Distribution and role of TET-mediated DNA and RNA hydroxymethylation in breast cancer and normal breast epithelium

Jinrui Wen

Division of Cancer

Department of Surgery and Cancer

Imperial College London

Supervisors:

Prof Justin Stebbing

Dr Biancastella Cereser

Thesis submitted for the degree of Doctor of Philosophy

Declaration of Originality

I, Jinrui Wen, declare that all research presented in this thesis is my own work unless otherwise stated. Contributions from others are clearly stated and fully acknowledged.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work.

This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Acknowledgements

First, I would like to thank my supervisor, Prof Justin Stebbing and my funder, Action Against Cancer, for generously funding my postgraduate research; I could not have finished my PhD without their support. Second, my sincere gratitude to my day-to-day supervisor, Dr Biancastella Cereser, who guided me, supervised me and taught me over the past five years. I wish to thank all the lab members in the 'Cancer Genetics' group for their kind suggestions and supports given to me all through my master and my PhD. A special thanks to Dr Lisa Del Bel Belluz (Research technician), who gave me lots of help on performing my experiments, Dr Nina Moderau (Research Associate), who gave me lots of suggestions on planning and troubleshooting and Dr Aleksandra Dabrowska, who accompanied me and gave me strength. I also thank my past and current colleagues Mr Pramodh Chandrasinghe, Dr Neha Tabassum, Sandesh Meyler, Michael Toeller, our previous Master student Liliane Eleid, Dr Silvia Ottaviani, Dr Sladjana Zagorac and Dr Angela Yiu for their support and partnership.

Apart from my colleagues, I also wish to thank my friends for the companionship and support they provided. Manjari Trivedi is my lucky star, who helped me to debug the confocal microscope and the program, which took me several months to troubleshoot. Hailing Fan helped me a lot on western blot and other experiments. My special thanks to Yingying Tan, Ning Wang, Sijia Yu for all the chats, laughter and support which made my life during my PhD much better.

I have faced lots of difficulties during my whole PhD life. Chinese is very different from English. Thus, my sincere thanks to all the teachers in the Centre of Academic English who taught my academic writing, especially, Hilary Glasman-Deal and Liz Chiu. These two teachers gave me new insights into writing, and their suggestions made my writing much easier for me though

I apologise for the basic mistakes in my thesis I have missed. I also wish to thank Sophie and Katie, who guided me to explore new ways to continue my studies smoothly.

Last but most important to my parents, my husband, Haiyang Yu, and my two cats, I cannot finish without their love and support. Thanks to all my friends who cheered me up and spent time with me.

This project used patient samples from Imperial College Tissue bank and St. Mark's Hospital pathology department. Sincere thanks to the late Prof Sami Sousha for providing histopathological assessment.

I am genuinely grateful for everyone's kindness and support throughout the years. Without any of you, I wouldn't be who I am today.

Abstract

Breast cancer (BC) is the most frequently diagnosed malignancy in women. Epigenetic modifications, involving DNA, RNA, and histone modulation, result in the alteration of gene expression. Among these epigenetic modifications, 5-hydroxymethylcytosine (5hmC) has emerged as being relevant in various biological and pathological processes. In particular, there is evidence that 5hmC is involved in cancer progression and may possibly function as a prognostic biomarker.

The first aim of this study was to validate 5hmC distribution and hypothesise its modulatory role in the transcription and translation of DNA and RNA in BC via immunohistochemistry (IHC) and immunocytochemistry (ICC). Through IHC and ICC, I detected cytoplasmic 5hmC for the first time in BC. RNA hydroxymethylation was detected by dot blot analysis of the human mammary cell lines MCF10A and BT474, a finding previously reported in other cells, such as colorectal cancer, but not in breast cancer. A stratification of the samples in different cancer subtypes showed that the proportion of positive cells of cytoplasmic 5hmC was significantly higher only in ER-negative cancer epithelium compared to the normal matched epithelium.

In further investigations, I examined whether 5hmC levels were associated with a specific TET protein in BC. Here, IHC experiments were performed on the same samples. The distribution of cytoplasmic TET2 showed correlations with the distribution of 5hmC and this may have prognostic significance. Following the knockdown of *TET* mRNA expression by siRNA (in BT474), down-regulation of the *TET2* transcript led to lower cell proliferation while loss of *TET1* or *TET3* expression was associated with increased proliferation.

In conclusion, I suggest that 5hmC may have a role in the development of certain sub-types of BC and TET2 is of particular relevance here.

Table of Contents

Declaration of Originality 2

Copyright Declaration 2

Acknowledgements 3

Abstract 5

Table of Contents 6

List of Tables 12

List of Figures 14

Abbreviations 17

1. Background 24

1.1 Breast cancer 24

1.1.1 Introduction and epidemiology 24

1.1.2 BC Progression 24

1.1.3 Diversity of breast cancers 26

1.1.3.1 Histopathological classification 27

1.1.3.2 Molecular classifications 28

1.1.4 Breast cancer treatment 30

1.1.5 Molecular markers in breast cancer 31

1.2 Epigenetic modifications 32

1.2.1 Methylation of cytosine 32

1.2.2 The DNA methyltransferases (DNMTs) 33

1.2.3 Post-transcriptional RNA modification and RNA hydroxymethylation 38

1.2.4 TET enzymes 39

1.2.5 Other oxidative derivatives of 5hmC 40

1.3 Epigenetic changes in cancer 41

1.4 Ten-eleven translocation protein in cancer 42

1.4.1 TET1 42

1.4.2 TET2 50

1.4.3 TET3 51

1.5 Hypothesis and Aims 52

2. Materials and methods 54

2.1 Patient Samples 54

2.2 Immunohistochemistry (IHC) 62

2.2.1 Optimisation for the protocols and antibodies 62

2.2.2 IHC of FFPE sections 64

2.2.3 Quantification of staining intensity and protein expression 65

2.2.4 Statistical Analysis 68

2.3 *In situ* Hybridisation (ISH) 68

2.3.1 Optimisation for RNAscope 68

2.3.2 Workflow of RNAscope 69

2.3.3 Quantifications and analysis 69

2.4 Cell culture 71

2.4.1 Cell lines 71

2.5 Immunocytochemistry (ICC) 73

2.5.1 Optimisations of antibodies 73

2.5.2 Workflow of ICC 74

2.6 Quantification of hydroxymethylated DNA 75

2.6.1 DNA extraction 75

2.6.2 Workflow of Colorimetric assay 75

2.6.3 Quantifications and analysis 76

2.7 Dot blot 77

2.7.1 RNA extraction 77

2.7.2 Large and small RNA isolation 78

2.7.3 mRNA isolation 79

2.7.4 Optimisations of the samples 80

2.7.5 Workflow of dot blot 81

2.7.6 Quantification and analysis 81

2.8 Functional assay 81

2.8.1 siRNA Reverse Transfections 82

2.8.1.1 Optimisations of reverse transfection 82

2.8.1.2 Workflow of siRNA transfection 84

2.8.2 Reverse Transcription and Real-time Quantitative Polymerase Chain Reaction (RT-qPCR) 86

2.8.2.1 Reverse Transcription and RT-qPCR 86

2.8.2.2 Analysis 88

2.8.3 Proliferation assay 88

2.9 Kaplan-Meier survival plots 89

3. Result 1: 91

The localisation of demethylation marks and the presence of 5hmC 91

3.1 Results 93

3.1.1 The breast cancer epithelium is characterised by a significantly higher proportion of positive cells of cytoplasmic 5hmC 93

3.1.2 The cancer epithelium from ER- patients is characterised by a significantly higher proportion of positive cells of cytoplasmic 5hmC 96

3.1.3 Higher proportion of positive cells of cytoplasmic 5hmC in colorectal cancer than quiescent epithelium 106

3.1.4 Cytoplasmic 5mC is significantly reduced in BC epithelium 108

3.1.5 Global levels of 5hmC are reduced in breast cancer cells compared with normal cells 111

3.1.6 The localisation of demethylation marks is differentially expressed in breast cell lines 114

3.1.7 5hmC distributes in DNA and RNAs in cell lines 117

3.2 Discussion 120

3.3 Conclusion 124

4. Result 2: 126

TET proteins and 5-hydroxymethylation in breast cells 126

4.1 Results 128

4.1.1 The breast cancer epithelium is characterised by a significantly higher level of cytoplasmic TET2 compared to the healthy breast epithelium 128

4.1.2 The association between TET proteins and 5hmC distributions in breast epithelium 134

4.1.3 TET proteins are expressed in both the nuclei and cytoplasm of breast cell lines. 138

4.1.4 The higher expression of *TET2* mRNA is positively correlated with the probability of higher BC survival rate 142

4.1.5 Distribution of TET2 mRNA and protein product in breast tissue 150

4.2 Discussion 154

4.3 Conclusion 158

5. Result 3: 159

The function of TETs in breast cell lines 159

5.1 Results 162

5.1.1 Optimisation of siRNAs 162

5.1.2 TET siRNAs significantly downregulated *TET* mRNA expression in BT474 cells.... 169

5.1.3 siRNAs mediated TET knockdown in MCF10A cells..... 174

5.1.4 Levels of 5hmC are not affected by *TET* knockdown for 96 hours..... 181

5.1.5 Increase in cell proliferation is associated with the knockdown of *TET*..... 183

5.1.6 Evaluation of gene expression following *TET* knockdown in BT474..... 186

5.2 Discussion..... 191

6. Discussion and Conclusions..... 195

6.1 Summary of findings..... 195

6.2 Limitations of the study..... 200

6.3 Future directions..... 202

6.4 Conclusions..... 206

References..... 207

Appendix..... 235

List of Tables

Table 1.1 TET proteins: known functions and research studies. 45

Table 2.1 Details of the FFPE sections of IDC tissue received from Imperial College Tissue Bank. 56

Table 2.2 Details of the FFPE sections of IDC tissue received from Imperial College Tissue Bank. 57

Table 2.3 Details of the FFPE sections of breast tissue received from Imperial College Tissue Bank. 59

Table 2.4 Details of the FFPE sections of CAC tissue received from St. Mark's Hospital. 61

Table 2.5 Details of primary antibodies used in IHC and ICC. 63

Table 2.6 Characteristic of cells and culture medium. 72

Table 2.7 Details of Secondary antibody used in ICC. 74

Table 2.8 Optimised transfection conditions. 83

Table 2.9 Target sequence of siRNAs. 85

Table 2.10 Primer Sequences for SYBR Green Real-Time qPCR. 87

Table 5.1 Significance between gene expression for TET1, TET2 and TET3 in Unt-, Scr- and Lipo - treated BT474 cells 170

Table 5.2 Significance between gene expression for POLR2A and RPLP0 in Unt-, Scr- and Lipo - treated BT474 cells 171

Table 5.3 The Ct values of each sample in BT474 untreated cells at 24 hours. 179

Table 5.4 The Ct values of each sample in MCF10A untreated cells at 24 hours. 179

Table 5.5 Significance between gene expression for TET1, TET2, TET3 in Unt-, Scr- and Lipo- in MCF10A..... 180

Table 5.6 Significance between gene expression for POLR2A and RPLP0 in Unt-, Scr- and Lipo - treated MCF10A cells..... 181

Table 5.7 Significance between gene expression for *DNMT1, DNMT3a, Ki67, MCM2, Bcl-2, Bcl-W, ESR1, GATA3* in Unt-, Scr- and Lipo- in BT474 188

List of Figures

Figure 1.1 The structure of the breast. 25

Figure 1.2 Example of breast cancer progression and the histopathology of DCIS. 26

Figure 1.3 Molecular subtypes of breast cancer. 29

Figure 1.4 Overview of the mechanisms of DNA demethylation (61). 35

Figure 1.5 The molecular structure changes during the active DNA demethylation pathways through 5hmC. 36

Figure 1.6 Known domain structures of TET proteins. 40

Figure 2.1 Immunohistochemistry staining and quantifications. 67

Figure 2.2. Quantification of *In situ* hybridisation (ISH) using the ImageJ software. 71

Figure 2.3 Standard curve calculated by logarithmic second order regression. 77

Figure 2.4 Dot blot. 80

Figure 3.1 5hmC immunohistochemistry staining. 95

Figure 3.2 (Previous page) The proportion of positive cells of 5hmC after stratification of the samples based on the hormonal receptor status of the cancer. 101

Figure 3.3 The proportion of positive cells of 5hmC after stratification of the samples according to the ER status of the cancer. 102

Figure 3.4 The proportion of positive cells of 5hmC after stratification of the samples according to the HER2 status of the cancer 103

Figure 3.5 The proportion of positive cells of 5hmC after stratification of the samples according to cancer grade. 104

Figure 3.6 The correlation between 5hmC immunohistochemistry staining and age in 29 patients. 105

Figure 3.7 5hmC immunochemistry staining in colorectal samples. 107

Figure 3.8 5mC immunohistochemistry staining. 110

Figure 3.9 Differences in the amount of 5hmC in the DNA of breast cell lines. 113

Figure 3.10 5mC, 5hmC, 5fC and 5caC immunocytochemistry staining. 116

Figure 3.11 5hmC levels in the DNA and RNA of human breast cell lines. DNA or RNAs are crosslinked to an Hybond-N+ membrane with different concentrations. 119

Figure 4.1 5hmC, TET1, TET2, and TET3 immunohistochemistry staining. 132

Figure 4.2 The association between the proportion of positive cells of 5hmC, TET1, TET2, and TET3 in six breast epithelial tissues. 136

Figure 4.3 The association between delta 5hmC, delta TET1, delta TET2, and delta TET3 in six breast epithelial tissues. 137

Figure 4.4 (Previous page) TET1, TET2, TET3 immunocytochemistry staining. 141

Figure 4.5 (Previous page) *TET2* is positively correlated with the probability of higher BC survival rate. 145

Figure 4.6 (Previous page) *TET1* and *TET3* are negatively correlated with the probability of higher ovarian cancer survival rate. 147

Figure 4.7 (Previous page) *TET2* mRNA expression is positively correlated with the probability of higher BC patient survival rate in subtypes..... 149

Figure 4.8 The quantification of the expression of *TET2* mRNA. 152

Figure 5.1 (Previous page) *TETs* mRNA relative expression decreased upon siRNA treatment in HEK293T cells (n=1). 165

Figure 5.2 (Previous page) *TETs* mRNA relative expression decreased upon siRNA treatment in HEK293T cells (n=1). 168

Figure 5.3 (Previous page) *TETs* mRNA expression significantly decreased upon siRNA treatment in BT474 cells (n=3). 174

Figure 5.4 (Previous page) siRNAs mediated TET knockdown in MCF10A cells (n=3). . 177

Figure 5.5 Delta Ct value for BT474 and MCF10A untreated cells at 24 hours. 179

Figure 5.6 The levels of 5hmC in the DNA are not affected by TET knockdown in BT474 cells (n=3)..... 182

Figure 5.7 (Previous page) Cell proliferation in BT474 cells following TET knockdown (n=5). 185

Abbreviations

0.1% PBST	0.1% Tween-20 in PBS
0.3%PBST	0.3% Triton X-100 in PBS
2-OG	2-Oxoglutarate
5-AZA	5-Azacytidine
5caC	5-Carboxylcytosine
5cadC	5-Carboxyl-2’deoxycytidine
5fC	5-Formylcytosine
5fdC	5-Formy-2’deoxycytidine
5fdU	5-Formyl-2'-deoxyuridine
5hmC	5-Hydroxymethylation
5hmdC	5-Hydroxymethyl-2’deoxycytidine
5hmdU	5-Hydroxymethyl-2’deoxyuridine
5mC	5-Methylcytosine
ACE-seq	APOBEC-coupled epigenetic sequencing
ADH	Atypical ductal hyperplasia
AML	Acute myeloid leukaemia
ATRA	All-trans retinoic acid
Bak	BCL2 Antagonist/Killer

Bax	Bcl-2 Associated X-protein
BC	Breast cancer
Bcl-2	B-cell lymphoma 2
BER	Base excision repair
BRCA1	Breast cancer type 1
BRD4	Bromodomain-containing protein 4
BS-seq	Bisulphite sequencing
CAC	Colitis-associated cancer
CDA	Cytidine deaminase
Cdkn	Cyclin dependent kinase inhibitor
CGI	CpG islands
CpG	5’-Cytosine-phosphate-guanine-3’
CXXC4	CXXC Finger Protein 4
Cys	Cysteine-rich
DAB	3,3'-Diaminobenzidine
DCIS	Ductal carcinoma *in situ*
DH	Ductal hyperplasia
DMEM	Dulbecco’s Modified Eagle’s Medium
DNA	Deoxyribonucleic acid

DNMT	DNA methyltransferase
DPBS	Dulbecco's phosphate buffered saline
DSBH	Double-strand β-helix
EDTA	Ethylenediaminetetraacetic acid
EMT	Epithelial-mesenchymal transition
ER	Oestrogen receptor
ErbB2	Erb-B2 Receptor Tyrosine Kinase 2
ESCs	Human embryonic stem cells
EZH2	Enhancer Of Zeste 2 Polycomb Repressive Complex
FBS	Foetal Bovine Serum
FCS	Foetal Calf Serum
FFPE	Formalin-fixed paraffin-embedded
FOA1	F-box and associated interaction domains-containing protein
FOXP1	Forkhead Box P1
GATA3	GATA binding protein 3
GS buffer	Goat serum buffer
H&E	Haematoxylin and Eosin
HBS	H-hue, B-brightness and S-saturation
HCl	Hydrochloric acid

hEGF / Hegf	Human Epidermal Growth Factor
HER2	Human epidermal growth factor receptor 2
HMGA2	High mobility group AT-hook 2
HOXA	Homeobox A
HRP	Horseradish peroxidase
HSK	Housekeeping
ICC	Immunocytochemistry
IDC	Invasive ductal carcinoma
IFN-γ	Interferon γ
IHC	Immunohistochemistry
ISH	*In situ* Hybridisation
JAK/STAT	The Janus kinase (JAK)-signal transducer and activator of transcription (STAT) pathway
LC-MS	Liquid chromatography–mass spectrometry
Lipo	Cells only transfected by lipofectamine reagent
lncRNA	Long non-coding RNA
M	Metastasis
MCM2	Minichromosome maintenance protein complex 2
miRs	MicroRNAs

MLL	Mixed-lineage leukemia
MMP9	Matrix metalloproteinase 9
mRNA	Messenger RNA
N	Lymph node involvement
NF-κB	Nuclear factor kappa B
OD	Optical Density
oxBS-seq	Oxidative bisulphite sequencing
PARPis	Poly (ADP-ribose) polymerase inhibitors
PARPis	Poly (ADP-ribose) polymerase inhibitors
PBS	Phosphate buffered saline
PD-L1	Programmed death-ligand 1
pKi67	Nuclear protein Ki67
POLR2A	RNA polymerase II subunit A
PR	Progesterone receptor
RNA	Ribonucleic acid
RPLPO	Hypoxanthine Ribosomal Protein Lateral Stalk Subunit P0
RT	Room temperature
RT-qPCR	Real-time polymerase chain reaction
S-2HG	(S)-2-hydroxyglutarate

SAM	S-adenosylmethionine
Scr	Cells transfected by scramble siRNA
SD	Standard deviation
siRNA	Small interfering RNA
T	Tumour
TAPS	TET-assisted pyridine borane sequencing
TDG	Thymine-DNA glycosylase
TET	Ten-eleven-translocation
$TET1^{FL}$	TET1 full-length
$TET2^{FL}$	TET2 full-length
$TET3^{FL}$	TET3 full-length
TMS1	Target of methylation-induced silencing
TNBC	Triple-negative breast cancer
TNM	Tumour node metastasis
TP53	Tumour Protein P53
tRNA	Transfer RNA
Unt	Untreated cells
UTX-1	Ubiquitously transcribed TPR on X
WNT	Wingless-Type MMTV Integration Site Family

WT1	Wilms’ tumour suppressor 1
Δ	Delta

1. Background

1.1 Breast cancer

1.1.1 Introduction and epidemiology

Breast cancer (BC) is currently the most commonly diagnosed type of cancer in women, surpassing lung cancer, and it has been estimated to cover 2.3 million new cases (11.7% of all cancer cases) worldwide in 2020 (1). Female BC is usually first in the incidence rate in most countries for cancer and for mortality in 110 countries, accounting for one in four cancer cases and one in six cancer deaths (1). Moreover, over 80% of female BC cases are diagnosed over the age of 50 in the UK (2). In contrast, only 0.5%-1% of breast cancer occurs in men (2). Thanks to the research into BC and the treatment opportunities, the survival rate of this disease has improved in the past decades; however, BC remains a significant issue globally.

1.1.2 BC Progression

Generally, the breast includes glandular tissue, fibrous stroma, and adipose tissue (Figure 1.1). The mammary gland is the functional part of the breast, consisting of 15 to 20 lobules, which produce milk, and a series of ducts that carry milk from lobules to nipples (3). Within the mammary lobule, the epithelium is formed by a single layer of luminal cells, surrounded by a layer of myoepithelial cells (Figure 1.2 Normal duct) (4). The term “myoepithelial” cell is also described as “basal” cells (5).

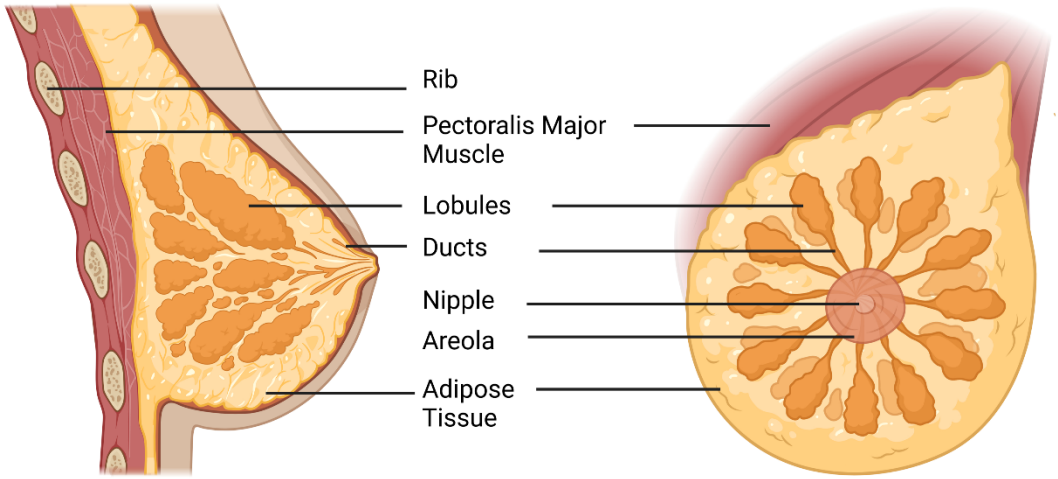


Figure 1.1 The structure of the breast. An anatomical schematic of the lobules and ductal system. Figure created with BioRender.com.

BC is a general term for a variety of malignant diseases in the breast, and its common feature is that it originates from the tissue, for example, ducts and lobules (6). These two subtypes comprise the majority of all breast cancers worldwide, with the ductal subtype accounting for 40-75% of all diagnosed cases (6). Moreover, BC can be classified consistent with the extent of disease (Figure 1.2): malignancies contained in the tissue of origin without any breach of the basal membrane are called carcinoma *in situ*, and those beyond are called invasive carcinoma (7). Epidemiologic evidence suggests that ductal carcinoma *in situ* (DCIS) is often initiated by ductal hyperplasia (DH) (8). Only a small subset of atypical ductal hyperplasia (ADH) and DCIS patients progressing to invasive ductal carcinoma (IDC) (8).

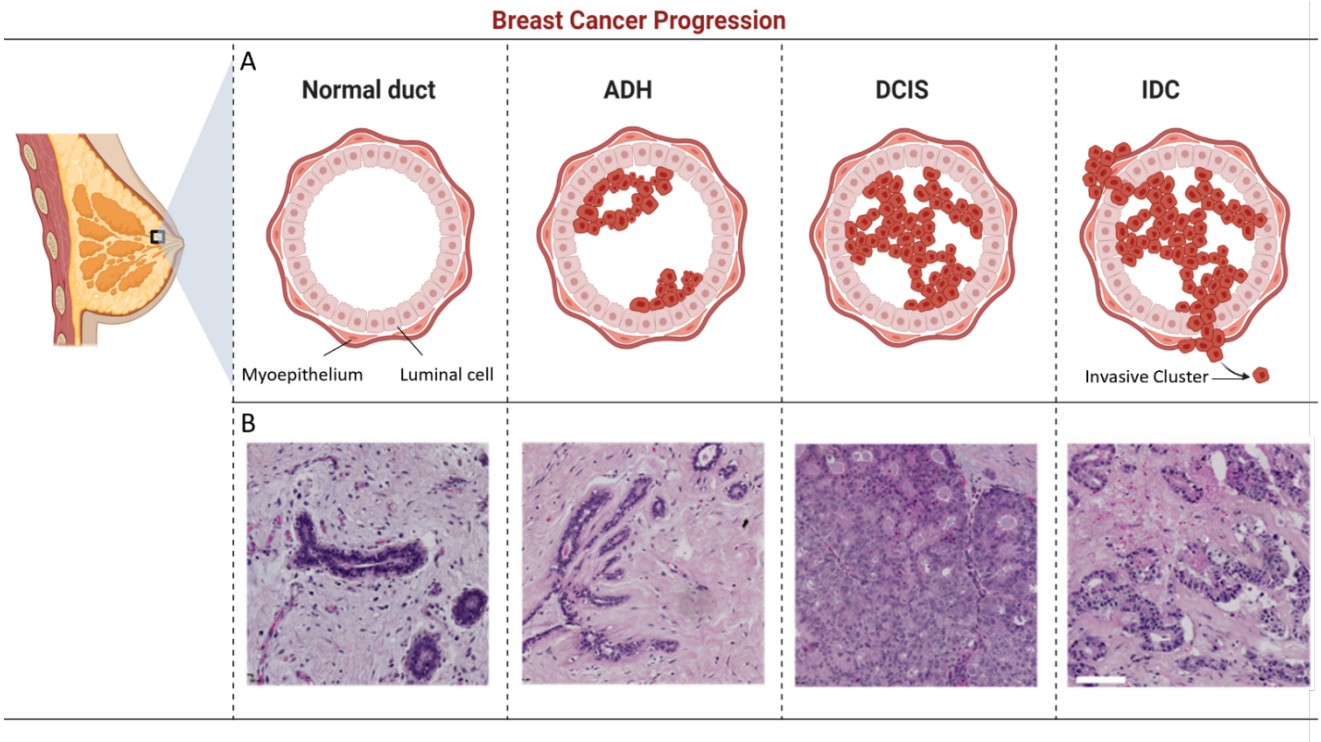


Figure 1.2 Example of breast cancer progression and the histopathology of DCIS. Panel A, Models of breast carcinoma progression in different stages. Panel B, H&E staining of breast tissue sections from the normal duct, ADH, DCIS and IDC (scale bar = 100 µm) (9). H&E Image reproduced with permission of the rights holder, MDPI. Figure created with BioRender.com.

1.1.3 Diversity of breast cancers

To improve the understanding of BC causation, prevention, diagnosis and treatment, it is helpful to classify BC into several subtypes according to the histopathological features and molecular characteristics. This subchapter introduces the main categories of BC (7).

1.1.3.1 Histopathological classification

Histopathology is a traditional way to classify the categories of BC based on its cellular structure and microanatomy of the tissue. According to the WHO classification of breast tumours published in 2019, breast tumour is further refined to classify up to 21 histological types based on cell morphology, growth, and architecture patterns (10-12). BC is a malignant disease under the category of breast tumour. Overall, invasive ductal carcinoma (IDC) is one of the most common types of breast cancer, which constitute 55% of BC upon diagnosis (13). This type of cancer often has a wide range of morphological variations and clinical behaviours, such as tumour size, and grade (13).

Histological grade is determined by the evaluation of the degree of tumour differentiation (tubule formation), nuclear polymorphism and proliferation (mitosis rate) (14). Each scheme is scored and counted together, and the total score represents the histological grade. Grade I means poorly proliferating and highly differentiated, grade II means moderately proliferating and differentiated, and grade III means highly proliferative and poorly differentiated (7). The grade provides a prediction for determining patient prognosis, where higher grade cancers have a worse prognosis (15).

Additionally, the other classification scheme is the stage of cancer based on “tumour node metastasis” (TNM) classification, determined by the size of the tumour (T), lymph node involvement (N) and metastasis (M) (16). According to the TNM staging system, BC are determined into 5 stages (17, 18). The earliest stage of BCs is stage 0, then ranges from stage I to stage IV (17). This TNM staging system can guide treatment and provide an indication of prognosis (19, 20).

1.1.3.2 Molecular classifications

Breast cancer is a heterogeneous disease with complex histological features, and it is commonly divided into different subgroups by the expression profile of the receptors: oestrogen receptor (*ER*), progesterone receptor (*PR*) and human epidermal growth factor receptor 2 (*HER2*) (14, 21). Most of the cases can be classified into four groups, luminal A, luminal B, HER2 positive (HER2+), basal-like/triple negative (Figure 1.3). Luminal A is ER/PR positive, or HER2 negative expressed in normal luminal epithelial cells, which has a good prognosis (13). Luminal B is ER and/or PR positive, HER2+ positive or negative expressed in luminal epithelial cells (13). The prognosis of luminal B is poorer than luminal A (13). HER2+ reflects Erb-B2 Receptor Tyrosine Kinase 2 (*ErbB2*) gene expression or HER2 protein amplification and overexpression, and ER, PR negative expression (13). This group of breast cancer is associated with a poorer prognosis. Moreover, “basal-like” or “triple-negative (TNBC)” reflects ER, PR, and HER2- negative expression in normal breast basal epithelial cells, which implies the worst prognosis (Figure 1.3) (13, 22).

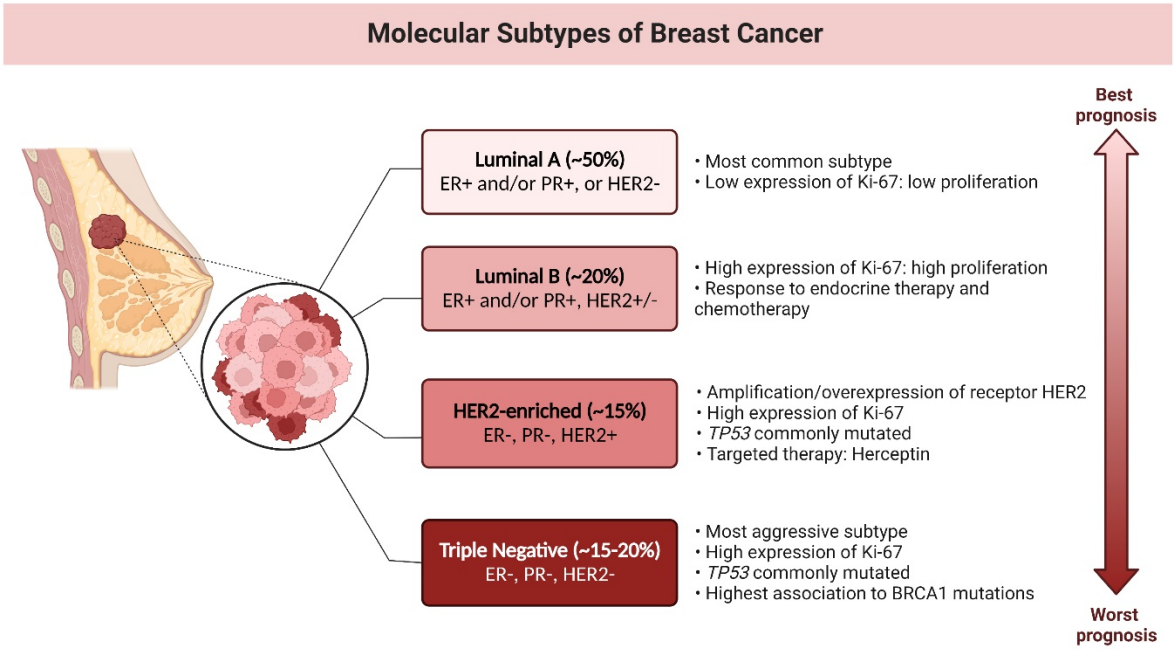


Figure 1.3 Molecular subtypes of breast cancer. Breast cancer can be summarised into four groups based on their hormonal receptor expression (ER, PR, and HER2+ expression). Subtypes from top to bottom have worse prognosis (Luminal A cancer has the best prognosis and the TNBC has the worst prognosis). The figure also includes the expression level of Ki67, the mutation level of TP53 and the suitable treatment options (13, 22). The Figure created with BioRender.com.

1.1.4 Breast cancer treatment

Nowadays, common types of BC treatment are surgery, radiotherapy, chemotherapy, hormone therapy, targeted therapy, and other systemic therapy (18, 23). The early stage of BCs are usually treated with surgery and radiation therapy, while higher stages of BC require other therapeutic options to treat patients and prolong their lives (24). Chemotherapy includes neoadjuvant chemotherapy and adjuvant chemotherapy (24). Neoadjuvant chemotherapy is used before surgery to down-size the tumour and/or reduce the risk of recurrence, and adjuvant chemotherapy is used after surgery (25). In triple-negative and HER2-positive early breast cancer, neoadjuvant therapy is typically used (26).

Hormone therapy, also known as endocrine therapy, is effective for most patients tested positive for ERs, and PRs (27-29). Hormone therapy aims to prevent interactions between oestrogen and oestrogen-dependant pathways of stimulating neoplastic cells through blocking the production of oestrogens and the oestrogens action upon cells (30). When hormone therapy is used alone or after chemotherapy, blocking this pathway can help prevent cancer recurrence and breast cancer death (23). Hormone-receptor-negative patients (TNBC or ER-, PR- and HER2-positive), do not respond to hormone therapy (29).

Besides, targeted therapy, also called precision medicine, uses drug or other substances to target specific molecules to prevent cancer cell to grow, divide and spread (31). Most targeted drugs are small-molecule drugs or monoclonal antibodies, some of which block the action of certain enzymes, proteins or other molecules involved in the growth of cancer cells, others of which help the immune system to kill cancer cells (32). HER2-targeted therapy has become an effective treatment for HER2 positive tumours. Trastuzumab was the first HER2-targeted therapeutic monoclonal antibody to treat metastatic HER2-positive breast cancer in 1998 (33).

1.1.5 Molecular markers in breast cancer

Apart from the expression of ER, PR and HER2, other biological markers predict a prognostic in breast cancer progression. The expression of biological markers is associated with the proliferation and migration of tumour cells. For example, the early stage of tumour development is increased proliferation, and the nuclear protein Ki67 (pKi67) and minichromosome maintenance protein complex 2 (MCM2) are considered to be biomarkers of proliferation in BC (7, 34, 35). Overexpression of KI67 is related to poor prognosis in breast cancer (Figure 1.3) (36, 37). Ki67 is expressed in all proliferating cells, reflecting the tumour's proliferation rate, and it has a different localisation and expression level throughout the cell-cycle (38). The expression of Ki-67 in luminal B BC and its histological grade is higher than in luminal A BC (13). On the other hand, MCM2 is a prognostic marker in BC, highly expressed in high histological grade cells, especially in luminal B, HER2-positive, and triple-negative BC (35, 39).

Based on the hallmarks of cancer, B-cell lymphoma 2 (Bcl-2) family members regulate apoptosis (*ie* programmed cell death) (40). Thus, dysregulation of Bcl-2 family proteins contributes to breast development, tumorigenesis, and therapeutic resistance. This family has two main subclasses: anti-apoptotic (i.e. BCL2 Apoptosis Regulator, Bcl-2, Bcl-w, and Bcl-xL) and pro-apoptotic (i.e. Bcl-2 Associated Agonist Of Cell Death, Bad, Bcl-2 Antagonist/Killer, Bak, and Bcl-2 Associated X-protein, Bax). The anti-apoptotic proteins prevent the regulation of pro-apoptotic proteins remaining the integrity of the mitochondrial outer membrane (41). Previous studies suggest overexpression of anti-apoptotic Bcl-2 proteins associated with BC

progression (42). For this reason, the Bcl-2 family protein could be a marker for identifying the evidence of apoptosis.

Furthermore, GATA binding protein 3 (GATA3) is identified as contributing to cell differentiation in multiple cell types and is considered an essential gene in BC development (43, 44). This positive correlation between GATA3 and ER-alpha expression is not unexpected, as they participate in a positive feedback loop (43). GATA3, as a transcription factor, directly regulates the expression of ER-alpha, thereby enhancing ER-alpha's activity within breast cancer cells. Conversely, ER-alpha, in turn, upregulates the expression of GATA3, forming a reinforcing feedback loop that further strengthens their association (43). This intricate interplay helps maintain the differentiated state of breast cells while also facilitating the normal functioning of ER-alpha. Moreover, GATA3 is highly expressed in luminal breast cells compared to the HER2 subtype and triple-negative BC (45-47). Therefore, GATA3 could be a luminal differentiation marker in breast cancer.

1.2 Epigenetic modifications

1.2.1 Methylation of cytosine

Epigenetic alterations have been rigorously investigated in the past few decades. The most common epigenetic modification is the methylation of cytosine, also known as “DNA methylation”, which adds a methyl group to the carbon-C5 position of cytosine, 5-methylcytosine (5mC). This modification is found in mammals, flowering plants, prokaryotes and in about 4% of the cytosines in the human genome (7, 48).

In humans, 5mC is rarely but globally distributed in the whole genome, which is predominantly occurs on cytosine-phosphate-guanine dinucleotide (CpGs) (49). Regions with

a high concentration of phosphate-linked cytosine-guanine pairs are called CpG islands (CGI) (50). While CGI are typically in promoters and unmethylated, other CpGs in the genome are often (60-80%) methylated (51). Through whole genome methylome profiles, 15%-21% of CpGs differ in the methylation status between human tissues (52). These CpG islands are specific DNA sequence regions enriched in CpGs and play crucial regulatory roles in organisms (53). They are commonly located in gene promoter regions and are essential for normal gene regulation (53). In contrast, CpGs in other parts of the genome can be subject to methylation, often associated with gene silencing or inactivity (54). This differential distribution of DNA methylation is instrumental in maintaining genomic stability and normal cellular functions, playing a pivotal role in gene regulation and epigenetics (54). DNA methylation contributes to gene expression, X-chromosome inactivation, genomic imprinting and tumorigenesis (55). 5mC is altered from unmodified cytosine during this process by DNA methyltransferases (DNMTs) (56).

1.2.2 The DNA methyltransferases (DNMTs)

DNA methyltransferases are a group of proteins, which contribute to DNA methylation. DNMTs catalyse the transfer of a methyl group ($-CH_3$) from S-adenosylmethionine (SAM) to the 5-carbon position of cytosine (57). This family consists of five members in humans, including DNMT1, DNMT2, DNMT3a, DNMT3b, and DNMT3l (58). Only three of them, DNMT1, DNMT3a, and DNMT3b, contribute to write and regulate methylation marks on the DNA (55, 58). In particular, DNMT1 is necessary to maintain the methylation state during DNA replication, whilst DNMT3a and DNMT3b are established for *de novo* methylation (55, 56, 58,

59). Apart from these, DNMT3l contributes to maternal imprints in the developing oocyte and acts as a cofactor for retrotransposon methylation in the male germline (56, 58, 60).

1.2.3 Hydroxymethylation of cytosine

DNA demethylation can act both “passively” and “actively” (Figure 1.4). Passive demethylation refers to the dilution of the methylated CpG sites during DNA replication (61). The methylated DNA double-strand are parental patterns. During DNA replication, newly incorporated cytosines remain unmethylated due to DNMT1 inhibition or dysfunction, thereby reducing overall methylation levels after each cell division (62). Therefore, both methylated and unmethylated cytosines on the parent chain pair with guanine, guanine on the parent chain pairs with unmethylated cytosine, and adenine pairs with thymine. After each round, the newly replicated daughter DNA strands are unmethylated, and methylated CpG sites become hemimethylated (63).

The active demethylation is achieved by the enzyme-mediated removal process. Methylated cytosine (5mC) is oxidised to hydroxymethylated cytosine (5-hydroxymethylation, 5hmC) and other derivatives (5-formylcytosine, 5fC and 5-carboxylcytosine, 5caC) (64). The reaction is catalysed by ten-eleven-translocation (TET) enzymes (64). These sites will then be depleted by thymine-DNA glycosylase (TDG) through the base excision repair (BER) pathways (63, 64).

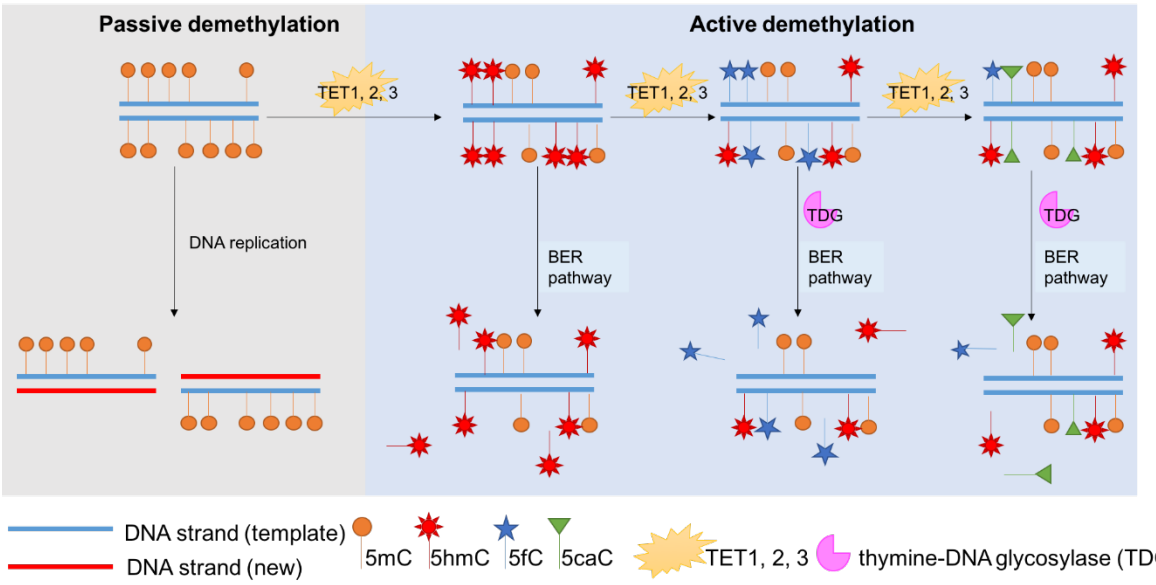


Figure 1.4 Overview of the mechanisms of DNA demethylation (63). DNA demethylation is achieved by passive demethylation (left) and active demethylation (right). The methylated strands are separated and replicated with newly unmethylated DNA strand through passive demethylation. The methylated CpG sites are oxidised to 5hmC, 5fC, 5caC by TET1, 2, and 3 and removed through the BER pathway. Adapted from "Handbook of epigenetics (Second Edition)," by Nicolas Veland, Taiping Chen, 2017, p12. (49). Copyright 2017 Elsevier Inc. Adapted with permission.

During active demethylation process (Figure 1.5), the three ten-eleven translocation (TET) enzyme family members, TET1, TET2 and TET3, together with cofactors (iron, α-ketoglutarate and oxygen), catalytically convert 5mC to 5hmC, 5fC and 5caC by oxidation. TET proteins are Fe(II)-dependent dioxygenases that oxidize DNA using molecular oxygen and α-ketoglutarate as co-substrates, generating oxidized DNA, succinate, and carbon dioxide (CO_2) as the co-products. Briefly, TET enzymes split molecular oxygen into oxygen atoms (with two unpaired electrons and a carbene-like propensity to intervene in other bonds). Insertion of one oxygen atom into succinic semialdehyde (formal product of decarboxylation of α-ketoglutaric acid) to form succinic acid and another oxygen atom shifted into the 5-position substituent of the

C-group converts 5mC to 5hmC, 5hmC to 5fC (due to equilibrium between jewel diol and aldehyde via dehydration/hydration), 5fC to 5caC (65).

Through the BER pathway, 5fC and 5caC can be removed from the sequence and replaced with cytosine by the TDG enzyme. TDG is a monofunctional DNA glycosylase active against T bases generated by deamination of 5mC with G mismatches (66). The 5fC and 5caC nucleotides have the same characteristics as damaged DNA nucleotides, and 5fC and 5caC nucleotides resemble BER substrates (65). Since the formyl and carboxyl groups are electron-withdrawing groups, 5fC and 5caC nucleotides have weakened glycosidic bonds (67, 68). 5caC:G is more weakly base-paired than C:G, which may lead to a higher rate of spontaneous flipping, like many damaged DNA bases (68, 69). Weakened glycosidic bonds and increased flipping rates allow BER glycosylases to erase 5fC and 5caC (65).

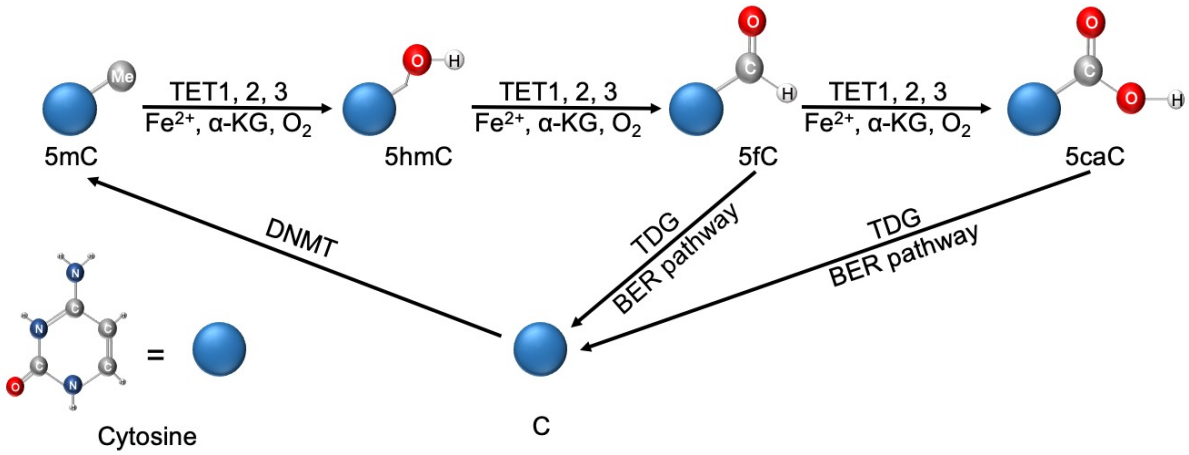


Figure 1.5 The molecular structure changes during the active DNA demethylation pathways through 5hmC. DNMT catalyses the cytosine to 5mC, and several 5mC are iteratively oxidised to 5hmC, 5fC and 5caC by TET enzymes together with Fe2+, α-KG, and O2. 5fC and 5caC can be further excised by the TDG via the BER pathway. The structure of cytosine is simplified as a blue ball. C, cytosine; 5mC, 5-methylcytosine; 5hmC, 5-hydroxymethylcytosine; 5-fC, 5-formylcytosine; 5-caC, 5-carboxylcytosine; DNMT, DNA methyltransferase; BER, base excision repair; TDG, thymine DNA glycosylase.

A derivative of 5mC, 5-hydroxymethylation (5hmC), is the first oxidation step in the multistep demethylation process, as described above (70). DNA hydroxymethylation is considered a marker of active gene expression, but the correlation between 5hmC and expression is varied (71). While 5mC generally exerts its function on CpG rich sequences in the gene promoter, 5hmC is mostly distributed in gene bodies, enhancers, and in sequences with lower CpG contents. This leads to the hypothesis that 5hmC could modify chromatin states to promote transcription, rather than silencing gene expression (72-75).

The levels of 5hmC differ in different tissues and cells. Whilst 5hmC is enriched in brain, liver, kidney and colorectal tissues (0.40% ~0.65% of total nucleotides), it is only quantified as 0.05% ~0.06% of total DNA in heart, breast and placenta tissues in human (76). In human cell lines, the percentage of 5hmC is even lower (0.007% ~0.009% of total nucleotides) (72). While 5hmC was first discovered in bacteriophages T2, T4 and T6 of E. *coli*, it is now considered a dynamic epigenetic change that can influence the gene expression in mammalian cells (77).

In the brain, hydroxymethylation has been implicated in several types of diseases, such as Alzheimer's (78, 79) and Parkinson's disease (80). In human embryonic stem cells (ESCs), the enrichment of 5hmC in gene bodies, promoters, enhancers, and some transcription factor binding sites suggests an active involvement in gene regulation (81, 82). Finally, the alterations of 5hmC levels have also been associated with B cells and T cells differentiation and development (83), and evidence suggests that TET-mediated demethylation process regulates the differentiation and function of lymphoid and myeloid cells (83).

1.2.3 Post-transcriptional RNA modification and RNA hydroxymethylation

Post-transcriptional modifications of RNA play an essential role in a wide variety of biological processes (84). RNA modification has been studied since the 1960s, and sequencing of the first biological RNA in 1965 from yeast (85). RNA modification on rRNA can lead to the formation of specific RNA structures and regulate RNA biogenesis; RNA modification on tRNA can refine molecular recognition; and RNA modifications on mRNA can encode dynamic regulatory features (86).

Methylation is the most common modification. Many different classes of RNA can be methylated using S-adenosylmethionine (SAM) as a methyl donor, for instance, mRNA, tRNA, rRNA, lncRNA, miRNA, small nuclear RNA (snRNA) and enhancer RNA (eRNA) (86, 87). Studies found that eukaryotic mRNA undergoes internal modifications, including additional methylation of adenosine to form N1-methyladenosine (m1A) and N6,2'-O-dimethyladenosine (m6Am), and the methylation of cytosine to form 5mC and 5hmC (85).

RNA methylation is involved in RNA metabolism pathways (88, 89). In the nucleus, pre-mRNAs undergo post-transcriptional modifications, such as capping, splicing, and polyadenylation (90). At this stage, many nuclear proteins are recruited to the mRNA to form RNA-protein complexes, and only a successfully processed mRNA forms a proper mRNA-protein complex to translocate to the cytoplasm through the nuclear pore complex (90). Some products, such as spliced introns, will be degraded by multi-subunit exosome complexes that may specialize in degrading RNA, while others, such as methylated tRNA and methylated rRNA, will move to the cytoplasm through nuclear pores (89). In the cytoplasm, the mature mRNA combines the rRNA and tRNA for protein synthesis (89, 90). The extent of 5mC in mRNA is still controversial,

while the methyltransferase NOL1/NOP2/SUN domain family 2(NSUN2) is suggested to responsible for installing 5mC on mRNA (89).

Like methylation patterns, hydroxymethylation affects DNA and RNA molecules, from wheat seedlings (91) to mammalian tissues (92-94). Similar to the demethylation process in DNA, 5mC in RNA can be oxidised to 5hmC by TET enzymes (85). The presence of 5hmC is also a feature of many classes of RNA, including tRNA in human cells and mouse brain (95), long non-coding RNA (lncRNA) in colorectal cancers in humans (96), and mRNA in Drosophila S2 cells (97). Therefore, the presence of 5hmC in the DNA and different RNA classes may support a possible regulatory role of hydroxymethylation in transcription and translation.

1.2.4 TET enzymes

TET family proteins, also known as TET methylcytosine dioxygenase, are a group of enzymes with similar structures and overlapping functions. In particular, TET proteins show similarity in their C-terminal catalytic domains but are characterised by differences in the N-terminal regulatory domains, ultimately reflecting some individual functions (Figure 1.6) (98, 99). On top of this, TET enzymes are characterised by a high number of alternative spliced variants, which can also contribute to the novel functions of each enzyme.

To date, twelve TET1 isoforms, twelve TET2 isoforms and fourteen TET3 isoforms are recorded in the NCBI database, out of which six major isoforms have been well studied, including TET1 full-length ($TET1^{FL}$), TET1 short ($TET1^{ALT}$), TET2, TET3 full-length ($TET3^{FL}$), TET3 ovarian (100). While these isoforms mainly appear in ESCs and neurons, here, I will focus on the isoforms found in cancer cells, $TET1^{FL}$, $TET1^{ALT}$, TET2, and $TET3^{FL}$.

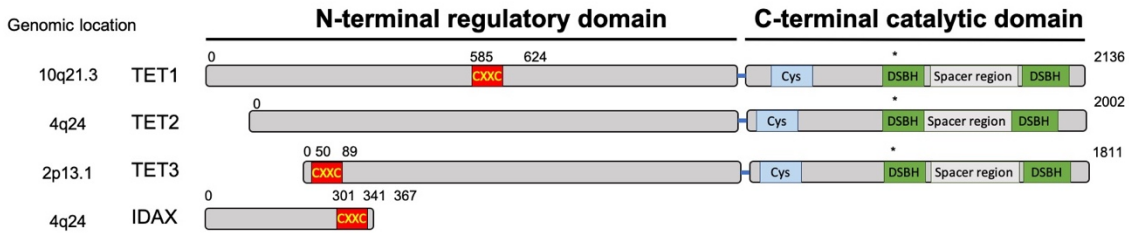


Figure 1.6 Known domain structures of TET proteins. TET proteins are similar in structure, with the C-terminal catalytic domains and different N-terminal regulatory domains. The full-length TET1 ($TET1^{FL}$) and TET3 ($TET3^{FL}$) proteins, but not TET2, contain a CXXC domain (DNA-binding site). The genomic location and the number of amino acids of the full-length protein and the CXXC domain are also shown. The isoform $TET1^{ALT}$ lacks the CXXC domain, but the remaining structure has not been characterised yet. The catalytic domain contains cysteine-rich (Cys), double-strand β-helix (DSBH) regions and spacer region. The DSBH domain includes a HxD motif that interacts with Fe^{2+} and α-KG.

Structurally, both $TET1^{FL}$ and $TET3^{FL}$ but not their truncated isoforms, contain the CXXC DNA-binding domain (101). The CXXC domain binds to unmethylated CpG islands, assisting TET recruitment to the same and protecting the DNA from aberrant methylation. On the other hand, the truncated isoforms have different regulatory domains, and they may have different binding sites to oxidase 5mC, 5hmC and 5fC (100). It is interesting to note how full-length TET2 ($TET2^{FL}$), on the contrary of TET1 and TET3, is also characterised by the absence of the CXXC domain but can interact with the neighbouring gene CXXC finger protein 4 (CXXC4), also known as IDAX, to bind to the DNA sequence (100).

1.2.5 Other oxidative derivatives of 5hmC

Several studies highlighted the contribution of both 5fC and 5caC to some human diseases and cancers, but these are rare changes. The global level of 5fC and 5caC over the total level

of all cytosines is around 10-1000 times less than 5hmC (102, 103). Not many studies focus on the function of 5fC and 5caC in cancer. One study found that 5caC levels are higher in BC than normal breast, which samples produced a low intensity of 5hmC staining (104).

1.3 Epigenetic changes in cancer

Methylated cytosine is commonly distributed in DNA, and it has been found in numerous RNA subtypes, including tRNA, rRNA, mRNA, ncRNA and enhance RNA (105). Since 5mC is a heritable change in gene expression, it is important to learn more about the role of methylation in genes regulating in cancer. Methylation at CpG enriched genes results in hypermethylation of tumour suppressor genes or hypomethylation of oncogenes, marks of cancer initiation and progression (8). On the one hand, 5mC is generally dysregulated in cancer (72, 75, 106-108). Hypomethylation occurs in the global DNA, while hypermethylation occurs in local CpG islands (109, 110). DNA hypomethylation play a causal role in tumour formation, possibly by promoting chromosomal instability (111). Hypermethylation can drive the critical tumour suppressor or silence regulatory regions in the genome, leading to dysregulation of cell growth or changes in response to cancer treatments (109). In BC, promoter hypermethylation causes the silencing of *breast cancer type 1 (BRCA1)*, *E-cadherin* and *target of methylation-induced silencing (TMS1)*, and the overexpression of *the Wilms' tumour suppressor 1 (WT1)* genes (112). On the other hand, several studies have demonstrated that 5mC modified in different RNA subtypes might have functions in cancer development (105). For example, in tRNA, 5mC may function in modulating cancer metabolism, and in mRNA and lncRNA, 5mC increases susceptibility to cancer which indicates a possible role in enhancing the stability of cancer-related transcripts (105, 113-115).

5-Hydroxymethylation occurs in gene bodies, promoters, and enhancers (72, 75, 116), suggesting that the synthesis of 5hmC can serve additional functions in the genome. As mentioned above, 5hmC is also enriched in different classes of RNA. An interesting example is the distribution of 5hmC at lncRNA loci, which positively correlates with lncRNA transcription in colorectal cancers (96). As discussed further below, evidence suggests that 5hmC may regulate gene transcription and translation, and ribosome regulation.

Interestingly, although 5hmC levels are generally lower in multiple human malignancies than in the normal tissues (117-119), this is not the case for all tumours. An example showed 5hmC also can be gained in genes in melanoma (120). However, the extent of enrichment of 5hmC in different cancers is still unknown and global 5hmC content does not seem to correlate with 5mC (76, 121, 122), highlighting the need for a more complete characterisation of 5hmC in cancer.

Therefore, analysis of 5hmC distribution in tissues or cells, outside of a disease context, may not be enough to help us determine its role in cancer onset and development. Since TET proteins act as demethylases in generating 5hmC in cancer, it is essential to review 5hmC expression concerning the distinct functions of each TET, which are directly responsible for modulating methylation signatures in their unique ways.

1.4 Ten-eleven translocation protein in cancer

1.4.1 TET1

Evidence suggests TET1 may have both oncogenic and tumour suppressor functions, and it covers several functions in cancer progression (Table 1.1). As a tumour suppressor, TET1 represses cell migration, invasion, proliferation and metastasis in breast, prostate, colorectal,

pancreatic, nasopharyngeal, lung, and renal carcinoma (123-130). In breast cancer, the high mobility group AT-hook 2 (HMGA2) leads to TET1 suppression and cause further the decreased expression of homeobox A (*HOXA*) genes, *HOXA7* and *HOXA9*, that stimulate tumour growth and metastasis (123). The expression of TET1 is downregulated in prostate cancer which induces cell invasion, tumour growth and cancer metastasis (124) and loss of *TET1* in colon cancer contributes to tumour cell growth through mediating the WNT signalling pathway (130). Like colon cancer, TET1 is also downregulated in pancreatic and nasopharyngeal carcinoma, and TET1 suppress the Wnt/β-catenin signalling pathway (126, 127). In renal carcinoma, low TET1 expression is associated with high proliferation and migration *in vitro* (129). Furthermore, loss of TET1 increases Enhancer Of Zeste 2 Polycomb Repressive Complex (EZH2) expression and reduces Ubiquitously transcribed TPR on X (UTX-1) expression, which decreases the 5hmC abundance on the promoter of E-cadherin (125). The resulting down-regulation of E-cadherin induces epithelial-mesenchymal transition (EMT) in colorectal cancer (125). On the other hand, in lung cancer cells, overexpression of TET1 results in the repression of metastasis *in vivo* via the regulation of cell-cell adhesion, and prevents tumour cell invasion ability by reducing the capacity of matrix metalloproteinase 9 (MMP9) (128). As an oncogene, high *TET1* expression results in the hypomethylation and activation of an oncogenic signalling pathway in triple-negative breast cancer (TNBC) (131) and up-regulates cell proliferation, inhibits apoptosis and cell differentiation in MLL - rearranged leukaemia (132). Finally, NF-κB family member p65 represses *TET1* expression in breast cancers by binding to its promoter in basal-like BC (133). The expression of mutated *TET1* could therefore be a prognostic biomarker for immune checkpoint inhibitor treatment in multiple cancer (134).

Alike TET2 and TET3, the variability in the functions of TET1 may also be linked to the distinct roles of its isoforms (101). Evidence suggests that $TET1^{ALT}$ is overexpressed in cancers, but it leads to a lower synthesis of 5hmC on DNA compared to $TET1^{FL}$ (101). The presence of $TET1^{ALT}$ isoform, characterised by different methylation and gene expression targets compared to the full-length variant, suggests TET1 and possibly other TETs could also change their oncogenic function by modulation of the expression of its isoforms.

Table 1.1 TET proteins: known functions and research studies.

Protein	Tumour type	Model used	Roles of TET protein	Techniques to quantify 5hmC	Ref
TET1	Breast cancer	- Human breast cancer cell lines - Mouse models	Tumour suppressor	Dot blot, IHC	(123)
	Breast cancer, Prostate cancer	- Human breast / prostate cancer cell lines -Human breast / prostate tissue samples - Mouse models	Tumour suppressor	Dot blot, MeDIP	(124)
	Colorectal cancer	-Human colon cancer cell lines	Tumour suppressor	GlucMS-qPCR analysis, IF	(125)
	Colorectal cancer	-Human colon cancer cell lines - Mouse models	Tumour suppressor	Dot blot, hMeDIP	(130)
	Pancreatic cancer	-Human pancreatic cancer cell lines	Tumour suppressor	Dot blot, GlucMS-qPCR	(126)

		-Human pancreatic tissue samples -Mouse models			
	Nasopharyng-eal cancer	-Human nasopharyngeal cancer cell lines -Human nasopharynx tissue samples	Tumour suppressor	Dot blot, IHC, MeDIP and hMeDIP	(127)
	Lung cancer	-Human lung cancer cell lines - Mouse models	Tumour suppressor	NA	(128)
	Renal cancer	-Human renal cancer cell lines -Human kidney tissue samples	Tumour suppressor	NA	(135)
	Acute myeloid leukaemia	-Human AML cell lines -Human AML samples - Mouse models	Oncogene	5hmC Labelling reaction, Dot blot, LC-MS/MS assay, IHC	(132)
	Breast cancer	-Human breast cancer cell lines	Oncogene	Slot blot analysis,	(131)

	Trophoblastic cancer	-Human trophoblastic tissue samples -Human trophoblastic cancer cell line	Oncogene	NA	(136)
TET2	Breast cancer	-Human breast cancer cell lines	Transcriptional coactivator	NA	(137)
	Breast cancer, Melanoma, Liver cancer, Lung cancer	-Human tumour cell lines -Mouse tumour cell lines -Mouse model	Tumour suppressor	RRHP sequencing	(138)
	Melanoma, Glioblastoma Colorectal cancer	-Human melanoma / glioblastoma / colorectal cancer cell lines -Human melanoma, glioblastoma, colorectal cancer and other cancers' tissue samples -Mouse model	-Oncogene -A drug target for SCCC elimination	IF, IHC, Tissue microarray, Dot blot,	(139)

	Acute myeloid leukaemia, Colon cancer	-Human AML cell line -Human colon cancer tissue samples -Mouse colon cancer cell line	Predictor for anti-PD-L1 immunotherapy	IHC, Dot blot	(140)
TET3	Head and neck cancer	-Human head and neck squamous cell cancer tissue samples -Human head and neck cancer and fibroblast cell lines	-Tumour suppressor -Predictor for recurrence in early-stage head and neck cancer	NA	(141)
TET2/ TET3	Melanoma	-Human melanoma cell lines -Mouse model	Tumour suppressor	Dot blot	(142)
TET2/ TET3	Pancreatic cancer	-Human pancreatic cancer cell lines -Mouse model	Oncogene	LC-MS, qPCR, hmC-seal, OxBS-seq and BS-seq, RNA-seq, ATAC-seq, HOMER	(143)

TET1/ TET2/ TET3	Early breast cancer	-Human early breast cancer tissue samples -Human glioblastoma cell line	Predictor for patients' prognosis	NA	(144)
TET1/ TET2/ TET3	Gastric cancer	-Human gastric cancer tissue samples -Human gastric cancer cell lines	Oncogene	IHC, Dot blot	(145)

1.4.2 TET2

Overall, most studies agree in suggesting that TET2 acts as a tumour suppressor (Table 1.1). Loss of TET2 might sensitise myeloid leukaemia cells to poly (ADP-ribose) polymerase inhibitors (PARPis) through inhibiting its partner proteins' function (such as BRCAs) in double-strand breaks DNA repair (146). The enzyme is a transcriptional coactivator, and its decrease leads to increased DNA methylation on enhancers of breast cancer cells (137).

Furthermore, in mouse models, TET2 regulates the 5hmC status on cell-cycle inhibitor genes *cyclin dependent kinase inhibitor 1A, Cdkn1a,* and *Cdkn1b*, resulting in an upregulation of p21 and p27, which leads to melanoma dormancy (138). Interestingly, also in colorectal cancer, the increase of TET2 expression and 5hmC has shown clinical relevance. TET2 is a critical factor in controlling the number of slow-cycling cancer cells, survival and tumour recurrence in mice model and colorectal cancer patients with higher expression of the protein relapse earlier after surgery and chemotherapy (139). Immunology studies also show that TET2 protein regulates IFN-γ/JAK/STAT signalling pathway which controls programmed death-ligand 1 (PD-L1) expression and the tumour response to anti-PD-L1/PD1 therapy, so the activity of TET2 could be a predictor for the efficacy of anti-PD-L1 immunotherapy (140).

TET2 consists of a full-length and eleven alternatively spliced isoforms, and the expression of some of those have already been detected in different tissues (100). Similar to the TET1 enzymes, TET2 is also frequently mutated in hematopoietic cancers compared to solid tumours (147-149). A recent study reported that TET2 mutation was correlated with a poor prognosis in TET2-mutated female cancers, including BC (147). This suggests that TET2 and its mutations may play important roles in BC.

1.4.3 TET3

While fewer studies focus on TET3, possible roles of this enzyme in cancer are still emerging (Table 1.1). TET3 act as a tumour suppressor in melanoma and its silencing induces EMT (142). Interestingly, TET3 displays a different expression level in various cancers: in head and neck squamous cell carcinomas and gliomas, expression of TET3 is low, while in ovarian and breast cancers, TET3 expression is higher, compared to normal tissue (141, 144, 150, 151). Clinically, these differences in expression are reflected in the fact that high TET3 expression predicts poor prognosis in ovarian cancer (150). This may be due to hormonal regulation, such as oestrogen and progestogen, but no conclusive research suggests a correlation between oestrogen or progestogen and TET3 in cancer.

Additionally, glioblastoma and breast cancer patients who have methylated TET3 have a lower survival rate compared to patients having unmethylated TET3 gene (144, 151). Finally, TET3 and bromodomain-containing protein 4 (BRD4) are overexpressed in pancreatic cancer (143). The 5hmC abundant at BRD4 enhancers and promoters regulates tumour growth, though there is no further study showed the relationship between TET3 overexpression and the increase of 5hmC on BRD4 (143).

Besides the unique expression of each TET enzyme, it is interesting to note how the levels of *TET* mRNAs can also be altered in several cancers without a respective difference in protein expression (145) and how *TET* mRNA expression is regulated by microRNAs (miRs). For instance, it has been shown that microRNAs, such as miR-22 in breast cancer or miR-29a in hepatocellular carcinoma, can in fact target *TET* mRNA, and subsequently reduce 5hmC levels (152, 153).

1.5 Hypothesis and Aims

Based on the studies mentioned above, 5hmC may have a functional role in BC progression. As the intermediates during the demethylation process, previous work pointed out that three TET proteins play an essential role in 5hmC synthesis. However, the roles of TET enzymes in cancer are controversial, so each TET protein respectively could play a different role in BC progression. Therefore, this research will validate the 5hmC levels in breast cancer cell lines and tissues. Moreover, to understand the molecular changes in hydroxymethylation, I will examine the 5hmC status of some breast cancer marker genes (i.e. *KI67, MCM2, BCL-2, BCL-W, ESR1, GATA3*) and the influence of hormone receptors on the TET protein family.

Hence the main hypotheses of this project are:

- There is a differential distribution of 5hmC changes between normal breast and breast cancer cells.
- The level of 5hmC is associated with a specific TET protein in BC. TET2 expression is directly proportional to hydroxymethylation levels.
- Dysregulation of TET expression changes the levels of 5hmC and regulations in breast cells.

To test my hypotheses, I, therefore, aim to:

1. Investigate the differential expression of 5hmC abundance and TET expression levels in normal and breast cancer samples. I aim to perform immunohistochemistry to quantify 5hmC and TET protein expression in breast tissues Moreover, immunocytochemistry was performed to determine the subcellular localisation of 5hmC and TET proteins in breast cell lines. I intend to determine the global level of

5hmC in breast cell lines by a colorimetric assay. (Chapter 3 for 5hmC and chapter 4 for TET)

2. Correlate the differential expression of *TET1*, *TET2* and *TET3* mRNA, and the abundance of 5hmC in breast cell lines. I aim to correlate the expression of three *TET* mRNAs with the probability of BC survival rate. Furthermore, I aim to investigate the *TET* mRNA expression between BC and normal samples through *in situ* hybridisation. (Chapter 4)
3. Investigate the possible role of TET proteins in breast cell lines. I aim to knockdown the *TET* mRNA expression in BC cells and non-tumorigenic breast cells and then to determine the global 5hmC changes, to measure changes of cell proliferations and to quantify the mRNA expression of selected marker genes. (Chapter 5)

2. Materials and methods

This chapter describes the materials and methods used. To answer my research questions, I perform the following experiments into two research objects, human tissue samples and human cell lines. This chapter first introduces the patient samples and the relevant experiments performed on tissue samples and then introduces the panel of cell lines and the experiments performed on cell lines.

2.1 Patient Samples

Formalin-fixed paraffin-embedded (FFPE) tissue sections were acquired from the pathological archives of Imperial College Tissue bank and St. Mark's Hospital pathology department, and a histopathologist selected the sections. Four different cohorts of tissue samples (Table 2.1, 2.2, 2.3 and 2.4) were used:

1) Eighteen FFPE of invasive ductal carcinoma (IDC) from 9 post-menopausal patients between 40 to 68 and their matched contralateral normal breast were collected from Imperial College Tissue Bank (REC number: 12/WA/0196).
2) Forty FFPE sections of IDC from 20 patients between 47 to 86 and their matched contralateral normal breast were collected from Imperial College Tissue Bank (REC number: 12/WA/0196)
3) Twenty FFPE samples from 12 pre-menopausal patients between 28 to 37 who have undergone breast surgery were selected by Prof Sami Shousha from the archives of Imperial College Tissue Bank (REC number: 12/WA/0196). Samples were divided into

two age-matched groups, consisting of eleven benign (adenomas) and nine cancer (IDC) samples.

4) Twenty-eight FFPE samples from 7 patients between 30 to 80 who have undergone colitis-associated cancer (CAC) were collected from St. Mark's Hospital pathology department (ethical approval from London North West NHS trust Research ref: RD/16/084). Samples included quiescent (no active inflammation, normal), dysplastic and cancer tissue.

Table 2.1 Details of the FFPE sections of IDC tissue received from Imperial College Tissue Bank.

Sample Number	Age at resection	Hormonal receptor status	Histological details
1	68	ER+ PR+ HER-	Grade 2 IDC
2	64	ER+ PR+ HER-	Grade 3 IDC with high-grade DCIS
3	40	ER+ PR+ HER-	Grade 3 IDC with intermediate-grade DCIS
4	49	ER- PR- HER+	Grade 3 IDC
5	50	ER- PR- HER2-	Grade 2 IDC
6	62	ER- PR- HER2-	Grade 3 IDC
7	41	ER- PR- HER2+	Grade 3 IDC with high-grade DCIS
8	66	ER- PR- HER+	Grade 3 IDC
9	55	ER- PR- HER-	Grade 2 IDC

Table 2.2 Details of the FFPE sections of IDC tissue received from Imperial College Tissue Bank.

Sample Number	Age at resection	Hormonal receptor status	Histological details
1	47	ER+ PR+ HER2+	3 Foci of grade 2 IDC
2	47	ER+ PR+ HER2+	Grade 3 IDC
3	52	ER+ PR+ HER2+	Grade 3 IDC
4	63	ER+ PR+ HER2+	Grade 2 mucinous carcinoma, adjacent intermediate-grade DCIS
5	59	ER- PR- HER2+	Grade 2 invasive mucinous carcinoma
6	68	ER- PR- HER2+	Grade 3 IDC
7	55	ER- PR- HER2+	Grade 2 IDC
8	76	ER- PR- HER2+	Grade 3 IDC, adjacent high-grade DCIS
9	71	ER- PR- HER2-	Grade 3 IDC
10	86	ER- PR- HER2-	Grade 3 IDC
11	75	ER- PR- HER2-	Grade 3 IDC
12	54	ER- PR- HER2-	Grade 3 IDC
13	74	ER+ PR+ HER2+	Grade 2 IDC
14	59	ER+ PR- HER2-	Grade 2 IDC

15	51	ER+ PR+ HER2+	Grade 2 IDC
16	51	ER+ PR+ HER2-	2 Foci of IDC
17	63	ER+ PR+ HER2+	2 Foci of grade 3 IDC
18	47	ER+ PR+ HER2+	Multifocal grade 2 IDC (x3); integral intermediate DCIS
19	68	ER- PR- HER2-	Grade 3 IDC
20	55	ER- PR- HER2-	Grade 3 IDC

Table 2.3 Details of the FFPE sections of breast tissue received from Imperial College Tissue Bank.

Sample number	Age at Diagnosis	Histological details
1	33	Grade 3 IDC
2	30	Grade 3 IDC
3	37	Grade 2 IDC
4	41	Grade 3 IDC
5	40	Grade 2 IDC with lymph node spread and DCIS
6	36	Grade 2 IDC
7	32	Invasive poorly differentiated lobular breast carcinoma
8	37	Grade 3 IDC with a large central necrotic area
9	37	IDC
10	36	Lactational change (B1)
11	32	Lactational change (B2)
12	28	Lactational change and fibrosis (B1)
13	31	Lactational change with focal lymphocytic infiltration (B1)
14	33	Lactational breast with focal fibrocystic and columnar cell change
15	32	Lactational adenoma

16	42	Lactational adenoma
17	36	Lactational adenoma
18	35	Infarcted lactational adenoma
19	33	Lactational induced change
20	36	Lactational adenoma

Table 2.4 Details of the FFPE sections of CAC tissue received from St. Mark's Hospital.

Sample number	Sex	Age at resection	Histological details
1	Female	53	• 3 cancer blocks • 3 dysplasia blocks • 1 quiescent block
2	Female	52	• 1 cancer block • 1 dysplasia block • 3 quiescent blocks
3	Male	44	• 1 cancer blocks • 1 dysplasia block • 1 quiescent block
4	Male	72	• 2 quiescent blocks
5	Male	37	• 1 cancer block • 1 quiescent block
6	Male	30	• 1 cancer block • 2 dysplasia blocks • 3 quiescent blocks
7	Male	49	• 1 cancer blocks • 1 dysplasia blocks • 2 quiescent blocks

2.2 Immunohistochemistry (IHC)

As mentioned above, a global loss of 5hmC was reported in BC tissue. To prove previous studies and understand where the demethylation markers work in human cells, IHC was performed to investigate the abundance and localisations of 5hmC and 5mC (result subchapter 3.1.1 and 3.1.4), and TET proteins (result subchapter 4.1.1) in human tissue samples.

2.2.1 Optimisation for the protocols and antibodies

Optimisation of the antibodies targeting 5hmC, 5mC, 5fC, 5caC, TET1, TET2 and TET3 was performed on human breast IDC tissue. Dr Lisa Del Bel Belluz previously performed the optimisation of the anti-5hmC antibody. I performed the optimisation of the anti-TET2 antibody together with previous MRes student Liliane Eleid, and of the other antibodies. After optimisation, a final concentration selected shows as below (Table 2.5), as the staining of the antibody in the cells could be detected with minimal background. However, both 5fC and 5caC antibodies gave negative staining with a final concentration of 1:100 for both two antibodies. Due to the availability of antibodies and budget constraints, only one antibody was used to test each antigen. This decision was based on limited experimental resources and financial constraints. The antibodies were used well validated in previous experiments and were already widely used in the laboratory. Characterization of the positive control samples was based on past research experience in our laboratory, which demonstrated the reliable properties of these positive control samples under our study conditions. Therefore, using positive control samples can effectively verify our experimental results. In addition, to ensure the accuracy of the experiments, the primary isotype antibody was used as a control for

negative staining in all the experiments, including both IHC and immunocytochemistry (ICC). This method helps eliminate possible false positive results and ensures the reliability of our experimental results. The positive controls were well-characterized based on previous experience in the laboratory. Samples were selected based on the research from past colleague Mr Pramodh Chandrasinghe, which has validated the robustness of the experimental procedures.

Table 2.5 Details of primary antibodies used in IHC and ICC.

Antibody	Catalogue number	Host species	Dilution factor for IHC	Dilution factor for ICC
5mC	CST28692	Rabbit	1:1000	1:200
5hmC	Active motif 39769	Rabbit	1:5000	1:200
5fC	CST 74178S	Rabbit	1:100	1:100
5caC	CST 36836	Rabbit	1:100	1:100
TET1	HPA019032	Rabbit	1:50	1:100
TET2	ab94580	Rabbit	1:100	1:100
TET3	SAB2700682	Rabbit	1:50	1:100
Beta-actin	ab6276	Mouse	-	1:500
Isotype	ab172730	Rabbit	-	1:100

Immunohistochemical staining of 5hmC was performed on 53 FFPE samples from 29 patients (cohort 1 and cohort 2) with 30 IDC sections and 23 matched contralateral sections. The 5hmC staining was also performed on 28 FFPE samples from 7 patients, including quiescent (no active inflammation, normal, n=13), dysplastic (n=8) and cancer tissues (n=8). IHC of 5mC was performed on 12 FFPE samples from 6 patients (cohort 1) with cancer and normal epithelial sections (n=6). In addition, the IHC staining of TET1, TET2, and TET3 was performed on 16 FFPE samples from 9 patients (cohort 1) with 6 IDC sections and 10 matched contralateral sections.

2.2.2 IHC of FFPE sections

FFPE sections of 5 μm thickness were deparaffinised in xylene twice and rehydrated in decreasing concentration of ethanol (twice in 100%, 90%, and 70%). The samples were immersed in antigen retrieval, sodium citrate buffer (10mM Sodium Citrate, 0.05% Tween 20, pH=6, all from Sigma-Aldrich, USA). To unmask cross-linking proteins, the sections were incubated with hydrochloric acid (HCl, 2M) for 15 minutes at room temperature (RT). The sections were then incubated in 10% goat serum buffer (10% GS buffer, 10% Goat Serum in PBS, Thermo Fisher Scientific) for an hour at RT in a humid chamber. All primary antibodies were optimised by using breast cancer tissue as positive controls. Both positive control and the samples were incubated overnight at +4˚C with the relative primary antibody from Table 2.5 diluted in 10% GS buffer, followed by incubating secondary antibody biotin (anti-rabbit immunoglobulins, Dako 0432) diluted 1:1000 in blocking buffer for an hour at RT. The samples were then incubated in 0.3% hydrogen peroxidase in methanol for 20 minutes, after which they were incubated for 40 minutes with streptavidin-HRP (Cell Signalling Technologies, USA)

diluted 1:200 in blocking buffer. The sections were developed with 3,3'-diaminobenzidine (DAB) substrate kit (Vector Laboratories, USA) at RT for 10 minutes and counterstained by Gills II Haematoxylin (Sigma-Aldrich, USA) for 3 minutes. Then the sections were dehydrated using serial dilutions of alcohol (70%, 90%, twice in 100% and twice in xylene) and mounted with Eukitt (Sigma, USA).

2.2.3 Quantification of staining intensity and protein expression

The intensity of staining of 5hmC, 5mC and TET proteins in the tissue was quantified with software (Visiopharm©, Horsham, Denmark) after digital slide acquisition on a Nanozoomer SQ slide scanner (Hamamatsu Photonics, Welwyn Garden City, UK). The software can distinguish brown nuclei or cytoplasm stained with DAB from the purple ones stained with haematoxylin after setting up the appropriate parameters (Figure 2.1). The epithelial cells of interest have to be selected manually, marking specific areas on the H&E-stained sections as control and to ensure lack of bias. After manually drawing a region of interest (ROI) based on pathological examination, cells were labelled based on the presence of Haematoxylin. The brown nucleus or cytoplasm is defined as a positive signal, and the purple nucleus or cytoplasm is defined as a negative signal. Nuclei (or cytoplasm) were separated from the background based on an intensity threshold set visually. The threshold was selected based on the evaluation of the staining intensity of the cells. The percentage of quantification of positive nucleus or cytoplasm against the total has been quantified by the programme. When detecting the staining of the nucleus, the size of the nucleus was set to 1.2um, the sensitivity was set to 77%, and the threshold was set to 61%. The ratio above this threshold was clarified as strong positive, and the ratio below this threshold was clarified as weak positive. When

detecting the degree of staining in the cytoplasm, use a similar method to evaluate and set the nucleus's size and the cytoplasm's width. The size of the nucleus was set to 2um, the width was set to 3um, the sensitivity was set to 60%, and the threshold was set to 73%. The ratio above this threshold was clarified as strong positive, and the ratio below this threshold was clarified as weak positive. Based on staining intensity, the results are shown as a number in a threshold from 0% to 100%.

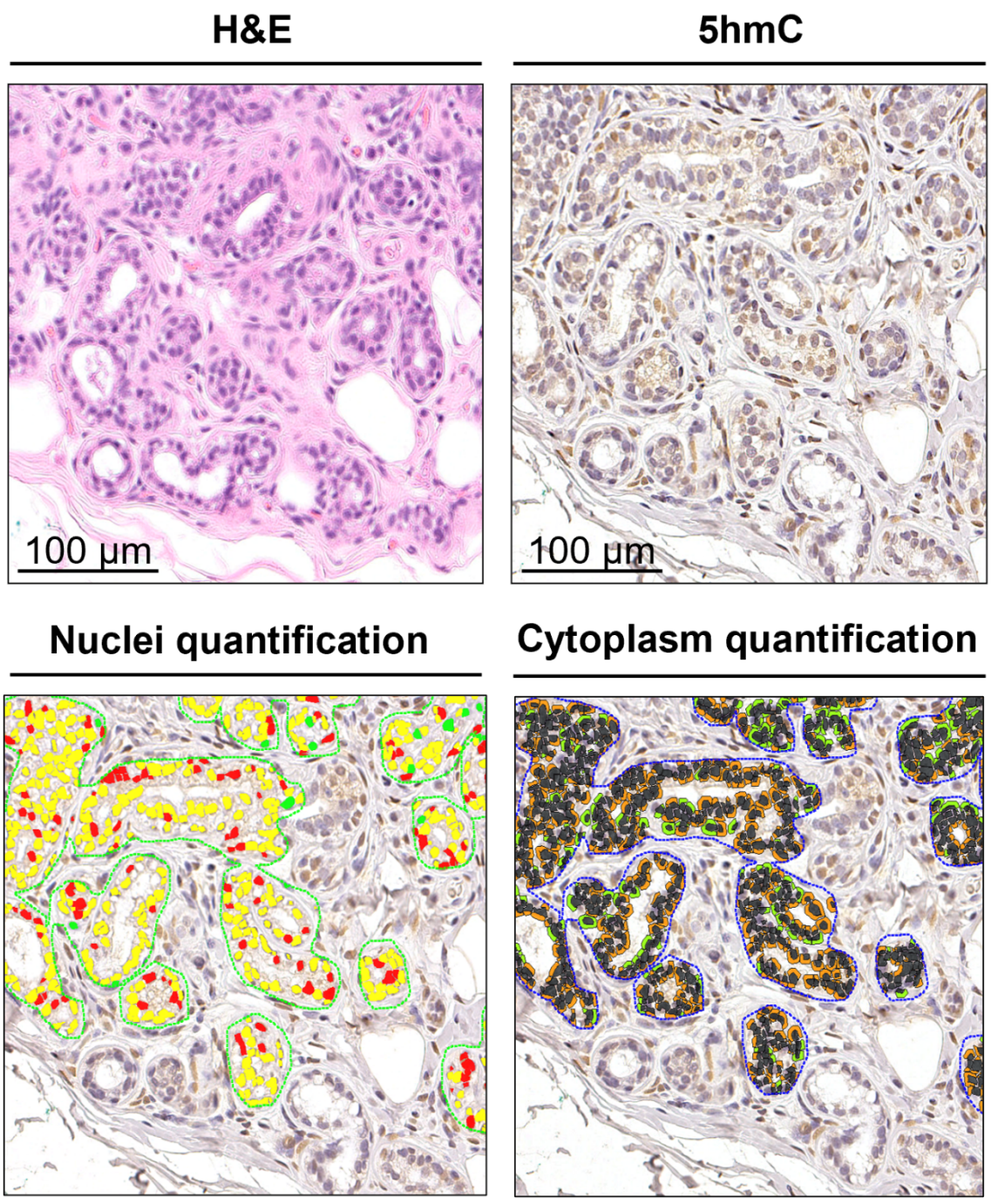


Figure 2.1 Immunohistochemistry staining and quantifications. Representative images of serial sections from a 55-year-old female patient, with areas of invasive breast cancer and normal epithelium. (scale bar = 100 µm). H&E, 5hmC expression, automated quantification nuclei of IHC, automated quantification cytoplasm of IHC are shown. Strong positive nuclei are represented in red, weak positive nuclei are in yellow, and negative nuclei are in green. Positive cytoplasm is represented in orange, and weak positive cytoplasm is in green.

2.2.4 Statistical Analysis

GraphPad Prism (GraphPad, USA) software was used for the statistical analysis. An unpaired T-test was used to compare the results between normal and cancer tissues. An unpaired T-test were used to compare TET protein expression in normal or cancer tissues.

The correlation matrix was analysed by GraphPad Prism. For this analysis, only patients with both paired cancer and normal epithelial sections were included (n=6). Data were analysed by Pearson correlation coefficients, as the data passed the normality test.

A p-value equal to or less than 0.05 was considered statistically significant and indicated with an asterisk (*p<0.05). Increasing statistical significance is indicated with more asterisk: *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

2.3 *In situ* Hybridisation (ISH)

This experiment aims to validate if *TET2* mRNA is expressed differently between normal and cancer tissue, therefore, *in situ* hybridisation (ISH) was performed to test the *TET2* mRNA expression on breast samples (cohort 4). This subchapter describes the optimisations and workflow of RNAscope.

2.3.1 Optimisation for RNAscope

The expression of *TET2* mRNA was measured on the tissue samples with RNAscope assay (Advanced Cell Diagnostics (ACD, USA) using the manufacturer's protocol. *TET2* mRNA expression was evaluated using ISH with RNAscope using a probe for peptidylprolyl isomerase B B (*PPIB*) as a positive control probe and a probe for the bacterial gene dapB as the negative

control probe. The experiment was performed on 20 FFPE samples from 12 pre-menopausal patients, comprising 11 benign and 9 cancer sections.

2.3.2 Workflow of RNAscope

Slides were baked in a dry oven at 60˚C for 1 hour, deparaffinized in xylene twice for 5 minutes each and 100% ethanol twice for 1 minute each, followed by an air dry for 5 minutes at RT. Tissue sections were incubated with hydrogen peroxide for 10 minutes at RT, boiled in 1x Target Retrieval buffer for 15 minutes and incubated with protease plus at 40˚C for 30 minutes. Sections were then incubated with a target Hybridise Probe. The selected RNAscope probes were Hs-TET2 probe (NM_017628.4, region 2684-3654, Cat No.420051), Hs-PPIB probe (positive-control probe, NM_000942.4, region 139-989, Cat No.313901) and dapB (negative control probe, EF191515, region 414–862, Cat No.310043). The target mRNA was amplified by the provided amplification probes, and the signal was developed using DAB A and B solutions provided. Sections were counterstained with 50% Gill's haematoxylin solution for 2 minutes at RT and serially dehydrated and mounted with Vectamount (Vector laboratories, UK).

2.3.3 Quantifications and analysis

The expression of *TET2* mRNA was quantified by manual counting using the ImageJ software (National Institute for Health, USA) after digital slide acquisition on a Nanozoomer SQ slide scanner. After scanning the tissue slides, the signal of mRNA was normalised to the signal of the positive control *PPIB* mRNA to evaluate the expression of *TET2* mRNA (Figure 2.2). First, epithelial regions of interest were randomly selected using the H&E and the positive control

slides as control and to ensure lack of bias. After selecting the epithelial cells, the values for HBS (H- hue, B- brightness and S-saturation) were adjusted so that the signal was specific to the brown dots or clusters, which represents individual or multiple mRNA copies, respectively. The numbers of pixels present in one single dot, representing one copy number of mRNA, were estimated by averaging the number of pixels from 30 individual brown dots. As one copy number corresponds to a single brown dot, a bigger dot or a cluster of dots represents more mRNA copies. In each selected region, the number of mRNA copies was calculated by measuring the total number of pixels divided by the average of pixels of one copy number of mRNA. The final ratio of mRNA dots/cell was calculated by the number of mRNA copies divided by the number of cells counted. The *TET2* mRNA ratios were normalised to the *PPIB* ratios to convert the normalized values to percentages. An unpaired T-test with Welch's correction was used to compare the results between cancer and benign tissues by Prism. Statistically significancy was measured as above (materials and methods 2.2.4).

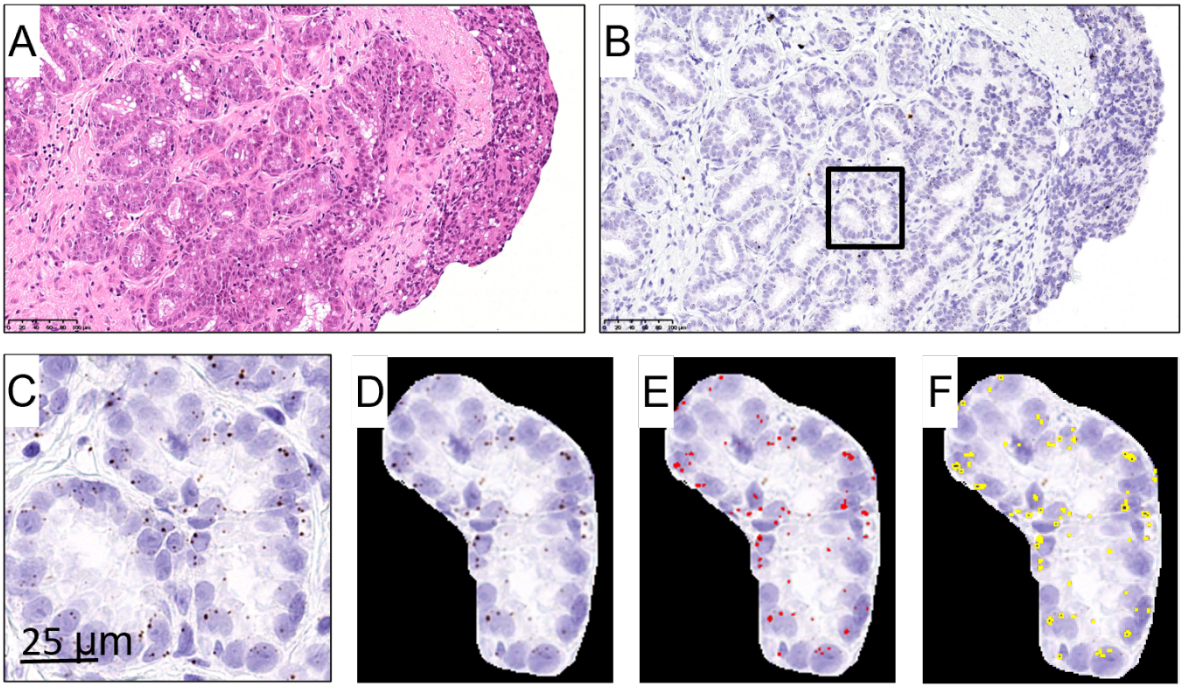


Figure 2.2. Quantification of *In situ* hybridisation (ISH) using the ImageJ software. A) H&E of benign breast tissue from a 31-year-old female patient. B) ISH of a serial section from the same lesion (selected area in-set) and C) magnified (insert = 25 µm) image of the selected area. D) The epithelial duct is selected as the region of interest. E) The brown dots are selected and marked as red. E) The red dots area is distinguished with a yellow border for the calculation of pixels.

2.4 Cell culture

This subchapter describes how to culture cell lines used. The cells cultured were collected for further experiments.

2.4.1 Cell lines

Five human mammary epithelial cell lines, one non-tumorigenic (MCF10), three breast cancer cell lines (MCF7, BT474, and SKBR3) and one human embryonic kidney cell line (HEK293T) have been used. All the cell lines were bought from ATCC, USA, and were kept in liquid

nitrogen and - 80 ˚C. The specific culture conditions for each cell line are detailed in Table 2.6. All cells were maintained in a 20% O_2, 5% CO_2, 37 °C cell culture incubator. All cells were passaged with 1% Trypsin (Sigma Life Science, UK) in 0.02% EDTA solution in 0.5mM Dulbecco's phosphate-buffered saline (DPBS) (Sigma) when they have reached 70%-80% confluence.

Table 2.6 Characteristic of cells and culture medium.

Cell lines	Genotype (154)	Medium	Supplements
MCF10A	ER- / PR- / HER2-	Dulbecco's Modified Eagle's Medium F12 (DMEM/F12; Sigma)	5% Horse Serum, 20 ng/ml, Human Epidermal Growth Factor (hEGF), 0.5 mg/ml hydrocortisone, 100 ng/ml cholera toxin, 10 µg/ml recombinant human insulin, and 1% penicillin-streptomycin (Invitrogen, USA)
BT474	ER- / PR+ / HER2+	Dulbecco's Modified Eagle's Medium (DMEM)	10% (v/v) Fetal Calf Serum (FCS) and 1% penicillin-streptomycin, 1% L-glutamine (Sigma).
HEK293T		DMEM	

2.5 Immunocytochemistry (ICC)

To validate the localisations of demethylation marks and TET proteins, I performed ICC on MCF10A and BT474 cells. This subchapter describes the optimisations of ICC protocol and workflow of this experiment, and the results show in result 3.1.6 and result 4.1.3.

2.5.1 Optimisations of antibodies

The general protocol was downloaded from the Abcam website (https://www.abcam.com/protocols/general-western-blot-protocol). The concentration of 0.1% or 0.3% Triton x-100 in PBS was first used to improve the penetration of the antibody in the cells. To break protein cross-links, the time for the incubation with 2M HCl at RT was first tested on the cells for 15 minutes, 30 minutes and 1 hour, respectively. The one-hour incubation time was selected for further experiments. Optimisation of the anti 5mC (1: 1600, 1:1000, 1:500, 1:200), 5hmC (1:1000, 1:200), 5fC (1:1000, 1:500, 1:200, 1:100), 5caC (1:1000, 1:500, 1:200, 1:100), TET1 (1:1000, 1:100, 1:50, 1:25), TET2 (1:1000, 1:100, 1:50, 1:25), TET3 (1:1000, 1:100, 1:50, 1:25) was serially diluted in 10% goat serum and incubated at RT.

After optimisation on human breast cell lines, the working concentration of Triton X-100 is 0.3% diluted in PBS and incubation an hour at RT was chosen for 2M HCl. The final concentration of each antibody was shown in Table 2.5, and the primary isotype antibody with the final concentration of 1:100 gave negative staining.

2.5.2 Workflow of ICC

Round cover glasses of 12mm thickness were coated by poly-D-lysine hydrobromide (0.1% (w/v) in water, Sigma-Aldrich, USA) for 5 minutes. After the cover glasses were completely dry, 10,000 cells were seeded into 24-well plates for two days. To allow the fixation of the cells on the cover glass, 4% paraformaldehyde diluted in 0.3% Triton X-100 in PBS (0.3%PBST) (pH 7.4, Thermo Fisher Scientific) was added into wells for 15 minutes at RT. Samples were incubated for 10 minutes with 0.3% Triton X-100 to permeabilize the cell membrane and then incubated in 2M HCl for 1 hour. Next, the samples were incubated with 10% goat serum buffer (10% Goat Serum in PBST, Thermo Fisher Scientific, USA) for 1 hour at RT to block the unspecific binding. The samples were then incubated in primary antibody (Table 2.5) diluted in 10% goat serum buffer for an hour at RT in a humid chamber. Followed by an hour blocking in 10% goat serum, samples were then incubated in the secondary antibody (Table 2.7) diluted in 10% goat serum buffer for an hour at RT. Samples were incubated in Hoechst 3325 (Sigma) for one minute and finally mounted with ProLong Gold antifade reagent (Thermo Fisher Scientific). The images were taken by Leica TCS SP5 confocal microscope.

Table 2.7 Details of Secondary antibody used in ICC.

Antibody	Catalogue number	Species	Dilution factor
Alexa Fluor 488	A11034	Rabbit	1:1000
Alexa Fluor 568	A11004	Mouse	1:1000

2.6 Quantification of hydroxymethylated DNA

To validate if the global 5hmC decrease in cancer cells, I performed a colorimetric assay to measure the level of global 5hmC in breast cell lines. This subchapter will describe how the DNA was extracted from human breast cell lines and how the colorimetric assay was used to test the hydroxymethylation level in human breast cell lines.

2.6.1 DNA extraction

Total DNA was prepared using DNeasy Blood & Tissue Kit (Qiagen, Germany), following the manufacturer's protocol, and quantified using NanoDrop ND1000 spectrophotometer (Thermo Fisher Scientific, USA). Briefly, an appropriate number of cells was lysed in 20 µl Proteinase K and 200 µl Buffer AL, mixed by vortexing and incubated at 56°C for 10 minutes. Cells were then mixed with 200 µl 100% ethanol by vortexing. Next, the mixture was applied into the DNeasy Mini spin column, followed by adding Buffer AW1 and AW2 and centrifuging to dry the membrane. Finally, DNA was eluted in 30 µl of DNase-free water after 1-minute incubation.

2.6.2 Workflow of Colorimetric assay

The global abundance of hydroxymethylation in the DNA was detected using MethylFlashTM Hydroxymethylated DNA Quantification Colorimetric Kit (Epigentek, USA), which was quantitated with a capture antibody a colorimetric reaction. According to the manufacturer, the detection limit of the input DNA can be 40 pg of hydroxymethylated DNA with a dynamic range from 40 pg to 20 ng, and the analytic sensitivity of the assay is about ten-fold higher

than that obtained by HPLC (0.008% vs 0.08%). Moreover, there is no cross-reactivity to methylcytosine and unmethylated cytosine. Briefly, samples and positive controls at different concentrations (0%-1%) were loaded into each well vertically, followed by 100 µl binding solution in each well. The solution was mixed, covered by parafilm, and incubated at 37°C for an hour. After incubation, the binding solution was removed and washed by 1X wash buffer (diluted in distilled water from 10X wash buffer) three times. The detection complex solution was then added to each well and incubated at RT for 50 minutes. After washing five times of 1X wash buffer, the developer solution was added to each well vertically and incubated at RT until the 1% positive control turning to deep blue. To stop the reaction, the stop solution was added into each well vertically. Absorbance at 450nm was read using a Sunrise™ microplate reader (Tecan, Switzerland). An unpaired T-test was used for comparing the results between normal and cancer cell lines.

2.6.3 Quantifications and analysis

Based on the protocol, 100ng genomic DNA was used for each sample. The levels of hydroxymethylated DNA were detected by colorimetry, and the amount of 5hmC, proportional to the Optical Density (OD) intensity, was measured in three independent biological repeats, with technical duplicates in each. The 5hmC amount was calculated using second-order logarithmic regression (Figure 2.3) based on the manufacturer's user guide. An unpaired T-test was used to compare normal (MCF10A) and cancer (BT474) cells.

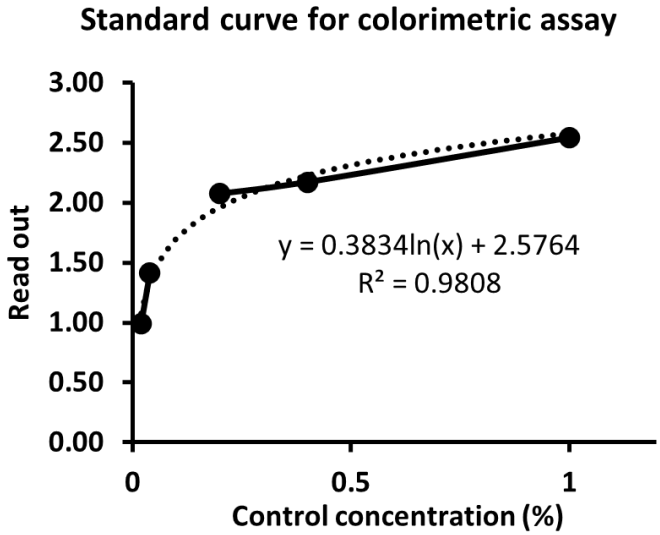


Figure 2.3 Standard curve calculated by logarithmic second order regression. Different concentrations of hydroxymethylated DNA (positive controls) were added into the assay wells and then measured with colorimetric kit following the protocol. The calibration curve for quantitation of the content of 5hmC is calculated by logarithmic second order regression. R square is shown in figure (R^2).

2.7 Dot blot

With IHC, cytoplasmic 5hmC was detected in breast and colon cancer cells. Therefore, to investigate if 5hmC was present in the RNA, I performed the dot blot experiments on different classes of RNA. This subchapter describes the workflow and optimisation of the dot blot and the preparation of different RNA samples.

2.7.1 RNA extraction

Total RNA was isolated using miRNeasy Mini Kit (Qiagen), according to the manufacturer's protocol and quantitated it using a NanoDrop ND1000 spectrophotometer (Thermo Fisher

Scientific). Briefly, cells were harvested and lysed by mixing with 700 µl Trizol (Qiagen) in a 1.5 mL microcentrifuge tube. The lysate was mixed with 140 µl chloroform and shaken vigorously for 15 seconds, then centrifuged for 20 minutes at 12,000 x g at 4°C. Next, the upper aqueous phase was transferred to a new tube and added with 1.5 volumes of 100% ethanol. The mixture was added to a miRNeasy Mini spin column and centrifuged at 8000 g for 15 seconds at RT. This step followed a single wash with 700 µl Buffer RWT and two washes with 500 µl RPE. Finally, RNA was eluted by 30 µl RNase-free water.

2.7.2 Large and small RNA isolation

Large RNAs (>200 nt) and small RNAs (<200 nt) were isolated using miRNeasy Mini Kit (Qiagen), according to the manufacturer's protocol and quantitated using Agilent 2200 TapeStation system (Agilent, USA).

Cells were harvested and lysed by mixing with 700 µl Trizol Reagent in a 1.5 mL microcentrifuge tube. The lysate was mixed with 140 µl chloroform and shaken vigorously for 15 seconds, then centrifuged for 20 minutes at 12,000 x g at 4°C. The upper aqueous phase was transferred to a new reaction tube mixed with 350 µl of 70% ethanol by vortexing. The sample was transferred into RNeasy Mini spin columns and centrifuged at ≥8000 x g for 15 seconds at RT. The flow-through, which contains small RNAs, was transferred into a new tube, and the large RNAs were bound on the RNeasy Mini Spin columns.

To purify the small RNAs (<200 nt), the flow-through was transferred into a new tube and added 0.65 volumes of 100% ethanol mixed thoroughly by vortexing. The sample was transferred into an RNeasy MinElute spin and centrifuged for 15 seconds at ≥8000 x g at RT. To wash the column, 700 µl Buffer RWT, 500 µl Buffer RPE and 500 µl of 80% ethanol was

added stepwise to the samples, followed by centrifuging 15 seconds at ≥8000 x g after each step. Small RNAs were eluted by 14 µl RNase-free water.

To purify large RNAs (>200 nt), the RNeasy Mini spin columns were placed into the new tubes followed by one wash with 700 µl Buffer RWT, two washes with 500 µl Buffer RPE, and after each wash centrifuged for 15 seconds at ≥8000 x g respectively. The large RNAs were eluted by 30 µl RNase-free water after centrifuging for 1 minute at ≥8000 x g.

2.7.3 mRNA isolation

MicroRNA was isolated using NEBNEXT Poly(A) mRNA magnetic isolation module kit (New England Biolabs, USA), according to the manufacturer's protocol and quantitated it using Agilent 2200 TapeStation system.

To wash the beads, 100 µl of RNA binding buffer was added to the well resuspended NEBNext Magnetic Oligo d(T)25 Beads and mixed thoroughly. The beads were washed twice and resuspended by 50 µl of RNA binding buffer. The total RNA sample was heated at 65°C for 5 minutes and held at 4°C to denature the RNA and facilitate the binding of the poly-A-RNA to the beads. The beads were resuspended by pipetting up and down to mix thoroughly, and the tubes were incubated at RT for 5 minutes to allow the RNA to bind to the beads. The sample was put on the magnetic rack at RT for 2 minutes, and the supernatant was removed. The beads were washed twice by a 200 µl wash buffer. The sample was heated at 80°C for 2 minutes, then held at 25°C. Next, the sample was mixed with 50 µl RNA Binding buffer and incubated at RT for 5 minutes twice, after discarding the supernatant by standing on the magnetic rack for 2 minutes at RT. The samples were washed with the 200 µl wash buffer and placed on the magnetic rack at RT for 2 minutes. To elute the poly-A mRNA from the beads,

the sample was added to 17 µl Tris Buffer and incubated at 80°C for 2 minutes, then held at 25°C. The samples were immediately placed on the magnetic rack for 2 minutes. The purified mRNA was collected by transferring the supernatant to a clean nuclease-free PCR Tube.

2.7.4 Optimisations of the samples

After optimisation, 50 ng and 25 ng of genomic DNA, 1000 ng of total RNA, 200 ng of large RNA (>200 nt), 200 ng of small RNAs (<200 nt), and 1000 ng of mRNA were used for the dot blot experiment (Figure 2.4). A hydroxymethylated polynucleotide (Epigentek, USA) was used as the 5hmC (positive) control, a methylated polynucleotide was used as the 5mC (negative) control, and an unmethylated polynucleotide was used as the unmethylated cytosine (negative) control. Each control has 0.15 ng, and the dot of the positive control contains 0.03 ng 5hmC.

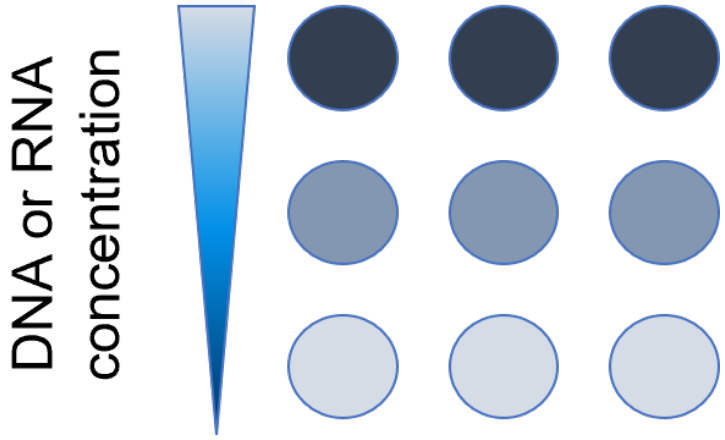


Figure 2.4 Dot blot. DNA or RNAs are crosslink to an Hybond-N+ membrane with two or three serial dilutions, triplicates each.

2.7.5 Workflow of dot blot

The protocol was optimised based on a previous study (155). The isolated DNA, RNA, large RNA, small RNAs and mRNA from cell samples were denatured at 95 °C for 3 minutes. The samples were crosslinked to an Amersham Hybond-N+ membrane (GE Healthcare, USA) by Stratalinker 2400 UV Crosslinker (Stratalinker, USA), followed by blocking in 5% non-fat milk (w/v) in PBS for 1 hour at RT. Next, the membrane was incubated in rabbit anti-5hmC primary antibody (Active Motif, USA, diluted 1:5000 in 5% non-fat milk) overnight at 4°C. Then, the membrane was incubated in anti-rabbit IgG-HRP (Dako, Denmark, diluted 1:5000 in 5% non-fat milk) for 1h at RT followed by exposure with Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific). The blot was developed by film or iBright CL1500 (Thermo Fisher Scientific).

2.7.6 Quantification and analysis

After development, the dot blot image was analysed by using ImageJ's gel analysis function. The relevant 5hmC intensity was analysed by comparing the dot with the positive control on the same membrane. Sidak's multiple comparisons test was used to compare the results between normal (MCF10A) and cancer (BT474) cells.

2.8 Functional assay

To investigate the regulatory role of 5hmC and TET proteins, cells were treated with siRNA to knockdown the TETs expression and were seeded for the following assays: qPCR to examine if *TET* mRNA and other target genes are downregulated, proliferation assay to test whether

downregulating *TET* mRNA expression affected cell proliferation and colorimetric assay (same as 2.6.2) to examine if the global level of 5hmC would change after *TET* mRNA downregulation. In this subchapter, the transfection of siRNA into cell lines is described alongside the functional assays.

2.8.1 siRNA Reverse Transfections

2.8.1.1 Optimisations of reverse transfection

Cells were grown in antibiotic-free medium and transfected with Lipofectamine 2000 or RNAiMAX transfection reagents (Invitrogen). To determine the highest transfection efficiency, the HEK293T cell line was used where the cells were transfected with different transfection reagents in 10 cm dish, 6 cm dish, 6-well plates shown below. Based on the manufacturer's protocol, HEK293T cells were transfected with 10 nM and 20 nM siRNA, respectively, to test the efficiency. HEK293T cells were seeded into 10 cm dishes with the final concentration of 10 µM siRNA (using Lipofectamine 2000) or 20 µM siRNA (using RNAiMAX) on day 0, then resuspended and seeded into 6 cm dishes, 6-well plates, and 96-well plates on day 1. Cells were harvested 48 hours, 72 hours, and 96 hours after transfecting siRNAs for future experiments. The condition which was used is shown below (Table 2.8).

Table 2.8 Optimised transfection conditions.

	Transfection reagent	OptiMEM	siRNA (20 µM stock)
10 cm dish	30 µl Lipofectamine 2000 / well	500 µl	5 µl (final 10 nM)
10 cm dish	20 µl Lipofectamine RNAiMAX / well	2000 µl	10 µl (final 20 nM)
6 cm dish	15 µl Lipofectamine 2000 / well	250 µl	2.5 µl (final 10 nM)
6 cm dish	10 µl Lipofectamine RNAiMAX / well	1000 µl	5 µl (final 20 nM)
6-well plate	7.5 µl Lipofectamine 2000 / well	125 µl	1.25 µl (final 10 nM)
6-well plate	5 µl Lipofectamine RNAiMAX / well	500 µl	2.5 µl (final 20 nM)
96-well plate	0.3 µl Lipofectamine 2000 / well	5 µl	0.05 µl (final 10 nM)
96-well plate	0.2 µl Lipofectamine RNAiMAX / well	20 µl	0.1 µl (final 20 nM)

2.8.1.2 Workflow of siRNA transfection

Predesigned ON-TARGET plus SMART pool siRNAs, negative control ON-TARGET non-targeting Pool siRNA were purchased from Dharmacon, USA. All target sequences are listed in Table 2.9. siRNA pellet was resuspended in 1x siRNA buffer (60mM KCl, 6mM HEPES-pH 7.5, and 0.2 mM $MgCl_2$). For 5 nmol of siRNA pellet and a 20 µM stock concentration, 250 µl 1x siRNA buffer was added and the solution was mixed with pipetting up and down 3-5 times. The solution was then placed on an orbital shaker (300rpm) for 30 minutes at RT.

siRNAs were transfected into HEK293T, BT474 and MCF10A cells at a final concentration of 20 nM in antibiotic-free conditions using Lipofectamine RNAiMAX (Thermo Fisher Scientific, USA) according to the manufacturer's reverse transfection protocol. Briefly, 20 µl stock siRNA (20 µM) was diluted in 3980 µl Opti-MEM I Medium without serum and mixed well. The mixture was then added 40 µl Lipofectamine RNAiMAX containing the diluted RNAi molecules and mixed well. The mixture was incubated for 1 hour at RT. While incubating, cells were seeded into two 10 cm dishes with 4 M cells in each plate. To each 10cm dish, the final RNA concentration is 20 nM in a final volume of 10 ml on day 0. Cells were incubated 24 hours at 37°C in a CO_2 incubator and then resuspended and seeded into 6 cm dishes, 6-well plates, and 96 well-plates on day 1. Cells were harvested or fixed on 24 hours, 48 hours, 72 hours, 96 hours, 120 hours, and 144 hours after transfecting siRNAs for further experiments.

Table 2.9 Target sequence of siRNAs.

Target	Target sequence (5'-3')
TET1	GAGAAUAGGUAUGGUCAAA
	CUGUCUUGAUCGAGUUAUA
	CCUUAACUCGAGAAGAUAA
	GCACGCAUGAAUUUGGAUA
TET2	ACAAGAAAGUAGAGGGUAU
	ACACCUAGUUUCAGAGAAU
	CCUCAGAAUAAUUGUGUGA
	CAGCAAAGGUACUUGAUAC
TET3	AAACAAAGCAAGUCGAGAA
	GGGCAAGAUUCCCGAGGAU
	CAUAUUAUAGCAACGGAAA
	ACGAGGAAAUAGCGAUUGA
Negative control **(Scramble)**	UGGUUUACAUGUCGACUAA
	UGGUUUACAUGUUGUGUGA
	UGGUUUACAUGUUUUCUGA
	UGGUUUACAUGUUUUCCUA

2.8.2 Reverse Transcription and Real-time Quantitative Polymerase Chain Reaction (RT-qPCR)

2.8.2.1 Reverse Transcription and RT-qPCR

To investigate the expression of mRNA, cDNA was first prepared from 1000 ng total RNA using Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Fisher Scientific) according to the manufacturer's instructions. 20 ng cDNAs were amplified using Fast SYBR Green Master Mix (Thermo Fisher Scientific) and 1 µM forward and reverse primers (see Table. 2.10) using the StepOne Plus Real-Time PCR System software (Applied Biosystems, USA) in 96-well plates.

Table 2.10 Primer Sequences for SYBR Green Real-Time qPCR

Genes	Forward Primer	Reverse Primer
RPLPO	5’ - GCAGCATCTACAACCCTGAAG - 3’	5’ - CACTGGCAACATTGCGGAC - 3’
POLR2A	5’ - GCAAATTCACCAAGAGAGACG - 3’	5’ - CACGTCGACAGGAACATCAG - 3’
TET1	5’ - AACGGAATCCAACGGGAAGG - 3’	5’ - TGCCATCACGTTAGCACACT - 3’
TET2	5’ - GATAGAACCAACCATGTTGAGGG - 3’	5’ - TGGAGCTTTGTAGCCAGAGGT - 3’
TET3	5’ - TCCAGCAACTCCTAGAACTGAG - 3’	5’ - AGGCCGCTTGAATACTGACTG - 3’
DNMT1	5’ - CCATCAGGCATTCTACCA - 3’	5’ - CGTTCTCCTTGTCTTCTCT - 3’
DNMT3a	5’ - TATTGATGAGCGCACAAGAGAGC - 3’	5’ - GGGTGTTCCAGGGTAACATTGAG - 3’
KI67	5‘ - TCCTTTGGTGGGCACCTAAGACCTG - 3’	5‘ - TGATGGTTGAGGTCGTTCCTTGATG - 3’
MCM2	5' - AGGCAGCATCCCCATTAC - 3'	5' - TCACATAGTCCCGCAGAT - 3'
BCL-2	5’ - GATGTGATGCCTCTGCGAAG - 3’	5’ - CATGCTGATGTCTCTGGAATCT - 3’
BCL-w	5’ - CATCTCTGCCCCCTCTGCTGA - 3’	5’ - GGATGACCTTGCCCACAGCCT - 3’
ESR1	5’ - GCTTACTGACCAACCTGGCAGA - 3’	5’ - GGATCTCTAGCCAGGCACATTC - 3’
GATA3	5’ - TTTTTTTGGAATTTTAGTTTTTTTT - 3’	5’ - CACCTAATACTTAAAACACTCTTTCTC - 3’

2.8.2.2 Analysis

Data were analysed by the delta-delta Ct method (the symbol Δ refers to delta): ΔCt is used to determine the amount of mRNA expression within individual groups calculated by Ct (target gene) - Ct (HSK). Two housekeeping genes were used: Ribosomal Protein Lateral Stalk Subunit P0 (*RPLPO*) (156) and RNA polymerase II subunit A (*POLR2A*) (157). The Ct (HSK) is defined as Ct mean of *RPLPO* and *POLR2A*. Expression fold change between untreated and siRNA treated groups was calculated with 2^-ΔΔCt, where ΔΔCt = ΔCt (siRNA treated) - ΔCt (Scramble). The two-way ANOVA was used to analyse the *TET* mRNA expression at different time points, and the one-way ANOVA was used to analyse the target gene expression in BT474 cells treaded 48 hours. Statistical significance is the same as 2.2.4.

2.8.3 Proliferation assay

Cells were seeded in five replicates in 96-well plates (3,000 cells/well) with antibiotic-free DMEM on day 1, transfected with siRNAs as above on day 0. The plate was then put into IncuCyte ZOOM (Essen BioScience, USA), and images of the vessel were collected every 4 hours from day 2 to day 7. Four regions of interest were taken for each well and saved in an image collection by clicking on the “Create or Add to Image Collection” link.

For Image analysis, the Incucyte software was used to determine cell confluence per image. The set-up of image analysis in brief: To analyse the software was trained specifically for the MCF10A and BT474 proliferation assay. “Basic Analyser” from the Analysis job Type dropdown menu was selected. The “New” was selected and assigned a new name when adding the first image of a collection.

The New Processing Definition was created to measure the cell confluence. The process started by clicking on the" New Processing Definition" link within the Vessel View and selecting the image Collection. Several images from different time points were selected and added to the Image Collection. The Basic Analyzer Processing Definition Editor defaulted to the image collection selected in the previous steps. The cells and backgrounds were differentiated by altering Segmentation Adjustment and Adjust Size in Training Image Collection. The training was performed until the cells from selected images were selected with parameters and recognised by the confluence mask. When the optimal settings were determined, they were used for analysis.

After analysis, the parameter of the percentage of Phase Object Confluence was chosen, and phase data were exported to an excel file. The value is the mean of the percentage of phase object confluence of each technical repeat. The Mann-Whitney test was used to determine the significance of the difference in cell proliferation compared to the untreated cells (negative control). Statistical significance is the same as 2.2.4.

2.9 Kaplan-Meier survival plots

To analyse the correlation between *TET* mRNA expression and patient survival , data were extracted from the Kaplan-Meier survival plot: http://kmplot.com/analysis/(158). Gene probes were selected as Jetset best probe set, which is the optimal probe set: TET1 228906_at, TET2 227624_at, TET3 235542_at. The patients were split by auto select best cutoff (values between lower and upper quartiles), which is the best performing threshold, and the overall survival of all these patients was measured (BC patient, n=943, ovarian cancer patient, n=655). Gene expression data and overall survival information were downloaded from GEO

(https://www.ncbi.nlm.nih.gov/geo/), EGA (https://www.ebi.ac.uk/ega/home) and TCGA (https://cancergenome.nih.gov/). To investigate the *TET* mRNAs expression and patient survival rate in different subtypes, intrinsic subtypes were categorised in the following major molecular subtypes: basal-like (ESR1-/HER2-), Luminal A (ESR1+/HER2-/MKI67 low), Luminal B (ESR1+/HER2-/MKI67 high and ESR1+/HER2+), HER2+ (ESR1-/HER2+).

3. Result 1:

The localisation of demethylation marks and the presence of 5hmC

Here the first and second research questions (whether the distribution of 5hmC changes between normal breast and breast cancer cells, and whether 5hmC plays any regulatory function through the methylation of different classes of RNA) are investigated. The aims of this chapter are to present data on: a) whether demethylation markers are differentially expressed in normal breast compared to BC in human tissues (3.1.1 and 3.1.4) and in cells (3.1.5 and 3.1.6), b) to verify if levels of 5hmC are related to the hormonal receptor status of breast cancer (3.1.2), and c) to investigate the 5hmC presence in the RNA (3.1.7). To achieve this, I aimed to validate the global 5hmC levels in breast cell lines, investigate the localisations of demethylation markers, and to identify the presence of 5hmC in DNA and RNA for further characterisation.

DNA methylation plays an essential role in regulating gene transcription and carcinogenesis in different cancers, including breast cancer (159, 160). During the demethylation process, the methylation of cytosine is converted in a stepwise sequence to 5hmC, 5fC, 5caC by the TET family of proteins. As discussed in the introduction, loss of 5hmC can be found in solid cancers, including breast cancer, compared to adjacent non-affected epithelia, and loss of 5hmC was shown as a poor prognostic factor in ER/PR- negative breast cancer patients (96, 161). These studies suggest the levels of 5hmC are altered in breast cancer progression, and they may also be related to hormonal receptor status. Moreover, the levels of 5caC were found increased in breast cancer tissue compared to normal breast tissue but not associated with the levels of 5hmC in cancer tissues (104). These studies suggest that all these

demethylation marks could have a regulative role in breast cancer progression. Therefore, it would be helpful to verify the abundance and distributions of these demethylation marks in breast cancer cells compared to normal cells and compare them in different hormonal receptor statuses.

Several studies hypothesised that 5hmC contributes to the regulation of transcription, as hydroxymethylation has been found in the RNA of mammalian cells, in the mRNA of *Drosophila* and mouse embryonic stem cells (96, 97, 162-164). However, up to now, there is lack of knowledge on whether 5hmC is present or regulated in the RNA of human breast cells. Thus, it would be valuable to investigate if 5hmC is present in any class of RNA in breast cell lines, as this will help us explore the functions of hydroxymethylation.

Additionally, another study found that modified nucleotides can undergo recycling processes from the DNA of dying cells within the cellular cytoplasm (165). 5-hydroxymethyl-2’deoxycytidine (5hmdC) and 5-formy-2’deoxycytidine (5fdC), but not 5-carboxyl-2’deoxycytidine (5cadC) exhibit susceptibility to cytidine deaminase (CDA) catalysis, resulting in the cytotoxic 5-hydroxymethyl-2’deoxyuridine (5hmdU) and 5-formyl-2'-deoxyuridine (5fdU) (165). Given the structural similarity between 5hmC and 5hmdC, it is conceivable that analogous recycling pathways may exist for 5hmC.

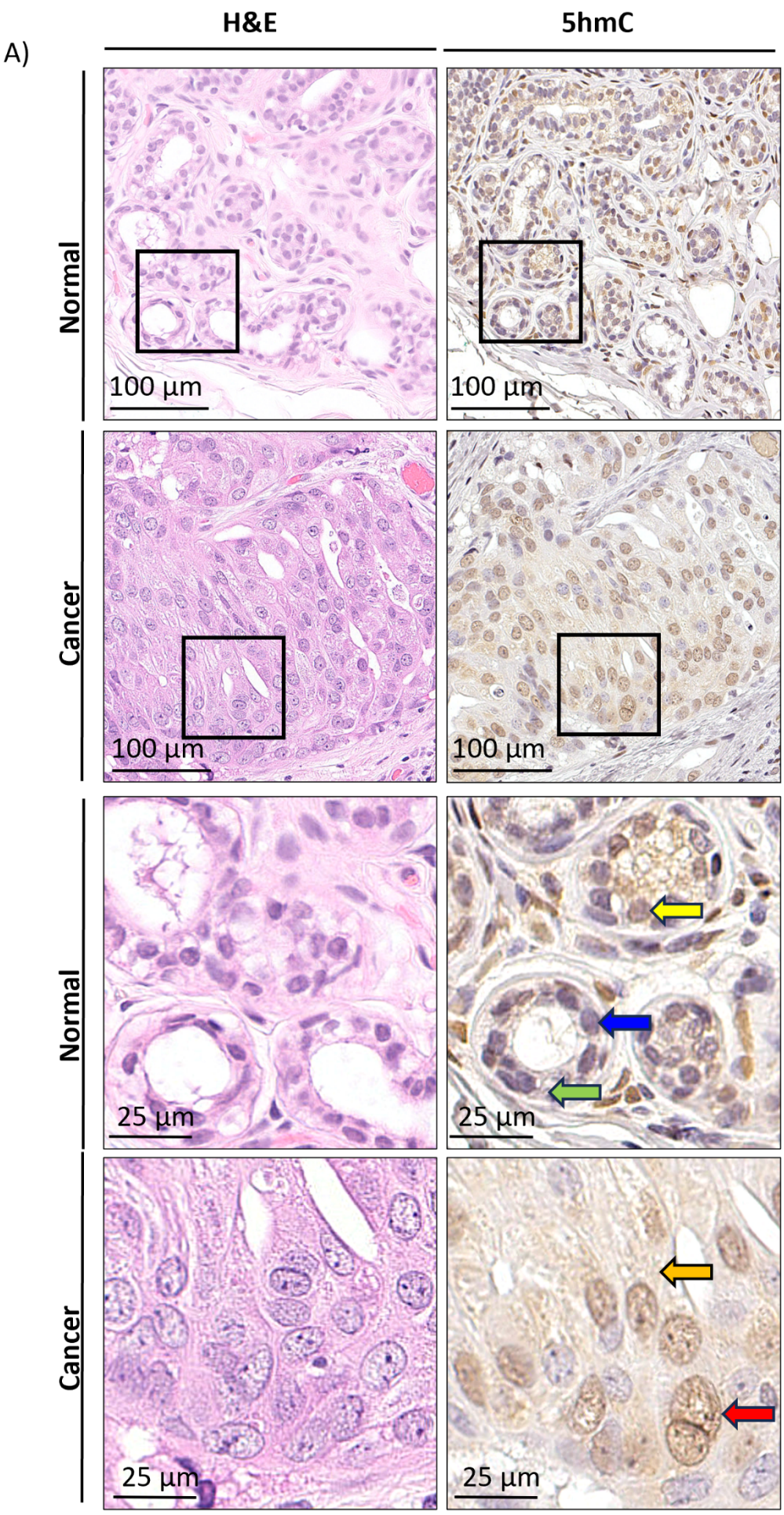
3.1 Results

3.1.1 The breast cancer epithelium is characterised by a significantly higher proportion of positive cells of cytoplasmic 5hmC

To verify how the level of 5hmC changes between normal and cancer cells in human tissues, IHC to detect 5hmC was performed together with Dr Lisa Del Bel Belluz on 53 FFPE samples from 29 patients (cohort 1 and cohort 2, Dr Del Bel Belluz performed on cohort 1 and I performed on cohort 2) with 30 IDC sections and 23 matched contralateral sections (cohort 1 and cohort 2) (see patient samples details in the subchapter 2.1, methods in the subchapter 2.2).

Previous studies only quantified the global cellular level and did not investigate specific subcellular localisations. It is interesting to consider the sublocalisation of 5hmC in the breast epithelium, because it may indicate the location of this modification (in DNA or RNAs). Thus, in this experiment, the intensities of nuclear and cytoplasmic staining were quantified separately.

The staining indicated that 5hmC was clearly expressed in the breast epithelium, not only in the nucleus but also in the cytoplasm (Figure 3.1). Interestingly, there was no significant difference in the proportion of positive cells of 5hmC in the nucleus between normal and cancer epithelium ($p=0.408$, Figure 3.1 B). However, the proportion of positive cells of cytoplasmic 5hmC was significantly more present in cancer tissues than in normal tissues ($p=0.012$, Figure 3.1 B), which has not been reported before.


A)
H&E
5hmC
Normal
Cancer
Normal
Cancer
100 µm
100 µm
100 µm
100 µm
25 µm
25 µm
25 µm
25 µm

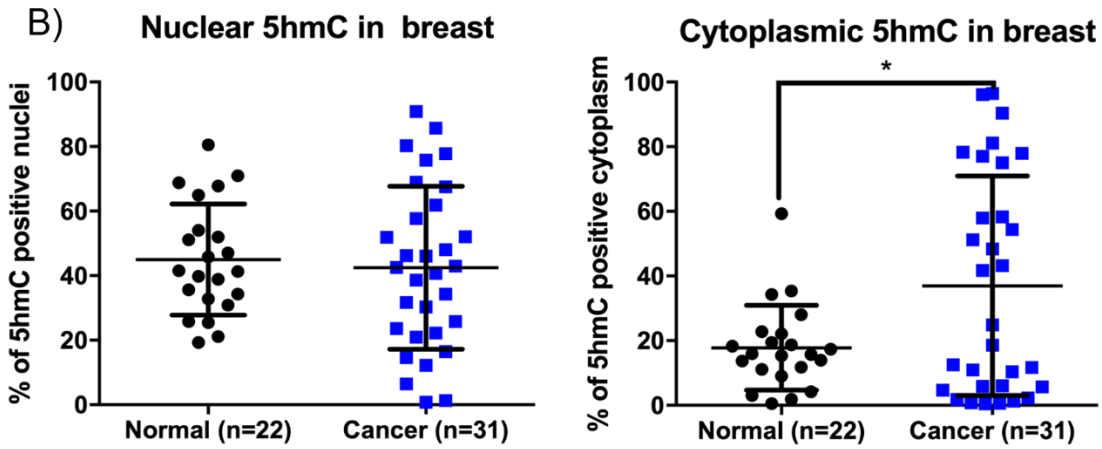


Figure 3.1 5hmC immunohistochemistry staining. Panel A) Representative images from serial sections from a 55-year-old female patient, with areas of invasive breast cancer and normal epithelium, showing H&E staining (left) and 5hmC levels (right, in brown). (Upper two scale bar = 100 µm, lower two scale bar = 25 µm). The figure in the first row is duplicated as the first row in Figure 2.1. The red arrow indicates strong positive staining in nuclei, the yellow arrow indicates weak positive staining in nuclei, and the blue arrow indicates negative staining in nuclei. The orange arrow indicates positive staining in cytoplasm and the green arrow indicates negative staining in cytoplasm. **Graph B)** Box plots showing the percentage of positive cells of 5hmC in the nuclei, or the cytoplasm as calculated with a digital software. The lines represent the mean, and the SD. Statistical analysis was performed with an unpaired T-test on each cohort and statistical significance is indicated with an asterisk: * = $p<0.05$. From the total of the 29 patients, normal tissue was missing from 7 patients (normal section n=22) and 2 patients provided more than one cancer section (cancer section n=31), thus not allowing a full paired comparison.

3.1.2 The cancer epithelium from ER- patients is characterised by a significantly higher proportion of positive cells of cytoplasmic 5hmC

As it was mentioned above, previous studies indicate that the levels of 5hmC may be associated with hormonal receptor status (161, 166). However, Tsai and colleagues only pointed out the association between the low level of 5hmC and a poor survival rate for ER/PR- subtype (161). They observed the different subcellular localisations of TET1 and TET2, but they did not mention the cytoplasmic 5hmC in breast tissue. For this reason, they did not investigate the association between subcellular 5hmC and breast cancer subtypes. I, therefore, analysed the same patient data from above by dividing the samples by hormonal receptor status. Samples were split into four groups: triple positive (ER+ PR+ HER+), triple negative (ER- PR- HER2-), ER+ PR+ HER2- and ER- PR- HER2+. Interestingly, the proportion of positive cells of 5hmC were not correlated with the combined status of the three receptors in neither the cellular compartments (Figure 3.2). While most samples were characterised by a high heterogeneity in the amount of 5hmC, the ER- PR- HER2+ patient group was particularly interesting. In this group, 83% (5/6) of the samples showed a higher proportion of positive cells of cytoplasmic 5hmC in the cancer epithelium compared to normal (Figure 3.2 F). In ER+ PR+ HER2- groups, two out of three samples have lower proportion of positive cells of cytoplasmic 5hmC in cancer compared to normal (Figure 3.2 H-K). Interestingly, the proportion of positive cells of nuclear 5hmC decreases from normal to cancer epithelium (Figure 3.2 G). This could be a tendency of decrease in ER+ PR+ HER2- patients, however, more repeats are needed to confirm this hypothesis.

The samples were therefore divided into ER positive and ER negative groups (Figure 3.3) and the correlation between the expression of the receptor and 5hmC was tested with both

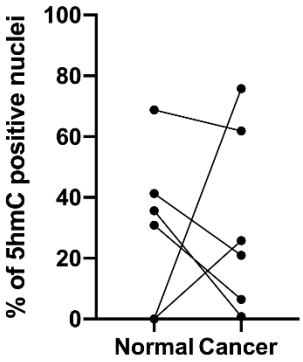
unpaired and paired T-test. While the expression of ER was not significantly correlated with differences in the proportion of positive cells of 5hmC in the samples (Figure 3.3 A-B), the absence of the ER was associated with a different proportion of positive cells of 5hmC. In particular, the proportion of positive cells of cytoplasmic 5hmC was significantly higher in the cancer epithelium of ER negative patients compared to the normal matched controls (Wilcoxon test, p=0.048, Figure 3.3 D-E). Finally, there was no significant difference in the proportion of positive cells of 5hmC between cancer and normal epithelium after HER2 stratification (Figure 3.4).

To determine whether the age of patients or the grade of BC is associated with an alteration for the proportion of positive cells of 5hmC, I examined the correlation between these three variables. Cancer samples and their matched normal sample were stratified according to the grade of the tumour (grade 2 or 3, Figure 3.5). For neither group, a correlation was observed. However, it was interesting to observe a strong division between high and low proportion of positive cells of 5hmC in the cancer of grade 3 samples, possibly due to the ER receptor levels. There were totally fourteen grade 3 patients, including three ER+ patients and eleven ER- patients. In eleven out of fourteen (78.6%) grade 3 patients, the proportion of positive cells of nuclear 5hmC and of cytoplasmic 5hmC showed the same correlation compared normal to cancer (Figure 3.5 C-D). In ER+ patients, two out of three (66.6%) had lower proportion of positive cells of nuclear 5hmC in cancer compared to normal (Figure 3.5 C). Meanwhile, same patients (66.6%) also had lower proportion of positive cells of cytoplasmic 5hmC in cancer compared to normal (Figure 3.5 D). In ER- patients, seven out of eleven (63.6%) patients had lower proportion of positive cells of nuclear 5hmC in cancer compared to normal (Figure 3.5 C), while seven out of eleven (63.6%) patients had higher proportion of positive cells of cytoplasmic 5hmC in cancer compared to normal (Figure 3.5 D). Moreover, eight out of eleven

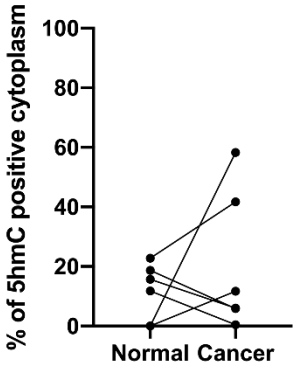
patients showed the same changes compared cancer to normal. Loss of ER may lead to an increase of 5hmC in grade 3 patients, which could be an explanation for this trend.

Besides, the samples in the different grades were not found to have any correlation between 5hmC and hormonal receptor status. Therefore, the proportion of positive cells of 5hmC were not correlated with the grade of breast cancer (Figure 3.5). In addition, I performed a Spearman r test between the proportion of positive cells of 5hmC staining and age. However, I did not detect any correlation between the age of resection and the proportion of positive cells of 5hmC in this cohort (Figure 3.6).

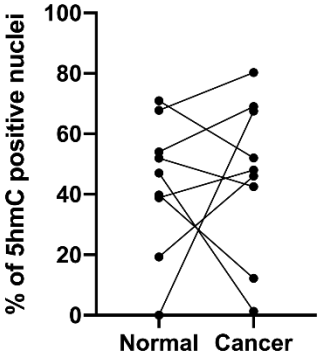
A) TP patient nuclei (n=6)



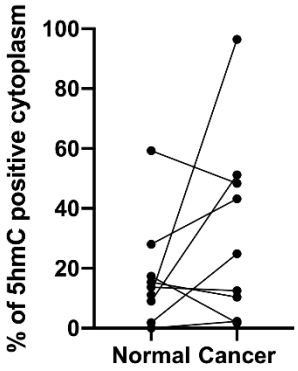
B) TP patient cytoplasm (n=6)



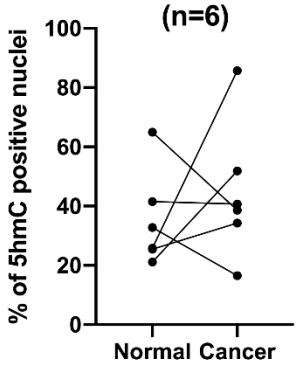
C) TN patient nuclei (n=9)



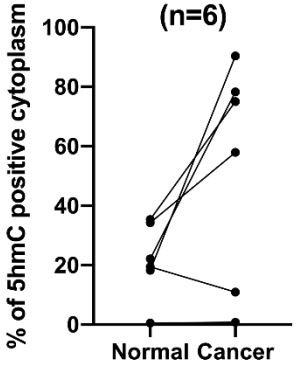
D) TN patient cytoplasm (n=9)



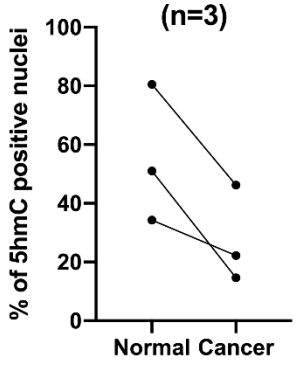
E) ER- PR- HER2+ patient nuclei (n=6)



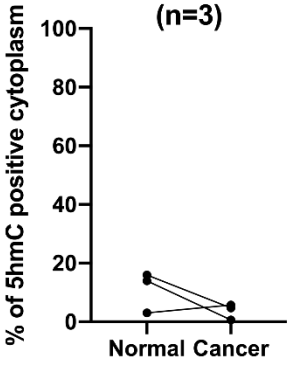
F) ER- PR- HER2+ patient cytoplasm (n=6)



G) ER+ PR+ HER2- patient nuclei (n=3)



H) ER+ PR+ HER2- patient cytoplasm (n=3)



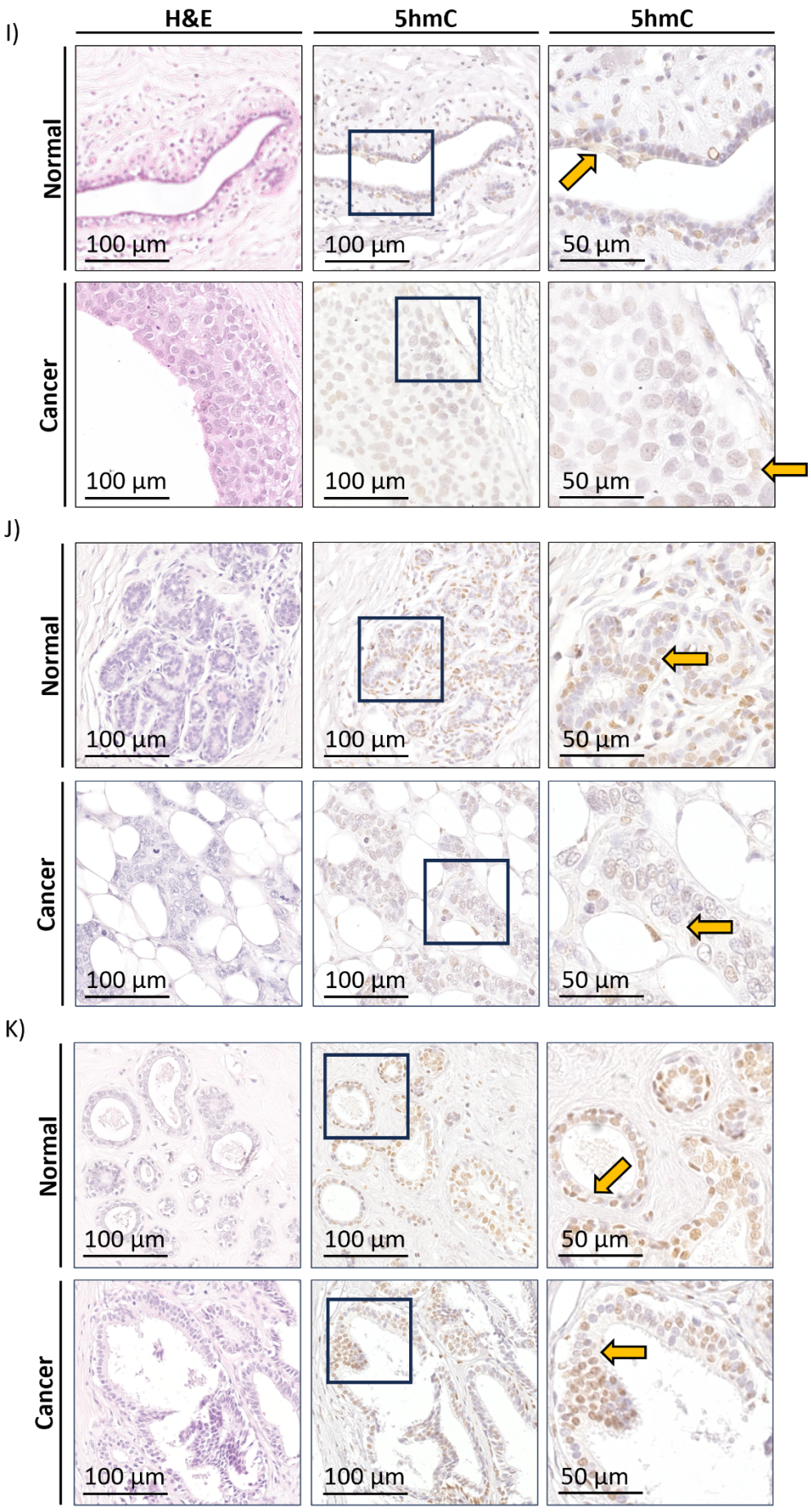
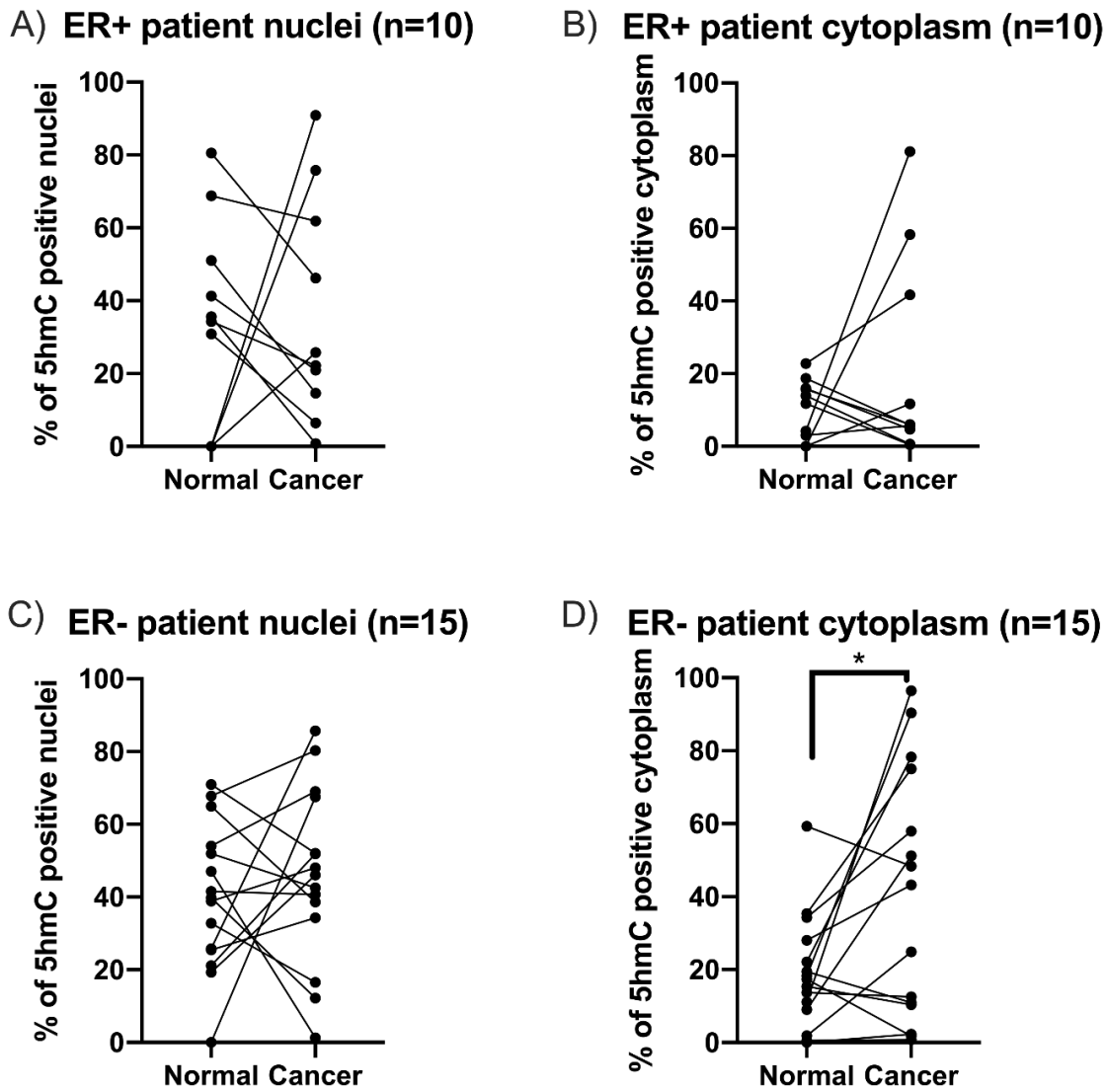

I)
H&E
5hmC
5hmC
Normal
Cancer
100 µm
50 µm
J)
Normal
Cancer
100 µm
50 µm
K)
Normal
Cancer
100 µm
50 µm

Figure 3.2 (Previous page) The proportion of positive cells of 5hmC after stratification of the samples based on the hormonal receptor status of the cancer. The graphs A-H) show the abundance of 5hmC (as percentage of positive cells) in the nuclei or the cytoplasm. Unpaired T-test was performed only on the samples where matched cancer and normal (non-affected) epithelium was provided. TP represents triple positive, and TN represents triple negative samples. **Panel I-K)** Representative images from serial sections from three ER+ PR+ HER2- patients, with areas of invasive breast cancer and normal epithelium, showing H&E staining (left), and 5hmC levels (in brown; middle, scale bar = 100 μm; right, scale bar = 50 μm). The orange arrow indicates positive staining in cytoplasm.



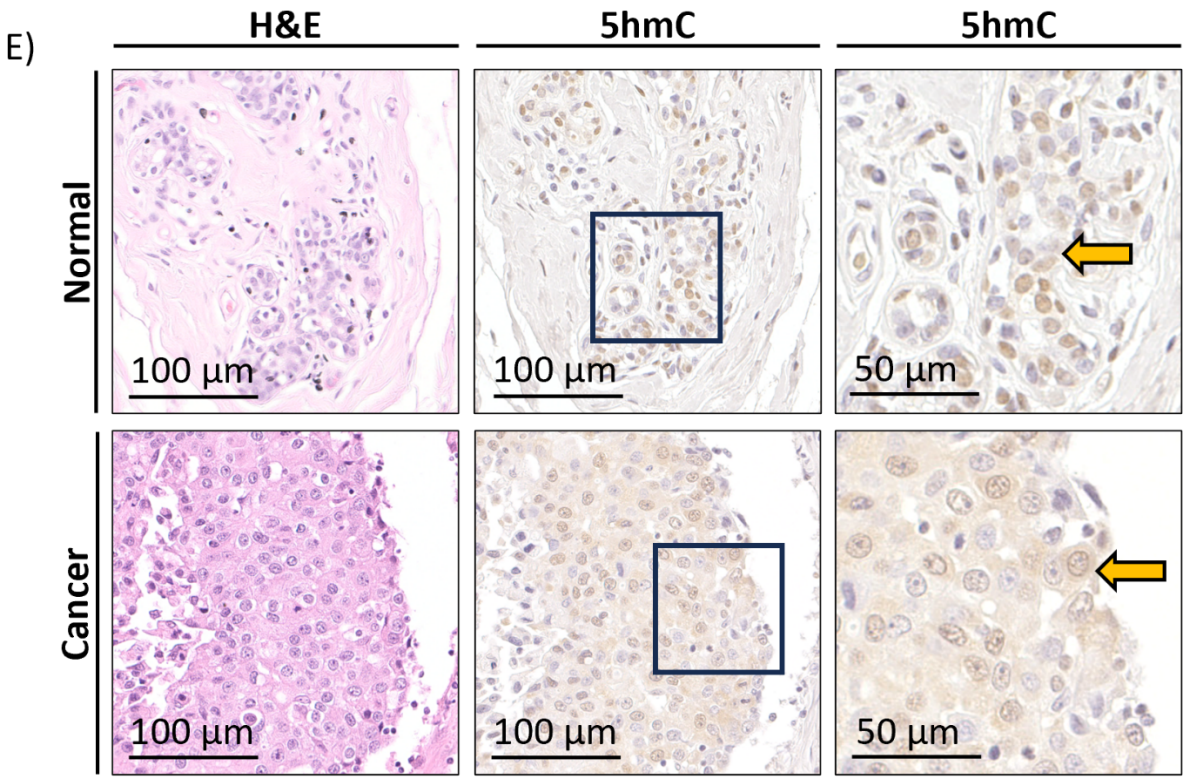


Figure 3.3 The proportion of positive cells of 5hmC after stratification of the samples according to the ER status of the cancer. Graph A) and B) are from patients with ER positive cancer. C) and D) are from patients with ER negative cancer. The graphs show the abundance of 5hmC (as percentage of positive cells) in the nuclei or the cytoplasm. Paired T-test was performed only on the samples where cancer and matched normal (non-affected) epithelium was provided. Significance is indicated with an asterisk: * = $p<0.05$. **Panel E)** Representative images from serial sections from a 55-year-old female patient (ER-, PR-, HER2+), with areas of invasive breast cancer and normal epithelium, showing H&E staining (left), and 5hmC levels (in brown; middle, scale bar = 100 µm; right, scale bar = 50 µm). The orange arrow indicates positive staining in cytoplasm.

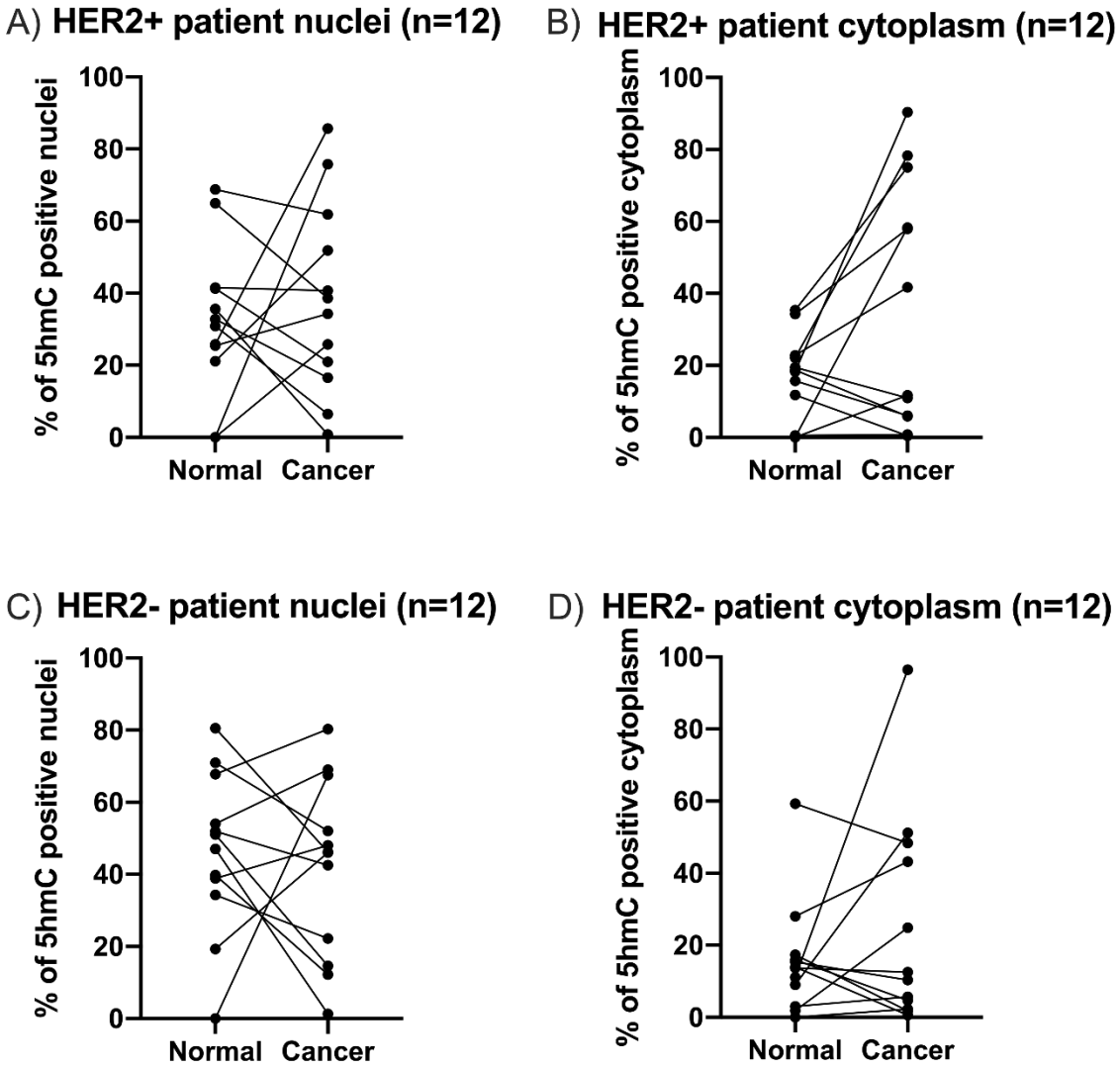


Figure 3.4 The proportion of positive cells of 5hmC after stratification of the samples according to the HER2 status of the cancer. A) and B) are from patients with HER2 positive cancer. C) and D) are from patients with HER2 negative cancer. The graphs show the abundance of 5hmC (as percentage of positive cells) in the nuclei or the cytoplasm. Paired T-test was performed only on the samples where matched cancer and normal (non-affected) epithelium was provided. Significance is indicated with an asterisk: * = $p<0.05$.

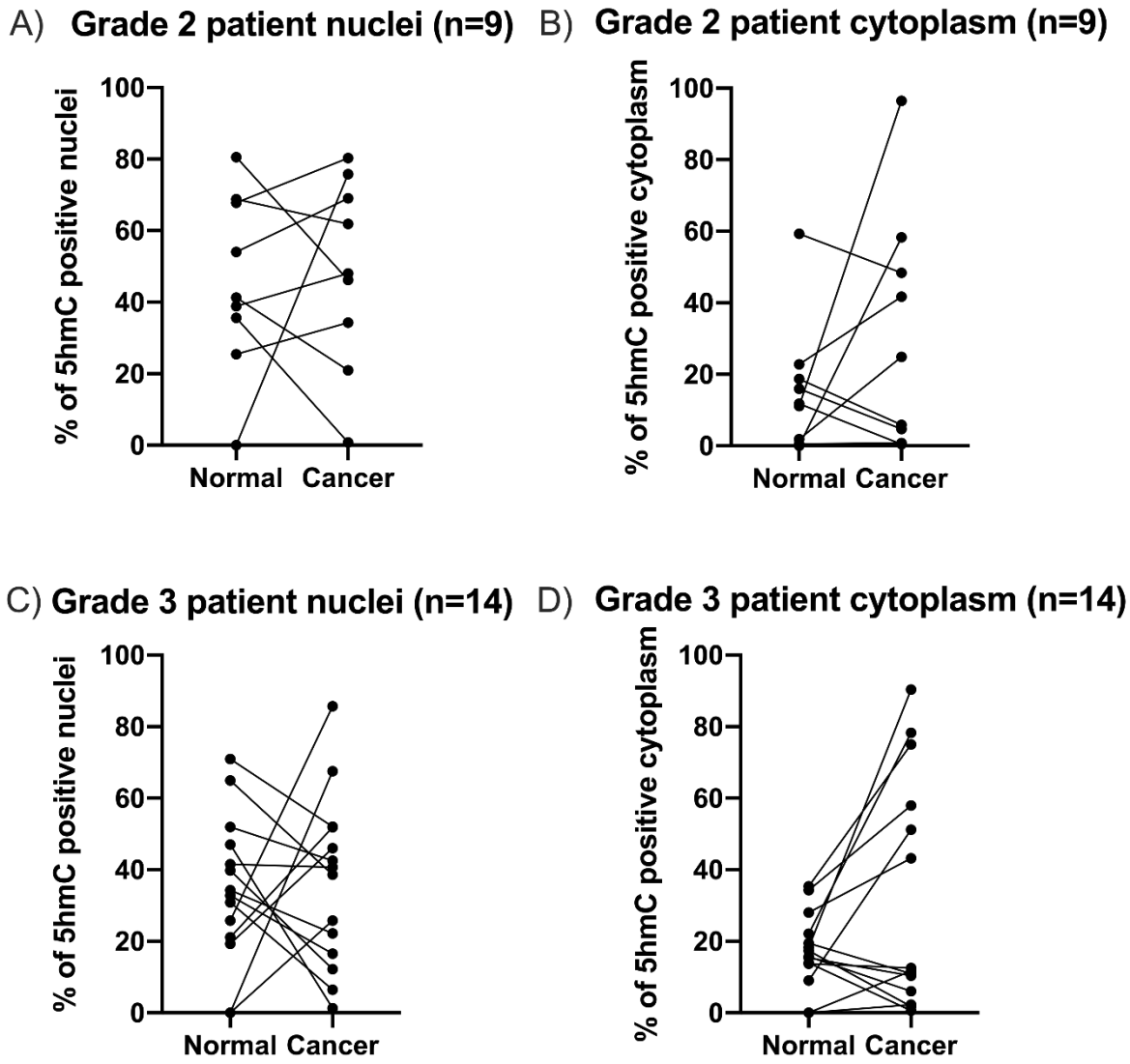


Figure 3.5 The proportion of positive cells of 5hmC after stratification of the samples according to cancer grade. A) and B) are from patients with cancer diagnosed as grade 2. C) and D) are from patients with cancer diagnosed as grade 3. The graphs show the abundance of 5hmC (as percentage of positive cells) in the nuclei or the cytoplasm. Paired T-test was performed only on the samples where matched cancer and normal (non-affected) epithelium was provided. Significance is indicated with an asterisk: * = $p<0.05$.

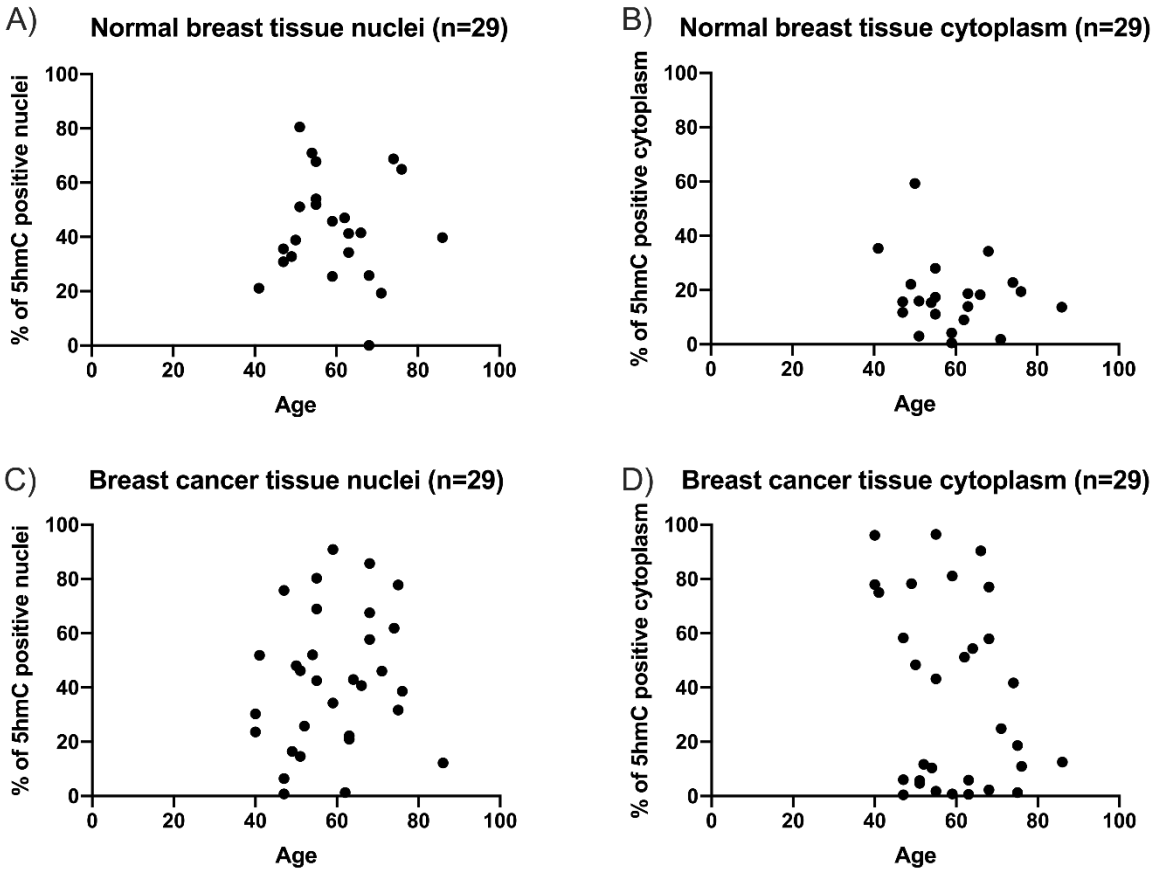


Figure 3.6 The correlation between 5hmC immunohistochemistry staining and age in 29 patients. A) and B) are from normal breast epithelium. C) and D) are from breast cancer epithelium. The graphs show the percentage of positive cells of 5hmC in the nuclei, or the cytoplasm in patients with different age. Statistical analysis was performed Spearman r test on cancer epithelium compared to normal.

3.1.3 Higher proportion of positive cells of cytoplasmic 5hmC in colorectal cancer than quiescent epithelium

To determine if cytoplasmic localisation of 5hmC is breast cancer-specific, IHC against 5hmC was then performed on a cohort of 28 FFPE colitis-associated cancer tissues with matched non-malignant epithelium (defined as quiescent; see patient samples details in the subchapter 2.1, cohort 4, methods in the subchapter 2.2) which were available in the laboratory. IHC staining indicated that 5hmC was distributed in both the nuclei and the cytoplasm in either quiescent or cancer in the colorectal epithelium. Therefore, cytoplasmic 5hmC is not breast cancer specific and this suggests maybe a similar mechanism of action in the demethylation process in colorectal cancer.

Similarly, when compared with breast tissues, the proportion of positive cells of 5hmC was less in quiescent colon tissues compared to cancer tissues. In particular, the proportion of positive cells of nuclear 5hmC was significantly different ($p=0.039$) between the quiescent and cancerous state, confirming the previous study (119, 167). On the other hand, the proportion of positive cells of cytoplasmic 5hmC in cancer was significantly ($p=0.004$) higher than in the quiescent epithelial cytoplasm (Figure 3.7), further confirming the result found in breast cancer.

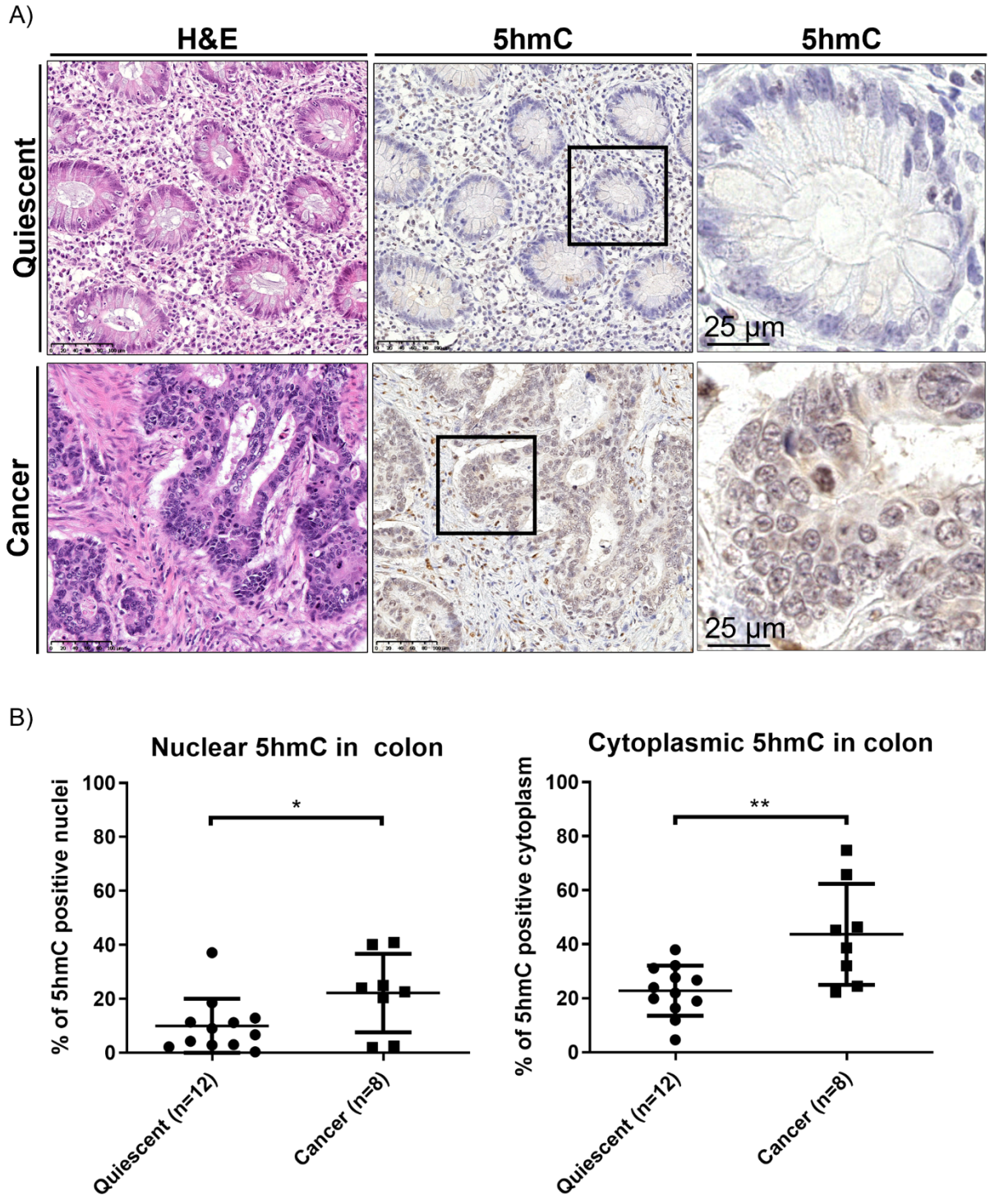


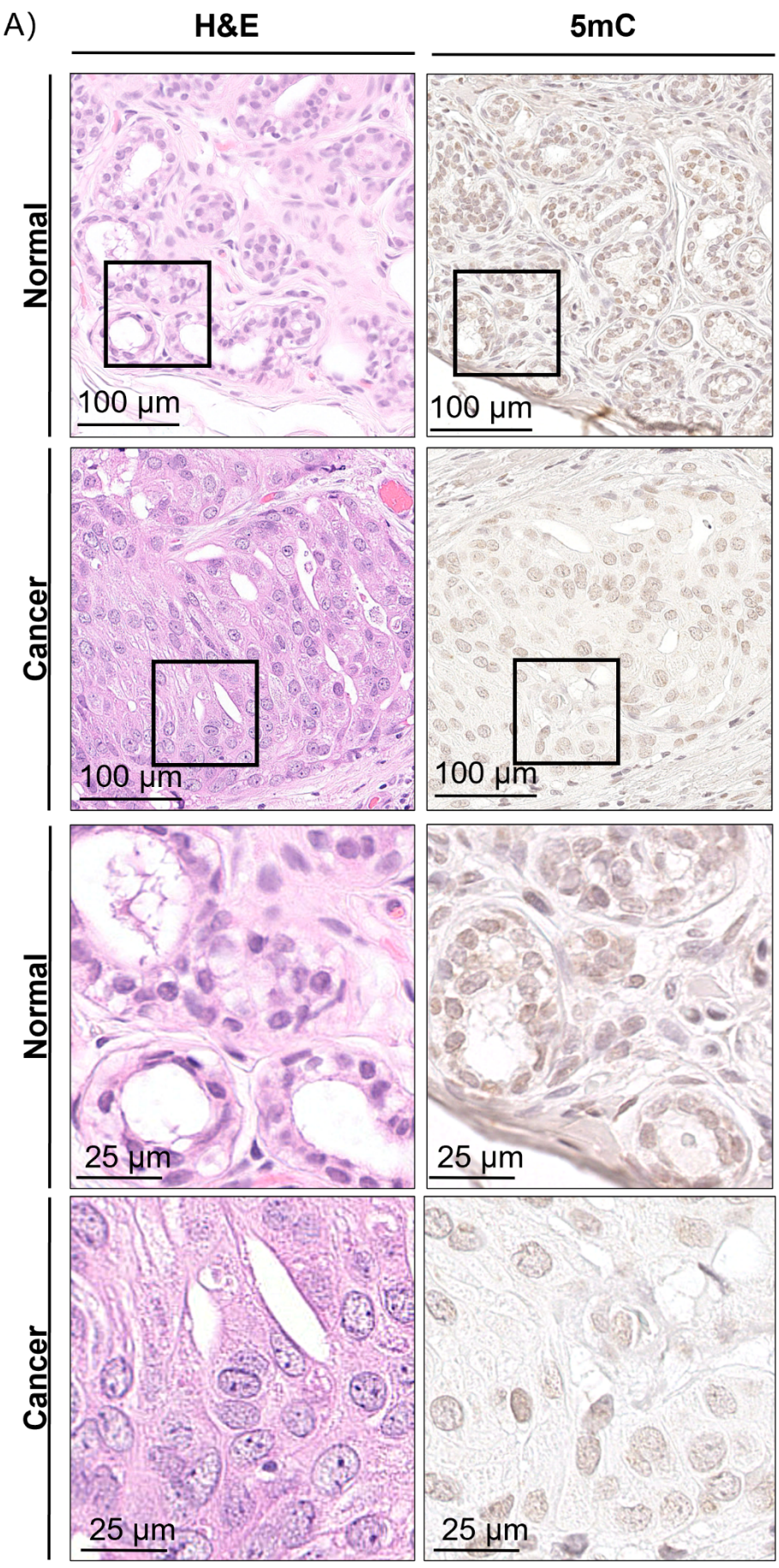
Figure 3.7 5hmC immunochemistry staining in colorectal samples. Panel A) Serial sections from a 39-year-old female patient, with areas of quiescent epithelium (top), and invasive cancer (bottom), (left and middle scale bar = 100 µm, right scale bar = 25 µm). **Graph B)** Scatter plots showing the percentage of positive cells of 5hmC in the nuclei, or the cytoplasm as

calculated with a digital software. The result shown represents the mean ± SD of this group. Statistical analysis was performed Unpaired T-test on each cohort compared to normal and it is indicated with an asterisk: * = $p<0.05$.

3.1.4 Cytoplasmic 5mC is significantly reduced in BC epithelium

Since both 5hmC and 5mC were reported to be reduced in cancer, and 5hmC is oxidised from 5mC by TETs, according to the previous study, the basal-like tumours have low 5mC level while Luminal B tumours display a high 5mC levels (102, 168). Thus, a further aim of this study was to verify the abundance and distributions of 5mC in the same cohort of patient tissues.

The abundance of 5mC was detected by IHC on the selected tissues from cohort 1. Due to the availability in the laboratory, all six samples are from ER negative patients which are matched with normal and cancer epithelium. However, because of a machine error, the matched samples from one patient cannot be quantified. The proportion of positive cells of nuclear 5mC was not significantly different in the cancer epithelium compared to the normal tissue, which contrasts with the findings of the previous study (102). On the other hand, the proportion of positive cells of cytoplasmic 5mC was significantly lower ($p=0.004$) in the cancer epithelium than the normal tissue (Figure 3.8). However, the proportion of cells showing positive 5mC staining in the cytoplasm remains at a very low level (less than 3%). Since 5mC is mainly present in nuclei, this suggests that hydroxymethylation operates with a different mechanism to regulate breast cancer progression.


A)
H&E
5mC
Normal
Cancer
Normal
Cancer
100 µm
100 µm
100 µm
100 µm
25 µm
25 µm
25 µm
25 µm

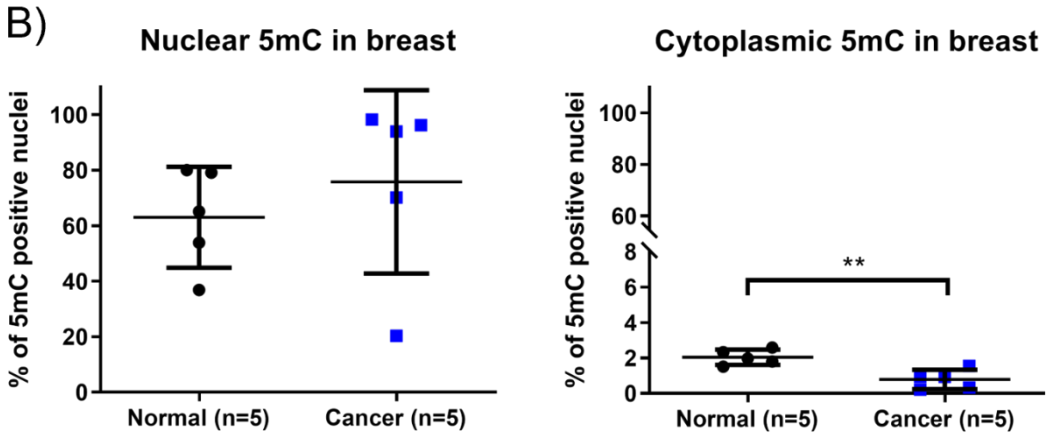


Figure 3.8 5mC immunohistochemistry staining. Panel A) Serial sections from a 55-year-old female patient (the same patient from Figure 3.1), with areas of invasive breast cancer and normal epithelium. (Scale bar = 100 µm). The H&E panel is duplicated as the panel in Figure 3.1 A). **Graph B)** Box plots showing the percentage of positive cells of 5mC in the nuclei, or the cytoplasm as calculated with a digital software. The result shown represents the mean ± SD of this group. Statistical analysis was performed Unpaired T-test on each cohort compared to normal and it is indicated with an asterisk: ** = $p<0.01$.

Finally, 5fC and 5caC are the next two products of the demethylation process. To determine whether 5fC and 5caC localised in the nucleus or the cytoplasm, IHC staining of 5fC and 5caC was also performed on the sample cohort 1. However, staining for 5fC and 5caC was unsuccessful, with no obvious expression detected in the breast tissue (data not shown). This could be because the samples from cohort 1 were sectioned over two years before the experiment, therefore, the quality of the tissue diminished over time. Another reason could be that the expression of 5fC and 5caC is extremely low, so the modification of 5fC and 5caC is undetectable with IHC. To determine the impact of 5fC and 5caC, the IHC staining on fresh tissues or newly cut FFPE sections should be repeated.

3.1.5 Global levels of 5hmC are reduced in breast cancer cells compared with normal cells

Previous studies suggest that the levels of 5hmC are reduced in breast cancer epithelia compared to normal tissues (119, 167, 169, 170). However, according to my previous IHC experiment (see subchapter 3.1.1), I hypothesised that 5hmC is not actually lost in the cancer cells, but it may be redirected to the cytoplasm. Based on the previous figure (Figure 3.1), the average proportion of 5hmC positive stained cells across all samples was calculated. The average of cytoplasmic 5hmC in normal epithelial cells is approximately 19%, while the average of cytoplasmic 5hmC in the cancer cells is approximately 40%. Additionally, the proportion of positive cells of nuclear 5hmC did not significantly change between normal and cancer tissues.

However, few studies compared the changes of 5hmC levels *in vitro*, between cancer cell lines to normal cell lines. It would be useful for further study if I can test the 5hmC changes in cell lines are the same as the changes in tissues, so that the cell lines can be used for functional studies. Moreover, based on the previous result (see subchapter 3.1.2), the proportion of positive cells of cytoplasmic 5hmC in ER negative patients is significantly higher ($p=0.048$) in cancer compared to normal epithelium. I hypothesised that the distribution of 5hmC in ER negative cell lines may be more stable than in other subtypes. Therefore, to investigate whether 5hmC levels change between normal and cancer breast cell lines, DNA was extracted from one non-tumorigenic epithelial cell line (MCF10A, triple negative cell line) and one breast cancer cell line (BT474, ER- cell line), and a colorimetric assay was performed, as described in the methods (see the methods in the subchapter 2.6). Because cytoplasmic 5hmC was significantly increased in tissues from ER- patients in terms of my previous experiment (see result subchapter 3.1.2) but the other subtypes did not show this difference. For this

reason, it would be interesting to first determine if the amount of 5hmC in DNA changes in ER- cell lines. As the ER- cell line, BT474 was selected for the colorimetric assay. In three independent biological experiments, the colorimetric assay was performed to measure the global 5hmC compared to the total cytosine (5hmC/C) content of the DNA using Sunrise™ microplate reader, and the data were analysed by the unpaired T-test on individual cell lines. According to the manufacturer's user guide, the standard curve was calculated by logarithmic second order regression (see subchapter 2.6.3):

$$y = 0.3834 \ln(x) + 2.5764$$

Note that the read out of observation was defined as y and the concentration of 5hmC was defined as x.

The concentration of 5hmC was calculated by using the following formulas:

$$x = e^{\frac{y-2.5764}{0.3834}} \times 100\%$$

The average global 5hmC/C content of DNA in MCF10A was 0.011%, significantly (p=0.018) higher than the global 5hmC/C content of DNA in BT474, which was 0.009% (Figure 3.9), confirming the previous study (72). Although, based on IHC (see subchapter 3.1.1 and 3.1.2), the levels of nuclear 5hmC did not show a significant decrease between normal breast and breast cancer tissues, the average of nuclear 5hmC level in normal is higher than in cancer samples. The levels of 5hmC tends to decrease in cancer, while the cytoplasmic 5hmC in cancer is higher than in normal epithelia. This result may indicate a balance between the hydroxymethylation in DNA and RNA.

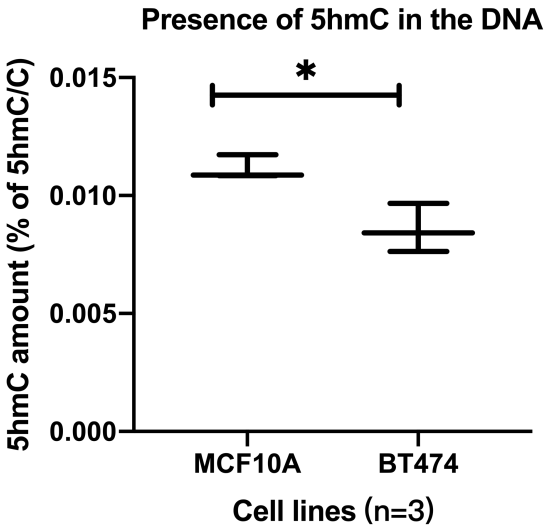


Figure 3.9 Differences in the amount of 5hmC in the DNA of breast cell lines. The amount of 5hmC, measured with a colorimetric assay, is significantly higher in MCF10A than in BT474. Data represent the mean and SD of three independent biological repeats performed in technical duplicates. Statistical analysis was performed with Unpaired T-test on individual cell lines, and it is indicated with an asterisk: * = $p<0.05$.

3.1.6 The localisation of demethylation marks is differentially expressed in breast cell lines

Through IHC experiments, 5hmC was observed in the nucleus and the cytoplasm of the breast epithelium (see subchapter 3.1.1). However, a limitation of the technique is the difficulty in visualising the subcellular localisation of the different modifications directly on the tissue, in particular of 5fC and 5caC, which are not easily detectable. For these reasons, I determined that immunocytochemistry (ICC) might be a more appropriate method to show the localisation of methylation and demethylation marks at a higher resolution, even if it may not allow the quantification of these marks. The reason for this relies on the variability of the integrated intensity related to the brightness and focus of the cells, and based on my previous observation, the variability of 5hmC abundance in the cell lines. Therefore, on this occasion, the ICC experiment was determined to be a qualitative experiment, not a quantitative experiment.

To determine the localisations of 5mC, 5hmC, 5fC and 5caC in the cell lines, I performed ICC on the breast cancer cell line BT474 and non-tumorigenic breast cell line MCF10A using the same antibodies mentioned above. Confirming the IHC results, 5hmC localised in both the nucleus and the cytoplasm in both cell lines, but cytoplasmic 5hmC showed a higher distribution in cancer cells compared to normal cells (Figure 3.10). The same distribution was observed for 5fC and 5caC. However, previous studies reported 5fC and 5cac are present in the nucleus in human embryonic stem cells, hepatic progenitors, and brain tumour cells (171, 172), suggesting the different localisation observed here is cell type specific. Although a previous study did not report the presence of 5caC in the cytoplasm, it showed that 5caC localised in the nucleus in normal breast tissue, and in both the nucleus and the cytoplasm in breast cancer tissue (104). In both normal and cancer cell lines, 5mC localises in the nucleus,

but 5mC localisation was restricted to the heterochromatin regions in the normal cells, confirming the previous study in mESCs (173).

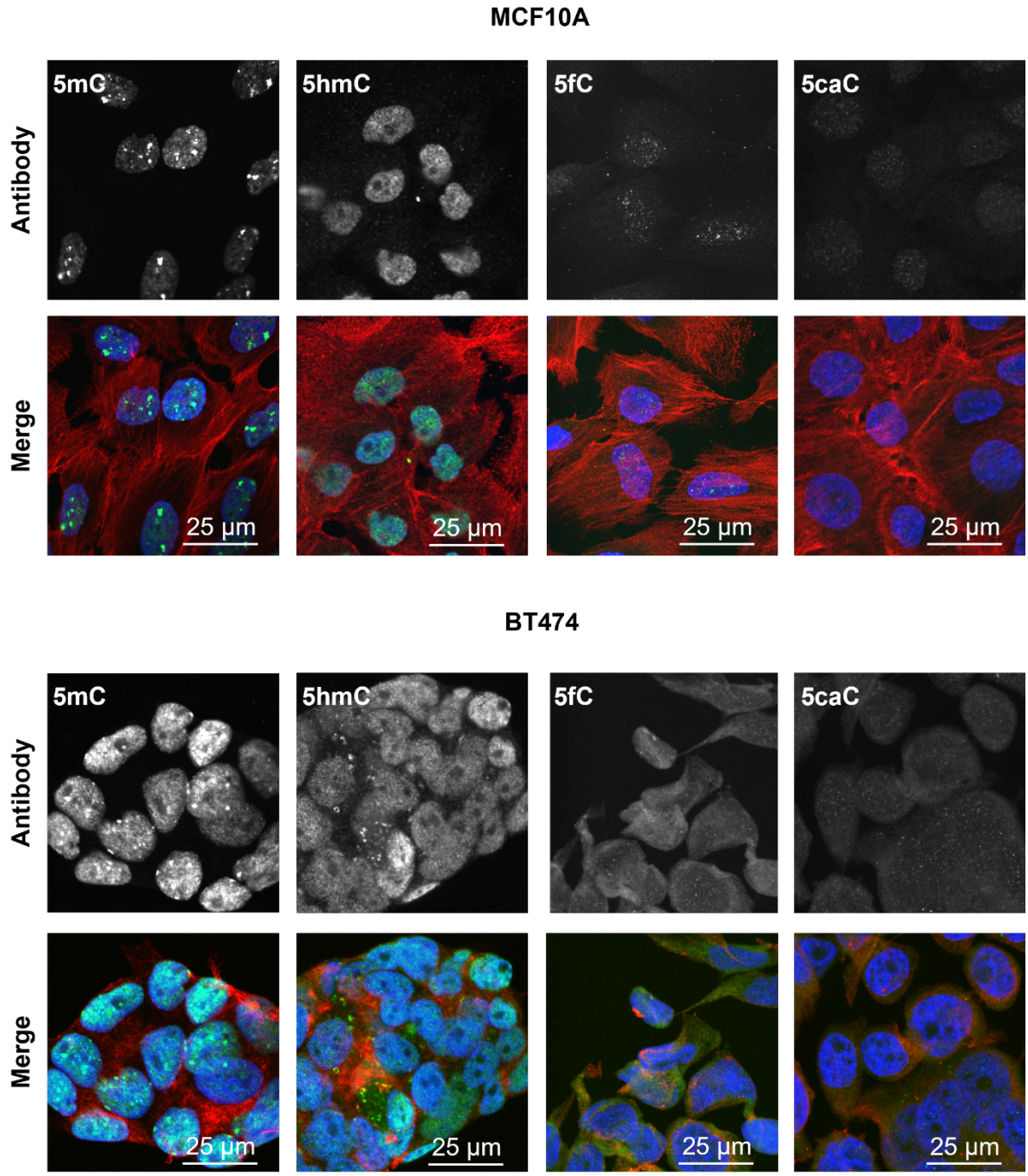


Figure 3.10 5mC, 5hmC, 5fC and 5caC immunocytochemistry staining. In the "Antibody" panel, representative images of 5mC, 5hmC, 5fC, 5caC staining, respectively. The "Merge" panel shows the combination of blue (Hoechst-stained DNA), green (corresponding antibody) and red (β–actin) signals in both MCF10A and BT474 cells. (Scale bar = 25 µm)

3.1.7 5hmC distributes in DNA and RNAs in cell lines

According to the experiments described above, the distribution of 5hmC in the breast epithelium indicates that 5hmC may regulate not only DNA in the nucleus, but also the RNA, which is present in the cytoplasm. While my previous colorimetric experiment was able to quantify the levels of 5hmC in the DNA *in vitro*, cell lines may not be an ideal model to quantify the presence of 5hmC accurately due to the variability of 5hmC amount, as described above and in the previously (174). However, I believe it could still be used to evaluate the presence of 5hmC in the DNA and RNA with a non-quantitative method.

To take this further, to verify whether 5hmC is a modification characteristic of different classes of RNAs, dot blots were performed on both DNA and different forms of RNA isolated from BT474 and MCF10A. Confirming this hypothesis, 5hmC was detected not only in the DNA, but also in the total RNA, in the long and short RNA (including mRNA) in both MCF10A and BT474 (Figure 3.11). This is the first time that 5hmC is found in RNA in breast cells.

However, after quantifying the relative 5hmC intensity by comparing the positive control, I did not observe a significant difference of the 5hmC amount between normal and cancer cell lines. The trend of the amount of 5hmC in BT474 is higher than in MCF10A in the dot blot, while the amount of 5hmC in DNA in BT474 was significantly lower than in MCF10A in the colorimetric assay (see subchapter 3.1.5). The inter- and intra- experimental variation may be due to the viability of 5hmC between the cell lines, which was previously found (174). Figure 3.11 B) shows the presence of 5hmC in 50 ng and 25 ng of DNA in MCF10A and BT474. The intensity of the dots of 12.5 ng was not possible to quantify. As Figure 3.11 C) showed, the signal of 1000 ng total RNA was also not quantifiable. Moreover, the distributions of long and short RNAs in Figure 3.11 C) and D) are different, I observed that short RNAs have more

intense dots than long RNAs in one repeat shown in Figure 3.11 C), whereas in another repeat shown in Figure 3.11 D), long RNAs have more intense dots than short RNAs. This result may prove the variability of 5hmC within the same cell line in biological replicates and possibly indicate an unfound mechanism behind this situation. I did not do further experiments to investigate the reason for this result due to time limitations.

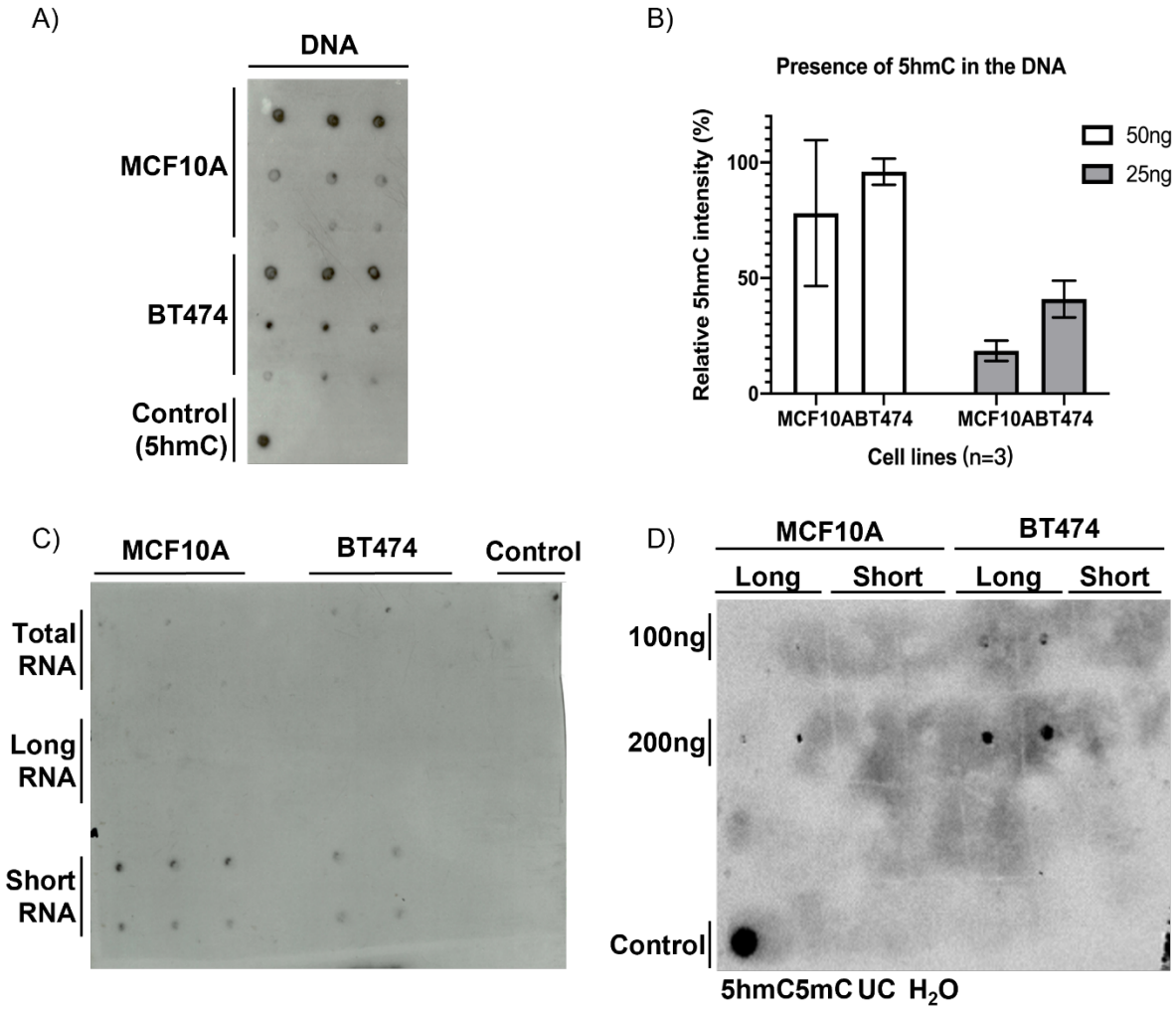


Figure 3.11 5hmC levels in the DNA and RNA of human breast cell lines. DNA or RNAs are crosslinked to an Hybond-N+ membrane with different concentrations. Panel A) Dot blot and quantification of 5hmC modification in 50 ng DNA, 25 ng DNA, 12.5 ng DNA of BT474 and MCF10A. Each dot represents hydroxymethylated DNA with three concentrations (50 ng, 25 ng, 12.5 ng DNA in vertical), triplicates at each concentration. **Graph B)** The amount of 5hmC in DNA, measured with dot blot, is not significantly different between two cell lines. Data represent the mean and SD of three independent biological repeats performed in technical triplicates. Statistical analysis was performed with an unpaired T-test on individual cell lines. **Panel C)** Dot blot of 5hmC modification in 1000 ng RNA, 200 ng long or short RNAs of BT474 and MCF10A, with triplicate at each concentration. **Panel D)** Dot blot of 5hmC modification in 100 ng and 200 ng (in vertical) long or short RNAs of BT474 and MCF10A, with duplicate at each concentration. 0.15ng 5hmC (positive control), 0.15 ng 5mC (negative control), 0.15 ng unmethylated cytosine (UC, negative control) and 0.15 ng H_20 (negative control) are shown.

3.2 Discussion

In this chapter, we show results of immunohistochemistry (results see subchapter 3.1.1 and 3.1.3) that was performed to quantify 5hmC in breast and colorectal tissues, and a colorimetric assay (results see subchapter 3.1.5) was undertaken to determine the global level of 5hmC in breast cell lines. Then, immunocytochemistry (results see subchapter 3.1.6) was performed to show the localisations of demethylation marks in cell lines. Finally, dot blot (results see subchapter 3.1.7) was used to detect the hydroxymethylation in DNA and different RNAs.

The global 5hmC/C content in the DNA, as measured by the colorimetric assay, reflects the same range reported by the previous study by Thienpont and colleagues, which indicates the colorimetric assay could be an equally sensitive method to replace the liquid chromatography-mass spectrometry (LC-MS) (175). However, this result differs from previous experiments performed in the laboratory in terms of variability (data not shown). Previous data showed a substantial variability amongst the levels of 5hmC in the biological repeats of individual cancer cell lines (BT474, MCF7, SKBR3) with different hormonal receptor status, and the amount of 5hmC in a normal cell line (MCF10A) did not show a significant ($p=0.084$) difference compared to any cancer cell lines. The difference between the two results suggests that 5hmC is variable amongst cell types, cell passage and cell cycle. Thus, cell lines may be an unreliability model for the study of 5hmC, confirming previous studies by Nector and colleagues (174).

Both 5hmC and 5mC are differently abundant and distributed in breast tissues, confirming a previous study by Zhang and colleagues (169, 176). Previous studies reported global 5hmC is reduced in breast cancer tissues compared to normal tissues, which is discordant with my

results. However, the limitation of these studies is not considering the sublocalisations of 5hmC.

Furthermore, the data presented in this chapter show that 5hmC is not only involved in modification of the DNA but could be involved in further regulatory pathways within the cytoplasm. Previous work has demonstrated that 5hmC is present in the cytoplasmic mRNA in *Drosophila* and in mouse embryonic stem cells (97, 177). Here I showed that 5hmC is present in the DNA, total RNA, long RNA, and short RNA. Interestingly, the presence of 5hmC in different classes of RNA is highly variable amongst three independent biological repeats. This observation further confirms that cell lines might not be a completely suitable model for the study of 5hmC (107, 174, 175). Also, it was not possible to detect the abundance of 5hmC in mRNA, which could possibly be explained by the lack of reliability of cell lines or the fact that 5hmC in mRNA in breast cells is very rare. In terms of the colorimetric assay, the amount of 5hmC in DNA is quantifiable using 100 ng DNA which amount is 4 times more than the dot blot experiments. Thus, I estimated that it may need more than 4000 ng RNA to quantify the amount of RNAs. Moreover, a suitable kit may still be in the development stage. To better detect and quantify the abundance of 5hmC, LC-MS can be a reliable method in the future. According to these data here, the changes of 5hmC may be correlated to ER status which could be a piece of evidence to support previous studies (161). The changes of 5hmC are not correlated to the age of surgery. Since the age of patients from cohort 1 and cohort 2 is mostly between 40 to 80, future works could add more samples from patients younger than 40 to confirm whether the age at the surgery is a factor. Moreover, lack of samples number in different grades, the current result did not show an association between 5hmC levels and grades. There is only a trend that the proportion of positive 5hmC stained cells depend on ER status in grade 3 tumours. Therefore, future experiments should add more samples from

different grades to confirm whether 5hmC changes are related to the grade of BC, in a larger sample set too including an independent validation set.

Comparing these data with the previous studies, my work has some advantages and limitations too. For the methodology, while previous studies used the adjacent tissue sample as the standard control, my experiment used the contralateral tissue as the standard control (96, 161). Although, Aran and their colleagues showed that the adjacent tissues can be used as the "normal" control, this could obscure some therapeutic targets in both adjacent and cancer tissue (161). Since the molecular mechanism of hydroxymethylation in cancer progression is still unknown, the contralateral tissue may be more appropriately categorised as "normal" than the adjacent tissue. Furthermore, the research here examined a sublocalisation of 5hmC and 5mC in breast epithelium, whereas previous studies only reported the global levels of 5hmC in tissues.

However, a limitation is that this experiment analysed the proportion of positive cells for global 5hmC in both nuclei and cytoplasm but did not analyse the overall abundance of 5hmC in whole cells. Therefore, it is difficult to directly compare if the 5hmC amount in total increases or decreases in cancer tissues compared to the normal contralateral tissue. Future work should involve IHC staining on more samples to increase the sample size to confirm these findings and should also involve the quantification of the 5hmC amount in total cells and also its sublocalisation.

From the ICC experiment, 5hmC distributes in both the nuclei and cytoplasm in the breast cancer cell line, BT474, but mostly in nuclei in the non-malignant breast cell line, MCF10A. Cytoplasmic 5hmC in both breast cancer tissues and cell lines suggests a hydroxymethylation in RNAs. This result should be further analysed, however, the small cytoplasm of BT474 cells

may have affected the staining strength and the interpretation of staining (Figure 3.10). There is less space for the target antigen in BT474, which may result in reduced staining intensity. Additionally, the limited cytoplasmic space may result in a more diffuse and less pronounced staining pattern. These challenges make it difficult to assess accurately the actual amount of the target molecule in the cytoplasm. Therefore, I did not perform an analysis of ICC staining. Another limitation of this experiment is that the expression of 5hmC or other demethylation marks should be blocked as a negative control to compare the changes of expression.

3.3 Conclusion

This chapter describes the distribution and localisations of 5hmC in both human breast and colonic epithelium, and breast cell lines. Using immunohistochemical staining, I found out a higher proportion of positive cells with cytoplasmic 5hmC staining in the cancer epithelium, which was not observed before. This is also the first study to determine hydroxymethylation in different classes of RNAs in breast cells using dot blot, which suggests a contribution of 5hmC in RNA processing and its clinical relevance.

Based on my results and the existing previous studies, as a modification in RNA, hydroxymethylated RNA may regulate the transcription and promote gene expression and differentiation. Lan and their colleagues suggested that TET-mediated hydroxymethylation contributes the embryonic stem cell differentiation (164). Broome and their colleagues reported that TET2 contributes to ER/GATA3 transcriptional activity and maintained 5hmC levels in ER+ mouse models (166). Moreover, Kim and their colleagues suggested TET2-FOXP1 complex mediates luminal cell differentiation, and the reduction of 5hmC is associated with *ESR1* repression (178). Previous studies concluded that the phenomenon of TET mediated hydroxymethylation in DNA or RNA may be related to ER status or *Esr1* genes. More evidence is needed to determine whether the changes in 5hmC levels are cause or consequence. Thus, I would hypothesise TET proteins, specially TET2, may act as maintenance of 5hmC in ER+ breast cancer to regulate the transcription of GATA3, FOA1, and other transcriptional factors.

Investigating the abundance and distribution of other methylation and demethylation marks is also valuable. In particular, I described how 5fC and 5caC are localised in both the nucleus and the cytoplasm, but this needs further experiments to be validated. Without the use of sequencing techniques or more accurate quantification methods, it is not possible to study

the role of 5fC and 5caC on the genomic level. However, since 5fC and 5caC are 10 to 100 times less than 5hmC (179), I believe that 5hmC, together with the more commonly studied 5mC, is the main important regulator in cancer progression. Thus, I decided to focus on examining the role of 5hmC rather than other demethylation marks in future chapters.

4. Result 2:

TET proteins and 5-hydroxymethylation in breast cells

Here we present results related to aim 2, where we aimed to test whether the variation in the levels of 5hmC in BC can be attributed to one specific TET or to all three TET proteins. The aims of this chapter are to present results related to work that aimed: a) to determine the expression of TET proteins in breast tissues and cell lines; b) to infer any correlation between the proportion of positive cells of 5hmC and the three TET proteins; c) to examine whether *TET* mRNA expression is related to a difference in survival in BC patients.

In the previous chapters, I provided evidence of the distribution and localisation of 5hmC and other demethylation marks. Considering the distribution of cytoplasmic 5hmC in the cancer epithelium and the presence of hydroxymethylation in different classes of RNAs in breast cells, the focus of this chapter is on 5hmC, rather than 5fC and 5caC.

To enhance comprehension of the role of 5hmC, the understanding of the contribution of the TET enzymes to the demethylation process is essential. As mentioned above, during DNA demethylation, TET proteins convert 5mC into 5hmC, 5fC and 5caC with the help of the cofactors Fe (II) and 2-oxoglutarate (2-OG). Therefore, the expression, regulation of activity and the interplay of each TET may affect the final levels of 5hmC.

In recent years, TET proteins were reported to have functions or mediating hydroxymethylation in different cancers, including BC. Zhang and colleagues reported a negative correlation ($p<0.05$) between 5hmC and TET2 expression in BC patients and a positive correlation ($p<0.001$) in patients carrying precancerous breast lesions (176). This

suggests that TET2 has different effects in patients with different pathological stages. Broome pointed out that depleting TET2 expression decreases the levels of 5hmC in ER+ BC, suggesting a possible mechanism for TET2 regulating 5hmC modification (180). Wu and colleagues showed that loss of nuclear TET2 and miR-200c expression is negatively correlated ($p<0.05$) with all-trans retinoic acid (ATRA) resistance in high tumour grade and aggressive BC (181). This result indicates microRNA may participate in the alteration of 5hmC and TET2 proteins. Besides, Good and colleagues discovered an isoform of TET1, $TET1^{ALT}$ (131). $TET1^{ALT}$ has been reported overexpressed in triple-negative BC, and knockdown of *TET1* resulted in the decrease of 5hmC (131). They suggested that TET1, TET2, and TET3 have different targets on the genome, and these targets may have distinctive effects on cancer (131). To sum up, several studies argued the possible functions of different TET proteins for 5hmC distribution and cancer progression, I would like to validate these findings.

4.1 Results

4.1.1 The breast cancer epithelium is characterised by a significantly higher level of cytoplasmic TET2 compared to the healthy breast epithelium

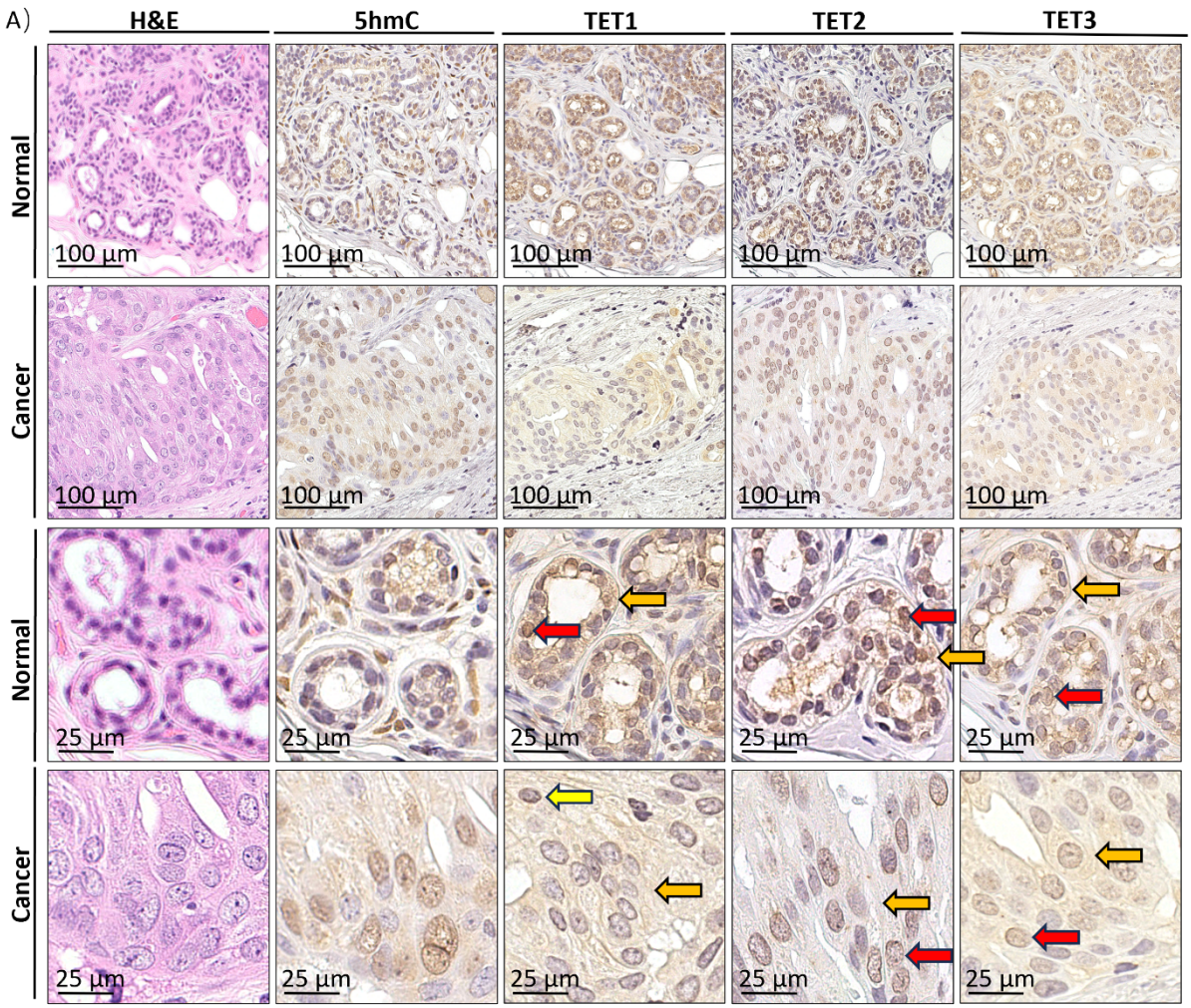
As shown in subchapter 3.1.1 and 3.1.3, a significant increase ($p=0.012$, Figure 3.1 B) of proportion of positive cells of 5hmC was detected in the cancer epithelial cytoplasm compared to the normal epithelial cytoplasm.

To investigate whether hydroxymethylation is associated with a specific TET protein, IHC was performed to detect the expression of TET1, TET2, and TET3 on 16 FFPE samples from 9 post-menopausal patients (cohort 1), a cohort comprised of 10 IDC sections and 6 matched contralateral sections. Because of tissue availability, the assay was performed in this subset of samples rather than the full cohort of 53 samples. MRes student Liliane Eleid contributed to the staining of TET2 under my training and supervision, as part of her practical project.

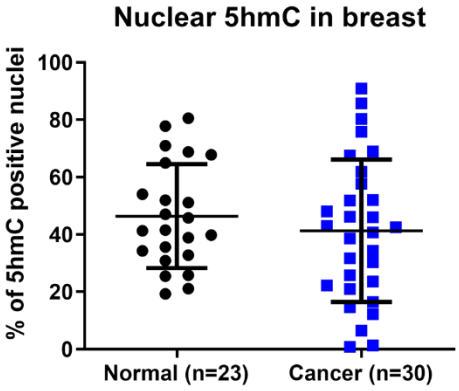
In both nuclei and cytoplasm, some levels of expression could be detected for all three TET proteins, but the expression pattern of TET1 and TET3 is different to the expression of TET2 on the same serial sections (Figure 4.1). The average proportion of positive cells of nuclear or cytoplasmic TET1 or TET3 in the cancer cells was not significantly different than in the normal counterpart. On the other hand, the average proportion of positive cells of cytoplasmic, but not nuclear, TET2 in cancer was significantly higher than in the normal breast tissues ($p<0.001$, Figure 4.1 B). As the changes of the proportion of positive cells of TET2 protein expression follow the same pattern of the proportion of positive cells of 5hmC modification, an association between the levels of 5hmC and the expression of TET2 protein can be hypothesised.

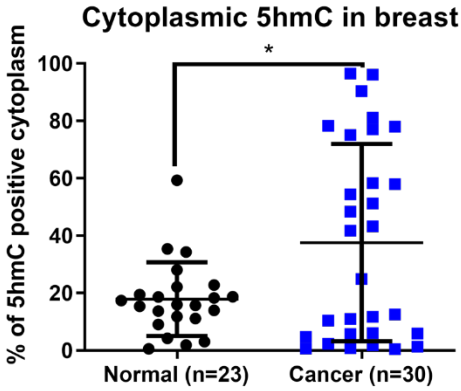
Moreover, multiple comparisons tests were performed to validate if there was any difference between the proportion of positive cells of each TET enzyme against each other. In the cancer sections, the only significant difference was the higher proportion of positive cells of cytoplasmic TET2 ($p<0.0001$, Figure 4.1 C) compared to the proportion of positive cells of cytoplasmic TET1. This result suggests that TET2 may play a more prominent role in cytoplasmic hydroxymethylation.

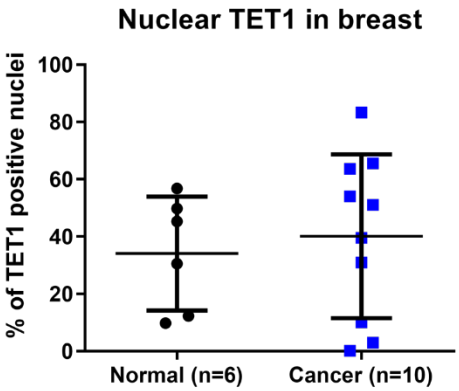
In the previous chapter I provided evidence that the distribution of 5hmC may be associated with the lack of expression of ER in BC tissue, a result of borderline statistical significance (p=0.048, Figure 3.3 D). To validate if the expression of TET proteins is also associated with the hormonal receptor status of the cancer, the samples were selected from ER negative patients. The correlation between the expression of the receptor and the proportion of positive cells for three TET proteins was tested using both unpaired and paired T-test (Figure 4.1 D). Interestingly, only the proportion of positive cell of cytoplasmic TET2 is significantly higher in cancer epithelium compared to normal epithelium in ER- patients. This may suggest in turn that TET2, rather than TET1 or TET3, regulates 5hmC levels in ER- patients. While we acknowledge the importance of examining TET expression in both nucleus and cytoplasm in ER+ tumour samples, the scope of our current study was primarily focused on ER- patients due to time constraints. As discussed in the previous chapter, ER- patients have a significantly higher proportion of positive cells of cytoplasmic 5hmC in cancer compared to normal epithelium. This finding suggests an unknown mechanism of hydroxymethylation of DNA or RNA in ER- patients. It is essential to note that ER+ and ER- breast cancers represent distinct molecular subtypes with different hormone receptor statuses. Future experiments should carry out IHC to test TET expression in more patient samples in different ER+ states/ hormone receptor levels, as this could provide valuable insights into its potential function in different hormonal contexts.

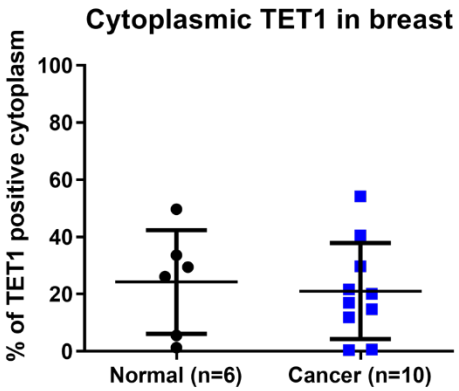

A)
H&E
5hmC
TET1
TET2
TET3
Normal
Cancer
Normal
Cancer
100 µm
25 µm

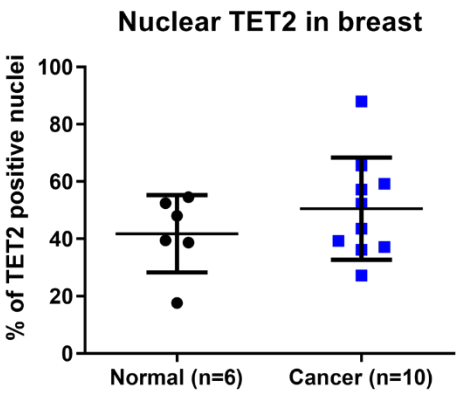
B)

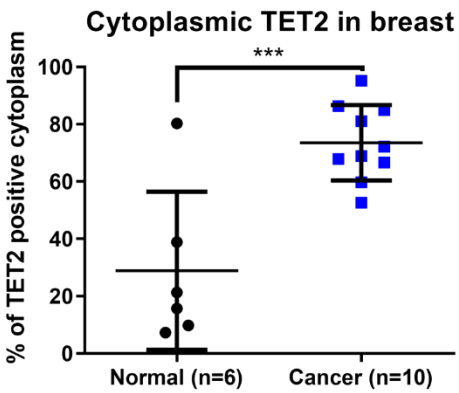

Nuclear 5hmC in breast
% of 5hmC positive nuclei
100
80
60
40
20
0
Normal (n=23)
Cancer (n=30)

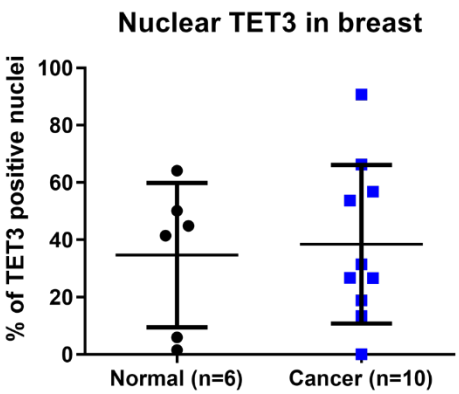

Cytoplasmic 5hmC in breast
*
% of 5hmC positive cytoplasm
100
80
60
40
20
0
Normal (n=23)
Cancer (n=30)

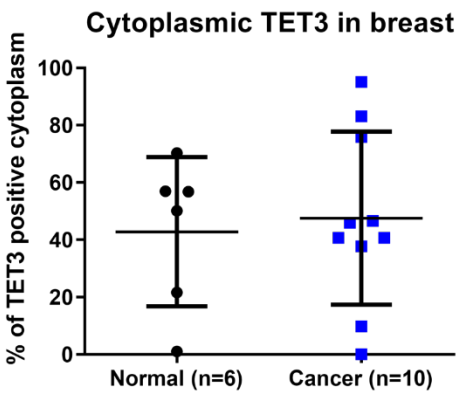

Nuclear TET1 in breast
% of TET1 positive nuclei
100
80
60
40
20
0
Normal (n=6)
Cancer (n=10)


Cytoplasmic TET1 in breast
% of TET1 positive cytoplasm
100
80
60
40
20
0
Normal (n=6)
Cancer (n=10)


Nuclear TET2 in breast
% of TET2 positive nuclei
100
80
60
40
20
0
Normal (n=6)
Cancer (n=10)


Cytoplasmic TET2 in breast
***
% of TET2 positive cytoplasm
100
80
60
40
20
0
Normal (n=6)
Cancer (n=10)


Nuclear TET3 in breast
% of TET3 positive nuclei
100
80
60
40
20
0
Normal (n=6)
Cancer (n=10)


Cytoplasmic TET3 in breast
% of TET3 positive cytoplasm
100
80
60
40
20
0
Normal (n=6)
Cancer (n=10)

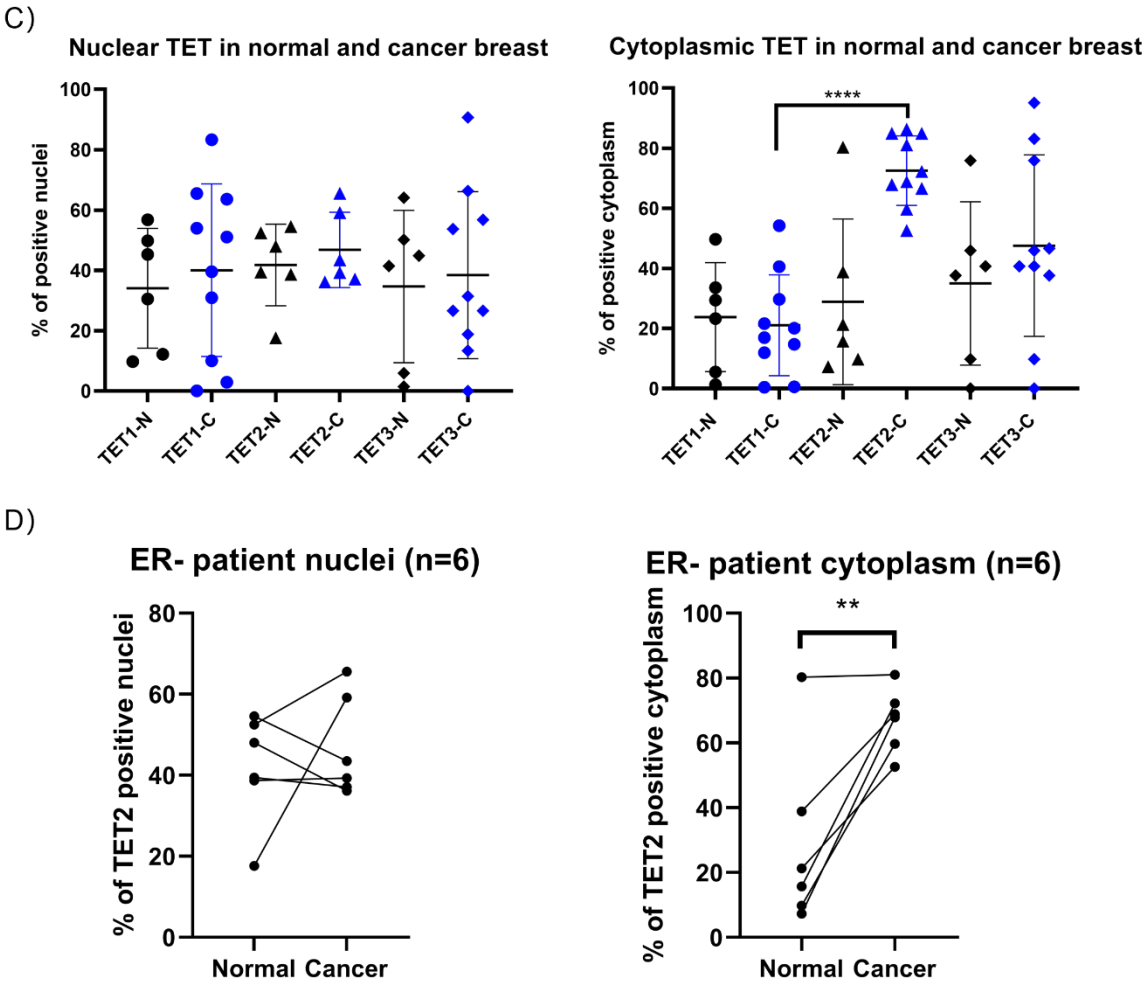


Figure 4.1 5hmC, TET1, TET2, and TET3 immunohistochemistry staining. Panel A) Serial sections from a 55-year-old female patient (the same patient from Figure 3.1), with areas of invasive BC and normal epithelium. Low magnification view (scale bar = 100 µm) and higher magnification view (scale bar = 25 µm). The H&E panel and 5hmC panel are the same as Figure 3.1 A). The red arrow indicates strong positive staining in nuclei and the yellow arrow indicates weak positive staining in nuclei. The orange arrow indicates positive staining in cytoplasm. **Graph B)** Box plots showing the proportion of positive cells of 5hmC (taken from the full cohort of 53 samples analysed in Figure 3.1) and three TET protein in the nuclei or the cytoplasm, as calculated with a digital software. The lines represent the mean and the SD. Statistical analysis was performed with an unpaired T-test on each cohort and statistical significance is indicated with an asterisk: * = $p<0.05$, ***$p<0.001$. From the total of the 9 patients, normal tissue was missing from 3 patients and 1 patient provided more than one

cancer section, thus not allowing a full paired comparison. **Graph C)** Statistical analysis was performed ordinary one-way ANOVA between each cohort and it is indicated with an asterisk: ***p<0.001, ****p<0.0001. N represents normal, and C represents cancer. **Graph D)** The graphs show the TET2 expression (as percentage of positive cells) in the nuclei or the cytoplasm in ER- patient samples. Paired T-test was performed only on the samples where cancer and matched normal (non-affected) epithelium was provided. Significance is indicated with an asterisk: ** = p<0.01.

4.1.2 The association between TET proteins and 5hmC distributions in breast epithelium

To demonstrate the association between hydroxymethylation and TET protein expression, I further analysed the correlation between the proportion of positive cells of 5hmC and the proportion of positive cells of TET proteins based on the IHC results. Data (12 samples including 6 normal and 6 cancer tissues) are aggregated from the results of IHC on 5hmC, TET1, TET2 and TET3 (see results in subchapter 3.1.1 and 4.1.1). Since the data passed the normality test, I performed a Pearson's Correlation Coefficient test on these data. I created a heat map of the correlation between the proportion of 5hmC positive cells and TET protein positive cells, both in the nucleus and the cytoplasm, across cancer and normal tissues (Figure 4.2).

In the normal breast epithelium, the proportion of positive cells of nuclear 5hmC are not significantly negatively correlated with the proportion of positive cells of three TETs. In comparison, the proportion of positive cells of cytoplasmic 5hmC are positively correlated with the proportion of positive cells of three TETs. In the cancer epithelium, the proportion of positive cells of nuclear 5hmC are significantly negatively correlated with the proportion of positive cells of TET2 (ρ=0.021) and not significantly correlated with the proportion of positive cells of TET1 (ρ=0.280). However, the proportion of positive cells of nuclear 5hmC are significantly positively correlated with the proportion of positive cells of TET3 (ρ=0.004) proteins. Moreover, the proportion of positive cells of cytoplasmic 5hmC is negatively correlated to the proportion of positive cells of TET2 (ρ=0.006) and not significantly negatively correlated to the proportion of positive cells of TET1 (ρ=0.584) while the proportion of positive cells of cytoplasmic 5hmC is not significantly positively correlated to the proportion of positive cells of TET3 (ρ=0.622) in cancer epithelial sections.

Moreover, only in the normal breast epithelium, the proportion of positive cells of nuclear TET1 and TET3 showed a significantly positive correlation (ρ=0.001). It suggested that nuclear TET1 and TET3 interact with each other and may have the same synthetic process. On other occasions, TET1, TET2, and TET3 did not show a significant correlation, which may indicate that these three proteins act differently in normal and cancer hydroxymethylation.

Furthermore, the data was analysed the differences between two states (paired normal and cancer) (Figure 4.3).

Delta 5hmC in the nuclei was calculated as:

% 5hmC positive nuclei in cancer – %5hmC positive nuclei in normal.

Delta 5hmC in the cytoplasm was calculated by:

% of 5hmC positive cytoplasm in cancer – % of 5hmC positive cytoplasm in normal.

Delta TET1/2/3 in the nuclei was calculated by:

% of TET1/2/3 positive nuclei in cancer – % of TET1/2/3 positive nuclei in normal.

Delta TET1/2/3 in the cytoplasm was calculated by:

% of TET1/2/3 positive cytoplasm in cancer – % of TET1/2/3 positive cytoplasm in normal. Only delta TET2 in the nuclei showed a significantly negatively correlation (ρ=0.039), while delta TET2 in the cytoplasm showed a non-significantly positively correlation (ρ=0.088). The correlation between delta 5hmC and delta TET2 in the cytoplasm confirmed that the changes of the proportion of 5hmC positive cells and TET2 positive cells were positively correlated in ER- patients. Confirming this hypothesis, 3 proteins contribute differently to the 5hmC alterations in breast cells. I need more data to validate how these three proteins work with hydroxymethylation in breast cells.

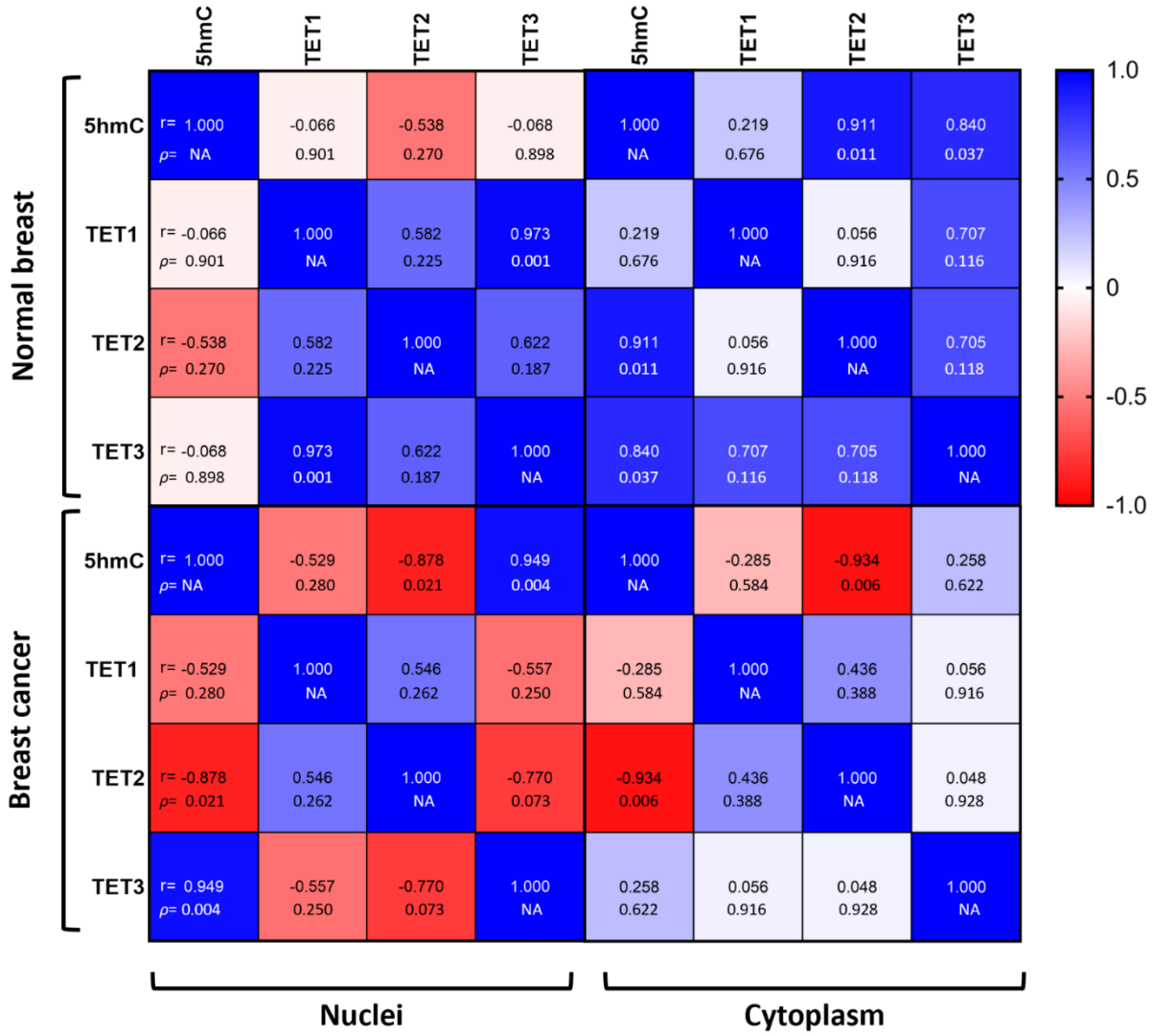


Figure 4.2 The association between the proportion of positive cells of 5hmC, TET1, TET2, and TET3 in six breast epithelial tissues. Heatmap shows the correlation between the proportion of positive cells of 5hmC and the proportion of positive cells of TET1, 2 and 3 in the normal breast epithelial nuclei and cytoplasm, breast cancer nuclei and cytoplasm. The intensity of the staining was calculated with digital software. The result shown represents the mean ± SD of each sample. Correlation matrix was performed by Pearson correlation coefficients. r represents the correlation coefficient, and *ρ* represents p value.

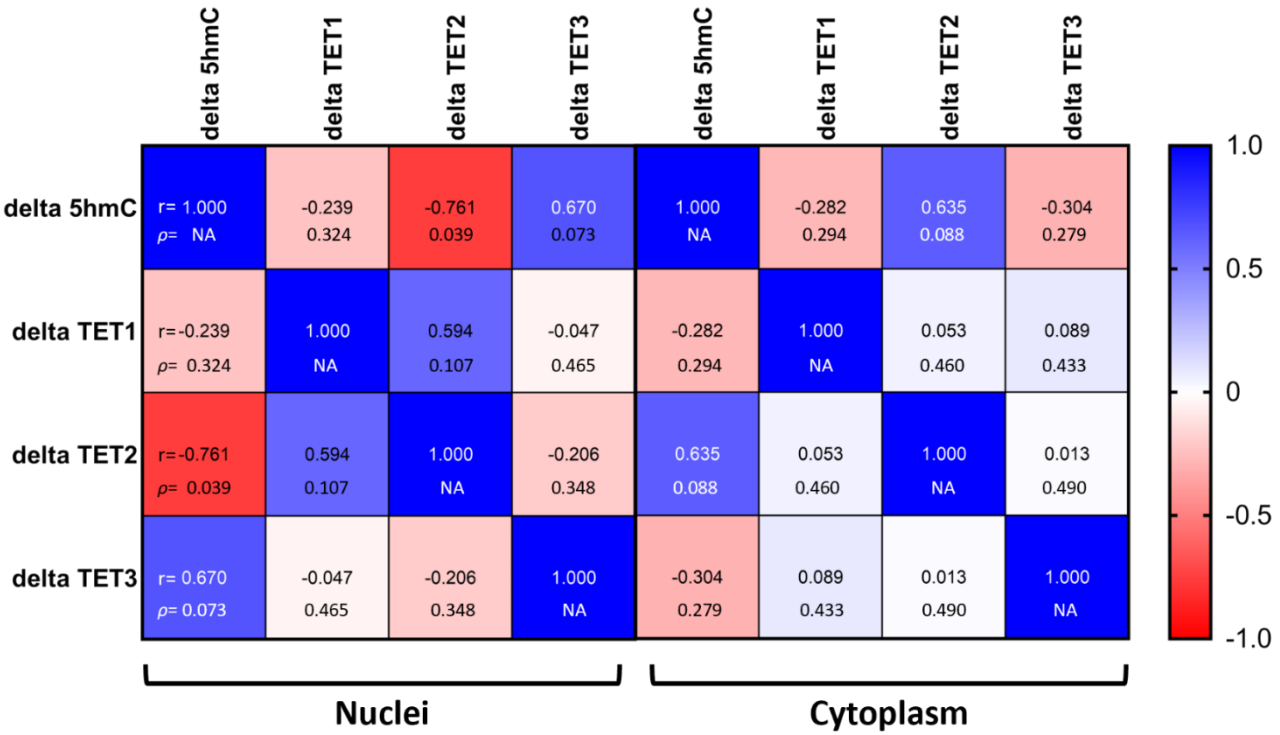


Figure 4.3 The association between delta 5hmC, delta TET1, delta TET2, and delta TET3 in six breast epithelial tissues. Heatmap shows the correlation between the proportion of positive cells of 5hmC and the proportion of positive cells of TET1, 2 and 3 in the normal breast epithelial nuclei and cytoplasm, breast cancer nuclei and cytoplasm. The intensity of the staining was calculated with digital software. The result shown represents the mean ± SD of each sample. Correlation matrix was performed by Pearson correlation coefficients. r represents correlation coefficients and ρ represents p value. delta 5hmC represents the changes in the proportion of positive cells showing 5hmC staining between normal and cancer tissues; delta TET1/2/3 represents the changes in the proportion of positive cells of TET1,2,3 between normal and cancer, individually.

4.1.3 TET proteins are expressed in both the nuclei and cytoplasm of breast cell lines.

In tissues (see subchapter 4.1.1), TET proteins were detected in the nucleus and cytoplasm of the breast epithelium. However, technical limitations of IHC do not allow visualisation of the subcellular localisation of each protein on the samples. As the cytoplasmic TET2 in ER negative BC patients was significantly higher ($p < 0.010$) in cancer compared to the normal epithelium, I hypothesised that the distribution of TET2 in ER negative breast cancer cell lines may also be elevated compared to non-cancer breast cells. Therefore, I performed ICC on a BC cell line, BT474 (ER- cell line), and a non-malignant breast cell line, MCF10A (triple negative cell line), to allow the visualisation of TET proteins at a higher resolution.

Like the results in human tissues (see subchapter 4.1.1), all three TET proteins are expressed in the nucleus and cytoplasm in both normal and cancer cells (Figure 4.4). TET1 is mainly expressed in the nucleus in normal breast cells, while TET1 expression increases in the cytoplasm in cancer cell lines in three biological repeats. The expression of TET2 showed a higher distribution in cancer cells in both the nucleus and cytoplasm compared to normal cells. The expression of TET3 showed a higher distribution in normal cells in both the nucleus and cytoplasm compared to cancer cells. To compare these further, the expression of cytoplasmic TET1 is lower compared to the cytoplasmic expression of both TET2 and TET3 in normal breast cells. On the other hand, the expression of both cytoplasmic TET1 and TET2 is higher in cancer cells compared to the expression of cytoplasmic TET3. However, the expression of nuclear TET1 shows a higher distribution in normal cells amongst all TETs. This result shows that the distributions of three TET proteins are different in normal and cancer cells, which suggests they may be differently associated with 5hmC.

Interestingly, compared to the IHC result, TET2 expression showed the same changes in the distribution. In tissue, the proportion of positive cells of cytoplasmic TET2 is significantly higher in cancer tissue compared to normal tissue, and in cell lines, the expression of cytoplasmic TET2 is higher in cancer cells compared to normal cells. This suggests that TET2 may play a more critical role in regulating hydroxymethylation in the BC cytoplasm. Moreover, the modification of 5hmC is mainly distributed in nuclei in normal cells while in both nuclei and cytoplasm in cancer cells, which is comparable to the expression of TET1 and TET2. Therefore, I hypothesise that TET1 and TET2 may be more important in 5hmC synthesis.

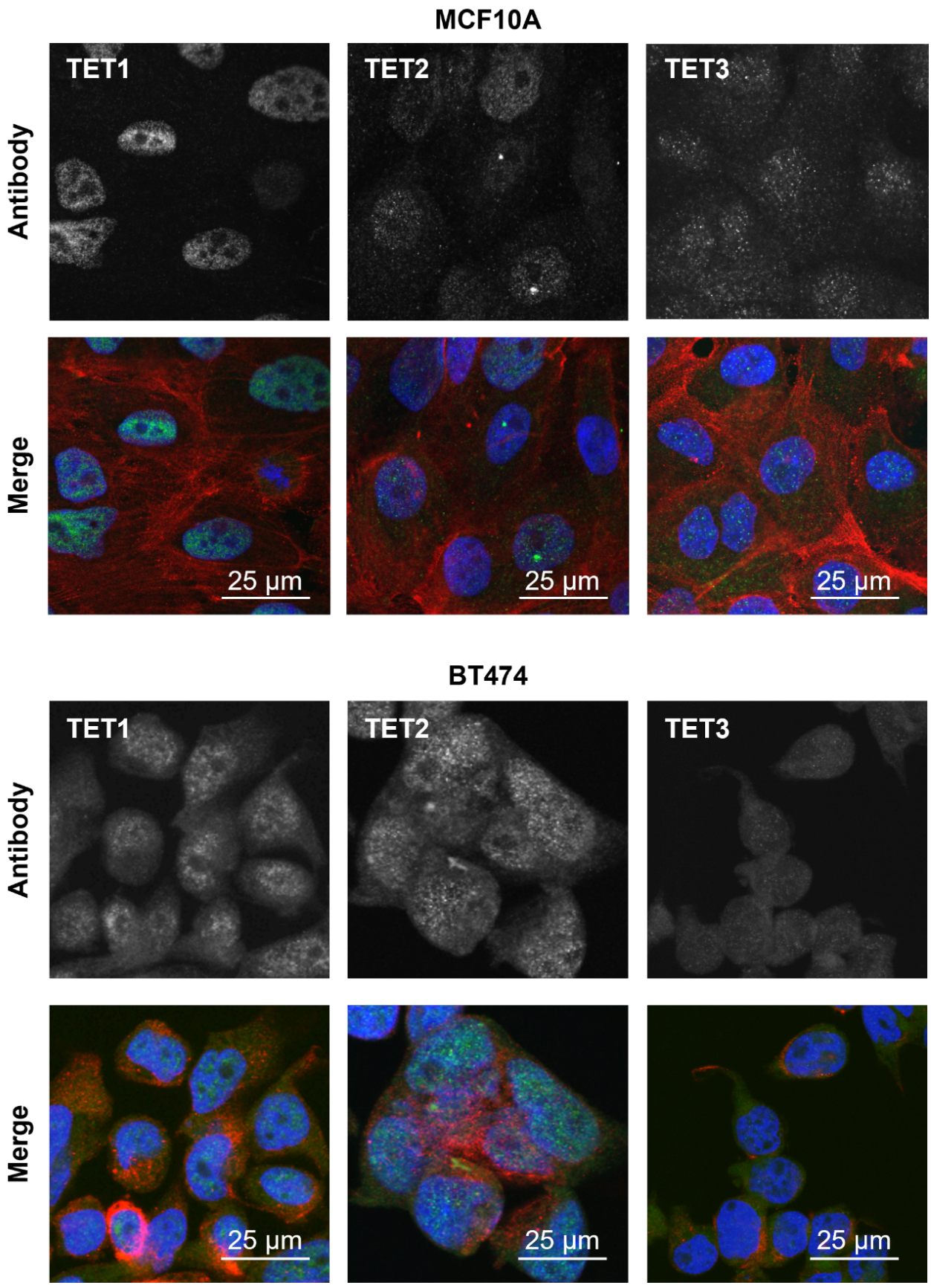

MCF10A
TET1
TET2
TET3
Antibody
Merge
25 µm
25 µm
25 µm
BT474
TET1
TET2
TET3
Antibody
Merge
25 µm
25 µm
25 µm

Figure 4.4 (Previous page) TET1, TET2, TET3 immunocytochemistry staining. The “Antibody” panel is representative images of TET1, TET2, TET3 expression measured with ICC, respectively. The “Merge” panel shows a combination of blue (Hoechst-stained DNA), green (Antibody) and red (beta-actin) signals in both MCF10A and BT474 cells. (Scale bar = 25 µm)

4.1.4 The higher expression of *TET2* mRNA is positively correlated with the probability of higher BC survival rate

My previous results indicated that three TET proteins might be differently associated with 5hmC formation, and 5hmC is associated with BC progression. In order to examine whether the expression of the three *TET* mRNAs is associated with BC patient survival, Kaplan-Meier Plotter, an online survival analysis tool, was employed to extract the data filtered by gene symbol, survival, probe set, and subtypes. Gene expression data (gene chip) and patient overall survival information were automatically collected from GEO, EGA and TCGA. Patients were divided into two groups, high *TET* mRNA expression and low *TET* mRNA expression, by the best performing threshold between upper and lower quartiles as the auto-select best cutoff. The results were automatically analysed by the tool using the Log-rank test.

In BC, only the expression of *TET2* mRNA, but not of *TET1* nor *TET3* mRNA, is positively correlated with the probability of the patient survival rate ($p<0.0001$, Figure. 4.5). To determine if this correlation is BC specific, it would be useful to perform the analysis in colorectal cancer because I observed a similar result in colorectal cancer (see subchapter 3.1.4). However, at the time of writing, the database used only includes breast, ovarian, lung and gastric cancer, without colorectal cancer. Thus, I chose another cancer to study whether *TET* mRNA expression is associated with the probability of patient survival rate. As ER and PR levels are positively associated with better clinical outcomes in ovarian cancer (182, 183), it could be interesting to study if there is a similar mechanism of action in the demethylation process in ovarian cancer. Therefore, the data were analysed for ovarian cancers and filtered with the same parameters. Interestingly, in ovarian cancer patients, *TET1* and *TET3*, but not *TET2* mRNA, are negatively correlated with the probability of patient survival rate ($p<0.0001$,

both, Figure 4.6). This indicates that the expression of *TET2* mRNA is specifically correlated with the BC survival rate.

To further investigate the association between the expression of *TET2* mRNA and the probability of BC survival amongst different subtypes, the data were filtered by different BC subtypes using the same settings. The expression of *TET2* mRNA was positively correlated with the probability of patient survival rate of the ER+ cohort (n=221, p=0.021, Figure 4.7 A), the ER- cohort (n=284, p<0.001, Figure 4.7 B), the HER2- cohort (n=720, p<0.0001, Figure 4.7 C), the basal-like cohort (n=278, p<0.01,Figure 4.7 D) and the luminal A cohort (n=377, p=0.016, Figure 4.7 E). This result indicates that *TET2* mRNA expression can potentially be a biomarker for predicting patient survival rates in these BC subtypes, confirming the previous study (147).

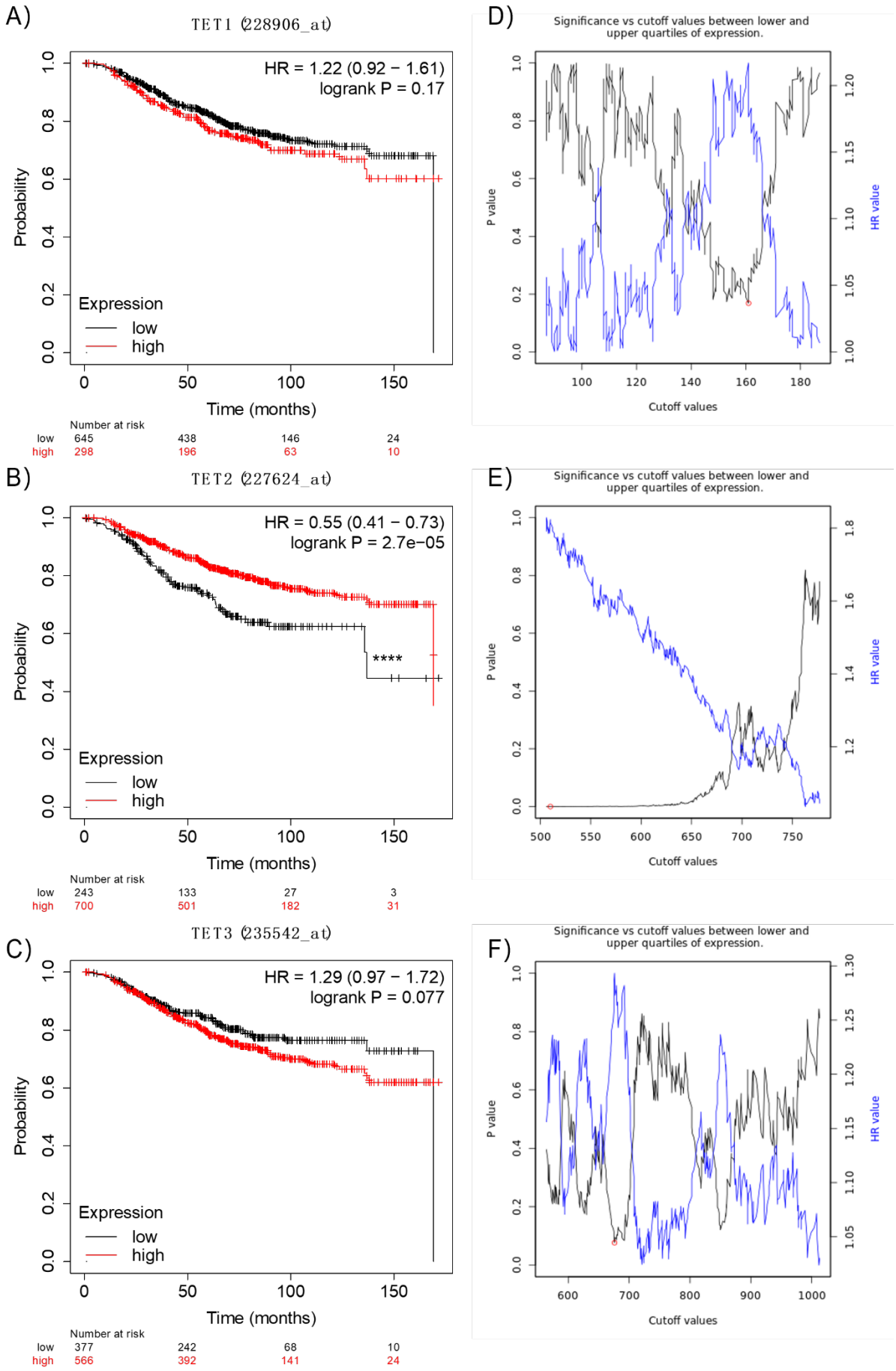
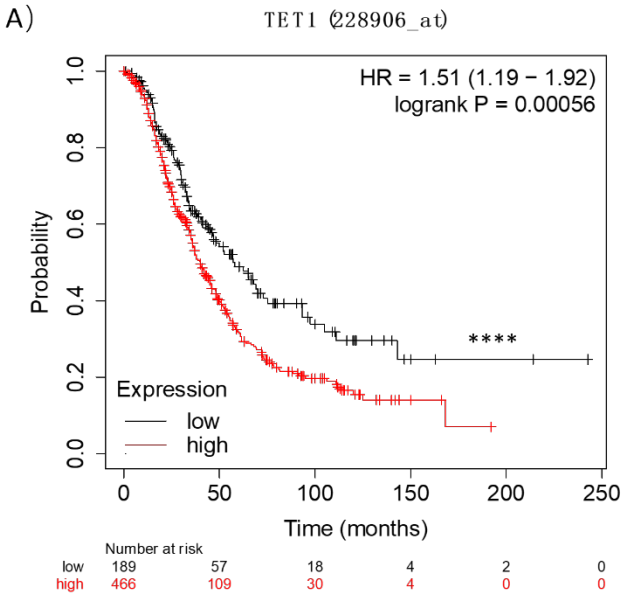
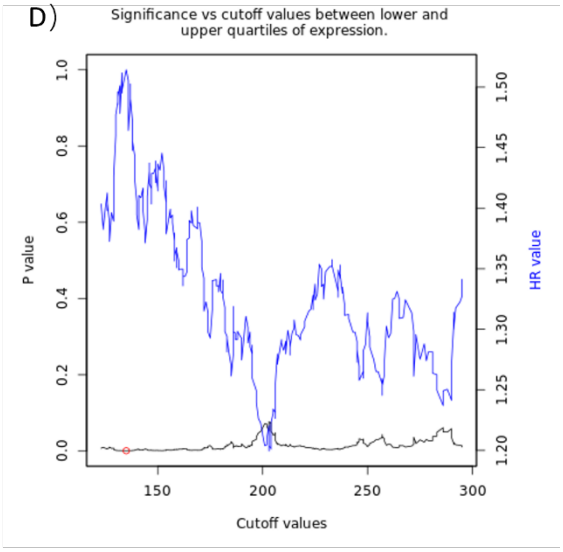
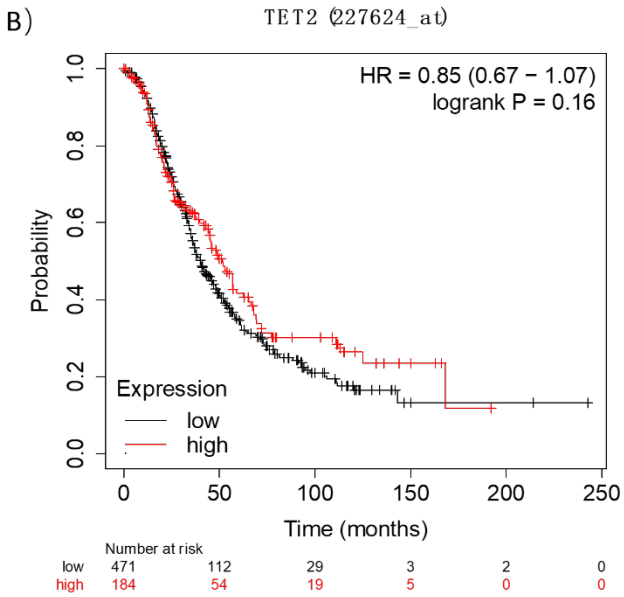
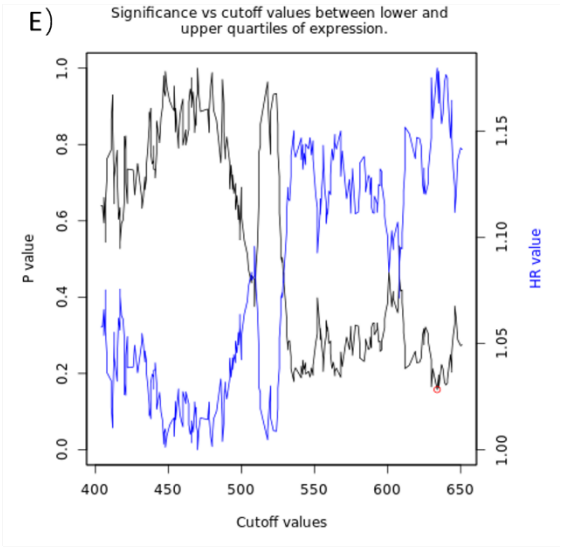

A)
TET1 (228906_at)
HR = 1.22 (0.92 − 1.61)
logrank P = 0.17
Probability
Expression
low
high
Time (months)
Number at risk
low 645 438 146 24
high 298 196 63 10
B)
TET2 (227624_at)
HR = 0.55 (0.41 − 0.73)
logrank P = 2.7e−05
****
Probability
Expression
low
high
Time (months)
Number at risk
low 243 133 27 3
high 700 501 182 31
C)
TET3 (235542_at)
HR = 1.29 (0.97 − 1.72)
logrank P = 0.077
Probability
Expression
low
high
Time (months)
Number at risk
low 377 242 68 10
high 566 392 141 24
D)
Significance vs cutoff values between lower and upper quartiles of expression.
P value
HR value
Cutoff values
E)
Significance vs cutoff values between lower and upper quartiles of expression.
P value
HR value
Cutoff values
F)
Significance vs cutoff values between lower and upper quartiles of expression.
P value
HR value
Cutoff values

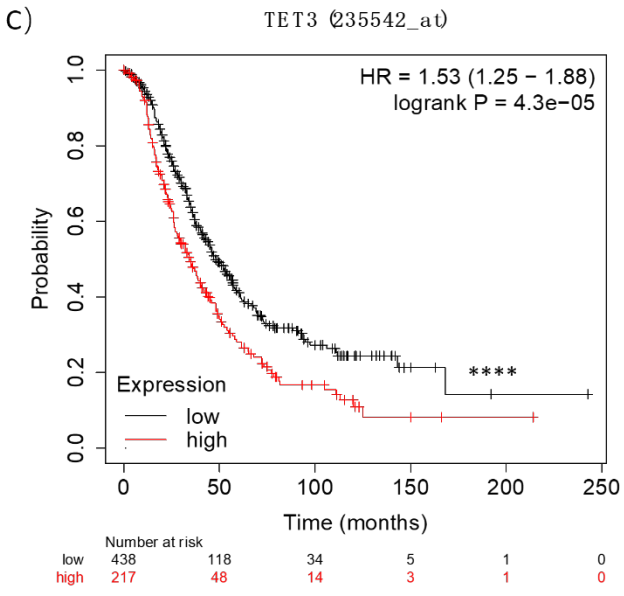
Figure 4.5 (Previous page) *TET2* is positively correlated with the probability of higher BC survival rate. The correlation between low or high expression of *TET* mRNA and the probability of BC patient survival rate (n=943) was analysed with Kaplan-Meier survival plots. **Graphs in (A)** and **(C)** show a lack of correlation between *TET1* and *TET3* mRNA levels and the probability of survival rates, respectively. **(B)** *TET2* mRNA is significantly correlated with the probability of a higher survival rate in BC patients. **Graphs in (D), (E), (F)** show the significance vs cutoff values between lower and upper quartiles of expression of *TET1*, *TET2*, and *TET3*, respectively. Statistically significant was calculated using the Log-rank test, and it is indicated with an asterisk: ****p<0.0001.


A)
TET1 (228906_at)
HR = 1.51 (1.19 − 1.92)
logrank P = 0.00056
****
Expression
low
high
Probability
Time (months)
Number at risk
low 189 57 18 4 2 0
high 466 109 30 4 0 0


D)
Significance vs cutoff values between lower and upper quartiles of expression.
P value
HR value
Cutoff values


B)
TET2 (227624_at)
HR = 0.85 (0.67 − 1.07)
logrank P = 0.16
Expression
low
high
Probability
Time (months)
Number at risk
low 471 112 29 3 2 0
high 184 54 19 5 0 0


E)
Significance vs cutoff values between lower and upper quartiles of expression.
P value
HR value
Cutoff values


C)
TET3 (235542_at)
HR = 1.53 (1.25 − 1.88)
logrank P = 4.3e−05
****
Expression
low
high
Probability
Time (months)
Number at risk
low 438 118 34 5 1 0
high 217 48 14 3 1 0

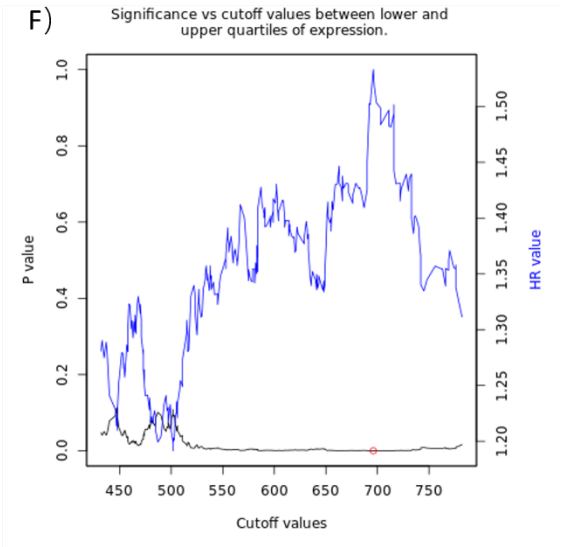
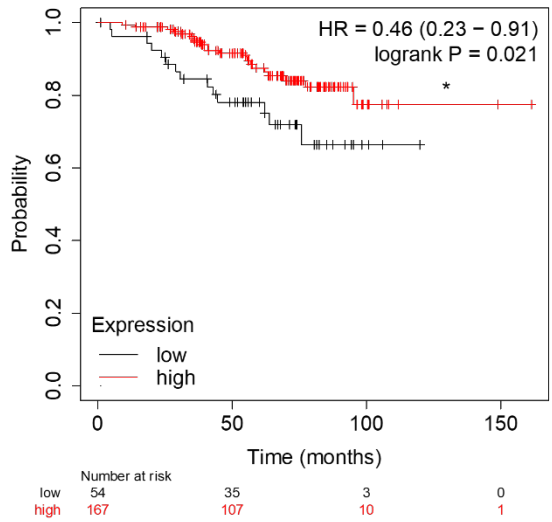
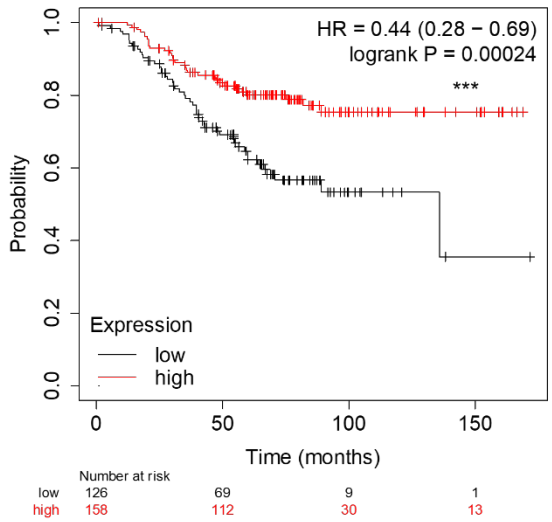

F)
Significance vs cutoff values between lower and upper quartiles of expression.
P value
HR value
Cutoff values

Figure 4.6 (Previous page) *TET1* and *TET3* are negatively correlated with the probability of higher ovarian cancer survival rate. The correlation between low or high expression of *TET* mRNA and the probability of ovarian cancer patient survival rate (n=655) was analysed with a Kaplan-Meier survival plot. **Graphs in (A)** and **(C)** show that *TET1* and *TET3* mRNA is significantly negatively correlated with the probability of a higher survival rate in ovarian cancer patients. **Graph (B)** shows a lack of correlations between *TET2* mRNA levels and the probability of ovarian cancer patient survival rate. **Graphs in (D), (E), (F)** show the significance vs cutoff values between lower and upper quartiles of expression of *TET1*, *TET2*, and *TET3*, respectively. Statistically significant was calculated using the Log-rank test, and it is indicated with an asterisk: ****p<0.0001.

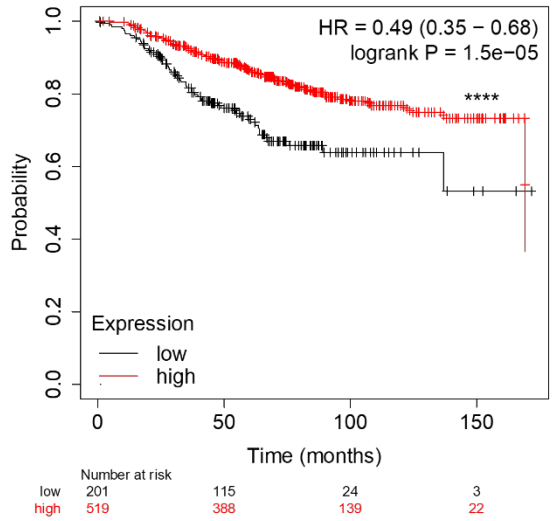
A)


TET2 (227624_at)
HR = 0.46 (0.23 − 0.91)
logrank P = 0.021
*
Probability
Expression
low
high
Time (months)
Number at risk
low 54 35 3 0
high 167 107 10 1

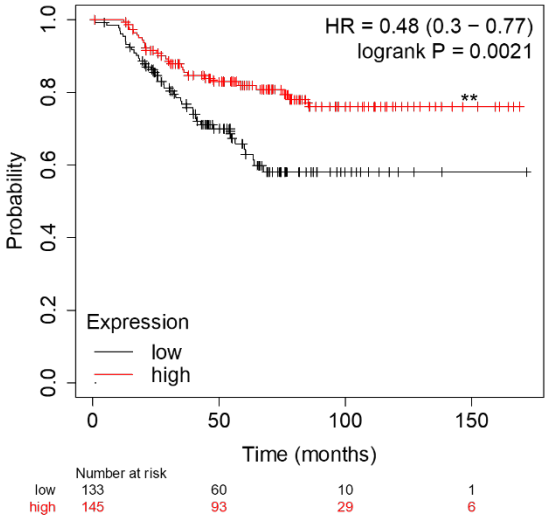
B)


TET2 (227624_at)
HR = 0.44 (0.28 − 0.69)
logrank P = 0.00024
***
Probability
Expression
low
high
Time (months)
Number at risk
low 126 69 9 1
high 158 112 30 13

C)


TET2 (227624_at)
HR = 0.49 (0.35 − 0.68)
logrank P = 1.5e−05
****
Probability
Expression
low
high
Time (months)
Number at risk
low 201 115 24 3
high 519 388 139 22

D)


TET2 (227624_at)
HR = 0.48 (0.3 − 0.77)
logrank P = 0.0021
**
Probability
Expression
low
high
Time (months)
Number at risk
low 133 60 10 1
high 145 93 29 6

E)

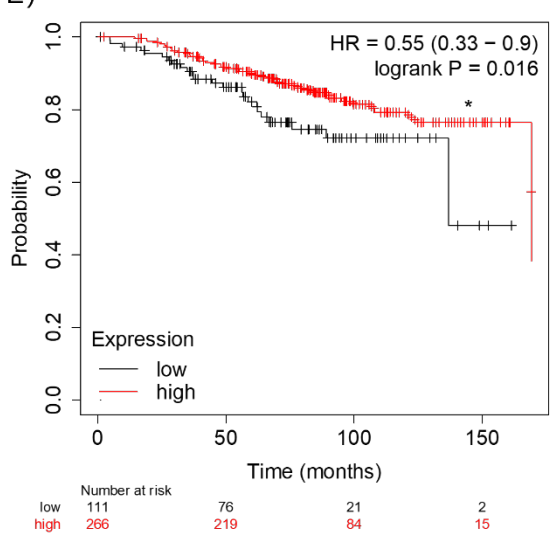

TET2 (227624_at)
HR = 0.55 (0.33 − 0.9)
logrank P = 0.016
*
Probability
Expression
low
high
Time (months)
Number at risk
low 111 76 21 2
high 266 219 84 15

Figure 4.7 (Previous page) *TET2* mRNA expression is positively correlated with the probability of higher BC patient survival rate in subtypes. The correlation between low or high expression of *TET* mRNA and the probability of BC patient survival rate was analysed with a Kaplan-Meier survival plot. **Graph A)** *TET2* mRNA is significantly correlated with the probability of a higher survival rate in ER+ BC patients. **Graph B)** *TET2* mRNA is significantly correlated with the probability of a higher survival rate in ER- BC patients. **Graph C)** *TET2* mRNA is significantly correlated with the probability of a higher survival rate in HER2- BC patients. **Graph D)** *TET2* mRNA is significantly correlated with the probability of a higher survival rate in basal-like BC patients. **Graph E)** *TET2* mRNA is significantly correlated with the probability of a higher survival rate in luminal A patients. Statistically significant was calculated using the Log-rank test, and it is indicated with an asterisk: *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

4.1.5 Distribution of TET2 mRNA and protein product in breast tissue

To further study the possible function of TET2 proteins, the expression levels of *TET2* mRNA in normal and BC cells should be investigated. Assuming the antibodies are specific for the protein of interest, the expression of *TET2* mRNA should correspond to the levels of TET2 proteins measured by IHC. According to the result of subchapter 4.1.1, the proportion of positive cells of cytoplasmic TET2 protein was significantly higher in cancer compared to normal breast epithelium, while the proportion of positive cells of nuclear TET protein did not show a significant difference. Therefore, one can hypothesise that only the copies of *TET2* mRNA in the cytoplasm, but not in the nucleus, would be higher in cancer epithelial cells than in normal epithelial cells.

To validate if *TET2* mRNA is expressed differently between normal and cancer tissue, *in situ* hybridisation (ISH) using RNAscope was performed on the samples (cohort 4) with the aid of the MRes student Liliane Eleid. Using this assay, each *TET2* mRNA is identified as a dot which reflects the amount and the localisations of *TET2* mRNA, since the target mRNA sequence is detected by a probe, amplified and visualised by DAB (Figure 4.8 B). Copies of *TET2* mRNA were found in both the nucleus and cytoplasm in normal and cancer epithelium, as expected (Figure 4.8 E). However, no significant difference in *TET2* mRNA expression was found amongst 12 pre-menopausal patients, consisting of 11 benign lesions and 9 IDC. I then selected breast cancer (n=5) and benign age-matched tissue controls (n=5) to investigate the subcellular localisations of *TET2* mRNA. After analysing the mean percentage of *TET2* positive cells, the expression of the cytoplasmic *TET2* mRNA is non-significantly higher in cancer compared to benign. However, the expression of the nuclear *TET2* mRNA is significantly

(p=0.001) lower in cancer compared to benign (three out of ten patients are ER+). This result is different from my hypothesis, which is probably because the samples are not all ER-.

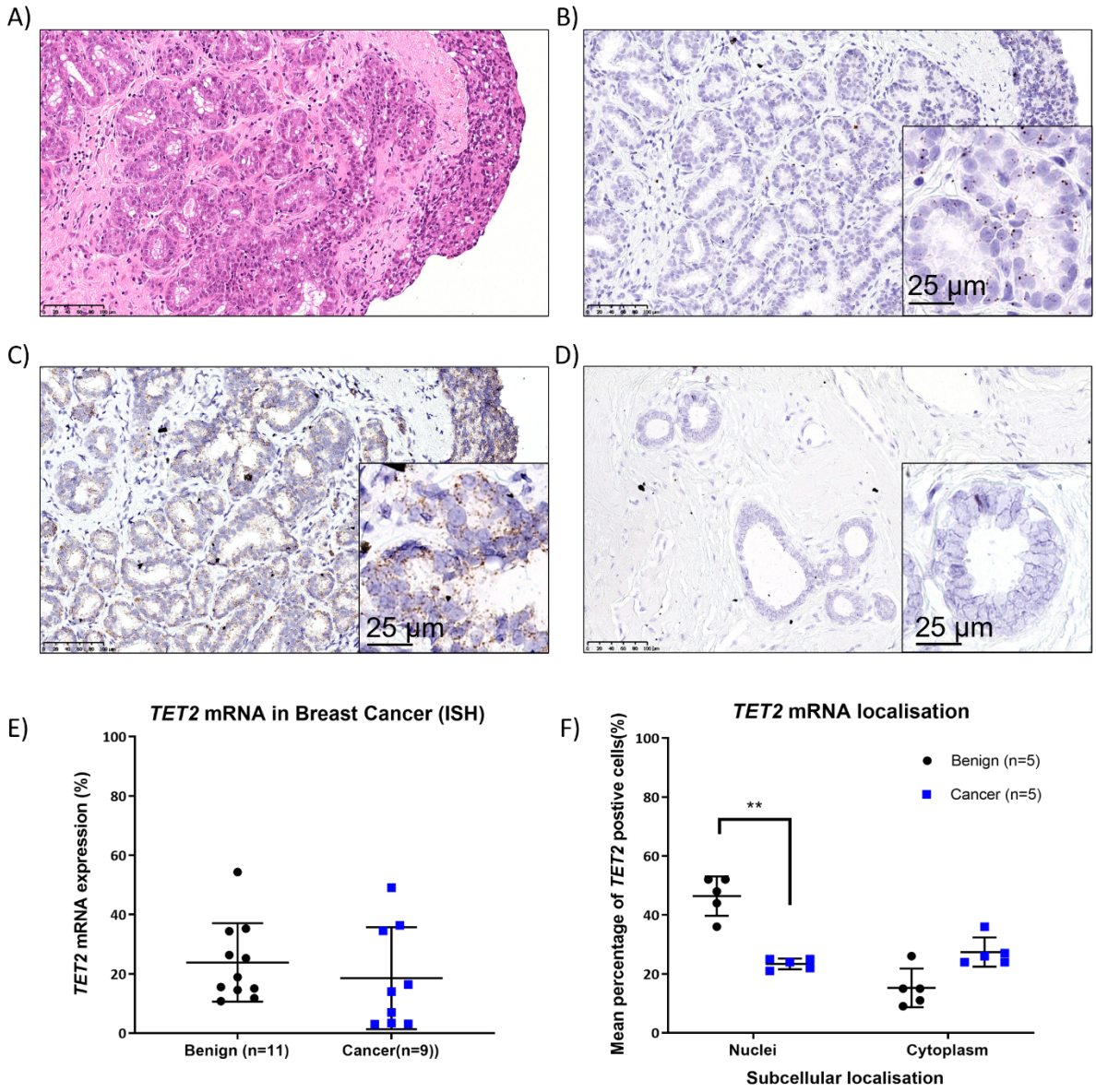


Figure 4.8 The quantification of the expression of *TET2* mRNA. (A-D) Due to the availability of the samples, the sections were not serial to the ones from the IHC staining (cohort 1 and 2). Serial sections are from a 31-year-old female patient, and the negative control section is from a 62-year-old female patient. In panels, H&E, *TET2* mRNA expression, PPIB mRNA expression (positive control) and dapB (negative control) are shown (scale bar = 100 µm, insert = 25 µm). **E)** Box plots presenting an average of *TET2* mRNA expression in benign tissue and in cancer. The expression of *TET2* mRNA in cancer is not significantly different from the expression in benign tissue. The lines represent the mean, and the SD. Statistical analysis was

performed with an unpaired T-test on each cohort compared to benign. F) Box plots presenting the mean percentage of nuclear *TET2* positive cells is lower in cancer compared to benign. The lines represent the mean, and the SD. Statistical analysis was performed using Two-Way Anova, and it is indicated with an asterisk, **p<0.005.

4.2 Discussion

In this chapter, we present results where IHC (result see subchapter 4.1.1) was performed to quantify three TET protein expression in the breast, and ICC (result see subchapter 4.1.3) to determine the subcellular localisation of three TET proteins. Then, ISH (result see subchapter 4.1.5) was performed to identify the localisation of *TET2* mRNA.

Both IHC and ICC results showed that the mRNA and the products of TET1, TET2 and TET3 are present in both the nucleus and cytoplasm. Using IHC staining, only cytoplasmic TET2, rather than TET1 or TET3, displays a significantly higher distribution in cancer epithelium than normal, with a similar pattern of distribution of 5hmC.

The proportion of positive cells of TET1 protein and 5hmC are positively correlated in the cytoplasm in normal breast epithelium but negatively correlated in the nucleus in normal breast and in BC epithelium (in both the nucleus and the cytoplasm). However, no significant correlation between 5hmC and TET1, and this may be due to the relatively small sample size. A previous study pointed out that 5hmC levels are significantly positively correlated with the expression of TET1 in BC, which is different from my result (184). While the study by Wu and colleagues did not examine the correlation between the 5hmC level and the expression of TET1 in subcellular localisation, this could be interesting if they perform further analysis. Future work could validate the correlation between TET1 and BC using more samples.

The proportion of positive cells of TET2 is negatively correlated to the proportion of positive cells of 5hmC in BC epithelium. Only the proportion of positive cells of cytoplasmic TET2 positively correlates to the proportion of positive cells of 5hmC in normal breast epithelium. On the contrary, a previous study shows a negative association between 5hmC and TET2 expression in precancerous breast lesions and a positive association between 5hmC and TET2

in BC patients (including 83 patients with breast ductal intraepithelial neoplasia, 20 patients with DCIS, and 10 patients with IDC) (176). Compared to their results, I performed the IHC staining on IDC patient samples so that it could be a negative correlation between 5hmC and TET2 in IDC patients. Moreover, Zhang *et al.*'s study did not distinguish the staining in the nucleus or the cytoplasm, so this could be another reason that the two results are different (176). The proportion of positive cells of TET3 and the proportion of positive cells of 5hmC are positively correlated in both normal breast and BC epithelium, which confirms Wu *et al*.'s study and Ramasamy *et al*.'s studies (184, 185).

Furthermore, the association between the expression of *TET* mRNA and the probability of BC survival rate was analysed. Based on the result from KM Plotter (result see subchapter 4.1.3), the expression of *TET2* mRNA is positively correlated with the probability of the patient survival rate in BC, not in ovarian cancer, confirming the previous study by Wan and colleagues (147). However, with the lack of patient data on colorectal cancer, I could not investigate the association between *TET* mRNA expression and the probability of CRC patient survival rate. Therefore, future work could validate if there is any correlation between *TET* mRNA expression and CRC patient survival rate. Moreover, Wan *et al*.'s study partly confirmed my results, in that high TET2 expression was correlated with a better prognosis in ER-, HER2- BC patients (147). The authors pointed out that the expression of TET2 is also associated with the prognosis in patients with different grades. In my study, I did not find any significant correlation between grades and *TET2* mRNA expression, and this could be due to the small sample size from the KM Plotter website or different cohorts being used. According to these findings, I hypothesised that *TET2* mRNA expression has the potential to serve as a

biomarker for predicting patient survival rate in these BC subtypes. Thus, it would be interesting to validate if *TET2* mRNA expression differs in tissue samples.

The ISH result (see subchapter 4.1.5) shows no significant difference in *TET2* mRNA expression between BC and benign samples. However, nuclear *TET2* mRNA was significantly lower in cancer than the normal epithelium, and cytoplasmic *TET2* mRNA did not show a significant difference between each group. The result could suggest that, in the benign samples, *TET2* transcripts may not only contribute to the demethylation process in the nuclei. *TET2* transcripts may act as a transcriptional activator or play an essential role in tumorigenesis. The difference in the localisation of the mRNA could also be due to the several alternative splice variants of *TET2*, which may be present at higher levels in the nuclei of normal cells but not be translated into functional proteins. Moreover, *TET2* can be regulated by miRNAs and lncRNAs in post-transcription regulation, and it can be degraded in post-translation regulation via four different pathways in normal cells (186). However, these processes may be blocked in BC cells. Hence, one possibility is that though lots of *TET2* mRNAs are made in normal breast cells, not many functional proteins are translated. In comparison, most of the *TET2* mRNA is translated into functional proteins in BC cells.

The disadvantage of this experiment is that the tissue samples of this experiment are from pre-menopausal patients, while the samples of the IHC experiment are from post-menopausal patients. This difference could lead to unexpected results. Moreover, the contralateral tissues should be used as the control rather than benign tissues. In the future, RNAscope on cohort 1 and cohort 2 could be performed to increase the numbers in the study and the addition of contralateral tissues as controls. Furthermore, in terms of the previous results (see subchapters 3.1.2 and 4.1.3), it would be interesting to investigate if the expression of *TET2* mRNA is correlated to the BC subtype. Therefore, future work could

determine the *TET2* mRNA expression in the division of the patients into different groups based on hormonal receptor status.

4.3 Conclusion

This chapter describes the expression of TET proteins in normal and cancer breast cells, the association between the proportion of positive cells of TET proteins and 5hmC in breast cells, the correlation between *TET* mRNA and the probability of BC patient survival rate, and the expression of *TET* mRNA in BC tissues.

Based on my results and the existing previous studies, the correlation between the levels of 5hmC and the expression of TET proteins is still controversial. The association between hydroxymethylation and TET protein may indicate an unknown regulatory pathway in BC. It is clear that all three TETs play crucial roles in the demethylation process. Previous studies worked on one of the TET proteins and investigated the function separately. However, this may hinder the study of the mutual regulation of the 3 proteins.

Furthermore, I aimed to confirm the importance of *TET* genes, specifically *TET2*, from a clinical point of view. Interestingly, TET2 expression, but not TET1 nor TET3, shows a significant positive correlation with patient survival rate in BC. Since the association between ovarian cancer patients and *TET* mRNAs differs, TET2 could be a specific biomarker for predicting BC patient survival rate. This study provides evidence for the clinical value of TET2 in BC. Moreover, Wan *et al.* suggested that TET2 expression significantly correlated with clinical prognosis in female cancers, including cervical and uterine cancer (147).

Through the above experiments, the role of TET proteins has become more relevant in breast cancer research. Moreover, TET proteins participate in cancer progression by regulating other genes. To confirm the specific regulatory role of TET proteins for 5hmC modification in the breast cells, I knockdown TET protein expression by siRNA pool and examine the levels of 5hmC and functional assays on breast cells, which I will describe in the next chapter.

5. Result 3:

The function of TETs in breast cell lines

This chapter attempts to present research I performed to answer the research question of whether altering TET expression changes the levels of 5hmC in breast cell lines, and whether any resulting functional changes can be observed. The results from the previous chapters in this thesis provide evidence that the 3 TET proteins may play different roles in regulating the levels of 5hmC in BC.

Therefore, the aims of the described experiments here were: a) to downregulate the expression of *TET* mRNA in breast cell lines; b) to investigate if the level of 5hmC changes after downregulating the TET expression; and c) to determine any change in cell behaviour. To achieve these, a siRNA reverse transfection was first performed to knockdown the expression of the three *TET* mRNAs. Then, after the validation of the knockdown by qPCR, the levels of 5hmC were determined using a colorimetric assay. Finally, the expression of the marker genes of cell proliferation, apoptosis, cell differentiation and DNA methylases were determined by qPCR, and a cell proliferation assay was performed.

To address the research questions above, a selection of cell line models was made. MCF7 is a well-established cell line derived from ER+ BC, which closely mirrors luminal A-type BC (187). SKBR3 represents the HER2-positive subtype of breast cancer, characterized by overexpression of the HER2+ gene (188). Based on prior experiments conducted in the laboratory (see subchapter 3.2), both MCF7 and SKBR3 showed a substantial variability of 5hmC within the cell lines. MDA-MB-231, simulating aggressive behaviour, represents the triple-negative subtype (189). Unfortunately, it was excluded from investigation due to its

unavailability in the laboratory. BT474 cell line is primarily characterized as a HER2-positive breast cancer cell line (154). Only BT474 demonstrated relatively stable 5hmC content compared to the others, among the cell lines (see subchapter 3.1.5). Additionally, MCF10A cell line is a model for normal breast epithelial cells, originating from benign proliferative breast tissue (190). This cell line is employed to emulate the normal cellular environment. HEK293T cells, derived from human embryonic kidney cells, are commonly employed as a model system for siRNA transfection (191, 192). HEK293T cells exhibit a high transfection efficiency, allowing for effective delivery of siRNA molecules into the cells (191). Before conducting experiments in breast cancer cell lines, HEK293T cells was employed to validate siRNA knockdown efficacy and observe initial effects on relevant pathways.

Previous studies suggested that TET proteins play different roles in cancer progression. TET1 acts as both oncogene and tumour suppressor functions in cancers. On the one hand, as a tumour suppressor, TET1 represses cell migration, induces cell apoptosis, and prevents invasion and metastasis in cancer cells (123-125, 128, 193). On the other hand, as an oncogene, TET1 reduces cellular proliferation in triple-negative breast cancer (131). Most studies regard TET2 as a tumour suppressor and describe how it regulates cell differentiation in luminal breast cancer (164). In comparison, the possible role of TET3 is still unknown, but a higher expression level in breast cancers, compared to normal tissues, has been observed (144). According to these findings, 3 TET proteins could affect the different hallmarks of cancer, so it would be helpful to demonstrate the possible function of TET proteins in cancer and normal breast cells.

To investigate the effects of three TET proteins, two marker genes for each function from the literature, including proliferation, apoptosis, and cell differentiation, were chosen. Since Ki67

is a biomarker of proliferation in solid cancers, overexpression of Ki67 is related to poor prognosis in breast cancer (36). Moreover, MCM2 is a family member of the minichromosomal maintenance protein complex and is reported as an alternative biomarker for measuring proliferation in breast cancer (35). For the marker genes of apoptosis, both BCL-2 (B-cell lymphoma) and BCL-w are anti-apoptotic proteins (40). Furthermore, *ESR1* and *GATA3* are key transcription factors often silenced by DNA methylation in basal-like breast cancer, including TNBC. These two genes were reported to mediate luminal cell differentiation, related to TET2 and 5hmC (178). Apart from these marker genes, the mRNA expression of DNA methylases *DNMT1* and *DNMT3a* was tested.

5.1 Results

5.1.1 Optimisation of siRNAs

HEK293T cells are commonly used cell lines in transfection experiments or to produce lentiviral and retroviral vectors, including an SV40 large T antigen (192). Transfection conditions were therefore optimised in this cell line. The final concentrations of 10 nM and 20 nM of siRNA were chosen based on previous experience in the laboratory and the manufacturer's recommendations.

The transfection efficiency was evaluated by measuring the reduction in the expression of *TET* mRNAs. Due to time constraints, only one repeat using Lipofectamine 2000 to transfect the siRNA with the concentration of 10 nM was performed (details see Table 2.8). The Ct values for all mRNA did not indicate any observable variation between the untreated cells (Unt), cells only transfected by Lipofectamine 2000 reagent (Lipo) and cells transfected by scramble siRNA (Scr). Because of human error, the sample transfected by Scr and incubated for 24 hours was lost. The mRNA expression of siRNA transfected cells was then normalised to the expression of Lipo.

After a 24-hour transfection (Figure 5.1 A-C), the expression of *TET1* mRNA was decreased to 51.5% compared to the negative control (Lipo) transfected with siTET1, and the expression of *TET2* mRNA was reduced to 56.8% compared to the negative control transfected with siTET2, and the expression of *TET3 mRNA* was reduced to 24.5% comparing to the negative control transfected with siTET3. However, I observed a decrease in *TET2* and *TET3* mRNA expression after transfecting the cells with siTET1, which declined to 70.3% and 55.8% at a 24-hour time point, respectively. As this result represents only one biological repeat, it could only show a tendency for a decrease in each *TET* mRNA expression pattern and is not conclusive. However,

the decrease in the expression of the other two mRNAs not targeted by each siRNA might indicate a balance between three TETs is present in HEK293 cells.

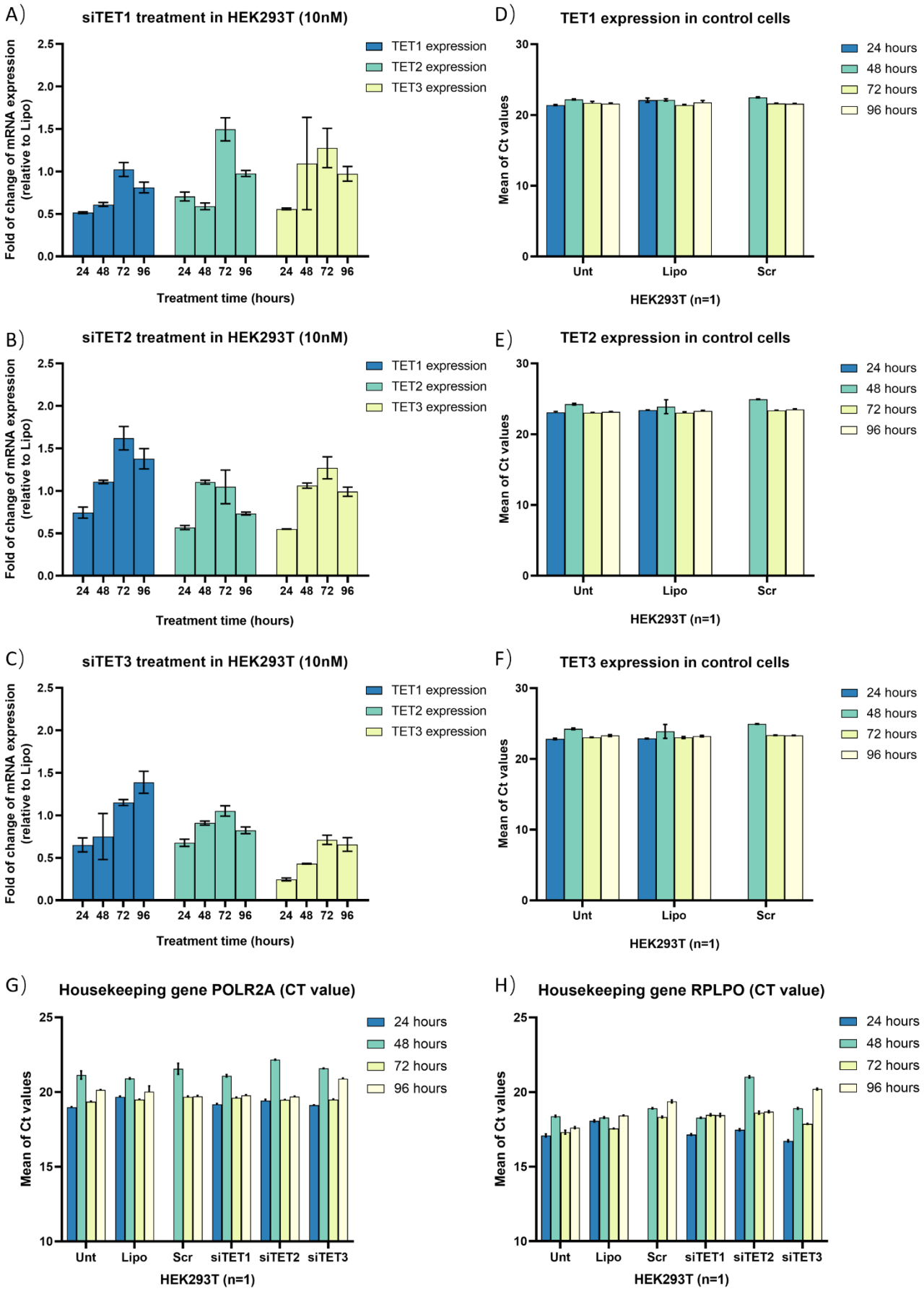

A)
siTET1 treatment in HEK293T (10nM)
Fold of change of mRNA expression (relative to Lipo)
Treatment time (hours)
TET1 expression
TET2 expression
TET3 expression
B)
siTET2 treatment in HEK293T (10nM)
C)
siTET3 treatment in HEK293T (10nM)
D)
TET1 expression in control cells
Mean of Ct values
HEK293T (n=1)
Unt
Lipo
Scr
24 hours
48 hours
72 hours
96 hours
E)
TET2 expression in control cells
F)
TET3 expression in control cells
G)
Housekeeping gene POLR2A (CT value)
siTET1
siTET2
siTET3
H)
Housekeeping gene RPLPO (CT value)

Figure 5.1 (Previous page) *TETs* mRNA relative expression decreased upon siRNA treatment in HEK293T cells (n=1). Graph A-C) TETs mRNA relative expression in HEK293T treated by siTET1, siTET2 or siTET3. HEK293T cells were seeded in a 10 cm dish mixed with siRNA (10 nM) and transfection reagent. On the following day, the cells were resuspended and seeded in 6-well plates for 48 hours, 72 hours and 96 hours and harvested for RT-qPCR for 24-hour time point of reverse transfection. Cells were resuspended and harvested for RT-qPCR after 48 hours, 72 hours and 96 hours of reverse transfection. **Graph D-F)** Mean of Ct values of control cells represent the Ct value of untreated cells (Unt), cells only treated with Lipofectamine 2000 (Lipo) and cells treated with scramble siRNA (Scr) for *TET1*, *TET2* and *TET3* mRNA. **Graph G-H)** Mean of Ct value for housekeeping genes. The mean and SD of three technical replicates is represented in the bar chart.

To improve the efficiency of the knockdown of each *TET* mRNA, the siRNA concentration was increased to 20 nM, and Lipofectamine RNAiMAX was used. As cationic liposomal transfection reagents, both Lipofectamine 2000 and Lipofectamine RNAiMAX are formulated for superior transfection efficiencies and are suitable for siRNA transfection. However, the commercial website regards Lipofectamine RNAiMAX as the most suitable reagent for hard-to-transfect cell lines, such as MCF10A and BT474 cells, which were the final desired cell lines.

Only one repeat was performed to test if the three *TET* mRNA expression was reduced in HEK293T cells. Based on the manufactural protocol, gene knockdown can be assayed after incubating the cells in the reagent for 24 hours. Moreover, the previous experiment (Figure 5.1) using Lipofectamine 2000 showed that the knockdown might lose effectiveness after 72 hours. For these reasons, I incubated the complexes for 24 hours and 48 hours to verify the reduction in the expression of *TET* mRNA. The Ct values were not discernibly different between Unt, cells only transfected by Lipofectamine RNAiMAX reagent (Lipo) and Scr. The

mRNA expression of siRNA transfected cells was therefore normalised to the expression of Lipo (Figure 5.2).

After a 24-hour transfection with siTET1, the expression of *TET1* mRNA was reduced to 57.9% compared to the negative control, but the expression of *TET3* mRNA increased to 148% after a 48-hour transfection (Figure 5.2 A). After 24-hour transfection with siTET2, the expression of *TET2* mRNA was decreased to 70.4% transfected siTET2, and further reduced to 54.5% after 48 hours (Figure 5.2 B). However, the expression of *TET1* mRNA was increased to 146% after 24-hour incubation, and the expression of *TET3* mRNA was increased to 143% after 48-hour incubation. After a 24-hour transfection with siTET3, the expression of *TET3* mRNA was reduced to 88.8%, while the expression levels of *TET1* mRNA and *TET2* mRNA increased (Figure 5.2 C). Since I did not observe any increase in the previous experiment at the 24-hour time point and 48-hour time point (Figure 5.1 A-C), the increase mentioned here (Figure 5.2 A-C) could be the variable of the experiments.

The two preliminary experiments indicate a higher transfection efficiency with the use of Lipofectamine 2000. However, Lipofectamine RNAiMAX at a concentration of 20nM siRNA was used to transfect the more challenging breast cell lines, as recommended by the manufacturer.

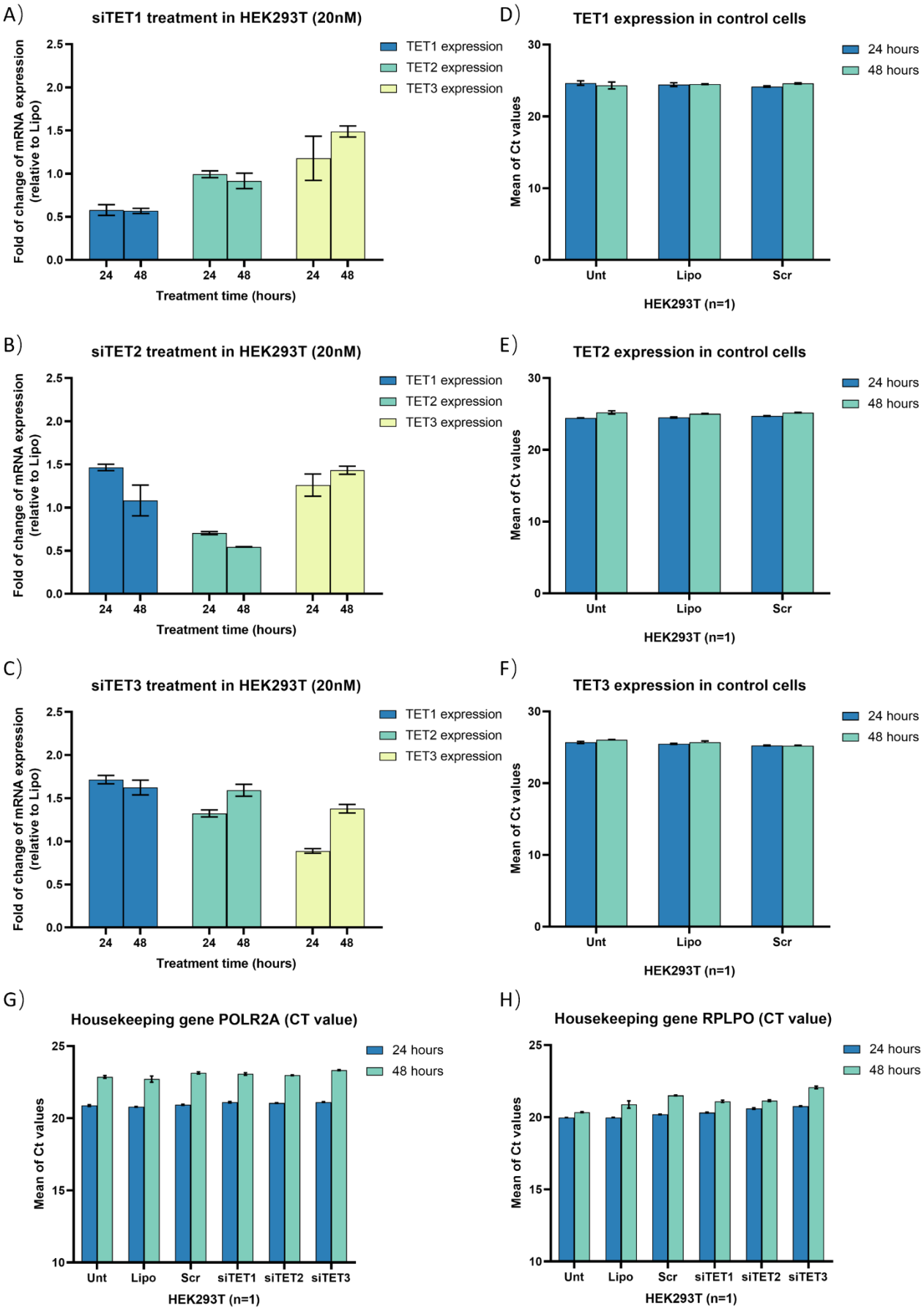

A)
siTET1 treatment in HEK293T (20nM)
Fold of change of mRNA expression (relative to Lipo)
Treatment time (hours)
TET1 expression
TET2 expression
TET3 expression
B)
siTET2 treatment in HEK293T (20nM)
C)
siTET3 treatment in HEK293T (20nM)
D)
TET1 expression in control cells
Mean of Ct values
HEK293T (n=1)
Unt
Lipo
Scr
24 hours
48 hours
E)
TET2 expression in control cells
F)
TET3 expression in control cells
G)
Housekeeping gene POLR2A (CT value)
siTET1
siTET2
siTET3
H)
Housekeeping gene RPLPO (CT value)

Figure 5.2 (Previous page) *TETs* mRNA relative expression decreased upon siRNA treatment in HEK293T cells (n=1). Graph A-C) TETs mRNA relative expression in HEK293T treated by siTET1, siTET2 or siTET3. HEK293T cells were seeded in a 10 cm dish mixed with siRNA (20 nM) and transfection reagent. The following day, the cells were resuspended and seeded in 6-well plates for 48 hours and harvested for RT-qPCR for a 24-hour time point of reverse transfection. Cells were then resuspended and harvested for RT-qPCR after 48 hours of reverse transfection. **Graph D-F)** Mean of Ct values of control cells represent the Ct value of untreated cells (Unt), cells only treated with Lipofectamine RNAiMAX (Lipo) and cells treated with scramble siRNA (SCR) for *TET1*, *TET2* and *TET3* mRNA. **Graph G-H)** Mean of Ct value for housekeeping genes. The mean and SD of three technical replicates is represented in the bar chart.

5.1.2 TET siRNAs significantly downregulated *TET* mRNA expression in BT474 cells

To gain insights into some of the functions of each TET in breast cell lines, the selected siRNA pools were used to downregulate the expression of the three *TET* mRNAs in the breast cancer cell line, BT474, together with Dr Lisa Del Bel Belluz in our department (Dr Del Bel Belluz performed RNA extraction on 120h samples and qPCR; I performed siRNA reverse transfections, RNA extraction of 24h and 48h samples, and all the data analysis). After optimisation, 20nM siTET1, siTET2 and siTET3 mixed with Lipofectamine RNAiMAX (condition shown in Table 2.8) was transfected into BT474 cells (Figure 5.3). The mRNA expression of siRNA transfected cells was normalised to the expression of Lipo.

The Ct values of three TET of Unt and Scr were analysed in relation to the Lipo at 24 hours, 48 hours and 120 hours (one-way ANOVA, p values shown in Table 5.1 below). There was no significant difference between the Ct value of Unt, Lipo and Scr. Moreover, the Ct values of housekeeping genes were also not significantly different between the groups at 24 hours, 48 hours and 120 hours (two-way ANOVA, p values shown in Table 5.2 below).

Table 5.1 Significance between gene expression for TET1, TET2 and TET3 in Unt-, Scr- and Lipo - treated BT474 cells

The p values of the Ct value of *TET1* mRNA expression between the control cells		
Time point	Lipo vs Unt	Lipo vs Scr
24 hours	p=0.791	p=0.890
48 hours	p=0.978	p=0.958
120 hours	p=0.872	p=0.999
The p values of the Ct value of *TET2* mRNA expression between the control cells		
Time point	Lipo vs Unt	Lipo vs Scr
24 hours	p=0.953	p=0.997
48 hours	p=0.997	p=0.999
120 hours	p=0965	p=0.997
The p values of the Ct value of *TET3* mRNA expression between the control cells		
Time point	Lipo vs Unt	Lipo vs Scr
24 hours	p=0.949	p=0.761
48 hours	p=0.954	p=0.683
120 hours	p=0.779	p=0.991

Table 5.2 Significance between gene expression for POLR2A and RPLP0 in Unt-, Scr- and Lipo - treated BT474 cells

The p values of the Ct value of *POLR2A* mRNA expression between the control cells					
Time point	Lipo vs Unt	Lipo vs Scr	Lipo vs siTET1	Lipo vs siTET2	Lipo vs siTET3
24 hours	p=0.931	p=0.979	p>0.999	p=0.607	p=0.793
48 hours	p>0.999	p=0.999	p>0.999	p=0.990	p=0.999
120 hours	p>0.999	p=0.997	p=0.871	p=0.738	p=0.570
The p values of the Ct value of *RPLPO* mRNA expression between the control cells					
Time point	Lipo vs Unt	Lipo vs Scr	Lipo vs siTET1	Lipo vs siTET2	Lipo vs siTET3
24 hours	p=0.999	p=0.826	p>0.999	p=0.999	p>0.999
48 hours	p=0.865	p>0.999	p>0.999	p=0.961	p=0.790
120 hours	p=0.792	p=0.895	p=0.945	p=0.980	p=0.497

The reduction of each *TET* mRNA was specific after the treatment with the individual targeted siRNA. After 24 hours of each siRNA treatment, the mRNA expression of *TET1* was significantly ($p=0.002$) reduced to 52.6%, and the downregulation remained consistently low until 120 hours ($p<0.001$, Figure 5.3 A). Furthermore, the mRNA expression of *TET2* mRNA was significantly ($p<0.0001$) reduced to 35.5%, remaining at 40% until 120 hours ($p=0.002$, Figure 5.3 B). Finally, the expression of *TET3* mRNA, was significantly ($p=0.005$) reduced to 41.7% after 24 hours and decreased to 30.1% ($p<0.001$) after 48 hours, to then increase to around 75.4% at 120-hour time point (Figure 5.3 C). This may be due to a higher loss of effectiveness of TET3 siRNA between 48 hours and 120 hours.

This experiment showed that the mRNA expression of *TET1, TET2* and *TET3* can be downregulated by up to 45.1%, 40% and 75.4%, respectively, by each siRNA until 120 hours in BT474 cells. Therefore, for future experiments, the expression of TET proteins expression and 5hmC levels will be measured within 120 hours.

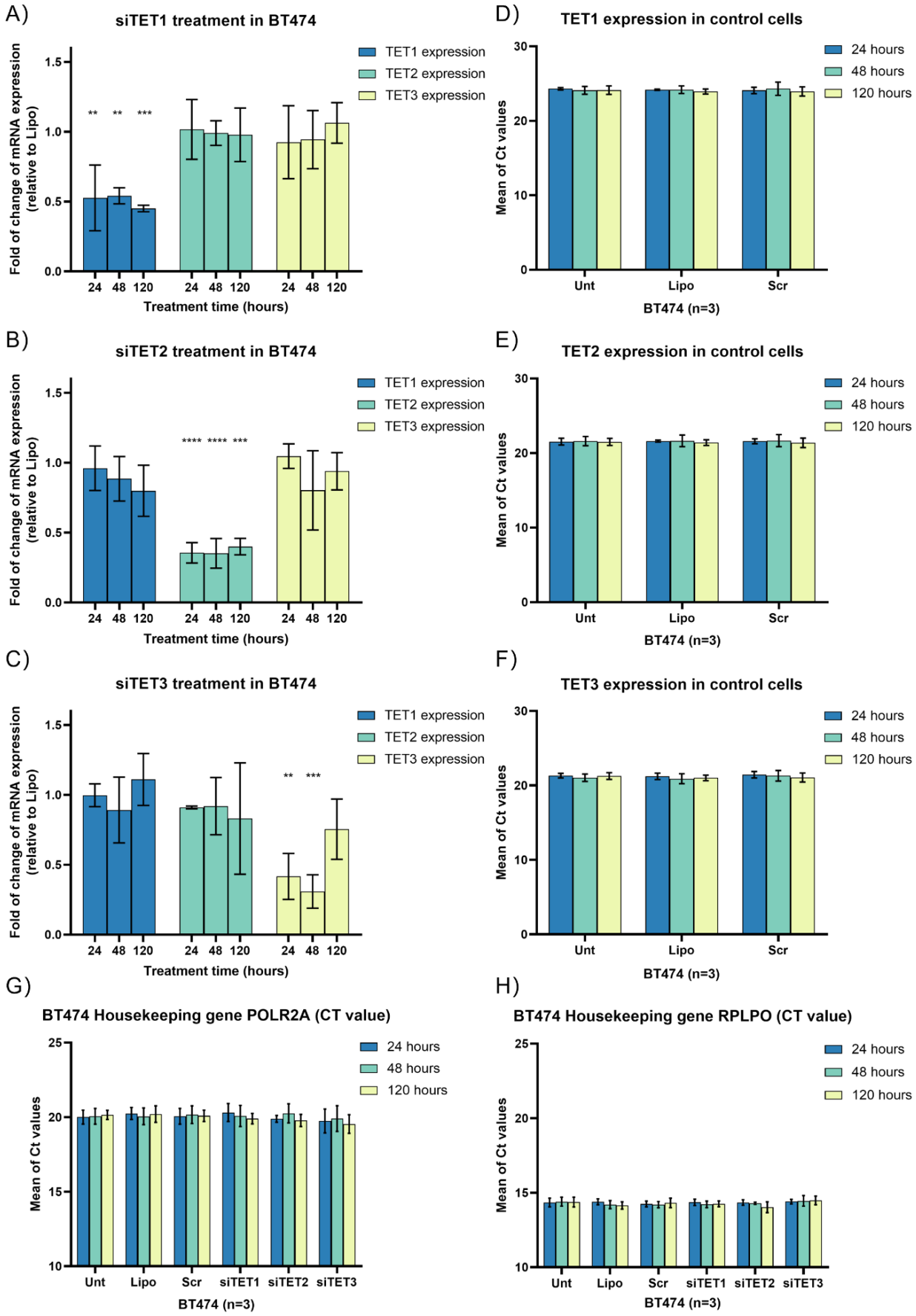

A)
siTET1 treatment in BT474
Fold of change of mRNA expression (relative to Lipo)
TET1 expression
TET2 expression
TET3 expression
Treatment time (hours)
B)
siTET2 treatment in BT474
C)
siTET3 treatment in BT474
D)
TET1 expression in control cells
Mean of Ct values
24 hours
48 hours
120 hours
Unt
Lipo
Scr
BT474 (n=3)
E)
TET2 expression in control cells
F)
TET3 expression in control cells
G)
BT474 Housekeeping gene POLR2A (CT value)
siTET1
siTET2
siTET3
H)
BT474 Housekeeping gene RPLPO (CT value)

Figure 5.3 (Previous page) *TETs* mRNA expression significantly decreased upon siRNA treatment in BT474 cells (n=3). Graph A-C) Relative expression of *TETs* mRNA in BT474 treated by siTET1, siTET2 or siTET3, respectively. BT474 cells were seeded in a 10 cm dish and mixed with siRNA and transfection reagent. On the following day, the cells were resuspended and harvested for RT-qPCR for 24 hour-time point of reverse transfection. The other cells were seeded in 6-well plates for 48 and 120 hours. Cells were then resuspended and harvested for RT-qPCR after 48 hours and 120 hours of reverse transfection. **Graph D-F)** The mean of the Ct values of control cells represents the Ct values of untreated cells (Unt), cells only treated Lipofectamine RNAiMAX (Lipo) and cells treated with scramble siRNA (SCR) for *TET1*, *TET2* and *TET3* mRNA, respectively. **Graph G-H)** Mean of Ct value for housekeeping genes. The result shown represents the mean ± SD of this group. Statistically significant was calculated using the two-way ANOVA, indicated with an asterisk: *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

5.1.3 siRNAs mediated TET knockdown in MCF10A cells

The downregulation of *TETs* mRNA with siRNA was also attempted on MCF10A cells. However, due to time and technical problems, not enough cells were available for the 120-hour time point. After both 24 and 48 hours of treatment with each siRNA, the mRNA expression of *TET1, TET2* and *TET3* did not significantly decrease, and a high amount of variation was observed in each biological replicate and group (Figure 5.4). However, a trend could be observed for each *TET* mRNA. After siTET1 treatment for 24 hours, the expression of *TET1* mRNA was non-significantly (p=0.944) reduced to 66.0%, and the trend in downregulation was constant (p=0.739) after 48 hours. While both *TET2* and *TET3* mRNA expression was not noticeably affected after 24 hours, a non-significant downregulation of both transcripts was observed at the later time point (to 53.6%, p=0.952 and 29.3%, p=0.998 for *TET2* and *TET3* mRNA, respectively).

Forty-eight hours were necessary to see a noticeable trend in downregulation of *TET2* mRNA (p=0.309) after siTET2 treatment. However, this downregulation was not specific for *TET2* mRNA, as a non-significant downregulation at 48 hours could be seen also for *TET1* mRNA (p=0.600) and *TET3* mRNA (p=0.945).

After siTET3 treatment for 24 hours, the expression of *TET3* mRNA was numerically but non-significantly (p=0.985) reduced to 68.4% after 24 hours and further decreased to 39.1% (p=0.823) after 48 hours. However, the expression levels of both *TET1* and *TET2* mRNA were also non-significantly decreased (p=0.729 and p=0.853, respectively) after siTET3 treatment for 48 hours.

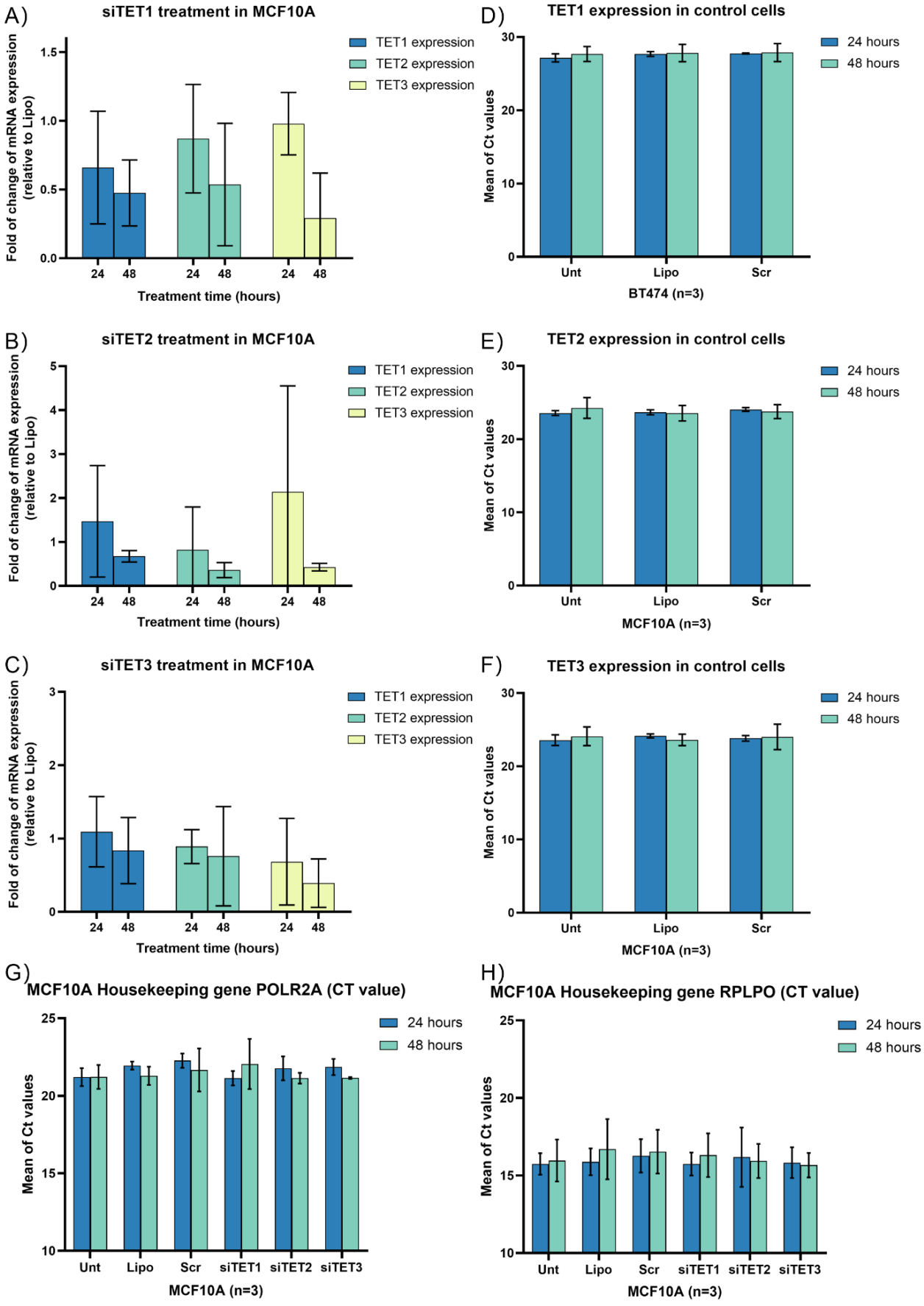

A)
siTET1 treatment in MCF10A
Fold of change of mRNA expression (relative to Lipo)
TET1 expression
TET2 expression
TET3 expression
Treatment time (hours)
B)
siTET2 treatment in MCF10A
C)
siTET3 treatment in MCF10A
D)
TET1 expression in control cells
Mean of Ct values
24 hours
48 hours
Unt
Lipo
Scr
BT474 (n=3)
E)
TET2 expression in control cells
MCF10A (n=3)
F)
TET3 expression in control cells
G)
MCF10A Housekeeping gene POLR2A (CT value)
siTET1
siTET2
siTET3
H)
MCF10A Housekeeping gene RPLPO (CT value)

Figure 5.4 (Previous page) siRNAs mediated TET knockdown in MCF10A cells (n=3). Graph A-C) Relative expression of *TETs* mRNA in MCF10A treated by siTET1, siTET2 or siTET3. MCF10A cells were seeded in a 10 cm dish and mixed with siRNA and transfection reagent. On the following day, the cells were resuspended and harvested for RT-qPCR for 24 hour-time point of reverse transfection. The other cells were seeded in 6-well plates for 48 and 120 hours. Cells were then resuspended and harvested for RT-qPCR after 48 hours of reverse transfection. **Graph D-F)** The mean of Ct values of control cells represents the Ct values of untreated cells (Unt), cells only treated lipofectamine (Lipo) and cells treated with scramble siRNA (Scr) for *TET1*, *TET2* and *TET3* mRNA. **Graph G-H)** Mean of Ct value for the housekeeping genes. The result shown represents the mean ± SD of this group. Statistically significant was calculated using the two-way ANOVA.

As the statistical test showed higher variability in MCF10A compared to BT474, a further analysis was performed to control for possible technical errors.

The Ct values of three TETs of Unt in BT474 and MCF10A were listed (see Table 5.3 and Table 5.4 below). There was no significant difference between the biological repeats, which means the expression of three TETs is stable in both BT474 and MCF10A. Moreover, the delta Ct value was analysed and showed in Figure 5.5. The result suggested that the expression of *TETs* were not affected by the experiment.

Both the Ct values of three TET of Unt and the Scr were analysed in relation to the Ct value of Lipo at 24 hours and 48 hours by one-way ANOVA (p values shown in Table 5.5 below). The Ct values of the housekeeping genes of samples were analysed by two-way ANOVA at 24 hours and 48 hours (p values shown in Table 5.6 below). There was no significant difference between each sample, indicating that this set of experiments was reproducible. Consequently, either the downregulation of *TETs* mRNA expression in MCF10A cells need more validation,

or there may be a still unknown biological reason behind this variation. Because of this, the non-malignant mammary cells were not used for further analysis.

Table 5.3 The Ct values of each sample in BT474 untreated cells at 24 hours.

Target name	Repeat 1	Repeat 2	Repeat 3	Average of Ct
TET1	24.312	24.129	24.443	24.295
TET2	21.697	21.028	21.881	21.535
TET3	21.354	20.993	21.604	21.317
Average of HSK	16.940	16.989	17.603	17.177

Table 5.4 The Ct values of each sample in MCF10A untreated cells at 24 hours.

Target name	Repeat 1	Repeat 2	Repeat 3	Average of Ct
TET1	27.365	27.548	26.518	27.144
TET2	23.302	23.934	23.434	23.557
TET3	22.861	24.303	23.517	23.560
Average of HSK	17.972	19.106	18.337	18.472

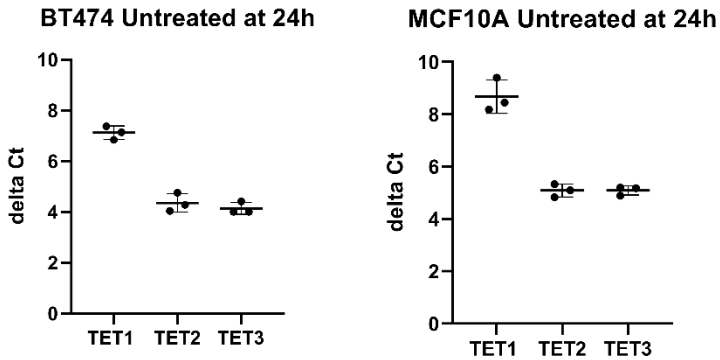


Figure 5.5 Delta Ct value for BT474 and MCF10A untreated cells at 24 hours. Scatter plots represent the mean with SD of Delta Ct per sample.

Table 5.5 Significance between gene expression for TET1, TET2, TET3 in Unt-, Scr- and Lipo- in MCF10A.

The p values of the Ct value of *TET1* mRNA expression between the control cells		
Time point	Lipo vs Unt	Lipo vs Scr
24 hours	p=0.208	p=0.973
48 hours	p=0.988	p=0.996
The p values of the Ct value of *TET2* mRNA expression between the control cells		
Time point	Lipo vs Unt	Lipo vs Scr
24 hours	p=0.877	p=0.286
48 hours	p=0.688	p=0.964
The p values of the Ct value of *TET3* mRNA expression between the control cells		
Time point	Lipo vs Unt	Lipo vs Scr
24 hours	p=0.328	p=0.648
48 hours	p=0.861	p=0.906

Table 5.6 Significance between gene expression for POLR2A and RPLP0 in Unt-, Scr- and Lipo - treated MCF10A cells

The p values of the Ct value of *POLR2A* mRNA expression between the control cells					
Time point	Lipo vs Unt	Lipo vs Scr	Lipo vs siTET1	Lipo vs siTET2	Lipo vs siTET3
24 hours	p=0.693	p=0.988	p>0.615	p=0.999	p>0.999
48 hours	p>0.999	p=0.993	p=0.903	p=0.993	p=0.998
The p values of the Ct value of *RPLPO* mRNA expression between the control cells					
Time point	Lipo vs Unt	Lipo vs Scr	Lipo vs siTET1	Lipo vs siTET2	Lipo vs siTET3
24 hours	p>0.999	p=0.995	p>0.999	p=0.999	p>0.999
48 hours	p=0.925	p>0.999	p=0.995	p=0.915	p=0.771

5.1.4 Levels of 5hmC are not affected by *TET* knockdown for 96 hours

As TET proteins are essential enzymes for the demethylation process, the downregulation of *TET* mRNA should consequentially reduce the levels of 5hmC. To test this, the level of 5hmC in the DNA was tested on cells treated with siRNA at 96 hours after transfection by a colorimetric assay, as described in the methods (see the methods in subchapter 2.6).

In the three independent biological experiments (Figure 5.6), the colorimetric assay was performed to measure the global 5hmC compared to the total cytosine (5hmC/C) content of DNA using Sunrise™ microplate reader. The data of this assay were analysed by the ordinary one-way ANOVA against the negative controls (Unt, Lipo and Scr). The global 5hmC/C content of DNA in BT474 in each group was in the range of 0.008% to 0.009%, and there was no

significant difference between each group and negative control (Lipo vs siTET1, p=0.993; Lipo vs siTET2, p= 0.991; Lipo vs siTET3, p=0.961). The results showed that the levels of 5hmC did not significantly change after 96 hours, because it may take some time to decrease the levels of TET proteins before the proteins act at the 5hmC level. It would be helpful to examine the levels of 5hmC in RNA and the expression of TET proteins in further experiments.

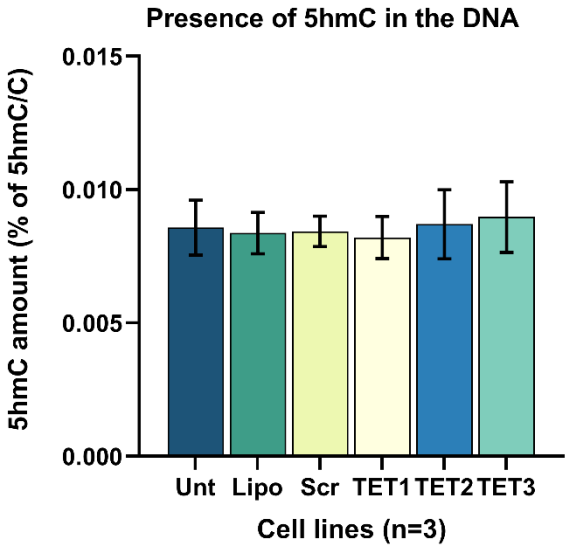


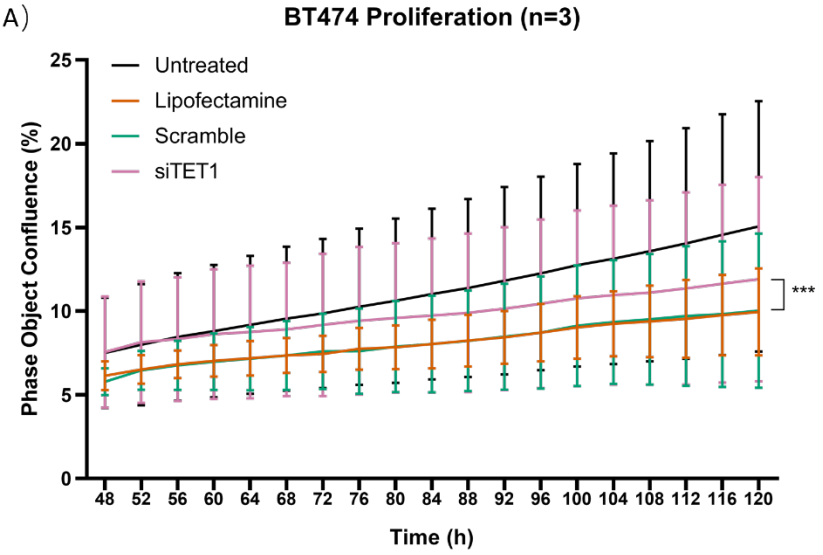
Figure 5.6 The levels of 5hmC in the DNA are not affected by TET knockdown in BT474 cells (n=3). BT474 cells were seeded in a 10 cm dish and mixed with siRNA and transfection reagent. On the following day, the cells were resuspended and seeded in 6-well plates. Cells were then harvested for colorimetric assay after 96 hours of reverse transfection. The result shown represents the mean ± SD of this group. Statistical significance was calculated by ordinary one-way ANOVA against Lipo.

5.1.5 Increase in cell proliferation is associated with the knockdown of *TET*

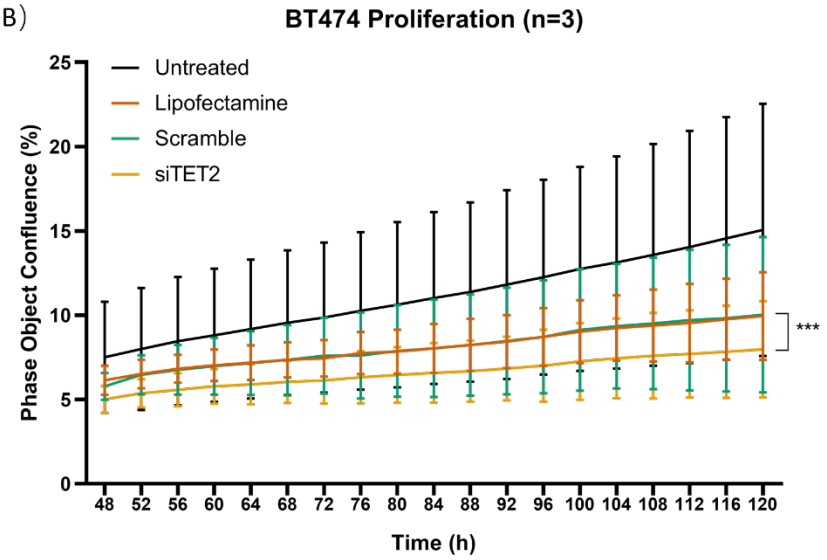
To investigate the proliferation of the cells transfected by siRNAs, a cell proliferation assay (see the methods in subchapter 2.8.3) was performed on the siRNA treated BT474 cells by monitoring it with the Incucyte ZOOM software. Cell proliferation was quantified by analysing the phase object confluence every 4 hours. The untreated cells showed a higher proliferation than cells treated with only Lipofectamine RNAiMAX reagent (Lipo) and cells treated with Lipofectamine RNAiMAX and scramble (Scr). This may be due to the technical aspect, since the transfection reagent is toxic to the cells and the cells become more sensitive after trypsinization. Interestingly, after 120 hours, it was demonstrated that knocking down *TET1* ($p<0.001$), and *TET3* ($p<0.001$) significantly increased the BT474 proliferation compared to Lipo control (Figure 5.7). In particular, *TET2* knockdown induces a significantly reduction ($p<0.0001$) of proliferation by 18.75% compared to Lipo.

Only the knockdown of *TET2* significantly decreased the cell proliferation compared to Lipo (Figure 5.7), supporting my hypothesis that three TET proteins work in different ways in cancer progression. TET2 may act as a tumour suppressor while TET1 and TET3 may act as oncogenes. Previous experiment of sublocalisation of TETs also support this hypothesis (see subchapter 4.1.1 and 4.1.3), I will discuss it in subchapter 5.2.

A)


BT474 Proliferation (n=3)
Untreated
Lipofectamine
Scramble
siTET1
Phase Object Confluence (%)
0
5
10
15
20
25
48 52 56 60 64 68 72 76 80 84 88 92 96 100 104 108 112 116 120
Time (h)
***

B)


BT474 Proliferation (n=3)
Untreated
Lipofectamine
Scramble
siTET2
Phase Object Confluence (%)
0
5
10
15
20
25
48 52 56 60 64 68 72 76 80 84 88 92 96 100 104 108 112 116 120
Time (h)
***

C)

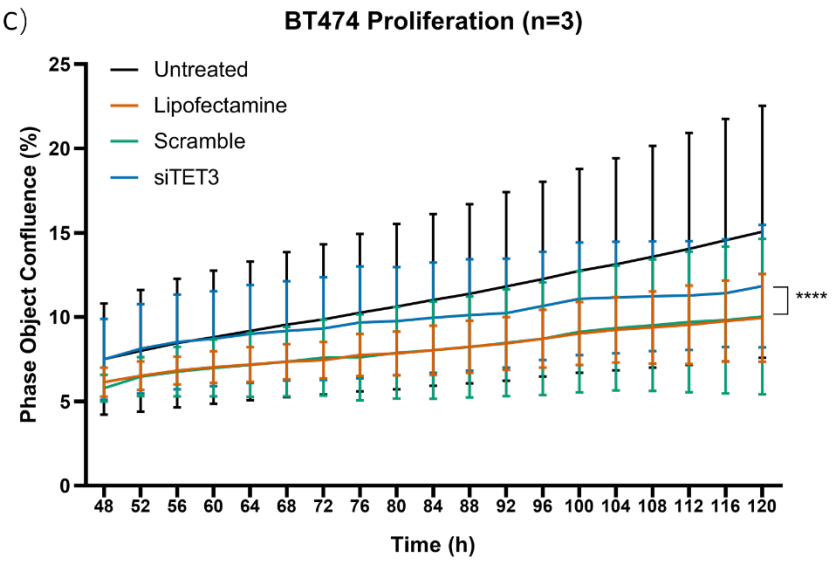

BT474 Proliferation (n=3)
Untreated
Lipofectamine
Scramble
siTET3
Phase Object Confluence (%)
0
5
10
15
20
25
48 52 56 60 64 68 72 76 80 84 88 92 96 100 104 108 112 116 120
Time (h)
****

Figure 5.7 (Previous page) Cell proliferation in BT474 cells following TET knockdown (n=5). **Graph A)** The proliferation of BT474 cells after treatment with siTET1 (20nM, pink) and negative controls (Unt, Lipo, Scr). **Graph B)** The proliferation of BT474 cells after treatment with siTET2 (20nM, yellow) and negative controls (Unt, Lipo, Scr). **Graph C)** The proliferation of BT474 cells after treatment with siTET3 (20nM, blue) and negative controls (Unt, Lipo, Scr).BT474 cells were seeded in 10 cm dishes and mixed with siRNA and transfection reagent. On the following day, the cells were resuspended and seeded in 96-well plates for cell proliferation analysis. The mean and SD for each condition is represented at each time point. Statistical significance was calculated by Mann-Whitney test against control, indicated with an asterisk: ***p<0.001, ****p<0.0001.

5.1.6 Evaluation of gene expression following *TET* knockdown in BT474

In the previous experiment (5.1.4), the proliferation of siRNA treated cells differed from the proliferation of control cells. *TET1* and *TET3* knockdown induce higher proliferation whilst *TET2* knockdown induces a lower proliferation. It would be very useful to investigate if *TET* siRNAs regulated any marker genes in breast cancer progression. Therefore, one of my aims is to investigate whether TET regulates any functional change in breast cells, qPCR was performed to test the expression of the target genes after *TET* knockdown. As I mentioned above, KI67 and MCM2 are biomarkers for measuring proliferation in breast cancer; BCL-2, and BCL-w are the marker genes of apoptosis; ESR1 and GATA3 are key transcription factors often silenced by DNA methylation in basal-like breast cancer; DNMT1 and DNMT3a are DNA methylases. Therefore, *DNMT1, DNMT3a, KI67, MCM2 BCL-2, BCL-w, ESR1* and *GATA3* were selected as target genes and their expression was analysed at 48 hours post transfection. According to my previous studies, I hypothesise that TET2 acts as a tumour suppressor while TET1 and TET3 act as oncogenes in breast cancer. If TET2 acts as a tumour suppressor, expression of Ki67 and MCM2 will increase, BCL-2, BCL-w, ESR1 and GATA3 will decrease after siTET2 transfection, on the contrary, if TET1 and TET3 protein act as an oncogene, the expression of these genes would change opposite.

Unfortunately, I lost the samples of Lipo control, so I could not normalise data to Lipo for this experiment. Based on previous experiments, the gene expression and the proliferation of Lipo and Scr had no significant change. Therefore, I normalised the qPCR data to Scr control for this experiment. However, there was no significant change (p values shown in Table 5.7 below) in the expression of these target genes (Figure 5.8). Thus, siTET did not change the mRNA

expression of these target genes. For the future experiment, the protein levels of these genes can be determined to evaluate if the post-translational regulation occurred.

Table 5.7 Significance between gene expression for *DNMT1, DNMT3a, Ki67, MCM2, Bcl-2, Bcl-W, ESR1, GATA3* in Unt-, Scr- and Lipo- in BT474

The p value of the Ct value of gene expression between siRNA treated cells and Scr				
Target gene	Scr vs Unt	Scr vs siTET1	Scr vs siTET2	Scr vs siTET3
DNMT1	p=0.701	p=0.685	p=0.845	p=0.522
DNMT3a	p=0.733	p=0.995	p=0.994	p=0.941
KI 67	p=0.398	p=0.994	p=0.997	p=0.905
MCM2	p=0.980	p=0.317	p=0.399	p=0.819
BCL-2	p=0.501	p=0.189	p=0.987	p=0.350
BCL-W	p=0.987	p=0.999	p=0.827	p>0.999
ESR1	p=0.995	p=0.936	p=0.995	p=0.863
GATA3	p=0.131	p=0.0.857	p=0.242	p=0.229

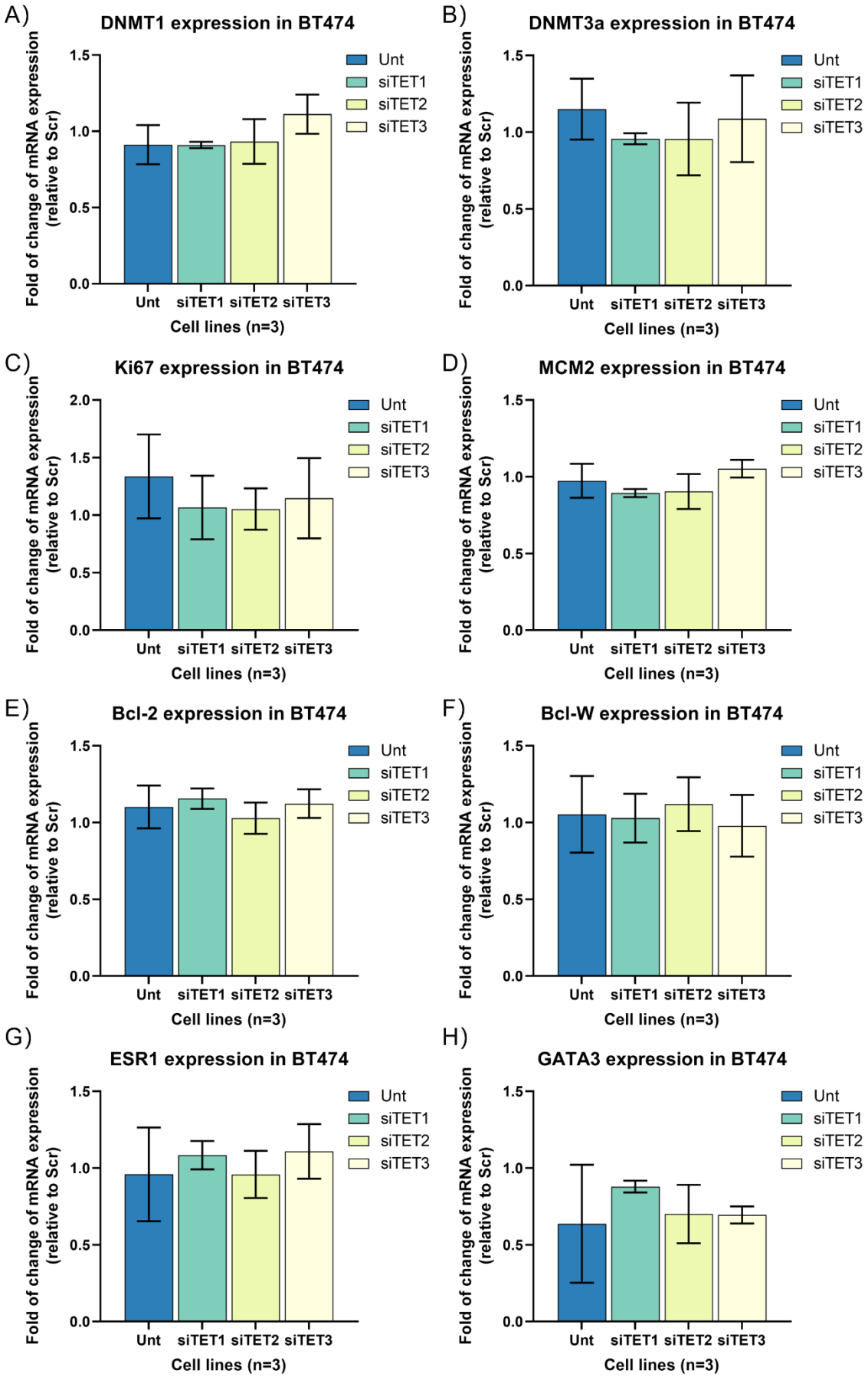

A)
DNMT1 expression in BT474
B)
DNMT3a expression in BT474
C)
Ki67 expression in BT474
D)
MCM2 expression in BT474
E)
Bcl-2 expression in BT474
F)
Bcl-W expression in BT474
G)
ESR1 expression in BT474
H)
GATA3 expression in BT474
Fold of change of mRNA expression (relative to Scr)
Unt
siTET1
siTET2
siTET3
Cell lines (n=3)

Figure 5.8 (Previous page) Target gene expression not significantly changes upon siRNA treatment in BT474 cells (n=3). BT474 cells were seeded in a 10 cm dish and mixed with siRNA and transfection reagent. On the following day, the cells were resuspended and seeded in 6-well plates. Cells were then harvested for RT-qPCR after 48 hours of reverse transfection. The mean and SD for each condition is represented in the bar chart. Statistically significant was calculated using the ordinary one-way ANOVA.

5.2 Discussion

This chapter reports the validation of the downregulation of *TET* mRNA expression in BC cells and non-tumorigenic breast cells (result see subchapter 5.1.2 and 5.1.3), the determination of the global 5hmC change in BT474 cells (result see subchapter 5.1.4), the measurement of cell proliferation of siRNA treated BT474 cells (result see subchapter 5.1.5) and the quantification of the mRNA expression of selected marker genes in BT474 cells (result see subchapter 5.1.6).

According to siRNA mediated knockdown experiments, the expression of three *TET* mRNAs was successfully reduced in BT474 cells individually, whilst in MCF10A, there is a considerable variation in each expression over time. The considerable variation could be because the number of cells not consistent, so the future experiment should seed a certain number of cells for transfection. Moreover, future work could use other non-malignancy breast cell lines, such as MCF12A and HB2, to validate the downregulation of *TET* mRNA expression. MCF12A is a non-tumorigenic, heterogeneous, ER-, mammary gland cell line, and HB2 is a triple positive mammary luminal epithelial cell line isolated from human milk (194, 195). Since results from my previous chapters mentioned that the levels of 5hmC increased in the cytoplasm in cancer compared to normal epithelium, I hypothesise that the distribution of 5hmC in ER negative cell lines may be more stable than in other subtypes. Therefore, MCF12A would be a better model to use.

To investigate the association of three TETs and 5hmC, I determined the levels of 5hmC by colorimetric assay after knockdown *TET* mRNA expression separately. The result of my colorimetric assay did not show a significant difference in the levels of 5hmC between the knockdown cell lines and the negative controls. This result could be because I did not find the

right time point. A previous study used the colorimetric assay and the immunofluorescence staining to show that silencing of TET1 and TET3 led to a decrease of 5hmC levels in uterine leiomyoma after siRNA transfection (196). However, this study did not mention when they quantified the 5hmC levels. Another study by Cakouros and colleagues measured the levels of 5hmC of human bone marrow-derived mesenchymal stem/stromal cells after being cultured under different conditions for two weeks (197). This study suggests that the levels of 5hmC need a long time to change. As the TET proteins are the enzymes for hydroxymethylation, the expression of TET proteins should be altered in advance, and then the levels of 5hmC will be altered later.

However, I did not perform the western blot experiments required to investigate the protein levels due to time limitations. Therefore, it is unknown when the TET proteins are reduced after the reverse transfection. Future work could validate the alteration of TET proteins followed by testing the levels of 5hmC.

According to the results from the previous chapters, the levels of 5hmC changed in the cytoplasm, which suggested the modifications of 5hmC may change in RNA. It would be helpful to validate if siTETs regulate the 5hmC levels in RNA. However, I did not perform any experiments to validate the levels of 5hmC in RNAs. Future work could validate the changes of 5hmC level in RNA in control cells and siRNA treated cells.

To provide evidence for the role of TETs in breast cancer, I performed functional studies after siRNA transfection. For functional studies, I hypothesised that *TET2* is a tumour suppressor gene, so I expected an increase of the expression of KI67 and MCM2 and a decrease of BCL-2, BCL-w, ESR1 and GATA3 after siTET2 transfection and opposite results after siTET1 and siTET3 transfections. However, there is no significant difference in marker genes' expression

between knockdown cell lines and negative controls. Same as the colorimetric assay, these gene expression could be altered after the reduction of TET proteins. Moreover, it would be helpful to perform western blots to examine the protein expression of these target genes. So that it could confirm if post-translational regulation occurred.

Interestingly, the cell proliferation assay showed the proliferation of BT474 was significantly increased after downregulating *TET1* and *TET3* mRNA. However, Navarro and colleagues suggested an opposite result that TET1 and TET3 knockdown decrease cell proliferation in leiomyoma (196). On the contrary, TET2 knockdown reduced cell proliferation in breast cancer cell lines, confirming the previous studies in mast cells by Montagner and colleagues, and in leukaemia cells by Wang and colleagues (198, 199). Another study suggests TET2-mediated DNA hydroxymethylation regulated cell proliferation by modulating bisphenol A/bisphenol S-stimulated pathway in ER+ breast cancer cells (200). Although BT474 is an ER- breast cancer cell line, it would be interesting to investigate if TET2 regulate cell proliferation in ER- cell lines through the same pathway in the future.

Based on the result from the previous chapter (see subchapter 4.1.1 and 4.1.3), the subcellular localisation of TET proteins does indeed suggest a tumour suppressive role of TET2. IHC result indicated that a higher proportion of positive stained cells of cytoplasmic TET2 in cancer compared to normal epithelial tissues. Moreover, the ICC result revealed a higher distribution of TET2 expression in both the nucleus and cytoplasm of cancer cells compared to normal cells. Additionally, from the previous *ISH* result (see subchapter 4.1.5), nuclear *TET2* mRNA is significantly lower in cancer compared to benign. DNA methylation and demethylation processes mainly occur within the nucleus, so the presence of TET2 in the cytoplasm in cancer may not directly participate in DNA demethylation regulation but in RNA

5hmC modification. The altered distribution of TET2 in cancer cells might suggest a disruption in its regulatory role in DNA demethylation, potentially contributing to tumorigenesis. However, the oncogenic roles of TET1 and TET3 require further investigation. Due to certain limitations, including time constraints and budget availability, I did not investigate the subcellular distribution of *TET1* and *TET3* mRNA on BC tissues. IHC results did not reveal a significant difference in TET1 or TET3 expression between cancer and normal tissues. In contrast, ICC result showed an increase in TET1 expression in the cytoplasm of cancer cells compared to normal cells. Since the lack of subcellular localisation of *TET1* mRNA, only the altered localisation of TET1 protein suggests that TET1 might have an oncogenic role, which supports the results of the proliferation assay. Conversely, the expression of TET3 showed a higher distribution in normal cells in both the nucleus and cytoplasm than in cancer cells in ICC experiment. Therefore, it is possible that TET3 may promote tumorigenesis, which is contrasts with the results of proliferation assay. Further study could perform *ISH* experiments rescue experiments, double knockdown experiments and other mechanistic studies to validate my hypothesis and fully understand the roles of TET proteins in BC.

Here are some limitations of my experiment. While I only reduced the expression of TET proteins, the TET proteins should be overexpressed to examine the alteration of 5hmC levels and functional studies. Furthermore, 5hmC levels are variable in breast cell lines, but I did not perform any effective measures to control the levels of 5hmC. This could lead to a lack of accuracy and repeatability. To control the levels of 5hmC, 5-azacytidine (5-AZA) could be a useful chemical compound, as 5-AZA inhibits DNA methylation, which is the income of 5hmC (201, 202). Otherwise, (S)-2-hydroxyglutarate (S-2HG), a competitive inhibitor of α-ketoglutarate, which inhibits the TET activity, could also reduce the 5hmC levels (203, 204).

6. Discussion and Conclusions

6.1 Summary of findings

Despite scientific advances, BC is still one of the most common cancers in women in Europe and in worldwide (1, 205). For the development of new targeted therapies, a complete understanding of how the cancer genome is regulated is essential.

Hydroxymethylation is one of the epigenetic changes occurring in the process of genome regulation. The loss of 5hmC is now considered a hallmark of melanoma (117), and it is being linked with progression and stages of several types of solid cancers. However, while a large number of 5hmC studies have sprung up, only a few research focused on breast cancer (104, 147, 161, 169, 170, 176, 184, 206, 207). Moreover, due to the technical limitations, the vast majority of the studies still focus on the investigation of 5hmC landscape of the human genome (208-210), while not many studies focus on the 5hmC functional study. Unlike previous studies, this work aimed to focus on the role of 5hmC specifically in breast cancer. Furthermore, the study also aimed to determine how the differential expression of the three ten-eleven translocation (TET) enzymes, catalysts of 5hmC synthesis, is altered in breast cancer compared to the non-diseased breast.

Therefore, the main aim of this study was to investigate the differential expression of 5hmC distribution and TET expression levels, the correlation between each TET and hydroxymethylation and the possible role of 5hmC and TET proteins in normal and breast cancer samples.

This study provided clear evidence that 5hmC is not only a modification of the DNA but could be involved in further regulatory pathways within the cytoplasm.

In chapter 3, the distribution of 5hmC in the breast was validated by IHC performed on non-malignant and cancer breast epithelium. I observed an interesting significant localisation of 5hmC in the cytoplasm of breast cancer cells, which has not been reported before. By validating this in colitis-associated cancers, I was also able to confirm that this novel finding is not a unique feature of breast cancer but could be a hallmark of other solid tumours. This study suggested a possibility that previous work ignored the existence of cytoplasmic 5hmC. For instance, the work from Pavlova and colleagues only quantified the 5hmC staining in nuclei but they did not analyse the cytoplasmic staining in melanoma patient tissues (211).

Moreover, the proportion of positive cell of cytoplasmic 5hmC was significantly higher in cancer compared to normal epithelium from ER- patients, which indicates the levels of 5hmC will probably relate to ER status in BC. Tsai and colleagues also suggest that a decrease of 5hmC could be a marker for poor prognostic factor in the ER/PR-negative breast cancer subtype (161). Taken together, the distribution of 5hmC may be subtype specific.

Since the levels of 5hmC are affected by the expression and function of TET proteins, it is beneficial to study the distribution and function of TETs. TET1 acts as both oncogene and tumour suppressor functions in cancers. As a tumour suppressor, TET1 represses cell migration, induces cell apoptosis, and prevents invasion and metastasis in cancer cells (123-125, 128, 193). Contradictorily, as an oncogene, TET1 reduces cellular proliferation in triple-negative breast cancer (131). Additionally, Good and colleagues discovered an isoform of TET1, $TET1^{ALT}$ which is overexpressed in triple-negative BC, and knockdown of *TET1* resulted in the decrease of 5hmC (131). Most studies regard TET2 as a tumour suppressor and describe how it regulates cell differentiation in luminal breast cancer (164). Zhang and colleagues reported a negative correlation ($p<0.05$) between 5hmC and TET2 expression in BC patients and a positive correlation ($p<0.001$) in patients carrying precancerous breast lesions (176).

This suggests that TET2 has different effects in patients with different pathological stages. Broome pointed out that depleting TET2 expression decreases the levels of 5hmC in ER+ BC, suggesting a possible mechanism for TET2 regulating 5hmC modification (180). Wu and colleagues showed that loss of nuclear TET2 and miR-200c expression is negatively correlated ($p<0.05$) with all-trans retinoic acid (ATRA) resistance in high tumour grade and aggressive BC (181). This result indicates microRNA may participate in the alteration of 5hmC and TET2 proteins. In comparison, the possible role of TET3 is still unknown, but a higher expression level in breast cancers, compared to normal tissues, has been observed (144). According to these findings, three TET proteins could affect the different hallmarks of cancer, so it would be helpful to demonstrate the possible function of TET proteins in cancer and normal breast cells.

Thus, the other aim of my study was to determine the expression of three TET proteins in BC. My results indicates that the distribution of TET2 showed the same trend of distribution of 5hmC, corresponding to a significant increase in the cancer cytoplasmic epithelium in ER- samples. This result supported an essential mechanism by Broome and colleague that TET2 co-bind with ER and GATA3 at enhancer elements which alters the regulation of 5hmC (166). This study suggests an oncogenic role of TET2 in ER+ breast cancer. On the contrary, Lyu and colleague suggests a tumour suppressor of TET2 in ER+ BC (212). From the result of KM plotter, the higher expression of *TET2* mRNA is positively correlated with the probability of higher survival rate in both ER+ and ER- BC patient samples. Taken together, it is possible that ER regulate or combine with TET2 and GATA3 to convert 5mC to 5hmC in breast cancer cells, but this hypothesis needs further validation.

A correlation analysis (only on 6 ER- samples) between the proportion of positive cells of 5hmC and the proportion of positive cells of three TET proteins was also performed in my study (Figure 4.2). The expression of TET2 was significantly negatively correlated to the 5hmC level in both the nuclei and the cytoplasm in cancer epithelium. Moreover, I analysed the changes of 5hmC and TETs between normal and cancer in each patient (Figure 4.3). There is a positive correlation between the changes of 5hmC and TET2 in the cytoplasm, while there is a negative correlation in the nucleus. Two correlations analysed data from different dimensions, one is cross-sectional data, and the other is the difference between two states (normal and cancer). Overall, the proportion of positive cells of TET2 and 5hmC were negatively correlated in between the patient, however, there was a significant increase of 5hmC and TET2 for each patient individually from normal breast cell to the BC cell.

Regarding the changes of 5hmC and TET2, I hypothesise that, in the cytoplasm, 5hmC is modified in RNA by the catalysis of TET2, followed by TET2 degradation. This hypothesis is based on evidence that the levels of 5hmC in tRNA were significantly decreased in TET2 knockdown in mouse ESCs (mESCs) compared with wild type mESCs (213). Moreover, it has been reported that TET2 is degraded via four different pathways (186).To validate my hypothesis, I performed the dot blots to detect the hydroxymethylation in DNA and different RNAs in breast cells.

To our knowledge, this study reported here determined a new for the first time the RNA hydroxymethylation in breast cells. My research determined that 5hmC distributes in the RNA, the long (>200 nt) and short (<200ng) RNAs, using the dot blot in MCF10A cells and BT474 cells. This could be further evidence supporting the notion that 5hmC may be a modification in multiple RNA types. This also confirmed some previous studies that 5hmC is present in the

RNA in mouse brain tissue, mRNA in Drosophila, and lncRNA in CRC (96, 97, 162, 163). This evidence also suggests that 5hmC may regulate the transcription or translational processes in the RNA in BC. Shen and colleagues supported this suggestion that TET-mediated tRNA hydroxymethylation promotes translation in mESCs and HEK293T cells (213). Besides, Lan and colleagues reported that RNA hydroxymethylation reduces the stability of pluripotency-promoting transcripts in mESCs and might also contribute to RNA splicing and mRNA metabolism (164). Therefore, it is assumed that 5hmC may regulate the stability of the mRNA. In addition, Zhou and colleagues pointed out that 5hmC-modified lncRNA genes play critical roles in cancer initiation, metastasis, and prognosis (214). Eventually, the levels of 5hmC in lncRNA could be a biomarker to predict cancer progressions.

My hypothesis is that, in the cytoplasm, 5hmC is added to the RNA by the catalysis of TET2, followed by TET2 degradation. To validate my hypothesis, it is useful to detect the distribution of *TET2* mRNA in benign and cancer tissues. There was no significant difference in the expression of *TET2* mRNA between cancer and benign samples. However, nuclear *TET2* was significantly lower in cancer than the normal epithelium, and cytoplasmic *TET2* did not show a significant difference between each group. A previous study reported that *TET2* can be regulated by miRNAs and lncRNAs in post-transcription regulation and it can be degraded in post-translation regulation via four different pathways in normal cells (186). Hence, it is possible that many *TET2* mRNAs are made in normal breast cells, but not many functional proteins are translated. In comparison, most of the *TET2* mRNA is translated into functional proteins in BC cells.

To further investigate the function of 5hmC and the TET enzymes, I performed functional studies after siRNA transfection. The mRNA expression of *TET1*, *TET2* and *TET3* was successfully downregulated in BT474. The level of 5hmC in knockdown cell lines was detected

by colorimetric assay, but it did not show a significant decrease between the knockdown cell lines and the negative controls, which was different from my expectation. As the TET proteins are the enzymes for hydroxymethylation, the expression of TET proteins should be altered in advance, and then the levels of 5hmC will be changed later. Further study should perform on the western blot to confirm the downregulation of the protein expression. To examine the effects of TET proteins in proliferation, apoptosis, and cell differentiation, the mRNA expression of *KI67*, *MCM2, BCL-2*, *BCL-w*, *ESR1* and *GATA3* were detected by qPCR in negative control and siRNA knockdown cell lines. However, I did not observe any significant difference in these marker genes. Like the colorimetric assay, these gene expression may be altered after the reduction of TET proteins. Interestingly, TET2 knockdown reduces cell proliferation of ER- BC cell lines, while TET1 and TET3 knockdown increase the cell proliferation of ER- BC cells. This result indicated that TET2 might act as a tumour suppressor. Li and colleagues suggest that TET2 works with ER and DNMTs regulated cell proliferation in ER+ BC cells (164). Moreover, Broom and colleagues also report that TET2 cooperates with ER to promote cell proliferation in ER+ BC cells (166). Thus, future studies could examine if TET2 cooperates with DNMTs to regulate cell proliferation in ER- BC cells.

6.2 Limitations of the study

The current research on the function of 5hmC remains obscure and controversial. Due to the lack of accurate research approaches, the research progress of 5hmC has been carried out slowly.

The first limitation of this study is that there is currently no accepted means of effectively controlling the variables of 5hmC in the BC cell lines, so I did not use the methods already

used in the current literature in my research. Due to the variability of 5hmC abundance in the cell lines, I did not perform the quantitative assays for the ICC experiment. Moreover, mRNA expression of *TET1*, *TET2* and *TET3* was not successfully downregulated in MCF10A. Though, previous studies showed that 5-AZA-2'-deoxycytidine treatment could increase the level of 5hmC in cell lines (202, 215). However, the length and the dose of this treatment need to be optimised. Due to time limitations, the challenge of the variability of 5hmC levels in cells was not overcome.

During the experiment, I observed that the 5hmC was mainly present in myoepithelium, not in luminal cells. Again, I did not perform any further analysis or experiments, due to the time limitation. Most BCs originate from the luminal epithelial cells involved in milk production which exhibit the most changes during morphogenesis (216). Instead, myoepithelial cells are thought to have only mostly contractile functions, helping to transport milk to the nipple for ejection (216). The myoepithelial cells not only have physiological functions like regulation of growth, differentiation and morphogenesis of neighbouring cells but also secrete growth factors and cytokines such as basic fibroblast growth factor (bFGF), transforming growth factor-a (TGF-a) and various cytokines. In BC progression, myoepithelial cells inhibit tumour cell growth, invasion, and angiogenesis, and prevent the invasion of tumour cells from the duct to the stroma. Moreover, they possibly alter the levels of the secretions resulting the changes of the tumour microenvironment (217). Therefore, 5hmC displays in myoepithelial cells, which may affect the tumour environment changes in the breast tissue. Further study should verify the amount of 5hmC and TET proteins in luminal and myoepithelial cells. If so, it would be interesting to study whether TET-mediated hydroxymethylation regulates the tumour microenvironment changes.

Moreover, since hypoxia is a hallmark of cancer, and oxygen is a necessary cofactor during the DNA modification process, this could be another reason for the different levels of 5hmC in two types of cells. At the beginning of my study, I hypothesised that low oxygen levels might influence TET proteins hydrolysis and 5hmC modification during the growth and progression of tumour cells. Not only hypoxia is a hallmark of cancer, but oxygen is a necessary cofactor during the DNA modification process. However, I did not observe any significant difference in 5hmC levels between normoxia and hypoxia. It is probable that 5hmC is variable amongst different cell types, which also agrees with the study by Nector and colleagues (174). Because of time limitations (and the pandemic), the challenge of the variability of 5hmC levels in cells was not overcome. Therefore, I did not continue to study the hypoxia condition.

Consequently, 5hmC in RNA was only observed in the dot blot experiments. Although, RNA hydroxymethylation was found in other cells (97, 164), it would be useful to perform other experiment to prove this novel finding, such as the colorimetric assay LC-MS, and hMeRIP-Seq. Notably, the colorimetric assay kit is available in our lab now, but it was not able to perform the experiment at the time of writing.

6.3 Future directions

As discussed, 5hmC is present in both DNA and RNAs, but the specific localisation of 5hmC in the nucleic acid sequences is still mostly elusive. It is imperative to develop a method to localise 5hmC on genome at a single-base resolution to take this further and to be able to correlate hydroxymethylation with the regulation of specific genes, as is currently possible for methylation studies. To identify the distribution and abundance of 5hmC, I used immunohistochemistry, dot blot and colorimetric assay. These qualitative or semi-

quantitative tests can only show the distributions of hydroxymethylation, but they cannot truly localise the 5hmC in cancer versus normal samples.

For mapping hydroxymethylation in the DNA, the gold standard method combines bisulphite sequencing (BS-seq) and oxidative bisulphite sequencing (oxBS-seq). In BS-seq, unmethylated cytosines are deaminated to uracils and then converted to thymines, whereas 5mC and 5hmC are not deaminated and can be sequenced as cytosine (218). In OxBS-seq, 5hmC is oxidised to 5fC which is sensitive to deamination by bisulphite, so that 5hmC and unmethylated cytosine read as thymine and only 5mC reads as cytosine after bisulphite treatment (218). The two sequencing methods run parallel on the same sample, to allow the final distinction between 5hmC and 5mC. However, as bisulphite conversion involves a harsh chemical conversion step, the majority of DNA is degraded after bisulphite treatment (219), and this process may be incompatible with RNA sequencing (218).

Two alternative methods, TET-assisted pyridine borane sequencing (TAPS) (220) and APOBEC-coupled epigenetic sequencing (ACE-seq) (221), are bisulphite-free methods and could provide few advantages on the classic BS/oxBS combination. Therefore, the advantage of these two methods is that they protect the hydroxymethyl group by glycosylation to be more suitable and refined choice methods for sequencing both DNA and RNA.

Finally, Oxford Nanopore Technologies has developed an Enzymatic Methyl-seq method based on APOBEC3A deamination, which can accurately identify 5hmC modifications on native DNA sequences (222) and adapted their own direct RNA sequencing to allow the identification of RNA modifications (223).

However, due to the limitation of techniques available, cost and time, the sequencing methods cannot be easily performed in the lab. Newer technique is still under development.

In this regard, for future studies, it would be very helpful if whole genome sequencing could be performed for hydroxymethylation on BC and normal breast samples.

As part of future work, it would be interesting to investigate the levels of 5hmC and the expression of TET proteins in luminal and myoepithelial cells. Previous works suggest that myoepithelial cells play an important role in the alteration of the microenvironment and tumour progression (217, 224). The tumour microenvironment is considered to be a major regulator of tumour progression (225). Moreover, previous work reported that macrophages, myoepithelial cells, and several ECM molecules play crucial roles in ductal morphogenesis (225). To start with, future work could investigate 5hmC levels in 3D organoids to mimic the breast duct. The level of 5hmC will be detected in 3D organoids derived from BT474 and MCF12A (both are ER- cell lines) to test whether the variability of 5hmC synthesis is dependent on cell types. If the levels of 5hmC are stable in 3D organoids, hypoxia markers, such as HIF1α, can also be examined in these 3D organoids to validate whether 5hmC synthesis is dependent on the tumour environment.

As mentioned in subchapter 3.2, the levels of 5hmC may be affected by the cell cycle. It would be helpful to study whether the level of 5hmC changes in cell cycle phases in breast cancer. Prikrylova and colleagues suggested that the amount of 5hmC is proportional to doubling time in Hela cells, and the amount of 5hmC showed a peak in the G1 phase and reduced through S and G2/M phases (226). I hypothesise that the 5hmC level may change in the cell cycle phase in BC. To validate the 5hmC level changes in cell cycle phases, future studies could synchronise cells in the G1/S phase using a double thymidine block and in the G2/M phase using a nocodazole block (226).

My research did not distinguish between TET isoforms ($TET1^{FL}$, $TET1^{ALT}$, TET2, $TET3^{FL}$), which were previously described in cancer cells. Based on Good and colleagues' work (131), the $TET1^{ALT}$ isoform characterised by different methylation and gene expression targets compared to the full-length variant. TET1 regulates 5hmC levels at gene promoters and exons through its CXXC domain. The balance between the expression of the full-length variant and of the CXXC-deficient $TET1^{ALT}$, which has been detected in a variety of tissues such as mice embryonic stem cells, mice brain tissues, but also in a variety of cancer cells lines, including breast, uterine and glioblastoma (101), may therefore serve as a regulation of methylation in a time or tissue-specific manners. Therefore, I hypothesise that different TET isoforms may act differently in breast cancer progression.

My study suggested that TET2 may be more important than TET1 and TET3 in this context. According to the NCBI database, *TET2* mRNA has 11 splice variants, which can produce 7 functional proteins (227). It would be helpful to investigate whether TET2 alternative splicing occurs in breast cells and if the activities of TET2 isoforms are different. Future work could also answer the question on whether alternative splicing is involved in regulating gene expression and in turn 5hmC modification. To quantify the expression of TET isoforms, real-time qPCR could be used to measure the transcription level of TET isoforms in different breast cell lines. Moreover, a new BaseScope Assay developed by ACD (Advanced Cell Diagnostics) could be used to identify the localisation of precise cells within a tissue expressing specific TET splice variants via *in situ* hybridisation. Thus, the expression and the sublocalisation could be validated by this experiment.

6.4 Conclusions

In summary, this study determined a novel localisation of 5hmC in the cytoplasm of cancer cells. The proportion of positive cell of cytoplasmic 5hmC was significantly higher in cancer compared to normal epithelium from ER- patients, which indicates the levels of 5hmC may be subtype specific. This study provided evidence that, in breast cells, 5hmC is a modification of both DNA and RNA, including long and short RNA molecules. The distribution of cytoplasmic TET2 showed correlations with the proportion of positive cell of 5hmC and this may have prognostic significance. Results presented here suggest that 5hmC may have a role in the development of certain sub-types of BC and TET2 is of particular relevance here.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021;71(3):209-49.

2. Breast Cancer Now. Facts and statistics 2021 2022 [Available from: https://breastcancernow.org/about-us/media/facts-statistics.

3. Ellis H, Mahadevan V. Anatomy and physiology of the breast. Surgery (Oxford). 2013;31(1):11-4.

4. Adriance MC, Inman JL, Petersen OW, Bissell MJ. Myoepithelial cells: good fences make good neighbors. Breast Cancer Research. 2005;7(5):190.

5. Gusterson B, Eaves CJ. Basal-like Breast Cancers: From Pathology to Biology and Back Again. Stem Cell Reports. 2018;10(6):1676-86.

6. Bombonati A, Sgroi DC. The molecular pathology of breast cancer progression. The Journal of Pathology. 2011;223(2):307-17.

7. Collignon E. Mining breast cancers with novel epigenetic modifications: Universite Libre de Bruxelles; 2017.

8. Casasent AK, Edgerton M, Navin NE. Genome evolution in ductal carcinoma in situ: invasion of the clones. The Journal of Pathology. 2017;241(2):208-18.

9. Poltavets V, Faulkner JW, Dhatrak D, Whitfield RJ, McColl SR, Kochetkova M. CXCR4-CCR7 Heterodimerization Is a Driver of Breast Cancer Progression. Life. 2021;11(10):1049.

10. Agarwal I BL. WHO classification.: PathologyOutlines.com website.; 2022 [updated November 3rd, 2022. Available from: https://www.pathologyoutlines.com/topic/breastmalignantwhoclassification.html. .

11. Dieci MV, Orvieto E, Dominici M, Conte P, Guarneri V. Rare breast cancer subtypes: histological, molecular, and clinical peculiarities. The OOncologist. 2014;19(8):805-13.

12. Tan P, Ellis I, Allison K, Brogi E, Fox S, Lakhani, Sr., et al. The 2019 WHO classification of tumours of the breast. Histopathology. 2020;77.

13. Makki J. Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. Clinical Medicine Insights: Pathology. 2015;8:23-31.

14. Eliyatkın N, Yalçın E, Zengel B, Aktaş S, Vardar E. Molecular Classification of Breast Carcinoma: From Traditional, Old-Fashioned Way to A New Age, and A New Way. The Journal of Breast Health. 2015;11(2):59-66.

15. Rakha EA, Reis-Filho JS, Baehner F, Dabbs DJ, Decker T, Eusebi V, et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. Breast Cancer Research. 2010;12(4):207.

16. Zurrida S, Veronesi U. A new TNM classification for breast cancer to meet the demands of the present and the challenges of the future. Womens Health (London). 2011;7(1):41-9.

17. Gradishar WJ, Moran MS, Abraham J, Aft R, Agnese D, Allison KH, et al. Breast Cancer, Version 3.2022, NCCN Clinical Practice Guidelines in Oncology. Journal of the National Comprehensive Cancer Network. 2022;20(6):691-722.

18. Waks AG, Winer EP. Breast Cancer Treatment: A Review. JAMA. 2019;321(3):288-300.

19. Cserni G, Chmielik E, Cserni B, Tot T. The new TNM-based staging of breast cancer. Virchows Archiv. 2018;472(5):697-703.

20. Benson JR. The TNM staging system and breast cancer. The Lancet Oncology. 2003;4(1):56-60.

21. Yersal O, Barutca S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. World Journal of Clinical Oncology. 2014;5(3):412-24.

22. Kumar P, Aggarwal R. An overview of triple-negative breast cancer. Archives of Gynecology and Obstetrics. 2016;293(2):247-69.

23. Norah Lynn Henry. Breast Cancer: Types of Treatment Cancer.Net2022 [Available from: https://www.cancer.net/cancer-types/breast-cancer/types-treatment.

24. Maughan KL, Lutterbie MA, Ham PS. Treatment of breast cancer. American Family Physician. 2010;81(11):1339-46.

25. Fisusi FA, Akala EO. Drug Combinations in Breast Cancer Therapy. Pharmaceutical Nanotechnology. 2019;7(1):3-23.

26. Harbeck N, Gnant M. Breast cancer. Lancet. 2017;389(10074):1134-50.

27. Early Breast Cancer Trialists' Collaborative G. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. The Lancet. 2011;378(9793):771-84.

28. Finn RS, Crown JP, Lang I, Boer K, Bondarenko IM, Kulyk SO, et al. The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): a randomised phase 2 study. The Lancet Oncology. 2015;16(1):25-35.

29. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California Cancer Registry. Cancer. 2007;109(9):1721-8.

30. Drăgănescu M, Carmocan C. Hormone Therapy in Breast Cancer. Chirurgia (Bucur). 2017;112(4):413-7.

31. Targeted Therapy to Treat Cancer National Cancer Institute2022 [Available from: https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies.

32. Peters GJ. From 'Targeted Therapy' to Targeted Therapy. Anticancer research. 2019;39(7):3341-5.

33. Shu M, Yan H, Xu C, Wu Y, Chi Z, Nian W, et al. A novel anti-HER2 antibody GB235 reverses Trastuzumab resistance in HER2-expressing tumor cells in vitro and in vivo. Scientific Reports. 2020;10(1):2986.

34. Juríková M, Danihel Ľ, Polák Š, Varga I. Ki67, PCNA, and MCM proteins: Markers of proliferation in the diagnosis of breast cancer. Acta Histochemica. 2016;118(5):544-52.

35. Yousef EM, Furrer D, Laperriere DL, Tahir MR, Mader S, Diorio C, et al. MCM2: An alternative to Ki-67 for measuring breast cancer cell proliferation. Modern Pathology. 2017;30(5):682-97.

36. Prihantono P, Hatta M, Binekada C, Sampepajung D, Haryasena H, Nelwan B, et al. Ki-67 Expression by Immunohistochemistry and Quantitative Real-Time Polymerase Chain Reaction as Predictor of Clinical Response to Neoadjuvant Chemotherapy in Locally Advanced Breast Cancer. J Oncol. 2017;2017:6209849.

37. Banin Hirata BK, Oda JMM, Losi Guembarovski R, Ariza CB, Oliveira CECd, Watanabe MAE. Molecular Markers for Breast Cancer: Prediction on Tumor Behavior. Disease Markers. 2014;2014:513158.

38. Sun X, Kaufman PD. Ki-67: more than a proliferation marker. Chromosoma. 2018;127(2):175-86.

39. Issac MSM, Yousef E, Tahir MR, Gaboury LA. MCM2, MCM4, and MCM6 in Breast Cancer: Clinical Utility in Diagnosis and Prognosis. Neoplasia. 2019;21(10):1015-35.

40. Williams MM, Cook RS. Bcl-2 family proteins in breast development and cancer: could Mcl-1 targeting overcome therapeutic resistance? Oncotarget. 2015;6(6):3519-30.

41. Carrington EM, Zhan Y, Brady JL, Zhang J-G, Sutherland RM, Anstee NS, et al. Anti-apoptotic proteins BCL-2, MCL-1 and A1 summate collectively to maintain survival of immune cell populations both in vitro and in vivo. Cell Death & Differentiation. 2017;24(5):878-88.

42. O'Brien MA, Kirby R. Apoptosis: A review of pro-apoptotic and anti-apoptotic pathways and dysregulation in disease. Journal of Veterinary Emergency and Critical Care (San Antonio). 2008;18(6):572-85.

43. Takaku M, Grimm SA, Wade PA. GATA3 in Breast Cancer: Tumor Suppressor or Oncogene? Gene Expression. 2015;16(4):163-8.

44. Afzaljavan F, Sadr AS, Savas S, Pasdar A. GATA3 somatic mutations are associated with clinicopathological features and expression profile in TCGA breast cancer patients. Scientific Reports. 2021;11(1):1679.

45. Voduc D, Cheang M, Nielsen T. GATA-3 Expression in Breast Cancer Has a Strong Association with Estrogen Receptor but Lacks Independent Prognostic Value. Cancer Epidemiology Biomarkers; Prevention. 2008;17(2):365-73.

46. Takaku M, Grimm SA, Roberts JD, Chrysovergis K, Bennett BD, Myers P, et al. GATA3 zinc finger 2 mutations reprogram the breast cancer transcriptional network. Nature Communications. 2018;9(1):1059.

47. McCleskey BC, Penedo TL, Zhang K, Hameed O, Siegal GP, Wei S. GATA3 Expression in Advanced Breast Cancer: Prognostic Value and Organ-Specific Relapse. American Journal of Clinical Pathology. 2015;144(5):756-63.

48. Breiling A, Lyko F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. Epigenetics & Chromatin. 2015;8:24.

49. Veland N, Chen T. Chapter 2 - Mechanisms of DNA Methylation and Demethylation During Mammalian Development. In: Tollefsbol TO, editor. Handbook of Epigenetics (Second Edition): Academic Press; 2017. p. 11-24.

50. Muñoz CC, Ares MS, López JRA. Chapter 49 - Colorectal carcinoma: From molecular pathology to clinical practice. In: Sierra AP, editor. Foundations of Colorectal Cancer: Academic Press; 2022. p. 551-61.

51. Vinson C, Chatterjee R. CG methylation. Epigenomics. 2012;4(6):655-63.

52. Luo C, Hajkova P, Ecker JR. Dynamic DNA methylation: In the right place at the right time. Science. 2018;361(6409):1336-40.

53. Cain JA, Montibus B, Oakey RJ. Intragenic CpG Islands and Their Impact on Gene Regulation. Frontiers in Cell and Developmental Biology. 2022;10.

54. Rose NR, Klose RJ. Understanding the relationship between DNA methylation and histone lysine methylation. Biochimica et biophysica acta. 2014;1839(12):1362-72.

55. Zhang W, Xu J. DNA methyltransferases and their roles in tumorigenesis. Biomarker Research. 2017;5(1):1.

56. Lyko F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. Nature Reviews Genetics. 2018;19(2):81-92.

57. Gibney ER, Nolan CM. Epigenetics and gene expression. Heredity. 2010;105(1):4-13.

58. Margot JB, Ehrenhofer-Murray AE, Leonhardt H. Interactions within the mammalian DNA methyltransferase family. BMC Molecular and Cell Biology. 2003;4:7.

59. Zhang TJ, Zhang LC, Xu ZJ, Zhou JD. Expression and prognosis analysis of DNMT family in acute myeloid leukemia. Aging (Albany NY). 2020;12(14):14677-90.

60. Patani H. Investigating cancer-like DNA methylation patterns and genome stability upon reprogramming to naïve human pluripotency: Queen Mary University of London; 2019.

61. Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. Nature. 2013;502(7472):472-9.

62. Moore LD, Le T, Fan G. DNA Methylation and Its Basic Function. Neuropsychopharmacology. 2013;38(1):23-38.

63. Tollefsbol T. Handbook of epigenetics: the new molecular and medical genetics: Academic Press; 2017.

64. Ohno R, Nakayama M, Naruse C, Okashita N, Takano O, Tachibana M, et al. A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells. Development. 2013;140(14):2892-903.

65. Bochtler M, Kolano A, Xu G-L. DNA demethylation pathways: Additional players and regulators. BioEssays. 2017;39(1):e201600178.

66. Bellacosa A, Drohat AC. Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites. DNA Repair (Amst). 2015;32:33-42.

67. Maiti A, Michelson AZ, Armwood CJ, Lee JK, Drohat AC. Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from DNA. Journal of the American Chemical Society. 2013;135(42):15813-22.

68. Dai Q, Sanstead PJ, Peng CS, Han D, He C, Tokmakoff A. Weakened N3 Hydrogen Bonding by 5-Formylcytosine and 5-Carboxylcytosine Reduces Their Base-Pairing Stability. ACS Chemical Biology. 2016;11(2):470-7.

69. Verdine GL, Bruner SD. How do DNA repair proteins locate damaged bases in the genome? Chemistry Biology. 1997;4(5):329-34.

70. Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. Genes & Development. 2016;30(7):733-50.

71. Song C-X, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nature Biotechnology. 2011;29(1):68-72.

72. Skvortsova K, Zotenko E, Luu PL, Gould CM, Nair SS, Clark SJ, et al. Comprehensive evaluation of genome-wide 5-hydroxymethylcytosine profiling approaches in human DNA. Epigenetics & Chromatin. 2017;10:16.

73. Yu M, Hon GC, Szulwach KE, Song C-X, Zhang L, Kim A, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell. 2012;149(6):1368-80.

74. Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J, et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature. 2011;473(7347):343.

75. Mariani CJ, Madzo J, Moen EL, Yesilkanal A, Godley LA. Alterations of 5-hydroxymethylcytosine in human cancers. Cancers. 2013;5(3):786-814.

76. Li W, Liu M. Distribution of 5-hydroxymethylcytosine in different human tissues. Journal of Nucleic Acids. 2011;2011.

77. Smeets E, Lynch AG, Prekovic S, Van den Broeck T, Moris L, Helsen C, et al. The role of TET-mediated DNA hydroxymethylation in prostate cancer. Molecular and Cellular Endocrinology. 2018;462(Pt A):41-55.

78. Madrid A, Chopra P, Alisch RS. Species-Specific 5 mC and 5 hmC Genomic Landscapes Indicate Epigenetic Contribution to Human Brain Evolution. Frontiers in Molecular Neuroscience. 2018;11(39).

79. Coppieters N, Dieriks BV, Lill C, Faull RLM, Curtis MA, Dragunow M. Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. Neurobiology of Aging. 2014;35(6):1334-44.

80. Stöger R, Scaife PJ, Shephard F, Chakrabarti L. Elevated 5hmC levels characterize DNA of the cerebellum in Parkinson’s disease. NPJ Parkinson's Disease. 2017;3(1):6.

81. Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biology. 2011;12(6):R54.

82. Szulwach KE, Li X, Li Y, Song C-X, Han JW, Kim S, et al. Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. PLoS Genetics. 2011;7(6):e1002154.

83. Lio C-WJ, Rao A. TET Enzymes and 5hmC in Adaptive and Innate Immune Systems. Frontiers in Immunology. 2019;10:210-.

84. Stojković V, Fujimori DG. Chapter Sixteen - Radical SAM-Mediated Methylation of Ribosomal RNA. In: He C, editor. Methods in Enzymology. 560: Academic Press; 2015. p. 355-76.

85. Roundtree IA, Evans ME, Pan T, He C. Dynamic RNA Modifications in Gene Expression Regulation. Cell. 2017;169(7):1187-200.

86. Chavali PL, Ramachandran R, Chavali S. Chapter 3 - Functional categories of RNA regulation. In: Pandey R, editor. Rna-Based Regulation in Human Health and Disease. 19: Academic Press; 2020. p. 59-83.

87. Song H, Zhang J, Liu B, Xu J, Cai B, Yang H, et al. Biological roles of RNA m5C modification and its implications in Cancer immunotherapy. Biomarker Research. 2022;10(1):15.

88. Popis MC, Blanco S, Frye M. Posttranscriptional methylation of transfer and ribosomal RNA in stress response pathways, cell differentiation, and cancer. Current Opinion in Oncology. 2016;28(1):65-71.

89. Bohnsack KE, Höbartner C, Bohnsack MT. Eukaryotic 5-methylcytosine (m^5C) RNA Methyltransferases: Mechanisms, Cellular Functions, and Links to Disease. Genes (Basel). 2019;10(2).

90. Bhagavan NV, Ha C-E. Chapter 24 - Regulation of Gene Expression. In: Bhagavan NV, Ha C-E, editors. Essentials of Medical Biochemistry (Second Edition). San Diego: Academic Press; 2015. p. 447-64.

91. Racz I, Kiraly I, Lasztily D. Effect of light on the nucleotide composition of rRNA of wheat seedlings. Planta. 1978;142(3):263-7.

92. Fu L, Guerrero CR, Zhong N, Amato NJ, Liu Y, Liu S, et al. Tet-Mediated Formation of 5-Hydroxymethylcytosine in RNA. Journal of the American Chemical Society. 2014;136(33):11582-5.

93. Huber SM, van Delft P, Mendil L, Bachman M, Smollett K, Werner F, et al. Formation and Abundance of 5-Hydroxymethylcytosine in RNA. ChemBioChem. 2015;16(5):752-5.

94. Zhang H-Y, Xiong J, Qi B-L, Feng Y-Q, Yuan B-F. The existence of 5-hydroxymethylcytosine and 5-formylcytosine in both DNA and RNA in mammals. Chemical Communications. 2016;52(4):737-40.

95. Huber SM, van Delft P, Tanpure A, Miska EA, Balasubramanian S. 2′-O-Methyl-5-hydroxymethylcytidine: A Second Oxidative Derivative of 5-Methylcytidine in RNA. Journal of the American Chemical Society. 2017;139(5):1766-9.

96. Hu H, Shu M, He L, Yu X, Liu X, Lu Y, et al. Epigenomic landscape of 5-hydroxymethylcytosine reveals its transcriptional regulation of lncRNAs in colorectal cancer. British Journal of Cancer. 2017;116(5):658-68.

97. Delatte B, Wang F, Ngoc LV, Collignon E, Bonvin E, Deplus R, et al. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. Science. 2016;351(6270):282-5.

98. Pastor WA, Aravind L, Rao A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nature Reviews Molecular Cell Biology. 2013;14(6):341-56.

99. Wu X, Zhang Y. TET-mediated active DNA demethylation: mechanism, function and beyond. Nature Reviews Genetics. 2017;18(9):517-34.

100. Melamed P, Yosefzon Y, David C, Tsukerman A, Pnueli L. Tet Enzymes, Variants, and Differential Effects on Function. Frontiers in Cell and Developmental Biology. 2018;6:22-.

101. Good CR, Madzo J, Patel B, Maegawa S, Engel N, Jelinek J, et al. A novel isoform of TET1 that lacks a CXXC domain is overexpressed in cancer. Nucleic Acids Research. 2017;45(14):8269-81.

102. Storebjerg TM, Strand SH, Høyer S, Lynnerup A-S, Borre M, Ørntoft TF, et al. Dysregulation and prognostic potential of 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) levels in prostate cancer. Clinical epigenetics. 2018;10(1):105.

103. Song C-X, Yi C, He C. Mapping recently identified nucleotide variants in the genome and transcriptome. Nature Biotechnology. 2012;30(11):1107-16.

104. Eleftheriou M, Pascual AJ, Wheldon LM, Perry C, Abakir A, Arora A, et al. 5-Carboxylcytosine levels are elevated in human breast cancers and gliomas. Clinical Epigenetics. 2015;7(1):88.

105. Song P, Tayier S, Cai Z, Jia G. RNA methylation in mammalian development and cancer. Cell Biology and Toxicology. 2021;37(6):811-31.

106. Jin B, Li Y, Robertson KD. DNA methylation: superior or subordinate in the epigenetic hierarchy? Genes & cancer. 2011;2(6):607-17.

107. Pfeifer GP, Xiong W, Hahn MA, Jin S-G. The role of 5-hydroxymethylcytosine in human cancer. Cell and Tissue Research. 2014;356(3):631-41.

108. Fleischer T, Frigessi A, Johnson KC, Edvardsen H, Touleimat N, Klajic J, et al. Genome-wide DNA methylation profiles in progression to in situ and invasive carcinoma of the breast with impact on gene transcription and prognosis. Genome Biology. 2014;15(8):435.

109. Locke WJ, Guanzon D, Ma C, Liew YJ, Duesing KR, Fung KYC, et al. DNA Methylation Cancer Biomarkers: Translation to the Clinic. Frontiers in Genetics. 2019;10(1150).

110. Pfeifer GP. Defining Driver DNA Methylation Changes in Human Cancer. International Journal of Molecular Sciences. 2018;19(4):1166.

111. Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, et al. Induction of Tumors in Mice by Genomic Hypomethylation. Science. 2003;300(5618):489-92.

112. de Almeida BP, Apolónio JD, Binnie A, Castelo-Branco P. Roadmap of DNA methylation in breast cancer identifies novel prognostic biomarkers. BMC Cancer. 2019;19(1):219.

113. Martinez FJ, Lee JH, Lee JE, Blanco S, Nickerson E, Gabriel S, et al. Whole exome sequencing identifies a splicing mutation in NSUN2 as a cause of a Dubowitz-like syndrome. Journal of Medical Genetics. 2012;49(6):380-5.

114. Li Y, Li J, Luo M, Zhou C, Shi X, Yang W, et al. Novel long noncoding RNA NMR promotes tumor progression via NSUN2 and BPTF in esophageal squamous cell carcinoma. Cancer Letters. 2018;430:57-66.

115. Chen X, Li A, Sun BF, Yang Y, Han YN, Yuan X, et al. 5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs. Nature Cell Biology. 2019;21(8):978-90.

116. Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell. 2012;149(6):1368-80.

117. Lian CG, Xu Y, Ceol C, Wu F, Larson A, Dresser K, et al. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. Cell. 2012;150(6):1135-46.

118. Johnson KC, Houseman EA, King JE, Von Herrmann KM, Fadul CE, Christensen BC. 5-Hydroxymethylcytosine localizes to enhancer elements and is associated with survival in glioblastoma patients. Nature Communications. 2016;7:13177.

119. Haffner MC, Chaux A, Meeker AK, Esopi DM, Gerber J, Pellakuru LG, et al. Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. Oncotarget. 2011;2(8):627-37.

120. Thomson JP, Meehan RR. The application of genome-wide 5-hydroxymethylcytosine studies in cancer research. Epigenomics. 2017;9(1):77-91.

121. Globisch D, Münzel M, Müller M, Michalakis S, Wagner M, Koch S, et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS ONE. 2010;5(12):e15367.

122. Nestor CE, Ottaviano R, Reddington J, Sproul D, Reinhardt D, Dunican D, et al. Tissue-type is a major modifier of the 5-hydroxymethylcytosine content of human genes. Genome Research. 2011:gr. 126417.111.

123. Sun M, Song C-X, Huang H, Frankenberger CA, Sankarasharma D, Gomes S, et al. HMGA2/TET1/HOXA9 signaling pathway regulates breast cancer growth and metastasis. Proceedings of the National Academy of Sciences. 2013;110(24):9920-5.

124. Hsu C-H, Peng K-L, Kang M-L, Chen Y-R, Yang Y-C, Tsai C-H, et al. TET1 Suppresses Cancer Invasion by Activating the Tissue Inhibitors of Metalloproteinases. Cell Reports. 2012;2(3):568-79.

125. Zhou Z, Zhang H-S, Liu Y, Zhang Z-G, Du G-Y, Li H, et al. Loss of TET1 facilitates DLD1 colon cancer cell migration via H3K27me3-mediated down-regulation of E-cadherin. Journal of Cellular Physiology. 2018;233(2):1359-69.

126. Wu J, Li H, Shi M, Zhu Y, Ma Y, Zhong Y, et al. TET1-mediated DNA hydroxymethylation activates inhibitors of the Wnt/β-catenin signaling pathway to suppress EMT in pancreatic tumor cells. Journal of Experimental & Clinical Cancer Research. 2019;38(1):348.

127. Fan J, Zhang Y, Mu J, He X, Shao B, Zhou D, et al. TET1 exerts its anti-tumor functions via demethylating DACT2 and SFRP2 to antagonize Wnt/β-catenin signaling pathway in nasopharyngeal carcinoma cells. Clinical Epigenetics. 2018;10(1):103.

128. Park SJ, Lee BR, Kim H-S, Ji YR, Sung YH, ShikChoi K, et al. Inhibition of Migration and Invasion by Tet-1 Overexpression in Human Lung Carcinoma H460 Cells. Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics. 2016;23(3):89-98.

129. Fan M, He X, Xu X. Restored expression levels of TET1 decrease the proliferation and migration of renal carcinoma cells. Molecular Medicine Reports. 2015;12(4):4837-42.

130. Neri F, Dettori D, Incarnato D, Krepelova A, Rapelli S, Maldotti M, et al. TET1 is a tumour suppressor that inhibits colon cancer growth by derepressing inhibitors of the WNT pathway. Oncogene. 2014;34:4168.

131. Good CR, Panjarian S, Kelly AD, Madzo J, Patel B, Jelinek J, et al. TET1-Mediated Hypomethylation Activates Oncogenic Signaling in Triple-Negative Breast Cancer. Cancer Research. 2018;78(15):4126-37.

132. Huang H, Jiang X, Li Z, Li Y, Song C-X, He C, et al. TET1 plays an essential oncogenic role in MLL-rearranged leukemia. Proceedings of the National Academy of Sciences. 2013;110(29):11994-9.

133. Collignon E, Canale A, Al Wardi C, Bizet M, Calonne E, Dedeurwaerder S, et al. Immunity drives TET1 regulation in cancer through NF-κB. Science Advances. 2018;4(6):eaap7309.

134. Wu H-X, Chen Y-X, Wang Z-X, Zhao Q, He M-M, Wang Y-N, et al. Alteration in TET1 as potential biomarker for immune checkpoint blockade in multiple cancers. Journal for ImmunoTherapy of Cancer. 2019;7(1):264.

135. Fan M, He X, Xu X. Restored expression levels of TET1 decrease the proliferation and migration of renal carcinoma cells. Molecular Medicine Reports. 2015;12(4):4837-42.

136. Zhu J, Wang K, Li T, Chen J, Xie D, Chang X, et al. Hypoxia-induced TET1 facilitates trophoblast cell migration and invasion through HIF1α signaling pathway. Scientific Reports. 2017;7(1):8077.

137. Wang L, Ozark PA, Smith ER, Zhao Z, Marshall SA, Rendleman EJ, et al. TET2 coactivates gene expression through demethylation of enhancers. Science Advances. 2018;4(11):eaau6986.

138. Liu Y, Lv J, Liang X, Yin X, Zhang L, Chen D, et al. Fibrin Stiffness Mediates Dormancy of Tumor-Repopulating Cells via a Cdc42-Driven Tet2 Epigenetic Program. Cancer Research. 2018;78(14):3926-37.

139. Puig I, Tenbaum SP, Chicote I, Arqués O, Martínez-Quintanilla J, Cuesta-Borrás E, et al. TET2 controls chemoresistant slow-cycling cancer cell survival and tumor recurrence. The Journal of Clinical Investigation. 2018;128(9).

140. Xu Y-p, Lv L, Liu Y, Smith MD, Li W-C, Tan X-m, et al. Tumor suppressor TET2 promotes cancer immunity and immunotherapy efficacy. The Journal of Clinical Investigation. 2019;129(10):4316-31.

141. Misawa K, Imai A, Mochizuki D, Mima M, Endo S, Misawa Y, et al. Association of TET3 epigenetic inactivation with head and neck cancer. Oncotarget. 2018;9(36):24480-93.

142. Gong F, Guo Y, Niu Y, Jin J, Zhang X, Shi X, et al. Epigenetic silencing of TET2 and TET3 induces an EMT-like process in melanoma. Oncotarget. 2017;8(1):315-28.

143. Bhattacharyya S, Pradhan K, Campbell N, Mazdo J, Vasantkumar A, Maqbool S, et al. Altered hydroxymethylation is seen at regulatory regions in pancreatic cancer and regulates oncogenic pathways. Genome Research. 2017;27(11):1830-42.

144. 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. 2015;10(7):e0133896-e.

145. Deng M, Zhang R, He Z, Qiu Q, Lu X, Yin J, et al. TET-Mediated Sequestration of miR-26 Drives EZH2 Expression and Gastric Carcinogenesis. Cancer Research. 2017;77(22):6069-82.

146. Feng Y, Li X, Cassady K, Zou Z, Zhang X. TET2 Function in Hematopoietic Malignancies, Immune Regulation, and DNA Repair. Frontiers in Oncology. 2019;9(210).

147. Wan F, Chen F, Fan Y, Chen D. Clinical Significance of TET2 in Female Cancers. Frontiers in Bioengineering and Biotechnology. 2022;10.

148. Yang SX, Hollingshead M, Rubinstein L, Nguyen D, Larenjeira ABA, Kinders RJ, et al. TET2 and DNMT3A mutations and exceptional response to 4'-thio-2'-deoxycytidine in human solid tumor models. Journal of Hematology & Oncology. 2021;14(1):83.

149. Koboldt DC, Kanchi KL, Gui B, Larson DE, Fulton RS, Isaacs WB, et al. Rare Variation in TET2 Is Associated with Clinically Relevant Prostate Carcinoma in African Americans. Cancer Epidemiology, Biomarkers & Prevention. 2016;25(11):1456-63.

150. Cao T, Pan W, Sun X, Shen H. Increased expression of TET3 predicts unfavorable prognosis in patients with ovarian cancer-a bioinformatics integrative analysis. Journal of Ovarian Research. 2019;12(1):101.

151. Carella A, Tejedor JR, Garcia MG, Urdinguio RG, Bayon GF, Sierra M, et al. Epigenetic downregulation of TET3 reduces genome-wide 5hmC levels and promotes glioblastoma tumorigenesis. International Journal of Cancer. 2020;146(2):373-87.

152. Song SJ, Poliseno L, Song MS, Ala U, Webster K, Ng C, et al. MicroRNA-antagonism regulates breast cancer stemness and metastasis via TET-family-dependent chromatin remodeling. Cell. 2013;154(2):311-24.

153. Chen Q, Yin D, Zhang Y, Yu L, Li X-D, Zhou Z-J, et al. MicroRNA-29a induces loss of 5-hydroxymethylcytosine and promotes metastasis of hepatocellular carcinoma through a TET-SOCS1-MMP9 signaling axis. Cell Death Disease. 2017;8(6):e2906-e.

154. Subik K, Lee J-F, Baxter L, Strzepek T, Costello D, Crowley P, et al. The expression patterns of ER, PR, HER2, CK5/6, EGFR, Ki-67 and AR by immunohistochemical analysis in breast cancer cell lines. Breast Cancer: Basic and Clinical Research. 2010;4:117822341000400004.

155. Shen L, Liang, Z. and Yu, H. Dot Blot Analysis of N6-methyladenosine RNA Modification Levels. Bio-protocol. 2017;7(1):e2095.

156. Lyng MB, Laenkholm A-V, Pallisgaard N, Ditzel HJ. Identification of genes for normalization of real-time RT-PCR data in breast carcinomas. BMC cancer. 2008;8:20-.

157. Zhan C, Zhang Y, Ma J, Wang L, Jiang W, Shi Y, et al. Identification of reference genes for qRT-PCR in human lung squamous-cell carcinoma by RNA-Seq. Acta Biochimica et Biophysica Sinica. 2014;46(4):330-7.

158. Győrffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast Cancer Research and Treatment. 2010;123(3):725-31.

159. Wajed SA, Laird PW, DeMeester TR. DNA methylation: an alternative pathway to cancer. Annals of Surgery. 2001;234(1):10-20.

160. Barciszewska A-M, Murawa D, Gawronska I, Murawa P, Nowak S, Barciszewska MZ. Analysis of 5-Methylcytosine in DNA of Breast and Colon Cancer Tissues. IUBMB life. 2007;59(12):765-70.

161. Tsai KW, Li GC, Chen CH, Yeh MH, Huang JS, Tseng HH, et al. Reduction of global 5-hydroxymethylcytosine is a poor prognostic factor in breast cancer patients, especially for an ER/PR-negative subtype. Breast Cancer Research and Treatment. 2015;153(1):219-34.

162. Miao Z, Xin N, Wei B, Hua X, Zhang G, Leng C, et al. 5-hydroxymethylcytosine is detected in RNA from mouse brain tissues. Brain research. 2016;1642:546-52.

163. Lin IH, Chen Y-F, Hsu M-T. Correlated 5-Hydroxymethylcytosine (5hmC) and Gene Expression Profiles Underpin Gene and Organ-Specific Epigenetic Regulation in Adult Mouse Brain and Liver. PLoS ONE. 2017;12(1):e0170779.

164. Lan J, Rajan N, Bizet M, Penning A, Singh NK, Guallar D, et al. Functional role of Tet-mediated RNA hydroxymethylcytosine in mouse ES cells and during differentiation. Nature Communications. 2020;11(1):4956.

165. Zauri M, Berridge G, Thézénas ML, Pugh KM, Goldin R, Kessler BM, et al. CDA directs metabolism of epigenetic nucleosides revealing a therapeutic window in cancer. Nature. 2015;524(7563):114-8.

166. Broome R, Chernukhin I, Jamieson S, Kishore K, Papachristou EK, Mao S-Q, et al. TET2 is a component of the estrogen receptor complex and controls 5mC to 5hmC conversion at estrogen receptor *cis* -regulatory regions. Cell Reports. 2021;34(8).

167. Haffner MC, Chaux A, Meeker AK, Esopi D, Gerber J, Pellakuru LG, et al. Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. Oncotarget. 2011;2(8).

168. Bediaga NG, Acha-Sagredo A, Guerra I, Viguri A, Albaina C, Ruiz Diaz I, et al. DNA methylation epigenotypes in breast cancer molecular subtypes. Breast Cancer Research. 2010;12(5):R77-R.

169. Yang H, Liu Y, Bai F, Zhang JY, Ma SH, Liu J, et al. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. Oncogene. 2013;32(5):663-9.

170. Yang H, Liu Y, Bai F, Zhang J, Ma S, Liu J, et al. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. Oncogene. 2013;32(5):663.

171. Wheldon Lee M, Abakir A, Ferjentsik Z, Dudnakova T, Strohbuecker S, Christie D, et al. Transient Accumulation of 5-Carboxylcytosine Indicates Involvement of Active Demethylation in Lineage Specification of Neural Stem Cells. Cell Reports. 2014;7(5):1353-61.

172. Au - Ramsawhook AH, Au - Lewis LC, Au - Eleftheriou M, Au - Abakir A, Au - Durczak P, Au - Markus R, et al. Immunostaining for DNA Modifications: Computational Analysis of Confocal Images. JoVE. 2017(127):e56318.

173. Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, et al. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Molecular Cell. 2011;42(4):451-64.

174. Nestor CE, Ottaviano R, Reinhardt D, Cruickshanks HA, Mjoseng HK, McPherson RC, et al. Rapid reprogramming of epigenetic and transcriptional profiles in mammalian culture systems. Genome Biology. 2015;16(1):11.

175. Thienpont B, Steinbacher J, Zhao H, D'Anna F, Kuchnio A, Ploumakis A, et al. Tumour hypoxia causes DNA hypermethylation by reducing TET activity. Nature. 2016;537:63.

176. Zhang Z, Jin Y, Zhang W, Chu C, Zhang K, Gao X, et al. Values of 5mC, 5hmC, and TET2 for identifying the presence and progression of breast precancerous lesion. Journal of Clinical Laboratory Analysis. 2020;34(5):e23162.

177. R Development Core Team: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. 2014, [http://www.R-project.org/].

178. Kim MR, Wu M-J, Zhang Y, Yang J-Y, Chang CJ. TET2 directs mammary luminal cell differentiation and endocrine response. Nature Communications. 2020;11(1):4642.

179. Shi D-Q, Ali I, Tang J, Yang W-C. New insights into 5hmC DNA modification: generation, distribution and function. Frontiers in Genetics. 2017;8:100.

180. Broome RE. Characterising the role of TET2 in oestrogen receptor positive breast cancer (Doctoral thesis): University of Cambridge; 2020.

181. Wu MJ, Kim MR, Chen YS, Yang JY, Chang CJ. Retinoic acid directs breast cancer cell state changes through regulation of TET2-PKCζ pathway. Oncogene. 2017;36(22):3193-206.

182. Tan J, Song C, Wang D, Hu Y, Liu D, Ma D, et al. Expression of hormone receptors predicts survival and platinum sensitivity of high-grade serous ovarian cancer. Bioscience Reports. 2021;41(5).

183. Chen S, Dai X, Gao Y, Shen F, Ding J, Chen Q. The positivity of estrogen receptor and progesterone receptor may not be associated with metastasis and recurrence in epithelial ovarian cancer. Scientific Reports. 2017;7(1):16922.

184. Wu S-L, Zhang X, Chang M, Huang C, Qian J, Li Q, et al. Genome-wide 5-Hydroxymethylcytosine Profiling Analysis Identifies MAP7D1 as A Novel Regulator of Lymph Node Metastasis in Breast Cancer. Genomics, Proteomics & Bioinformatics. 2021;19(1):64-79.

185. Ramasamy D, Rao AKDM, Balaiah M, Rangan AV, Sundersingh S, Veluswami S, et al. Locus-specific enrichment analysis of 5-hydroxymethylcytosine reveals novel genes associated with breast carcinogenesis. bioRxiv. 2022:2022.06.17.496656.

186. Cong B, Zhang Q, Cao X. The function and regulation of TET2 in innate immunity and inflammation. Protein & Cell. 2021;12(3):165-73.

187. ŞErban C, Anca Maria C, Marius R. The Story of MCF-7 Breast Cancer Cell Line: 40 years of Experience in Research. Anticancer Research. 2015;35(6):3147.

188. Fogh J, Fogh JM, Orfeo T. One hundred and twenty-seven cultured human tumor cell lines producing tumors in nude mice. Journal of the National Cancer Institute. 1977;59(1):221-6.

189. Chavez KJ, Garimella SV, Lipkowitz S. Triple negative breast cancer cell lines: one tool in the search for better treatment of triple negative breast cancer. Breast Disease. 2010;32(1-2):35-48.

190. Qu Y, Han B, Yu Y, Yao W, Bose S, Karlan BY, et al. Evaluation of MCF10A as a Reliable Model for Normal Human Mammary Epithelial Cells. PLoS ONE. 2015;10(7):e0131285.

191. Tan E, Chin CSH, Lim ZFS, Ng SK. HEK293 Cell Line as a Platform to Produce Recombinant Proteins and Viral Vectors. Frontiers in Bioengineering and Biotechnology. 2021;9:796991.

192. Yuan J, Xu WW, Jiang S, Yu H, Poon HF. The scattered twelve tribes of HEK293. Biomedical and Pharmacology Journal. 2018;11(2):621-3.

193. Teng S, Ma C, Yu Y, Yi C. Hydroxyurea promotes TET1 expression and induces apoptosis in osteosarcoma cells. Bioscience Reports. 2019;39(5):BSR20190456.

194. Sweeney MF, Sonnenschein C, Soto AM. Characterization of MCF-12A cell phenotype, response to estrogens, and growth in 3D. Cancer Cell International. 2018;18:43.

195. Caradonna F, Luparello C. Cytogenetic characterization of HB2 epithelial cells from the human breast. In Vitro Cellular & Developmental Biology - Animal. 2014;50(1):48-55.

196. Navarro A, Yin P, Ono M, Monsivais D, Moravek MB, Coon JSt, et al. 5-Hydroxymethylcytosine promotes proliferation of human uterine leiomyoma: a biological link to a new epigenetic modification in benign tumors. The Journal of Clinical Endocrinology & Metabolism. 2014;99(11):E2437-45.

197. Cakouros D, Hemming S, Gronthos K, Liu R, Zannettino A, Shi S, et al. Specific functions of TET1 and TET2 in regulating mesenchymal cell lineage determination. Epigenetics & chromatin. 2019;12(1):3.

198. Montagner S, Leoni C, Emming S, Della Chiara G, Balestrieri C, Barozzi I, et al. TET2 Regulates Mast Cell Differentiation and Proliferation through Catalytic and Non-catalytic Activities. Cell Reports. 2016;15(7):1566-79.

199. Wang Y, Xiao M, Chen X, Chen L, Xu Y, Lv L, et al. WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation. Molecular Cell. 2015;57(4):662-73.

200. Li Z, Lyu C, Ren Y, Wang H. Role of TET Dioxygenases and DNA Hydroxymethylation in Bisphenols-Stimulated Proliferation of Breast Cancer Cells. Environmental Health Perspectives. 2020;128(2):27008.

201. Christman JK. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene. 2002;21(35):5483-95.

202. Sajadian SO, Ehnert S, Vakilian H, Koutsouraki E, Damm G, Seehofer D, et al. Induction of active demethylation and 5hmC formation by 5-azacytidine is TET2 dependent and suggests new treatment strategies against hepatocellular carcinoma. Clinical Epigenetics. 2015;7(1):98.

203. Yamashita AS, da Costa Rosa M, Borodovsky A, Festuccia WT, Chan T, Riggins GJ. Demethylation and epigenetic modification with 5-azacytidine reduces IDH1 mutant glioma growth in combination with temozolomide. Neuro Oncology. 2019;21(2):189-200.

204. Chowdhury B, McGovern A, Cui Y, Choudhury SR, Cho I-H, Cooper B, et al. The hypomethylating agent Decitabine causes a paradoxical increase in 5-hydroxymethylcytosine in human leukemia cells. Scientific Reports. 2015;5(1):9281.

205. Guttery DS, Page K, Hills A, Woodley L, Marchese SD, Rghebi B, et al. Noninvasive Detection of Activating Estrogen Receptor 1 Mutations in Estrogen Receptor–Positive Metastatic Breast Cancer. Clinical Chemistry. 2015;61(7):974-82.

206. Liu N-T, Perng C-L, Chou Y-C, Ko P-S, Lin Y-J, Lin Y-C, et al. Loss of ten-eleven translocation 1 (TET1) expression as a diagnostic and prognostic biomarker of endometrial carcinoma. PLoS ONE. 2021;16(11):e0259330.

207. Jia J, Shi Y, Chen L, Lai W, Yan B, Jiang Y, et al. Decrease in Lymphoid Specific Helicase and 5-hydroxymethylcytosine Is Associated with Metastasis and Genome Instability. Theranostics. 2017;7:3920-32.

208. Han X, Guo J, Wang M, Zhang N, Ren J, Yang Y, et al. Dynamic DNA 5-hydroxylmethylcytosine and RNA 5-methycytosine reprogramming during early human development. Genomics, Proteomics & Bioinformatics. 2022.

209. He B, Zhang C, Zhang X, Fan Y, Zeng H, Liu Je, et al. Tissue-specific 5-hydroxymethylcytosine landscape of the human genome. Nature Communications. 2021;12(1):4249.

210. Cui X-L, Nie J, Ku J, Dougherty U, West-Szymanski DC, Collin F, et al. A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. Nature Communications. 2020;11(1):6161.

211. Pavlova O, Fraitag S, Hohl D. 5-Hydroxymethylcytosine Expression in Proliferative Nodules Arising within Congenital Nevi Allows Differentiation from Malignant Melanoma. Journal of Investigative Dermatology. 2016;136(12):2453-61.

212. Lyu R, Zhu X, Shen Y, Xiong L, Liu L, Liu H, et al. Tumour suppressor TET2 safeguards enhancers from aberrant DNA methylation and epigenetic reprogramming in ERα-positive breast cancer cells. Epigenetics. 2021:1-15.

213. Shen H, Ontiveros RJ, Owens MC, Liu MY, Ghanty U, Kohli RM, et al. TET-mediated 5-methylcytosine oxidation in tRNA promotes translation. Journal of Biological Chemistry. 2021;296:100087.

214. Zhou M, Hou P, Yan C, Chen L, Li K, Wang Y, et al. Cell-free DNA 5-hydroxymethylcytosine profiles of long non-coding RNA genes enable early detection and progression monitoring of human cancers. Clinical Epigenetics. 2021;13(1):197.

215. Rocha MA, Veronezi GMB, Felisbino MB, Gatti MSV, Tamashiro WMSC, Mello MLS. Sodium valproate and 5-aza-2'-deoxycytidine differentially modulate DNA demethylation in G1 phase-arrested and proliferative HeLa cells. Scientific Reports. 2019;9(1):18236.

216. Lakhani SR, O'Hare MJ. The mammary myoepithelial cell - Cinderella or ugly sister? Breast Cancer Research. 2000;3(1):1.

217. Pandey PR, Saidou J, Watabe K. Role of myoepithelial cells in breast tumor progression. Frontiers in Bioscience-Landmark. 2010;15(1):226-36.

218. Booth MJ, Ost TWB, Beraldi D, Bell NM, Branco MR, Reik W, et al. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. Nature Protocols. 2013;8(10):1841-51.

219. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proceedings of the National Academy of Sciences. 1992;89(5):1827-31.

220. Liu Y, Siejka-Zielińska P, Velikova G, Bi Y, Yuan F, Tomkova M, et al. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. Nature Biotechnology. 2019;37(4):424-9.

221. Schutsky EK, DeNizio JE, Hu P, Liu MY, Nabel CS, Fabyanic EB, et al. Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. Nature Biotechnology. 2018;36:1083.

222. Sun Z, Vaisvila R, Yan B, Baum C, Saleh L, Samaranayake M, et al. Non-destructive enzymatic deamination enables single molecule long read sequencing for the determination of 5-methylcytosine and 5-hydroxymethylcytosine at single base resolution. BioRxiv. 2019:2019.12.20.885061.

223. Parker MT, Knop K, Sherwood AV, Schurch NJ, Mackinnon K, Gould PD, et al. Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m6A modification. eLife. 2020;9:e49658.

224. Polyak K, Hu M. Do Myoepithelial Cells Hold the Key for Breast Tumor Progression? Journal of Mammary Gland Biology and Neoplasia. 2005;10(3):231-47.

225. Place AE, Jin Huh S, Polyak K. The microenvironment in breast cancer progression: biology and implications for treatment. Breast Cancer Research. 2011;13(6):227.

226. Prikrylova T, Robertson J, Ferrucci F, Konorska D, Aanes H, Manaf A, et al. 5-hydroxymethylcytosine Marks Mammalian Origins Acting as a Barrier to Replication. Scientific Reports. 2019;9(1):11065.

227. TET2 tet methylcytosine dioxygenase 2 [Homo sapiens (human)] [updated 25.09.2022. Available from: https://www.ncbi.nlm.nih.gov/gene/54790.

Appendix

In Figure 1.2, I used the figure from Poltavets and colleagues' work. According to MDPI open access Information and Policy (see below), no special permission is required for re-using the figures.

