-value for CG site in KIT in Cluster 1, Cluster 2 and normal breast (right). **(C)** RNA-seq (RSEM) for ITGA10 expression in Cluster 1 and Cluster 2 (left) and DNA methylation beta-value for CG site in IGTA10 in Cluster 1, Cluster 2 and normal breast (right).

|  | Cluster 1 | Cluster 2 | p value |
|---|---|---|---|
| Number of TNBC Patients | N=27 | N=38 |  |
| Age at Diagnosis | 55 | 55 | n.s. |
| Stage I (%) | 22 | 14 | n.s. |
| Stage II (%) | 59 | 62 | n.s. |
| Stage III (%) | 15 | 24 | n.s. |
| Stage IV (%) | 4 | 0 | n.s. |
| Asian (%) | 11 | 9 | n.s. |
| White (%) | 70 | 66 | n.s. |
| Black (%) | 19 | 26 | n.s. |
| TET1 Z score expression (AVG) | 3.2 | 0.2 | 0.0002 |
| Mutation in PIK3CA or PTEN (%) | 0 | 21 | 0.01 |
| BRCA1 mutation (%) | 7.4 | 7.4 | n.s. |
| BRCA2 mutation (%) | 7.4 | 0 | n.s. |
| TP53 mutation (%) | 77.8 | 65.8 | n.s. |

**Table 7. Clinical characteristics of patients in Cluster 1 and Cluster 2.**

PIK3CA or PTEN mutations, whereas 21% of Cluster 2 patients have mutations in the pathway (Fisher's exact test, p=0.01). Thus, in TNBC, there is an inverse correlation between TET1 levels and PI3K-AKT pathway mutations, raising the possibility that the two molecular events are not co-selected because they are equivalent ways of activating the same oncogenic pathway. This is consistent with DNA methylation and gene expression studies in Cluster 1 (see above).

### 3.9.2 Phenotype of TET1 KO

To test whether TET1 directly regulates the PI3K pathway, we generated CRISPR Cas9 TET1 KO cells in the TNBC cell line MDA-MB-231. Single cell clones were generated from pooled cells with three different sgRNAs targeting different exons: exon 6 (KO-1), exon 3 (KO-2) and exon 11 (KO-3). KO was confirmed by western blot analysis using a TET1 antibody (Figure 46A). Of note, the largest effect of the TET1 KO was observed at 162 kDa, the size of TET1$^{ALT}$. Upon overexposure, TET1$^{FL}$ is also reduced in the TET1 KOs (data not shown). Loss of TET1 resulted in a loss of phospho-4EBP1 (Thr37/46) (Figure 46B) and significantly reduced cellular migration (Figure 46C, D) and proliferation (Figure 46E).

### 3.9.3 Vulnerabilities imparted by TET1 expression in breast cancer

If TET1 is important for maintaining PI3K/mTOR activation, it is expected that cell lines with high TET1 will be more sensitive to inhibitors targeting this pathway. To determine potential corresponding vulnerabilities in a panel of 50 breast cancer cell lines, we computed Spearman correlations between TET1 expression and drug sensitivity (IC50) for 265 drugs in a public database (Genomics of Drug Sensitivity) [121]. Multiple drugs showed high negative correlations with TET1 levels, including PI3K pathway targeting drugs (ERK, MAPK etc.), consistent with our hypothesis (Table 8). Of note, several high-correlation drugs target mitosis, and some of the genes affected are themselves regulated by the PI3K pathway (e.g. PLK1, regulated by PIP3/PDK1 interactions) [122]. These drugs constitute a potential entry towards targeted therapy for TNBC. Three representative drugs are plotted in Figure 47. XMD8-85 (targets ERK, r= -0.85) and VX-680 (targets AURK, r= -0.86) show breast cancer cell lines with high TET1 having the lowest IC50. Further, TGX-221 (targets PI3Kβ) also shows a strong negative correlation (r= -0.51) although not significant.

### 3.10 Gene expression targets of TET1 in vitro

Next we performed RNA-seq in the TET1 KO cell lines generated above, to investigate whether TET1 expression is important for maintaining oncogenic signaling pathways (including PI3K) in TNBC. Compared to empty vector control, TET1 KO cells showed dramatic changes in expression as indicated by principal component analysis (Figure 48A). We identified 566 genes that are downregulated (FC<0.05, p<0.001) in at least 2 of the 3 KO clones (Figure 48B). Interestingly, we performed a pathway analysis and found several overlapping pathways (indicated by red circle) with the pathways identified as TET1 hypomethylated targets in Figure 42, including oncogenic pathways that feed into PI3K (EGFR, VEGF, Integrins, etc.) (Figure 48C). Next, we analyzed genes that are upregulated upon loss of TET1 (FC>2, p<0.001) (Figure 48D). We identified 343 genes upregulated in at least 2 of the 3 KO clones, and found they are enriched for pathways involved with immune system function (Figure 48E). This agrees with the differential gene expression analysis in the TCGA datasets (immune system genes downregulated in Cluster 1). Because high TET1 expression in TNBC is associated with low expression of immune genes and KO of TET1 resulted in the upregulation of immune genes, TET1 may be involved in suppressing an immune response in TNBCs.

Next, we overlapped the differentially expressed genes identified in the TCGA datasets in Figure 44A with the differentially expressed genes in the TET1 KO cells, and found 50 genes that are differentially expressed in both datasets (Chi-squared p<0.0001, when compared to genes not differentially expressed in TET1 KO). This list includes several genes in the PI3K-Akt pathway including THBS3, ANGPT1, PIK3CG, MAP2, and IL7R. ANGPT1 (a glycoprotein that activates the TEK receptor tyrosine kinase) and THBS3 (a glycoprotein that activates ITGA/ITGB integrin receptors) are both upregulated in Cluster 1 TCGA patients, and are downregulated in the TET1 KO cells (Figure 49A, B). Interestingly, both proteins have been implicated in facilitating tumor growth [123] [124].

### 3.11 Methylation targets of TET1 in vitro

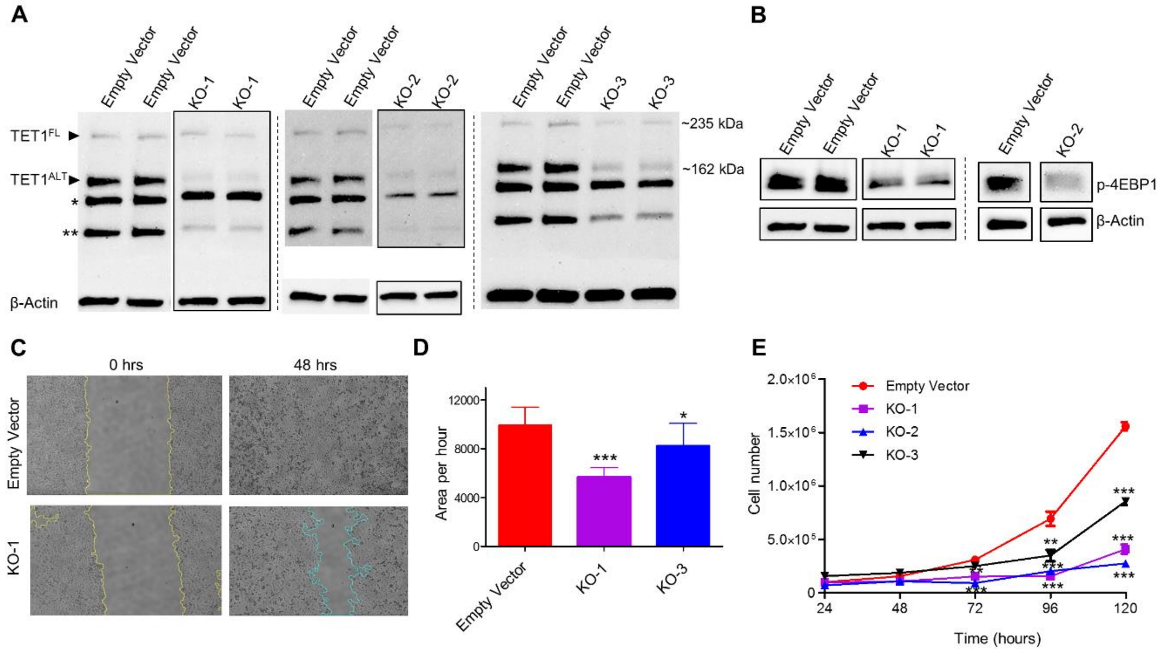


**Figure 46. Loss of TET1 leads to loss of PI3K/mTOR signaling and reduced cellular migration and proliferation**

**(A)** Western blot (in duplicate) of empty vector and TET1 knockout single clones. β-actin used as a loading control. (* denotes putative non-specific band, ** denotes another potential truncated isoform of TET1) **(B)** Western blot of phospho-4EBP1 (Thr37/46) in TET1 KO cells. KO-1 is in duplicate, KO-2 is single experiment. β-actin used as a loading control. **(C)** Representative images of wound-healing assays in empty vector and TET1 KO-1 cells at 0 and 48 hours. **(D)** Quantification of migration for empty vector, KO-1 and KO-3, experiment performed in triplicate. **(E)** Cell proliferation assay for empty vector, KO-1, KO-2 and KO-3. Cells were counted (in triplicate, each sample counted twice) at 24, 48, 72, 96 and 120 hours. X axis is time, Y axis is total cell number.

| Drug name | Targets | Pathway | Spearman | P-value |
|---|---|---|---|---|
| VX-680 | Multi-kinase (AURK*) | mitosis | -0.86 | 0.007 |
| XMD8-85 | ERK5 (MK07) | other | -0.85 | 0.004 |
| BI-2536 | PLK1, PLK2, PLK3 | mitosis | -0.81 | 0.015 |
| AZ628 | BRAF | ERK, MAPK | -0.78 | 0.008 |
| Pyrimethamine | DHFR | DNA replication | -0.76 | 0.011 |
| GW843682X | PLK1 | mitosis | -0.71 | 0.047 |
| S-Trityl-L-cysteine | KIF11 | mitosis | -0.71 | 0.047 |

**Table 8. Top correlated drugs in breast cancer.**

Top correlated drugs from Spearman analysis of TET1 expression vs IC50 drug sensitivity in breast cancer cell lines.

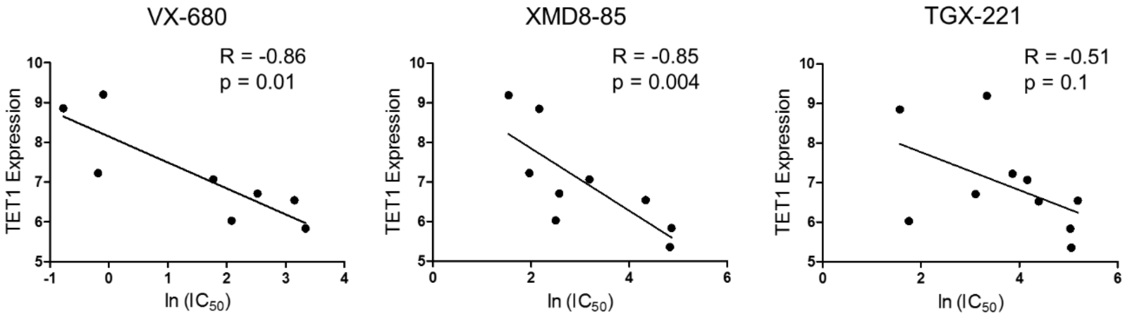


**Figure 47. Drug sensitivity vs TET1 expression in breast cancer cell lines.**

VX-680, XMD8-85, and TGX221 correlations in breast cancer cell lines. Y axis is TET1 expression (microarray) and X axis is ln(IC50) for each drug.

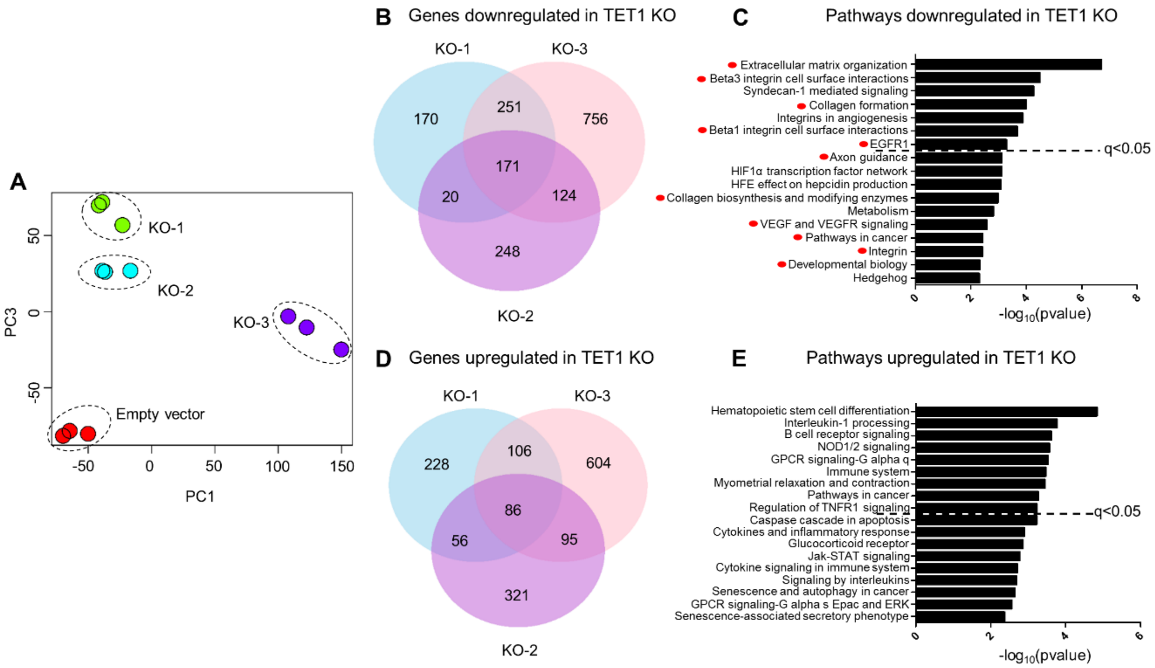


**Figure 48. Gene expression targets of TET1 in vitro.**

**(A)** Principal component analysis of RNA-seq data for empty vector, TET1 KO-1, KO-2 and KO-3 (in triplicate) in MDA-MB-231 cells. **(B)** Venn diagram overlap of genes downregulated in the TET1 KO cells (FC<0.5, p<0.001). **(C)** Pathway analysis of genes downregulated in at least 2 of 3 TET1 KO clones. X axis –$\log_{10}$(pvalue). **(D)** Venn diagram overlap of genes upregulated in the TET1 KO cells (FC>2, p<0.001). **(E)** Pathway analysis of genes upregulated in at least 2 of 3 TET1 KO clones. X axis –$\log_{10}$(pvalue).

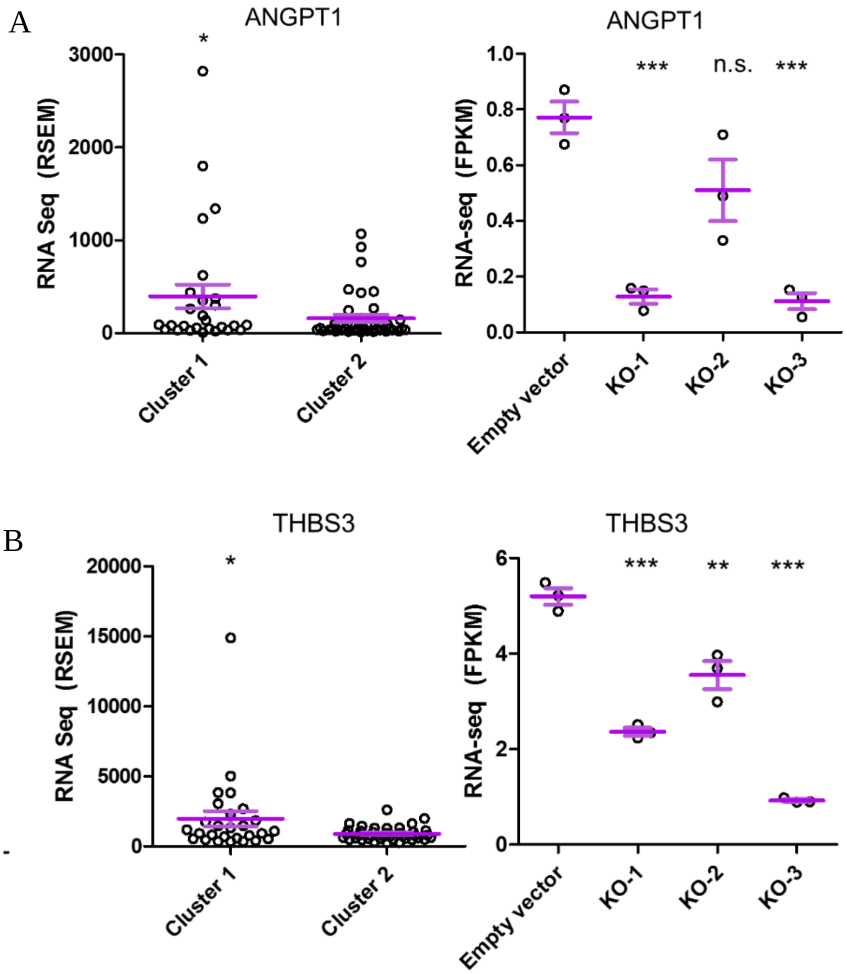


**Figure 49. ANGPT1 and THBS3 expression in TCGA clusters and TET1 KO cells.**

**(A)** RNA-seq (RSEM) for ANGPT1 expression in Cluster 1 and Cluster 2 (left) and RNA-seq (FPKM) in empty vector, TET1 KO-1, KO-2 and KO-3 (right). **(B)** RNA-seq (RSEM) for THBS3 expression in Cluster 1 and Cluster 2 (left) and RNA-seq (FPKM) in empty vector, TET1 KO-1, KO-2 and KO-3 (right).

Next, we sought to investigate if the changes in expression are dependent or independent of changes in methylation. We performed DREAM, a genome-wide DNA methylation assay, in the parental cell line MDA-MB-231, empty vector control pooled cells, empty vector single clones and the TET1 KO cells (all in triplicate except the MDA-MB-231 parental line). A principal component analysis of the data revealed that the parent line and pooled cells are very similar, however, there is a minor drift in methylation in the empty vector single clone cells, indicating minor cell to cell heterogeneity in methylation in the parental cells (Figure 50A). By contrast, the methylome of the TET1 KO cells was dramatically different from all control cells. To analyze these changes in more detail, we first controlled for the methylation drift in the single clones by excluding sites that have a standard deviation of greater than 3% between all control cells (parental, pooled and single clone empty vector). This method identified the CG sites that are the least variable in the controls (N=31,070 CG sites).

Because we were interested in methylation changes that correlated with changes in gene expression, we next filtered for CG sites within promoter regions (+/- 2000 bp from the TSS), which left 15,221 CG sites corresponding to 6,715 genes. We identified gene promoters that gain methylation >10% compared to empty vector single clones in each TET1 KO cell line (Figure 50B). In this genomic compartment, the predominant change was a gain in methylation in TET1 KO cells (with two fold more sites gaining methylation rather than losing), as expected. Overall, 212 promoters gained methylation in at least 2 of the 3 TET1 KO clones. Pathway analysis of these hypermethylated genes revealed enrichment for several pathways that feed into the PI3K/mTOR pathway, including GPCR and insulin signaling (Figure 50C). Next, we overlapped the genes that gain methylation upon loss of TET1 with the TET1 methylation targets identified in the TCGA datasets, and found that 47% of the genes that gain methylation in TET1 KO are also TET1 methylation targets identified in TNBC patient samples. Several of the overlapping genes are upstream regulators of PI3K, including FGF19, INSR, GHR and THBS4, of which methylation values from both datasets can be found in Figure 51A, B. A few of the methylation targets identified also lose expression in the TET1 KO cells and are overexpressed in Cluster 1 patients, including INSR (Figure 52A-D).

## 3.12 TET1 mediated hypomethylation in serous ovarian cancer

### 3.12.1 Identification of TET1 targets in ovarian cancer

To determine if TET1 mediated hypomethylation is found in other cancer types, we downloaded methylation and gene expression data for all serous ovarian cystadenocarcinoma cases (N=304) from the TCGA. We examined serous ovarian cancers because this type has similar TET1 expression levels as TNBC, unlike other cancers such as pancreatic cancer where TET1 expression is very low (Figure 53A). In addition, promoter hypomethylation has been previously reported to activate oncogenes and associate with reduced overall survival in epithelial ovarian cancer [125]. Further, serous ovarian cancer has been reported to have very similar molecular features to TNBC [42]. First, we identified TET1 targets by computing Spearman correlation coefficients between TET1 expression and CG methylation for 27,000 CG sites. The distribution of the r values showed a skewing to the left indicating negative r values as expected (Figure

53B). Cluster analysis of the negatively correlated CG sites (r<-0.3) revealed three methylation clusters, with Cluster 3 being the most hypermethylated and Cluster 1 being the most hypomethylated (data not shown). To identify potential ovarian cancer drivers affected by TET1-mediated hypomethylation, we filtered for sites that had average methylation >40% in Cluster 3 (because we did not have normal ovary tissue) and clustered the methylation values for all patients (Figure 53C). Cluster 1 was the most hypomethylated and these patients significantly overexpressed TET1 (p<0.001, Figure 53D). Cluster 3 was the most methylated and had the lowest amount of TET1 expression; Cluster 2 was in between. The TET1 targets that are hypomethylated are enriched for genes in the GPCR pathway, a prominent activator of PI3K/Akt (Figure 53E). These data suggest TET1 mediated hypomethylation may be a more wide-spread mechanism utilized by cancer cells. Importantly, we showed high expression of TET1 in serous ovarian cancer is associated with a worse overall survival (see above).

### 3.12.2 Vulnerabilities imparted by TET1 expression in ovarian cancer

To identify if ovarian cancer cell lines with high TET1 expression are sensitive to PI3K/Akt/mTOR inhibitors, we performed a Spearman correlation analysis between drug sensitivity in ovarian cancer cell lines (IC50) vs TET1 expression, similar to the analysis performed previously in breast cancer. Consistent with our hypothesis, of the top eight negatively correlated drugs (r<-0.3), four inhibitors directly affect the PI3K or mTOR pathway, including two IGF1R inhibitors (BMS-754807, Lisitinib), a receptor tyrosine kinase inhibitor (Axitinib) and an ERK/MAPK inhibitor (HG-6-64-1) (Table 9).

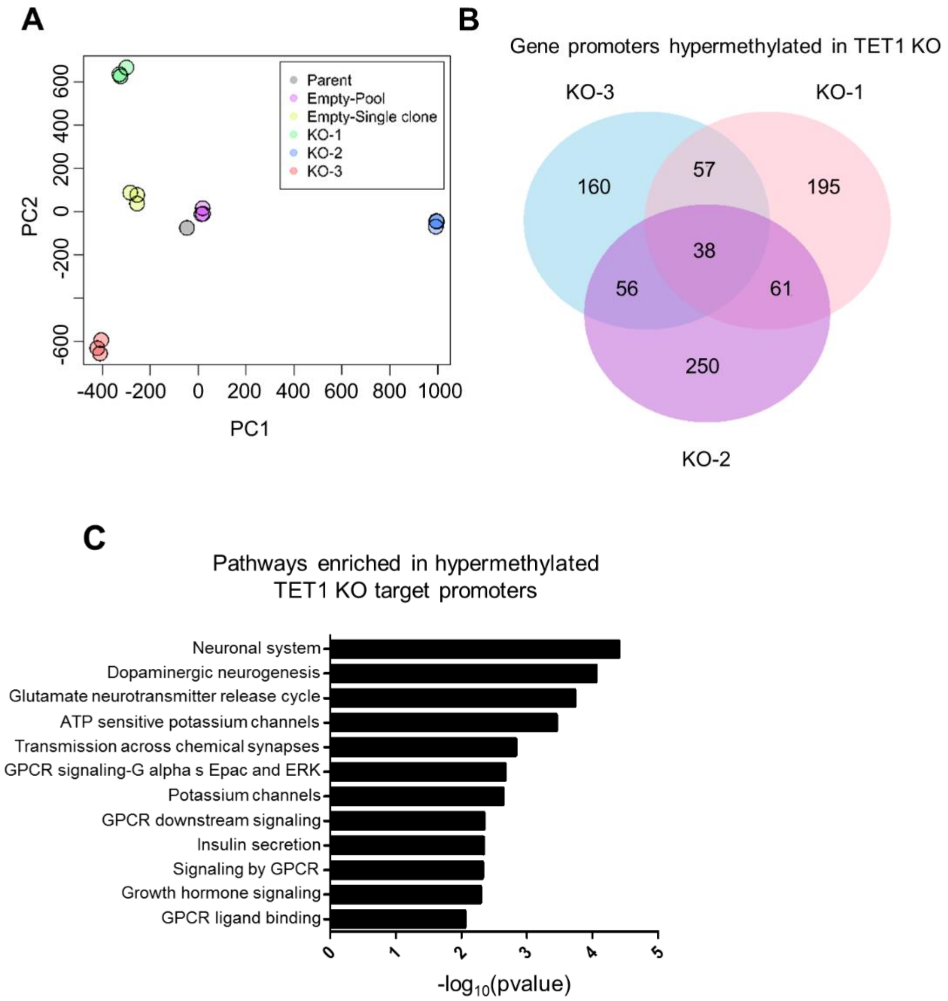


**Figure 50. DNA methylation targets of TET1 in vitro.**

**(A)** Principal component analysis of DREAM methylation data for MDA-MB-231 parental cell line, empty vector pooled cells, empty vector single clone, TET1 KO-1, KO-2 and KO-3 single clones (in triplicate). **(B)** Venn diagram overlap of genes with promoters than gain methylation (>10%, p<0.05) in the TET1 KO clones, compared to empty vector single clones. Sites were filtered for stdev< 3 (MDA-MB-231, empty vector pooled and empty vector single clone) and for CG sites +/- 2000bp from the TSS. **(C)** Pathway analysis of genes with hypermethylated promoters in at least 2 of 3 TET1 KO clones. X axis –$\log_{10}$(pvalue).

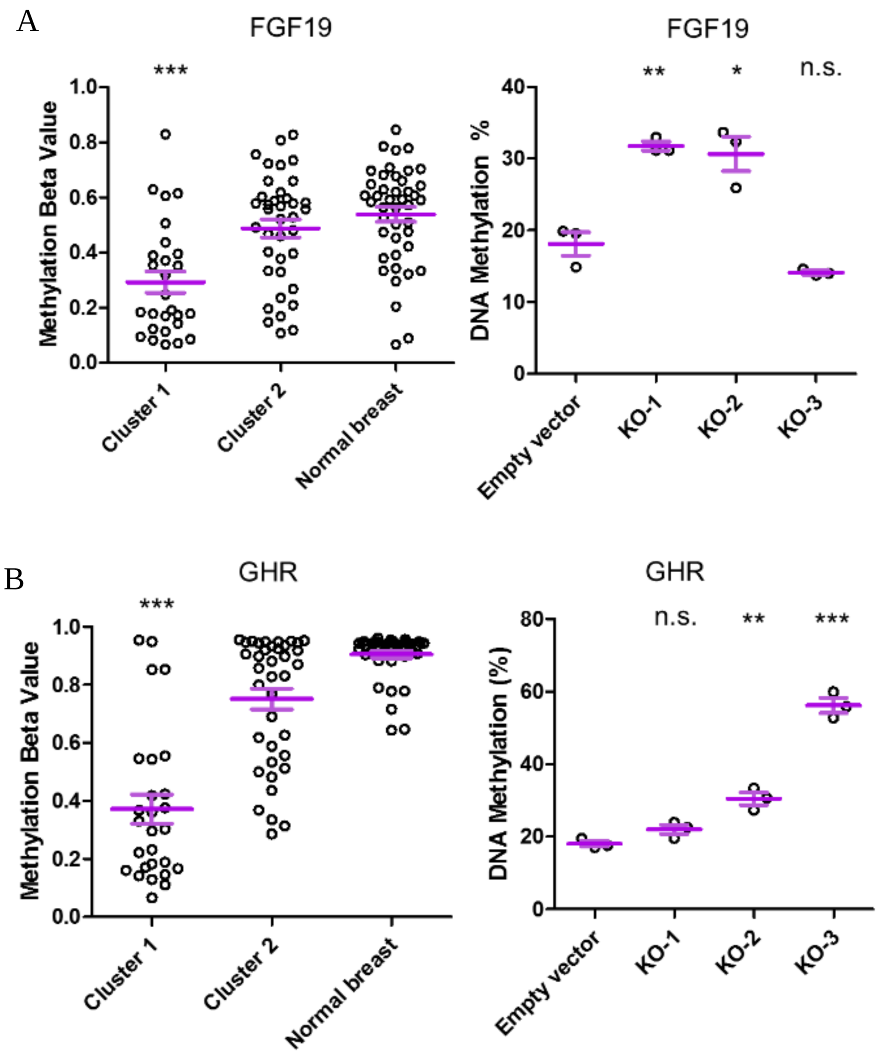


**Figure 51. Methylation and gene expression targets in TCGA and TET1 KO cells.**

**(A)** DNA methylation beta-value (450K, TCGA) for FGF19 in Cluster 1, Cluster 2 and normal breast (left) and DNA methylation % (DREAM) at the FGF19 promoter in empty vector, TET1 KO-1, KO-2 and KO-3 (right). **(B)** DNA methylation beta-value (450K, TCGA) for GHR in Cluster 1, Cluster 2 and normal breast (left) and DNA methylation % (DREAM) at the FGF19 promoter in empty vector, TET1 KO-1, KO-2 and KO-3 (right).

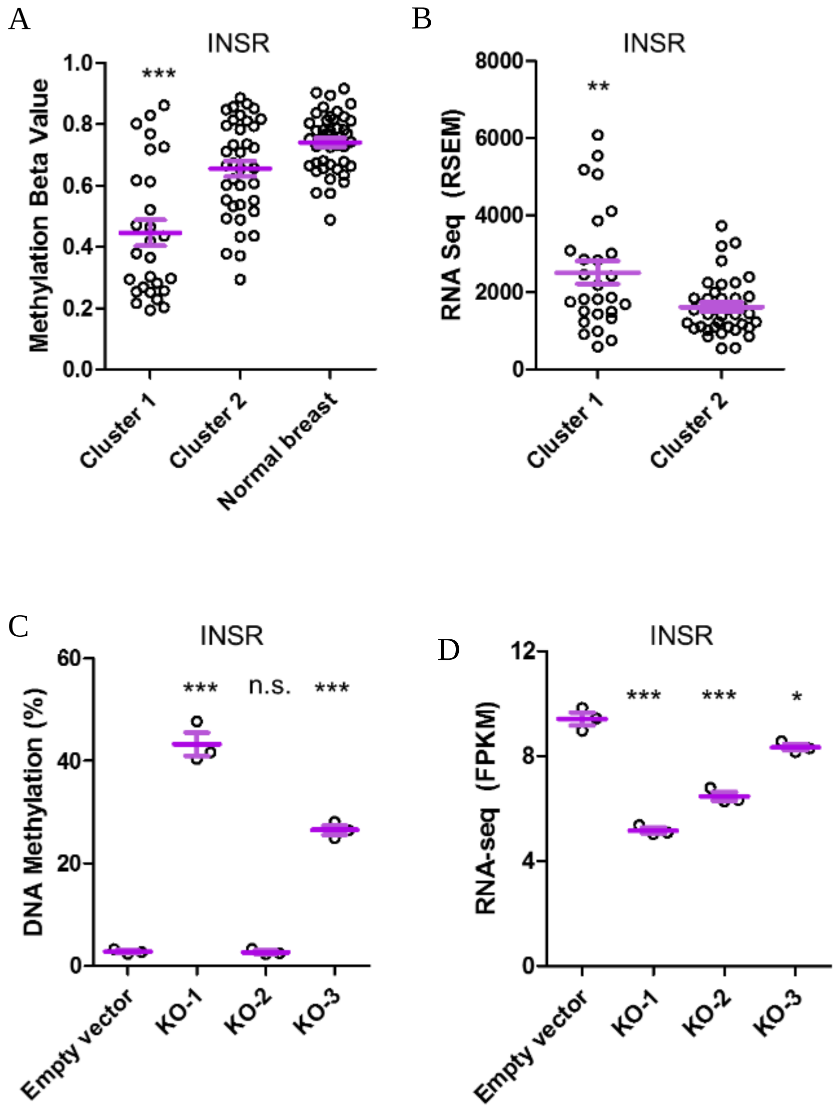


**Figure 52. INSR gene expression and methylation in TCGA and TET1 KO cells.**

**(A)** DNA methylation beta-value (450K, TCGA) for INSR in Cluster 1, Cluster 2 and normal breast. **(B)** RNA-seq (RSEM, TCGA) for INSR in Cluster 1 and Cluster 2. **(C)** DNA methylation % (DREAM) at the INSR promoter in empty vector, TET1 KO-1, KO-2 and KO-3. **(D)** RNA-seq (FPKM) expression of INSR in empty vector, TET1 KO-1, KO-2 and KO-3.

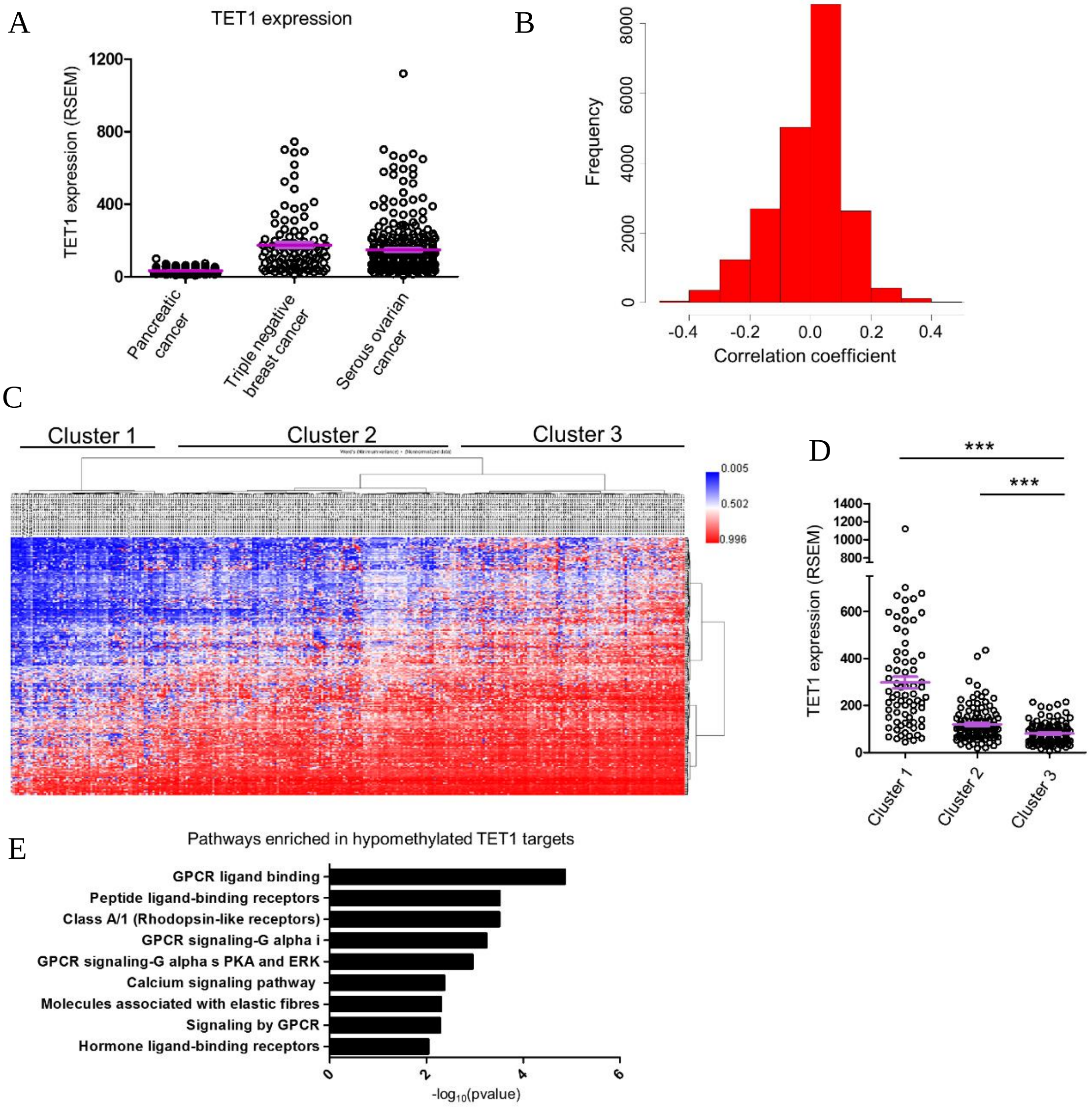


**Figure 53. TET1 mediated hypomethylation in ovarian cancer.**

**(A)** TET1 expression (RNA-seq, RSEM values) for pancreatic cancer (N=145), TNBC (N=100), and ovarian cancer (N=307). **(B)** Frequency distribution of Spearman correlation r values (TET1 expression vs DNA methylation 27K array values) in ovarian cancer patients. X axis (correlation coefficient) and Y axis (frequency). **(C)** Unsupervised cluster analysis of TET1 methylation targets in ovarian cancer for sites that have Spearman r<-0.3 and filtered for methylation >40% in Cluster 3 patients. **(D)** TET1 expression (RSEM) in ovarian cancer for each methylation cluster. **(E)** Pathway analysis of TET1 targets in ovarian cancer. X axis –$\log_{10}$(pvalue).

| Drug name | Targets | Pathway | Spearman | P-value |
|---|---|---|---|---|
| BMS-754807 | IGF1R | IGFR signaling | -0.434 | 0.049 |
| Elesclomol | HSP70 | other | -0.404 | 0.033 |
| Lisitinib | IGF1R | IGFR signaling | -0.394 | 0.070 |
| Axitinib | PDGFR, KIT, VEGFR | RTK signaling | -0.344 | 0.073 |
| Vorinostat | HDAC inhibitor Class I, IIa, IIb, IV | chromain histone acetylation | -0.319 | 0.099 |
| HG-6-64-1 | BRAFV600E, TAK, MAP4K5 | ERK MAPK signaling | -0.311 | 0.148 |
| Gemcitabine | DNA replication | DNA replication | -0.306 | 0.155 |
| EHT 1864 | Rac GTPases | cytoskeleton | -0.305 | 0.114 |

**Table 9. Top correlated drugs in ovarian cancer.**

Top correlated drugs from Spearman analysis of TET1 expression vs IC50 drug sensitivity in ovarian cancer cell lines.

# CHAPTER 4

# DISCUSSION

In this work, we sought to identify proteins that regulate the balance between hyper and hypomethylation in cancer and to understand why methylation is prognostic in breast cancer. Through our analyses, we found that TET1, and a truncated isoform of TET1 we discovered (TET1$^{ALT}$) are dysregulated in cancer, leading to aberrant methylation and activation of oncogenic signaling pathways. We further showed that TET1 and TET1$^{ALT}$ are functionally distinct proteins that serve different purposes in different cellular contexts.

A major finding of our work is the discovery of a novel isoform of TET1 (TET1$^{ALT}$) that lacks the CXXC DNA binding domain, but retains its catalytic activity. The canonical TET1 protein (TET1$^{FL}$) is the only isoform expressed in ESCs while TET1$^{ALT}$ is expressed in most adult and cancer cells, suggesting a different function for TET1 in ESCs vs. adult cells. Previous reports found TET1 to bind to CGIs and protect them from gains of methylation [30]. This is an important mechanism in ESCs to protect CGIs during waves of de novo methylation (via TET1$^{FL}$). However, TET1$^{FL}$ and TET1$^{ALT}$ are dramatically different proteins, as the 671 amino acid truncation of TET1$^{ALT}$ causes the loss of important regulatory domains including its DNA binding domain. Our findings suggest there is no role for TET1$^{ALT}$ in ESCs, and that instead TET1$^{ALT}$ serves as a dynamic regulator in adult cells where it is likely recruited to DNA by specific co-factors as previously shown for TET1 [79 82]. This would allow for the precise control of methylation in a tissue specific manner. It is important to note that much of the literature on TET1 has focused on ESCs, which may not be as relevant to how TET1 functions in adult cells.

The TET1$^{ALT}$ promoter is highly enriched for the active promoter mark H3K4me$^{3}$ and is bound by a multitude of transcription factors in multiple cell types. We show that tethering an activator domain to the TET1$^{ALT}$ promoter increases transcription specifically for the TET1$^{ALT}$ isoform. Overexpression of TET1$^{ALT}$ yields a truncated protein at ~162 kDa and results in production of 5-hydroxymethylcytosine, suggesting the protein is catalytically active. Other published work found the TET1$^{ALT}$ promoter to be an enhancer in human ESCs [126]. The investigators show the TET1$^{ALT}$ promoter region to be marked by H3K4me1 and H3K27Ac and bound by OCT3/4, MYC and NANOG. Because TET1$^{ALT}$ is not expressed in ESCs, it is possible that the TET1$^{ALT}$ promoter serves as an enhancer in ESCs but switches to a promoter during differentiation. This phenomena has been observed in the literature where intragenic enhancers act as alternative tissue specific promoters, allowing for transcription to occur in a developmental and cell type-specific manner [127].

TET1$^{FL}$ has been shown by us and others to both activate and repress gene transcription both dependent and independent of its demethylase activity [81 30]. We find that TET1$^{ALT}$ has fewer effects on gene expression than TET1$^{FL}$. One possible explanation is that the co-regulatory proteins targeting TET1$^{ALT}$ to DNA are not expressed in HEK293T cells or that the co-regulatory proteins must be co-expressed with TET1$^{ALT}$ to see robust changes in gene expression. Another possibility is that TET1$^{ALT}$ has few target genes in HEK293T cells. In the future, it will be important to identify the co-regulatory proteins recruiting TET1$^{ALT}$ to DNA. There are several examples of TET1 being targeted to DNA, such as via HIF1A where it affects methylation at hypoxic response genes [82]. In addition, TET1 can be targeted to DNA via FOXA1, where these proteins work together to modify the epigenetic signature at linage-specific enhancers [79]. A closer look at the published data suggest that TET1$^{ALT}$, not TET1$^{FL}$, is the interacting factor because when the investigators pulled down endogenous FOXA1 and probed with a TET1 antibody, the band was visible at ~150 kDa (the approximate size of TET1$^{ALT}$) [79].

We and others have shown that TET1$^{FL}$ is unable to induce global demethylation because its CXXC domain limits its ability to bind outside of CGIs [30]. Our data suggest that TET1$^{ALT}$, even though it lacks the CXXC domain, is still unable to induce wide-spread hypomethylation. We believe this is likely due to TET1$^{ALT}$ being restricted to specific DNA sites by co-regulatory proteins. We observed that TET1$^{ALT}$ demethylates ~6-fold more CG sites at non CGIs, compared to CGIs. We assume this preference for non CGIs is due to its lack of the CXXC domain. This also provides a rationale as to why there is minimal overlap between TET1$^{ALT}$ and TET1$^{FL}$ gene expression and methylation targets. TET1$^{FL}$ is targeted to CGIs by its CXXC domain and could be targeted to non CGIs by proteins that interact with its CXXC domain. However, because TET1$^{ALT}$ lacks this domain, it is likely regulated by another set of proteins that bind elsewhere in the TET1$^{ALT}$ protein.

It is important to note that our data do not comment on the physiological differences between TET1$^{FL}$ and TET1$^{ALT}$, but simply demonstrate that the isoforms have different methylation and gene expression targets and are thus distinct. Recently, another group reported that a truncated isoform of TET1 (TET1s) fails to erase imprints in primordial germ cells, suggesting physiological differences between the two isoforms [128]. In their study and our study, overexpression experiments were used to investigate the isoforms function. This is because TET1$^{FL}$ and TET1$^{ALT}$ share the same coding exons and it remains a challenge to cleanly knockout each isoform independently. For now, we believe overexpression is the cleanest way to study the proteins, however, isoform specific knockouts should be the focus of future experimentation.

Although we are in the early stages of understanding the molecular functions of TET1$^{ALT}$, our results indicate that TET1$^{ALT}$ is activated in cancer. This includes activation at the DNA level (gains of transcription start site peaks at the TET1$^{ALT}$ promoter), and at the RNA and protein level. TET1$^{ALT}$ appears to be the dominant isoform overexpressed in cancer, as few activation marks are found at the TET1$^{FL}$ promoter and TET1$^{FL}$ protein expression is unchanged between the normal and breast cancer cell lines. Because TET1 is frequently amplified in cancer and is associated with a worse overall survival in breast, uterine and ovarian cancers, we believe that TET1, specifically TET1$^{ALT}$, could play a more oncogenic role in cancers found predominantly in women.

To explore the roles of TET1 and TET1$^{ALT}$ in cancer, we looked across multiple TCGA datasets to determine if TET1 expression associates with changes in methylation. We identified a previously uncharacterized role for TET1 in TNBC, where it acts as an oncogene leading to hypomethylation and activation of oncogenic signaling pathways. Further, we provide evidence to suggest TET1 and TET1$^{ALT}$ are important regulators in the balance of hyper/hypomethylation in cancer. Approximately 42% of TNBCs overexpress TET1, where high levels of expression associate with a worse overall survival. On the contrary, remarkable downregulation of TET1 is found in HRBCs. Indeed, several studies have reported TET1 as a tumor suppressor in breast cancer, with one study finding that overexpression of TET1 led to reduced tumor volume in mice [60]

[129]. This contradictory evidence raises the interesting possibility of TET1 being both an oncogene and a tumor suppressor, a phenomenon that has been observed with other epigenetic regulators including EZH2 and DNMTs. This goes along with the Goldilocks principle where too much or too little of a protein is detrimental for the cell, and an optimal amount of expression is required for normal function. Disruption of EZH2 in mice is enough to cause T-acute lymphoblastic leukemia, suggesting a tumor suppressor function [130]. On the contrary, EZH2 is overexpressed in several cancers including melanoma, breast, and endometrial cancer where its expression is associated with disease progression [131]. In addition, caveolin-1 (CAV1) acts as a tumor suppressor in HRBCs (where it is downregulated) but an oncogene in TNBCs (where it is upregulated) [132]. The difference in CAV1 expression between TNBCs and HRBCs is in part due to CGI shore hypomethylation in the TNBCs [132], which may be TET1 mediated as we identified CAV1 as a TET1 target through our Spearman analysis in TNBC (r= 0.31, p=0.01). The opposing roles of TET1 add an additional layer of complexity in breast cancer and pose the question of whether TET1 should be inhibited or induced. For example, Vitamin C is a potent activator of the TET enzymes [58]. Should TNBC patients avoid Vitamin C and HRBC patients take Vitamin C as part of their treatment regimen? More work is required to answer these questions, and future studies should parse out the varying roles of TET1 in different cellular contexts.

One possible explanation for the varying roles of TET1 is different interacting partners. The interacting partners of TET1 and TET1$^{ALT}$ could vary in the different subtypes of breast cancer, where in TNBC a specific interacting partner could allow TET1 to hypomethylate and activate cancer specific pathways. For example, HIF1a and

XBP1 have been shown to be activated in TNBC [89]. XBP1 drives TNBC tumorigenicity by assembling a transcriptional complex with HIF1a that leads to the recruitment of RNA polymerase II at HIF1a target genes [89]. Another study found in neuroblastoma cells that under hypoxic conditions, HIF1a induces TET1 expression, leading to hypomethylation and activation of HIF1a target genes [133]. Therefore, TET1 upregulation may be an early response to hypoxic stress conditions, but as a consequence, leads to the demethylation of both hypoxic response genes and oncogenic signaling pathways, ultimately driving malignant transformation. Another possibility is that the opposing roles of TET1 may be isoform specific; full length TET1 may function as a tumor suppressor and the truncated $\text{TET1}^{\text{ALT}}$ may function as an oncogene.

Through the analysis of publicly available data from the TCGA, CCLE, and the genomics of drug sensitivity database, we provide evidence to suggest that a subset of TNBC and ovarian cancer patients may benefit from treatment with PI3K/Akt/mTOR inhibitors. So far, clinical results have been equivocal for using PI3K inhibitors in breast cancer and more focus on patient selection is needed to yield better results [120]. We propose that TET1 high patients may be particularly vulnerable to this type of therapy. Through careful KO experiments, we confirmed that TET1 is important for maintaining activation of the PI3K/mTOR pathway as loss of TET1 resulted in decreased phosphorylation of 4EBP1 and decreased gene expression of upstream regulators of the PI3K pathway, including EGFR, VEGFR, and integrin signaling. In addition, we observed decreased cellular proliferation and migration in the TET1 KO cells, further evidence that TET1 may be playing a more oncogenic role. Overlap of the genes differentially expressed in TCGA and the genes differentially expressed in the TET1 KO cells revealed numerous genes deregulated in both datasets, including the PI3K pathway genes INSR, THBS3, ANGPT1, PIK3CG, MAP2, and IL7R.

It remains unclear whether the downregulation of the PI3K genes is due to methylation dependent or independent effects of TET1. We showed that select genes, such as the insulin receptor, have decreased expression and CGI shore promoter hypermethylation in TET1 KO cells. Interestingly, the insulin receptor is also upregulated and hypomethylated in TET1 high Cluster 1 patients. However, not all of our differentially expressed PI3K genes had promoter methylation data available through the DREAM assay. For example, ANGPT1, a gene involved in the PI3K pathway [134], is upregulated and hypomethylated in Cluster 1 patients and is downregulated upon TET1 KO. However, the ANGPT1 promoter is not covered in the DREAM assay so we are unable to say if the downregulation is dependent or independent of TET1's demethylase activity. Thus, a limitation of our study includes the low number of CGI shores sites covered by the DREAM assay, especially because we showed that TET1 methylation targets are enriched in CGI shores. Indeed, several reports have linked CGI shore methylation changes in cancer to changes in expression. For example, CAV1 is upregulated in basal-like tumors following promoter hypomethylation specifically in CGI shores, where its expression is associated with a worse overall survival [132]. In the future, a technique that covers more CG sites outside of CGIs should be used to discern the methylation targets of TET1, such as bisulfite sequencing.

Another interesting finding is that TET1 may be a suppressor of the immune system. TNBC patients with high TET1 have decreased expression of immune pathway genes, and upon KO of TET1, immune genes are upregulated. In contrast, TNBCs with low TET1 have upregulated immune mediators and therefore could be sensitive to checkpoint inhibitors. Indeed, early results from several phase 1 clinical trials have shown success in using PD-L1 inhibitors in TNBC, with one study reporting a 19% response rate in women with heavily pretreated TNBCs [135]. If our hypothesis is correct, TNBC patients with low TET1 will be more likely to respond to PD-L1 treatment, and TET1 expression may be a way to pre-screen TNBCs to determine likelihood of response to immune checkpoint therapy. Thus, for cancer patients with high TET1 expression, one might envision combining TET1 inhibitors with immune checkpoint inhibitors such as PD-1 or PD-L1 in the future.

Our work addresses a critical gap in knowledge of how and why methylation is prognostic in breast cancer, as loss of methylation predominatly occurs at genes associated with oncogenic pathways. Further, we shed light on how methylation and TET1 expression levels can be used to stratify TNBC patients for targeted therapy. We also provide evidence that TET1 mediated hypomethylation occurs not only in TNBC, but also in serous ovarian cancer. Therefore, an oncogenic role for TET1 may be a more widespread phenomena utilized by cancers cells to hijack signaling pathways. Lastly, our work establishes TET1 as an oncogene that could serve as a novel druggable target for therapeutic intervention in TNBC, ovarian cancer and beyond.

# REFERENCES CITED

1. Jenuwein, T. & Allis, D. Translating the Histone Code. *Science (80-. ).* **293,** 1074–1081 (2001).

2. Messerschmidt, D. M., Knowles, B. B. & Solter, D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev.* **28,** 812–828 (2014).

3. Estécio, M. R. H. & Issa, J.-P. J. Dissecting DNA hypermethylation in cancer. *FEBS Lett.* **585,** 2078–86 (2011).

4. Elsheikh, S. E. *et al.* Global histone modifications in breast cancer correlate with tumor phenotypes, prognostic factors, and patient outcome. *Cancer Res.* **69,** 3802–9 (2009).

5. Chen, K. *et al.* Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat. Genet.* **47,** 1149–1157 (2015).

6. Pulakanti, K. *et al.* Enhancer transcribed RNAs arise from hypomethylated, Tet-occupied genomic regions. *Epigenetics* **8,** 1303–1320 (2013).

7. Wu, T. P. *et al.* DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature* **532,** 329–333 (2016).

8. Robertson, K. DNA methylation and human disease. *Nat. Rev. Genet.* **6,** 597–610 (2005).

9. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16,** 6–21 (2002).

10. Wu, S. C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* **11,** 750–750 (2010).

11. Stevens, M. *et al.* Estimating absolute methylation levels at single-CpG resolution from methylation enrichment and restriction enzyme sequencing methods. *Genome Res.* **23,** 1541–1553 (2013).

12. Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* (2017). doi:10.1038/nrg.2017.33

13. Deaton, A. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25,** 1010–1022 (2011).

14. Issa, J.-P. CpG island methylator phenotype in cancer. *Nat. Rev. Cancer* **4,** 988–993 (2004).

15. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13,** 484–492 (2012).

16. Schübeler, D. Function and information content of DNA methylation. *Nature* **517,** 321–326 (2015).

17. Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502,** 472–9 (2013).

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

19. Scourzic, L., Mouly, E. & Bernard, O. A. TET proteins and the control of cytosine demethylation in cancer. 1–16 (2015). doi:10.1186/s13073-015-0134-6

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

21. Shukla, V. *et al.* BRCA1 affects global DNA methylation through regulation of DNMT1. *Cell Res.* **20,** 1201–1215 (2010).

22. Ludwig, A. K., Zhang, P. & Cardoso, M. C. Modifiers and readers of DNA modifications and their impact on genome structure, expression, and stability in disease. *Front. Genet.* **7,** 1–24 (2016).

23. Jeltsch, A. & Jurkowska, R. Z. New concepts in DNA methylation. *Trends Biochem. Sci.* **39,** 310–318 (2014).

24. Bashtrykov, P. *et al.* Specificity of dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain. *Chem. Biol.* **19,** 572–578 (2012).

25. Bostick, M. *et al.* UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science (80-. ).* **317,** 1760–1764 (2007).

26. Spada, F. *et al.* DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *J. Cell Biol.* **176,** 565–571 (2007).

27. Abdel-wahab, O. & Levine, R. L. Mutations in epigenetic modifiers in the pathogenesis and therapy of acute myeloid leukemia. *Epigenetics Hematol.* **121,** 3563–3572 (2013).

28. Yang, X. *et al.* Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* **26,** 577–590 (2014).

29. Putiri, E. L. *et al.* Distinct and overlapping control of 5-methylcytosine and 5-hydroxymethylcytosine by the TET proteins in human cancer cells. *Genome Biol.* **15,** R81 (2014).

30. Jin, C. *et al.* TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. *Nucleic Acids Res.* **42,** 6956–71 (2014).

31. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* **41,** 178–186 (2009).

32. Gonzalez-Gomez, P. *et al.* CpG island methylation status and mutation analysis of the RB1 gene essential promoter region and protein-binding pocket domain in nervous system tumours. *Br. J. Cancer* **88,** 109–14 (2003).

33. Zhang, L. & Long, X. Association of BRCA1 promoter methylation with sporadic breast cancers: Evidence from 40 studies. *Sci. Rep.* **5,** 17869 (2015).

34. Fang, F. *et al.* Breast cancer methylomes establish an epigenomic foundation for metastasis. *Sci. Transl. Med.* **3,** 75ra25 (2011).

35. Kelly, A. D. & Issa, J. P. J. The promise of epigenetic therapy: reprogramming the cancer epigenome. *Curr. Opin. Genet. Dev.* **42,** 68–77 (2017).

36. Tahara, T. *et al.* Colorectal carcinomas with CpG island methylator phenotype 1 frequently contain mutations in chromatin regulators. *Gastroenterology* **146,** 530–38.e5 (2014).

37. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell* **18,** 553–567 (2010).

38. Kelly, A. D. *et al.* A CpG island methylator phenotype in acute myeloid leukemia independent of IDH mutations and associated with a favorable outcome. *Leukemia* 1–9 (2017). doi:10.1038/leu.2017.12

39. Hunter, W. N., Brown, T. & Kennard, O. The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res.* **38,** ii (2010).

40. Huang, Y. & Rao, A. Connections between TET proteins and aberrant DNA modification in cancer. *Trends Genet.* **30,** 464–474 (2014).

41. Ehrlich, M. DNA hypomethylation in cancer cells. *Epigenomics* **1,** 239–259 (2010).

42. Koboldt, D. C. *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2012).

43. Stefansson, O. a. *et al.* A DNA methylation-based definition of biologically distinct breast cancer subtypes. *Mol. Oncol.* **9,** 555–568 (2014).

44. Lee, J. S. *et al.* Basal-like breast cancer displays distinct patterns of promoter methylation. *Cancer Biol. Ther.* **9,** 1017–1024 (2010).

45. Holm, K. *et al.* Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. *Breast Cancer Res.* **12,** R36 (2010).

46. Xu, J. *et al.* Methylation of HIN-1, RASSF1A, RIL and CDH13 in breast cancer is associated with clinical characteristics, but only RASSF1A methylation is associated with outcome. *BMC Cancer* **12,** 243 (2012).

47. Tsai, H. & Baylin, S. B. Cancer epigenetics : linking basic biology to clinical medicine. *Nat. Publ. Gr.* **21,** 502–517 (2011).

48. Maegawa, S. *et al.* Age-related epigenetic drift in the pathogenesis of MDS and AML. *Genome Res.* **24,** 580–591 (2014).

49. Lorsbach, R. B. *et al.* TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* **17,** 637–641 (2003).

50. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science (80-. ).* **333,** 1300–3 (2011).

51. Guo, J. U., Su, Y., Zhong, C., Ming, G. L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145,** 423–434 (2011).

52. Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein- mediated 5-methylcytosine oxidation. *Genes Dev.* **25,** 2436–2452 (2011).

53. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–5 (2009).

54. Iyer, L. M., Tahiliani, M., Rao, A. & Aravind, L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8,** 1698–1710 (2009).

55. Pastor, W. a, Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* **14,** 341–56 (2013).

56. Ko, M. *et al.* Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* **497,** 122–6 (2013).

57. Fan, J. *et al.* Human phosphoglycerate dehydrogenase produces the oncometabolite D-2-hydroxyglutarate. *ACS Chem. Biol.* **10,** 510–516 (2015).

58. Blaschke, K. *et al.* Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500,** 222–226 (2013).

59. Dawlaty, M. M. *et al.* Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* **9,** 166–175 (2011).

60. Hsu, C.-H. *et al.* TET1 suppresses cancer invasion by activating the tissue inhibitors of metalloproteinases. *Cell Rep.* **2,** 568–79 (2012).

61. Li, L. *et al.* Epigenetic inactivation of the CpG demethylase TET1 as a DNA methylation feedback loop in human cancers. *Sci. Rep.* **6,** 26591 (2016).

62. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466,** 1129–33 (2010).

63. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473,** 343–8 (2011).

64. Dawlaty, M. M. *et al.* Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development. *Dev. Cell* **24,** 310–323 (2013).

65. Delatte, B., Deplus, R. & Fuks, F. Playing TETris with DNA modifications. *EMBO J.* **33,** 1198–211 (2014).

66. Song, S. J. *et al.* MicroRNA-antagonism regulates breast cancer stemness and metastasis via TET-family-dependent chromatin remodeling. *Cell* **154,** 311–24 (2013).

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

68. Teif, V. B. *et al.* Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. *Genome Res.* **24,** 1285–1295 (2014).

69. Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149,** 1368–1380 (2012).

70. Costa, Y. *et al.* NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* **495,** 370–374 (2013).

71. Shi, F.-T. *et al.* Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells. *J. Biol. Chem.* **288,** 20776–84 (2013).

72. Zhang, Q. *et al.* Differential regulation of the ten-eleven translocation (TET) family of dioxygenases by O-linked β-N-acetylglucosamine transferase (OGT). *J. Biol. Chem.* **289,** 5986–96 (2014).

73. Deplus, R. *et al.* TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J.* **32,** 645–55 (2013).

74. Vella, P. *et al.* Tet Proteins Connect the O-Linked N-acetylglucosamine Transferase Ogt to Chromatin in Embryonic Stem Cells. *Mol. Cell* **49,** 645–656 (2013).

75. Neri, F. *et al.* Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. *Genome Biol.* **14,** R91 (2013).

76. Okashita, N. *et al.* PRDM14 promotes active DNA demethylation through the Ten-eleven translocation (TET)-mediated base excision repair pathway in embryonic stem cells. *Development* **141,** 269–280 (2014).

77. Zeng, Y. *et al.* Lin28A Binds Active Promoters and Recruits Tet1 to Regulate Gene Expression. *Mol. Cell* **61,** 153–160 (2016).

78. Cartron, P.-F. *et al.* Identification of TET1 Partners That Control Its DNA-Demethylating Function. *Genes Cancer* **4,** 235–41 (2013).

79. Yang, Y. A. *et al.* FOXA1 potentiates lineage-specific enhancer activation through modulating TET1 expression and function. *Nucleic Acids Res.* **44,** 8153–64 (2016).

80. Müller, U., Bauer, C., Siegl, M., Rottach, A. & Leonhardt, H. TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res.* 1–13 (2014). doi:10.1093/nar/gku552

81. Wu, H. *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473,** 389–93 (2011).

82. Tsai, Y.-P. *et al.* TET1 regulates hypoxia-induced epithelial-mesenchymal transition by acting as a co-activator. *Genome Biol.* **15,** 513 (2014).

83. Sun, M. *et al.* HMGA2/TET1/HOXA9 signaling pathway regulates breast cancer growth and metastasis. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 9920–5 (2013).

84. Yang, L., Yu, S.-J., Hong, Q., Yang, Y. & Shao, Z.-M. Reduced Expression of TET1, TET2, TET3 and TDG mRNAs Are Associated with Poor Prognosis of Patients with Early Breast Cancer. *PLoS One* **10,** e0133896 (2015).

85. Huang, H. *et al.* TET1 plays an essential oncogenic role in MLL-rearranged leukemia. *Proc. Natl. Acad. Sci.* **110,** 11994–11999 (2013).

86. Martin, H. L., Smith, L. & Tomlinson, D. C. Multidrug-resistant breast cancer: current perspectives. *Breast cancer (Dove Med. Press.* **6,** 1–13 (2014).

87. Stirzaker, C. *et al.* Methylome sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value. *Nat. Commun.* **6,** 1–11 (2015).

88. Gordon, V. & Banerji, S. Molecular pathways: PI3K pathway targets in triple-negative breast cancers. *Clin. Cancer Res.* **19,** 3738–3744 (2013).

89. Chen, X. *et al.* XBP1 promotes triple-negative breast cancer by controlling the HIF1alpha pathway. *Nature* **508,** 103–107 (2014).

90. Fruman, D. a. Phosphoinositide Kinases. *Annu. Rev. Biochem.* **67,** 481–507 (1998).

91. Hemmings, B. A. & Restuccia, D. F. PI3K-PKB / Akt Pathway. 1–4 (2012). doi:10.1101/cshperspect.a011189

92. Coleman, M. L., Marshall, C. J. & Olson, M. F. RAS and RHO GTPases in G1-phase cell-cycle regulation. *Nat. Rev. Mol. Cell Biol.* **5,** 355–366 (2004).

93. Cantley, L. C. The phosphoinositide 3-kinase pathway. *Science* **296,** 1655–1657 (2002).

94. Majchrzak, A., Witkowska, M. & Smolewski, P. Inhibition of the PI3K/Akt/mTOR signaling pathway in diffuse large B-cell lymphoma: Current knowledge and clinical significance. *Molecules* **19,** 14304–14315 (2014).

95. Thorpe, L. M., Yuzugullu, H. & Zhao, J. J. PI3K in cancer: divergent roles of isoforms, modes of activation and therapeutic targeting. *Nat. Rev. Cancer* **15,** 7–24

(2014).

96. Elenbaas, B. *et al.* Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev.* **15,** 50–65 (2001).

97. Jelinek, J. *et al.* Conserved DNA methylation patterns in healthy blood cells and extensive changes in leukemia measured by a new quantitative technique. *Epigenetics* **7,** 1368–1378 (2012).

98. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).

99. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28,** 511–515 (2010).

100. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11,** 783–784 (2014).

101. Kearns, N. A. *et al.* Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Development* **141,** 219–223 (2014).

102. Yang, A. S. *et al.* A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res.* **32,** e38 (2004).

103. Dunham, I. T. E. P. C. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

104. Pereira, B. *et al.* The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat. Commun.* **7,** 11479 (2016).

105. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486,** 346–52 (2012).

106. Getz, G. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497,** 67–73 (2013).

107. Voigt, P. & Reinberg, D. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia The Cancer Genome Atlas Research Network. *N. Engl. J. Med.* **368,** 2059–74 (2013).

108. Kamburov, A., Wierling, C., Lehrach, H. & Herwig, R. ConsensusPathDB - A database for integrating human functional interaction networks. *Nucleic Acids Res.* **37,** 623–628 (2009).

109. Djebali S, Merkel A, Lassmann T, Tanzer A, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, B, Gingeras TR., Dav, Dobin A, Mortazavi A, L. J. Landscape of transcription in human cells. *Nature* **489,** 101–108 (2012).

110. Stothard, P. The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences. *BioTchniques* **28,** (2000).

111. Rosenbloom, K. R. *et al.* ENCODE Data in the UCSC Genome Browser: Year 5 update. *Nucleic Acids Res.* **41,** 56–63 (2013).

112. Suzuki, A. *et al.* DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data. *Nucleic Acids Res.* **43,** D87–D91 (2015).

113. Ardlie, K. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science (80-. ).* **348,** 648–660 (2015).

114. Eirew, P. *et al.* Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* **518,** 422–6 (2015).

115. Baca, S. C. *et al.* Punctuated evolution of prostate cancer genomes. *Cell* **153,** 666–677 (2013).

116. Kumar, A. *et al.* Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. *Nat. Med.* **22,** 369–78 (2016).

117. Grasso, C. S. *et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487,** 239–43 (2012).

118. Beltran, H. *et al.* Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat. Med.* **22,** 298–305 (2016).

119. Uhlen, M. *et al.* Tissue-based map of the human proteome. *Science (80-. ).* **347,** 1260419–1260419 (2015).

120. Fruman, D. A. & Rommel, C. PI3K and cancer: lessons, challenges and opportunities. *Nat. Rev. Drug Discov.* **13,** 140–156 (2014).

121. Yang, W. *et al.* Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41,** D955-61 (2013).

122. Kasahara, K. *et al.* PI 3-kinase-dependent phosphorylation of Plk1-Ser99 promotes association with 14-3-3γ and is required for metaphase-anaphase transition. *Nat. Commun.* **4,** 1882 (2013).

123. Flores-Pérez, A. *et al.* Dual targeting of ANGPT1 and TGFBR2 genes by miR-204 controls angiogenesis in breast cancer. *Sci. Rep.* **6,** 34504 (2016).

124. Dalla-Torre, C. A. *et al.* Effects of THBS3, SPARC and SPP1 expression on biological behavior and survival in patients with osteosarcoma. *BMC Cancer* **6,** 237 (2006).

125. Zhang, W. *et al.* DNA hypomethylation-mediated activation of Cancer/Testis Antigen 45 ( CT45 ) genes is associated with disease progression and reduced survival in epithelial ovarian cancer. *Epigenetics* **10,** 00–00 (2015).

126. Neri, F. *et al.* TET1 is controlled by pluripotency-associated factors in ESCs and downmodulated by PRC2 in differentiated cells and tissues. *Nucleic Acids Res.* **43,** 6814–26 (2015).

127. Kowalczyk, M. S. *et al.* Intragenic Enhancers Act as Alternative Promoters. *Mol. Cell* **45,** 447–458 (2012).

128. Zhang, W. *et al.* Isoform Switch of TET1 Regulates DNA Demethylation and Mouse Development. *Mol. Cell* **64,** 1062–1073 (2016).

129. Yang, H. *et al.* Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* **32,** 663–9 (2013).

130. Hock, H. A complex Polycomb issue: The two faces of EZH2 in cancer. *Genes Dev.* **26,** 751–755 (2012).

131. Kim, K. H. & Roberts, C. W. M. Targeting EZH2 in cancer. *Nat. Med.* **22,** 128–134 (2016).

132. Rao, X. *et al.* CpG island shore methylation regulates caveolin-1 expression in breast cancer. *Oncogene* **32,** 4519–4528 (2013).

133. Mariani, C. J. *et al.* TET1-Mediated Hydroxymethylation Facilitates Hypoxic Gene Induction in Neuroblastoma. *Cell Rep.* **1,** 1–10 (2014).

134. Huang, H., Bhat, A., Woodnutt, G. & Lappe, R. Targeting the ANGPT–TIE2 pathway in malignancy. *Nat. Rev. Cancer* **10,** 575–585 (2010).

135. McArthur, H. L. Checkpoint inhibitors in breast cancer: Hype or promise? *Clin. Adv. Hematol. Oncol.* **14,** 392–395 (2016).