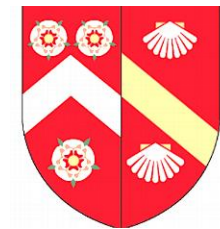


Heterogeneity within Colorectal Cancer Cell Lines and Epigenetic Regulation of *CD24*

Mustak Ibn Ayub



Wadham College

This thesis has been submitted to the Division of Medical Sciences, University of
Oxford, in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Oncology

Michaelmas Term, 2016
Weatherall Institute of Molecular Medicine



Declaration

I, Mustak Ibn Ayub, hereby declare that the work presented in this thesis is original and has been carried out by me as part of my doctoral degree requirements. This work will not be presented for any other degree in this University or elsewhere. I obtained research supports from other lab members at various stages of this work, which has been acknowledged duly. I particularly want to mention that the drug study during this thesis was designed and implemented by Dr. Djamila Ouaret of the Bodmer Laboratory where I played a supportive role. DNA sequence analyses were conducted by Dr. Kamila Koprowska of the Bodmer Laboratory and Dr. Simon McGowan of CBRG, WIMM, Oxford. Throughout the research period, I got intellectual support from Professor Sir Walter Bodmer (my supervisor), Dr. Jennifer Wilding and Dr. Djamila Ouaret.

Dedication

When I returned home in Bangladesh after finishing my DPhil research in the University of Oxford, holding my hands my ailing father said in an evening - My son, I have been praying to God so that He keeps me alive to see your return; now as you have come back, I don't think I will live for too long." My father passed away within a month after this conversation.

I dedicate this work to my late father **Md Ayub Ali** who loved me more than anything else in the world and to my mother **Mrs Sheherunnesa** whose care and love still inspire me to live in this world.

I want to include here **Professor Sir Walter F. Bodmer**, who became a fatherly figure away from my home and patiently supervised me with prudence, passion and affection from the beginning to the end of this research work.

Acknowledgements

When I started my DPhil, I did not envision the ending I have had here.

I open this acknowledgement by thanking my supervisor Professor Sir Walter Bodmer, who facilitated a wonderful research environment for me and guided me to the right direction with his knowledge, passion and experience. I also acknowledge the vital contribution of Dr. Jennifer Wilding who helped me eloquently throughout my research.

I want to extend special thanks to Dr. Djamila Ouaret for her relentless support and affection inside and outside the lab. Anything I have learnt about research planning and experiment set up- it came from her.

I have been hugely benefitted from Dr. Kamila Koprowska's expertise and efforts. I acknowledge her contribution in my research. I want to thank Dr. Neil Ashley, who, in the beginning helped to boost my confidence in research. Here, I also remember the encouragement of Dr. Shazad Ashraf who inspired me to work hard. I want to thank Dr. Trevor Yeung, whose seminal work on CD24 and CD44 became an important background support for my research.

I remember the contributions of previous DPhil students in the Bodmer Laboratory- Dr. Matthew F. Jones, Dr. Linting Hsia, Dr. Ruchi Patel and Jens Puschhof for their friendship and help.

I extend my heartfelt appreciations to the current lab members- Haoyu Liu, Laura Colling, Marahaini Musa, Peter Kalugin, Thomas Barnes and Shuhui Lui for their compassion and support. I fondly remember the contribution of Tammy Dey who gave us all a sense of family here in Oxford.

During my entire DPhil time, I got unreserved love from my friends who helped me to go through this journey. Among them I want to particularly thank Bahar Tuncgenc, Adam Berrington, Brendan Fong, Stephanie Lin and Olivia Waring. Bahar, Adam and Brendan have provided extraordinary support by giving candid suggestions on my manuscripts. I shall carry their love and friendship with me for my whole life. In this place I shall also mention my lifelong friend S.M Mahbubur Rashid and my new found friend in Oxford Dr. Tania S. Khaleque for their support.

My research was funded by Commonwealth Scholarship Commission, UK. Without this, I could not pursue my study here. They have made a huge contribution in my life and I promise to return the favour by disseminating the knowledge I have gathered here among others. I am grateful to my first college in Oxford, Hertford College for giving me a chance to discover my strength and nurture them by providing a wonderful supporting environment. I am extremely grateful to my current college, Wadham College for providing me with vital supports in the last years of my Oxford life. I am forever indebted to them.

I have been extremely lucky to have a very supportive family. I am forever grateful to my parents Md Ayub Ali and Mrs Sheherunnesa and my siblings Nesar Uddin Ayub, M Ansar Uddin and Md Tazuddin. Without their contribution and sacrifice, I could not be in Oxford. My life was crafted by the careful plan of my eldest brother Nesar Uddin Auyb. His thoughtful and often strict guidance was vital for my success. Here I thank him for his selfless dedication to shape up my career. In this place, I also remember the contributions of my teachers and mentors from all the previous stages, who enriched my life with their knowledge and wisdom.

This journey of my DPhil was a challenging one, not only from the research perspective, but also for many extra academic issues. I have learned from all of them. I shall cherish all the memories and friendship from this period of my life. To me, those are equally important as to my research achievements.

Mustak Ayub

4th January, 2017, Oxford

Note to the examiners/reviewers

For most part in this thesis, several clones and subclones have been studied. They are summarised here for quick tracking.

Primary name of the clone	Alias	Source	Methods of generation
Clone 1	: Clone 1 ^{CD24+} : LS174T_Clone 1 ^{CD24+}	LS174T	Single cell sorting of the bulk LS174T cells by FACS
Clone 2	: Clone 2 ^{CD24-} : LS174T_Clone 2 ^{CD24-}	LS174T	Single cell sorting of the bulk LS174T cells by FACS
Sub 3	: Subclone 3	Clone 2	Single cell sorting of the CD24 ⁺ cells of Clone 2 of LS174T
Sub 5	: Subclone 5	Clone 2	Single cell sorting of the CD24 ⁺ cells of Clone 2 of LS174T
Clone 2_Aza-1 Clone 2_Aza-2 Clone 2_Aza-3 Clone 2_Aza-4 Clone 2_Aza-5 Clone 2_Aza-6		Clone 2	Subclones developed from 5-aza-2'-deoxycytidine treated CD24 ⁺ cells of Clone 2 by FACS sorting
Clone C1	: Clone C2 ^{CD24+} :SW480_Clone C2 ^{CD24+}	SW480	Single cell sorting of the bulk by FACS
Clone D1	: Clone D1 ^{CD24-} : SW480_Clone D1 ^{CD24-}	SW480	Single cell sorting of the bulk by FACS

Abstract

Understanding the mechanisms and nature of tumour heterogeneity is a key focus of current cancer research. Tumour heterogeneity can arise from clonal evolution and/or differentiation. This thesis investigated the role of methylation in dynamic regulation of CD24 based heterogeneity in colorectal cancer cell lines. First, E-Cadherin variation between the cell lines LS174T and LS180 was investigated to find out whether E-Cadherin had any causal role in the difference in lumen formation between these two cell lines derived from the same tumour. These studies found no evidence of a causal role of E-Cadherin. However, morphological heterogeneity in LS174T was observed during the E-Cadherin study, suggesting that this cell line might be a mixture of two different clones. Single cell sorting by FACS allowed to isolate and establish these clones which were stable in the culture for long enough (until passage ~30) to allow their characterisations. Between the two clones, the CD24⁺ clone (named the LS174T_Clone 1^{CD24+}) was found to have shorter doubling time (23 hours) than the CD24⁻ clone (named the LS174T_Clone 2^{CD24-}; 30 hours). When cultured in matrigel, the LS174T_Clone 1^{CD24+} showed higher clonogenicity (Chapter 4). Microarray analysis further revealed differences in gene expression including *LAPTM4B*, *CXCR4*, *TGFIB* and *IL8* between these clones. Interestingly, when maintained for a long time in culture (around passage 50, which is

equivalent to ~7 months), CD24 expression went through a gradual change in these clones, which became more evident from subclones of LS174T_Clone 2^{CD24-} (Chapter 6) and clones from another CRC cell line SW480 (chapter 5 and 6). To understand the mechanism of this dynamic change in the expression of CD24, it was first shown that no mutation could be responsible for this phenomenon. This suggested that promoter methylation of *CD24* might be the mechanism of the observed dynamic changes in CD24 expression. Bisulfite (BS) modification of DNA from the LS174T clones and CD24 differentially expressing CRC cell lines (such as CD24⁻ cell lines: CC20, RKO vs CD24⁺ cell lines: DLD1, NCIH716) followed by Sanger sequencing showed that direct methylation of seven CpG positions in the *CD24* promoter region Chr6:106975560-106975834 is strongly correlated with the expression of CD24. Further subcloning and sequencing revealed that changes in the methylation of only two out of the seven CpG positions might be the main contributor to the CD24 expression differences. This is the first evidence of direct methylation-mediated regulation of *CD24*, showing, more broadly, how methylation can contribute to and maintain dynamic heterogeneity in cancer cells. Finally, a mixed culture experiment with the CD24⁺ and CD24⁻ clones was conducted to test a simple mathematical model, which aimed to explain the interaction between the clones that are stably present in the LS174T cell line (Chapter 7). Altogether, these experiments suggest that genes such as *REG1A* (an inducer of angiogenesis) might be expressed because of synergistic interactions between the clones, whereas CXCR4 and TFF2 might be involved in a receptor-ligand complementary relationship. These findings have set a ground for future studies to confirm such interactions between co-existing subpopulations within a heterogeneous milieu of cancer cells.

(506 words)

Key words: Colorectal cancer; Tumour heterogeneity; Cancer cell lines; Methylation regulation; Dynamic control; CD24

Table of Contents

Declaration	ii
Dedication	iii
Acknowledgments	iv
Note to the examiners/reviewers	vi
Abstract	vii
Contents outline	x-xviii
List of figures	xix-xxii
List of tables	xxiii
Abbreviations	xxiv
Thesis chapters	1-200
Appendix	201-212
List of references	213-226

Contents outline

Chapter 1

Introduction and Literature review

(page 1-29)

1.1 Chapter Introduction	2
1.2 Cancer: Definition, Epidemiology and Etiology	3
1.3 Epithelial to Mesenchymal Transition (EMT) and its role in metastasis	4
1.3.1 Loss of E-Cadherin as the marker of EMT	6
1.4 Heterogeneity in Cancer	6
1.4.1 Inter-tumour heterogeneity	7
1.4.2 Intra-tumour heterogeneity	7
1.4.2.1 Origin and evolution of intra-tumour heterogeneity	8
1.4.2.2 Causes of Intra-tumour heterogeneity: Genetic factors	9
1.4.2.3 Non-genetic factors of heterogeneity: Epigenetic contributions	10
1.4.2.4 Source of intratumour heterogeneity: Genotypic heterogeneity and phenotypic heterogeneity	10
1.4.3 Interaction among clones in a heterogeneous environment	14
1.4.4 Consequence of tumour heterogeneity	14
1.5 Colorectal Cancer as a model for studying cancer	14
1.5.1 Colorectal Cancer cell lines as the models of CRC	16

1.5.1.1 3D culture: making the most from the study of colorectal cancer cell lines	20
1.5.1.2 Relationship between lumen formation and cancer stem cell	23
1.6 Colorectal cancer stem cell markers	23
1.6.1 CD24: Structure and function	25
1.6.2 Plasticity in the CD24 expression	26
1.7 Scope of this thesis work	29

Chapter 2

Materials and Methods

(Page 30-50)

2.1 Chapter introduction	31
2.2 Cell culture techniques	31
2.2.1 Cell lines	31
2.2.2 Cell culture	32
2.2.3 Transient cell transfection	32
2.2.3.1 siRNA transfection of LS180 cells	32
2.2.3.2 E-Cadherin plasmid transfection in LS174T cells	33
2.2.4 Cell culture in Matrigel	33
2.2.4.1 F-actin staining assay	34
2.3 DNA methods	34
2.3.1 DNA isolation	35
2.3.2 Validation of the mutation status of the cell lines	35

2.3.2.1 Primer design and PCR	35
2.3.2.2 Gel Electrophoresis	36
2.3.2.3 Purification of PCR products and sequencing	37
2.3.3 Bisulfite sequencing	38
2.3.3.1 DNA isolation and Bisulfite conversion	38
2.3.3.1.1 Purification of the bisulfite converted DNA	40
2.3.3.1.2 PCR amplification of purified bisulfite converted DNA	40
2.3.3.1.3 Analysing bisulfite converted sequence for CpG methylation	43
2.4 RNA methods	44
2.4.1 RNA Isolation	44
2.4.2 Quantitative RT-PCR (qRT-PCR)	45
2.5 Protein methods	46
2.5.1 Protein Isolation	46
2.5.2 Western Blot	47
2.5.3 Flow cytometry	48
2.5.3.1 Preparation of cells (scrapping and washing cells)	49
2.5.3.2 Antibody staining	49
2.5.3.3 Flow cytometry analysis	49
2.5.4 SRB assay	50
2.6 Microarray Data Analysis using Partek Genome Suite	50
2.7 Statistical test	50

Chapter-3

LS174T: A model for studying clonal heterogeneity and E-Cadherin

(page 51-69)

3.1 Chapter Introduction	52
3.2 Molecular characterisation of LS174T	53
3.3 LS174T and LS180: A duo with differences	54
3.3.1 Morphological differences	54
3.3.2. E-Cadherin expression difference between LS174T and LS180	56
3.4 LS174T: A model cell line to study the role of E-Cadherin in the regulation of differentiation	57
3.5 Knockdown of E-Cadherin in LS180	58
3.5.1: E-Cadherin knockdown efficiency in LS180	58
3.5.2. Effect of E-Cadherin knockdown on differentiation	60
3.6 E-Cadherin knock-in in LS174T and its effects	61
3.6.1. E-Cadherin knock-in using plasmid	61
3.6.2. Effects of E-Cadherin expression on LS174T differentiation	61
3.7 Loss of E-Cadherin due to mutation versus E-cadherin silencing due to EMT have different consequences	63
3.8 Chapter Discussion	67

Chapter 4

CD24 based Clonal Heterogeneity in LS174T

(page 70-101)

4.1 Chapter Introduction	71
4.2 Morphological heterogeneity of LS174T is separable as stable clones	72
4.3 Confirmation of the cellular origins of Clone 1 and Clone 2 in LS174T	74
4.4 Differences between Clone 1 and Clone 2 of LS174T	76
4.4.1 Clones have different growth rates	76
4.4.2 Clonogenicity and colony morphology differences between Clone 1 and Clone 2 of LS174T	77
4.5 Clonal heterogeneity in LS174T can be linked to differential expression of CD24	80
4.6 Clone 1 and Clone 2 did not differ in other stem cell markers	84
4.7 Proof of stability: When subcloned, Clone 1 and 2 reproduced the same CD24 expression profiles	85
4.8 Microarray analysis revealed expression differences of key genes between the clones of LS174T	85
4.9 LS174T clones have genetic differences	90
4.10 Drug response in clones	96
4.11 Chapter discussion	99

Chapter- 5

CD24 Based Heterogeneity in CRC Cell Lines

(page 102-118)

5.1 Chapter Introduction	103
5.2 Isolation and characterization of morphologically distinct clones with differential expression of CD24 in SW480	104
5.2.1 SW480 is a cell line with morphologically distinct subpopulations	104
5.2.2 CD24 expression difference between SW480_Clone C2 and SW480_Clone D1	108
5.2.3. SW480_Clone C2 and SW480_Clone D1 differ in clonogenicity and differentiation in matrigel	110
5.2.4. SW480_Clone C2 changes the expression of CD24 over time but SW480_Clone D1 remains stable	113
5.3 Chapter discussion	114
5.3.1 CD24 based heterogeneity in CRC cell lines	114
5.3.2 Using FACS profile as the suggestion for heterogeneous cell lines	118

Chapter 6

Methylation Mediated Regulation of CD24 in CRC Cell Lines

(page 119-168)

6.1 Chapter introduction	120
6.2 Mutation is not responsible for down/up regulation of <i>CD24</i> in the LS174T clones	121
6.3 Dynamic change in CD24 expression in continuous culture	123
6.3.1 If grown separately for a long time, clones tend to approach their parental profiles	123
6.3.2. CD24 expression change is bidirectional	128
6.4 Response to 5-aza-2'-deoxycytidine treatment indicates methylation mediated regulation of CD24 expression	131
6.4.1 Treatment with 5-aza-2'-deoxycytidine increases CD24 expression in CD24 ⁻ clones	131
6.4.2 5-aza-2'-deoxycytidine treatment in CD24 ⁻ cell lines increases CD24 expression	138
6.5 Evidence of direct regulation of <i>CD24</i> by promoter methylation	140
6.5.1 UCSC genome browser shows large CpG islands on 5' region of <i>CD24</i>	142
6.5.2 Bisulfite (BS) sequencing showed evidence of direct methylation of <i>CD24</i> promoter region	143
6.5.2.1 Bisulfite (BS) modification and sequencing of Clone 1 ^{CD24+} and Clone 2 ^{CD24-} of LS174T confirmed direct methylation of <i>CD24</i>	143

6.5.2.2 BS modification and sequencing of other CD24 ⁺ and CD24 ⁻ cell lines confirmed CpG position methylation observed in CD24 ⁺ and CD24 ⁻ clones of the LS174T	146
6.5.2.3 BS modification and sequencing of 5-aza-2'-deoxycytidine modified Clone 2 ^{CD24-} shows direct link of CpG positions with CD24 expression	147
6.6 Methylations of CpG-2 and CpG-5 show strong association with CD24 suppression indicating 'seeding' effect of these CpG positions	154
6.6.1 Methylation in CpG-2 and CpG-5 is mirrored by the expression of CD24	157
6.6.2 Relative importance of CpG-2 over CpG-5 to control CD24 expression	158
6.7 Chapter Discussion	160

Chapter-7

Evidence of Cooperation between Clones in Cell Line

(page 169-191)

7.1 Chapter Introduction	170
7.2 Mathematical model supports positive interactions between clones	173
7.2.1 Modelling cooperation between Clone 1 and Clone 2 of LS174T: mixed culture experiment	174
7.3 Evidence of positive interactions between the clones of LS174T	182
7.3.1 Evidence of synergism: microarray analysis shows high expression of genes in the LS174T which are low in individual clones.	182
7.3.2. Temporal influence on REG1A expression	185

7.3.3 Complementation between clones: receptor-ligand relationship of CXCR4 and TFF2	187
7.4 Chapter Discussion	190

Chapter 8

General Discussion & Future direction

(page 192-200)

8.1 Summary	193
8.2 Relevance of using cell line models for this work	193
8.3 Heterogeneity in CD24 expression in cancer cell lines and its implications	194
8.4 Methylation of a small part of a CpG island regulates CD24 expression	196
8.5 <i>CD24</i> : Is it a marker to identify stem cells with differentiation ability?	197
8.6 Morphology is not related to CD24 expression	198
8.7 Limitations of using cell lines in this work	199
8.8 Concluding remarks and future direction	199

List of figures

Fig 1.1: The regulators and markers of EMT -----	5
Fig 1.2: Successive accumulation of changes eventually	8
Fig 1.3: Two sources of intra-tumour heterogeneity.....	12
Fig 1.4: Stages from adenoma to colorectal carcinoma.....	16
Fig 1.5: Types of mutations for 70 human colorectal -----	18
Fig 1.6: C80 and HCT116 cell lines in matrigel.....	22
Fig 1.7: Sodium Bisulfite treatment converts an	27
Fig 1.8: Bisulfite conversion results in 2 different strands	28
Fig 2.1: Primer pairs suggested by Methprimer program.....	42
Fig 3.1: LS174T Karyotype by multicolour chromosome	53
Fig 3.2: Difference between LS180 and LS174T cell lines.....	55
Fig 3.3: E-Cadherin mutation difference between LS174T.....	57
Fig 3.4: E-cadherin siRNA knockdown efficiency.....	59
Fig 3.5: siRNA ^{E-cadherin} transfected (left) and scramble transfected.....	60
Fig 3.6: Western blot of E-Cadherin in LS174T	61
Fig 3.7: Effect of E-Cadherin expression on lumen	62
Fig 3.8: mRNA expression of CDH1, VIM and ZEB1 in	63
Fig 3.9: Effects of mutations on E-Cadherin expression.....	64
Fig 3.10: Heterogeneous appearance of LS174T in	69
Fig 4.1: Bright field images (20×) of 12 randomly	73
Fig 4.2: Morphological differences between Clone 1 and	74
Fig 4.3: Confirmation of the origins of clones	76
Fig 4.4: Clonogenicity difference between Clone 1	78

Fig 4.5: Colony number and morphology of Clone 1	78
Fig 4.6: Colonies of Clone 1 and Clone 2 and LS180	79
Fig 4.7: CD44/CD24 levels are related to clonogenicity	80
Fig 4.8: Assessment of CD44 and CD24 expression	82
Fig 4.9: Live cell imaging of Clone 1 and.....	83
Fig 4.10: Delta CT values for CD24 in Clone 1 and	84
Fig 4.11: Reproducibility of Clones. Subclones	85
Fig 4.12: Heterozygous (T>C) change in <i>BRAF</i> . In place	92
Fig 4.13: Confirmation of T>C change in <i>BRAF</i>	93
Fig 4.14: A. In the LS174T bulk sequence, <i>PIK32C2B</i>	94
Fig 4.15: Sanger sequence results for <i>TP53</i> variation.....	95
Fig 4.16: Responses to drug treatments: Clone 2	98
Fig: 5.1: Heterogeneous appearance of	103
Fig 5.2: CD44 and CD24 expression assessments	105
Fig 5.3: Bright field images of SW40 and	106
Fig 5.4: Bright field image (5x) to show morphology	107
Fig 5.5: Morphology of selected clones. A. SW480_Clone C2.....	108
Fig 5.6: Flow Cytometry profile of CD24 and CD44	109
Fig 5.7: Fluorescence microscopy of SW480_Clone C2	109
Fig 5.8: Bright field image of colonies of SW480_Clone C2 and	111
Fig 5.9: TRITC-Phalloidin staining of colonies	112
Fig 5.10: Colonies in matrigel after week 3.....	112
Fig 5.11 Change of CD24 expression in SW480_Clone C2.	114
Fig 5.12: Bimodality of CD24 expression in cell lines.....	116
Fig: 5.13: Homogenous expression of CD24 in	116

Fig 5.14: Heterogeneity in CD24 ⁻ cell lines. Like.....	117
Fig 5.15: OXCO1 was homogeneously CD24 ⁻ , unlike	117
Fig 6.1: Change in CD24 expression in late.....	124
Fig 6.2: Increase of CD24 ⁺ cell populations in Clone 2.....	125
Fig 6.3 The change of CD24 in SW480 Clone C2.....	127
Fig 6.4: CD24 ⁺ cells in LS174T Clone 2. A small.....	128
Fig 6.5: Flow cytometry histograms of CD24 expression	129
Fig 6.6: Subclone 3 and subclone 5 of LS174T clone 2 ^{CD24-}	130
Fig 6.7: Effect of 5-aza-2'-deoxycytidine on Clone 2.	133
Fig 6.8: Effect of 5-aza-2'-deoxycytidine on the.....	135
Fig 6.9: Effect of 5-aza-2'-deoxycytidine on CD24	136
Fig 6.10: Enhanced expression of CD24 in Clone 2	137
Fig 6.11: qRT-PCR of 5-aza-2'-deoxycytidine.....	138
Fig 6.12: Two possible modes of 5-aza-2'-deoxycytidine	141
Fig 6.13: <i>CD24</i> is predicted to have a large CpG island.....	142
Fig 6.14: Amplicon from primer pair 4. A region between.....	144
Fig 6.15: Clone 2 shows significant level of methylation.....	145
Fig 6.16: Methylation status in CD24 ⁻ and CD24 ⁺ cell.....	147
Fig 6.17: Effect of 5-aza-2'-deoxycytidine on	148
Fig 6.18: FACS profile of the subclones developed.....	149
Fig 6.19: Selection of the 'unmethylated state' after	151
Fig 6.20: Difference between 5-aza-2'-deoxycytidine	153
Fig 6.21: Role of a few CpG positions for methylation	155
Fig 6.22: BS sequence results of Sub3 ^{CD24-} and	156
Fig 6.23: Higher degree methylation of CpG-2.....	157

Fig 6.24: Clone C2 and D1 from SW480 shows that.....	158
Fig 6.25: CpG-2 shows a consistently higher degree of	159
Fig 6.26: The amplicon of primer 4 covers 21 CpGs)	162
Fig 6.27: Average methylation percentage of the seven CpG	165
Fig 6.28: Positions of the seven CpGs in relation to the	166
Fig 6.29: TFs binding prediction at the 5'UTR- flanking.....	167
Fig 7.1: Positive interactions among cancer cells. A. In case	171
Fig 7.2: Clones of LS174T have distinguishable CD24	174
Fig 7.3: FACS profiles of clones LS174T_Clone 1 ^{CD24+} and	176
Fig 7.4: FACS profiles of the mixed cultures after 10 days.....	178
Fig 7.5: FACS profiles of the mixed cultures after 41 days.	180
Fig 7.6: CD24 expression profiles of LS174T_Clone 1 ^{CD24+} and	181
Fig 7.7: Fold difference in <i>REG1A</i> in qRT-PCR compared to <i>UBC</i>	184
Fig 7.8: Expression of <i>REG1A</i> over time: LS174T_Clone 1 ^{CD24+}	186
Fig 7.9: Mutually exclusive expressions of <i>CXCR4</i> and <i>TFF2</i>	189
Fig 7.10: No bias from the attachment ability of clones used in the	190
Fig 8.1: Possible route of heterogeneity in cell lines: if two	195
Fig Ax.3: Flow cytometry analysis of 11 makers in Clone 1.....	204
Fig Ax.6: The changes in CD24 expression in DLD1 clone.....	207
Ax.7: Methylations in the CpGs of which are downstream to the Exon 1.....	208
Fig Ax.8: All E-Cadherin low cell lines in the panel of 95 cell lines.....	211
Fig Ax.9: Expression of <i>TTL</i> in CRC cell line panel.....	212

List of tables

Table 1.1: Deaths caused by leading cancers.....	3
Table 1.2 Components of basement membrane	20
Table 1.3: Morphology of cell lines as observed in	22
Table 2.1: Primer information for E-Cadherin mutation detection	36
Table 2.2: Recipe for bisulfite conversion reaction mix	39
Table 2.3: Composition of a typical bisulfite conversion reaction	39
Table 2.4: Conditions for bisulfite conversion reaction	39
Table 2.5: Carrier RNA addition during the purification of.....	40
Table 2.6: Recipe for PCR wit bisulfite	43
Table 2.7: Master mix for cDNA synthesis	45
Table 2.8: PCR conditions for cDNA synthesis	46
Table 3.1: EMT associated gene mRNA expression.....	58
Table 3.2: List of cell lines selected for analysis of E-Cadherin	65
Table 3.3 : List of genes that are upregulated in the Mesenchymal	66
Table 3.4: Chi-square test for E-Cadherin expression	68
Table 4.1: Clone 1 and Clone 2 showed differences	79
Table 4.2: Microarray analysis shows differential	86
Table 4.3-4.9: Genet Ontology (GO) analysis	87
Table 4.10: Genetic differences between Clone 1 and Clone 2.	91
Table 4.11: List of 10 drugs used in CRC treatment.	96
Table 5.1: Linear value of CD24 and CD44 message in SW40	104
Table 6.1: Changes in the up and downstream region of	122
Table 6.2: Basal level of CD24 mRNA in 10 cell lines	139
Table 6.3: List of cell lines and clones used in this study	165
Table 7.1: Doubling time and division number of	173
Table 7.2: Genes more highly expressed in LS174T compared	183
Table 8.1: Correlations between CD24 and differentiation regulators.....	198
Ax.4: List of genes that are more highly expressed.....	205
Ax.5: List of genes compared between Clone 1 and	206

List of Abbreviations

BS	Bisulfite
BSA	Bovine Serum Albumin
Bp	Base Pairs
cDNA	Complementary DNA
ChIP	Chromatin Immunoprecipitation
CSC	Cancer Stem Cell
Del	Deletion
DNA	Deoxyribonucleic Acid
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulfoxide
dNTP	Deoxyribonucleotide Triphosphate
ECL	Enhanced Chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EMT	Epithelial to Mesenchymal Transition
FBS	Fetal Bovine Serum
FACS	Fluorescence-Activated Cell Sorting
Het	Heterozygous
kb	Kilobase
mAb	Monoclonal Antibody
MET	Mesenchymal to Epithelial Transition
MSI	Microsatellite Instability
miRNA	microRNA
mL	Milliliter
μ L	Microliter
μ M	Micromolar
mut	Mutant
ng	Nanogram
nM	Nanomolar
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline

PCR	Polymerase Chain Reaction
pM	Picomolar
PMSF	Phenylmethanesulfonyl Fluoride
PRC1, 2	Polycomb Repressive Complex 1, 2
PVDF	Polyvinylidene Difluoride
qRT-PCR	Quantitative Reverse-Transcription Polymerase Chain Reaction
RNA	Ribonucleic Acid
RNAse	Ribonuclease
RPM	Revolutions Per Minute
RT	Reverse Transcription
SDS	Sodium Dodecyl Sulfate
siRNA	Small Interfering RNA
SNP	Single Nucleotide Polymorphism
TBS	Tris-Buffered Saline
TBST	Tris-Buffered Saline plus Tween-20
TEMED	N,N,N,N-Tetramethylethylenediamine
UTR	Untranslated Region
UV	Ultraviolet
wt	Wild type

Chapter 1

Introduction and Literature Review

Chapter 1

1.1 Chapter introduction

A major challenge in the battles against cancer is to understand the nature and evolution of heterogeneity that characterise them. Genetic and epigenetic factors, plasticity in gene expression and microenvironmental differences contribute to the establishment of the heterogeneous milieu of cancer (Sun & Yu, 2015). Driver mutations, somatic mutations with selective advantage, chromosomal instability etc. are the major genetic factors that cause continuously changing cell populations within a tumour. A subset of cells within a tumour is maintained as stem cell population that can differentiate into other cell types, as in colorectal cancer, a single stem cell has shown to be able to create colonies comprising of enterocytes, enteroendocrine cells and goblet cells (Yeung, Gandhi, Wilding, Muschel, & Bodmer, 2010).

In the past two decades, exclusive roles of cancer stem cell (CSC) markers and other driver genes in cancer initiation, progression and prognosis have been studied extensively. A key understanding about cancer stem cells from recent researches is the plastic change in their gene expression (Marjanovic, Weinberg, & Chaffer, 2013; Meacham & Morrison, 2013) that can contribute to the heterogeneity within and between tumours. Apart from mutations, epigenetic mechanisms, especially the methylation of CpG positions in the promoter regions can regulate the expression of these genes.

This introductory chapter sets the background for the study of clonal heterogeneity on the basis of the dynamic regulation of an important CSC marker (CD24) by methylation in colorectal cancer (CRC).

1.2 Cancer: Definition, epidemiology and etiology

Cancer is a somatic evolutionary process (Bodmer, 2006) which results in abnormal cell growth and spread in the body with diverse etiology, symptoms, prognosis and outcome. Cancer does not mean only one disease condition; it rather represents more than hundred different diseases. One of the major challenges in tackling cancer is the diversity it displays in disease manifestations and underlying molecular cues.

Cancer is a leading cause of deaths in the industrial world (David & Zimmerman, 2010) and by 2030, in the US, it is predicted to surpass cardiovascular disease as the nation's leading killer (Wishart, 2015). In 2012 there were 14.1 million new cases of cancer worldwide and 8.2 million deaths from cancer, which was 13% of the total deaths of the year (Ferlay et al., 2015). The number of new cancer cases may rise by about 70% in the next two decades (mainly because of increases in the developing countries) (World Cancer Report, 2014).

Recent advances in diagnosis and treatment have increased the number of cancer survivors world-wide. In 2012, approximately 32.6 million people were living with cancer within 5 years of the diagnosis (Ferlay et al., 2015). It has been projected that from 2010 to 2040, one million cancer survivors will be added per decade in the UK, most of them will be aged over 65 (Maddams, Utley, & Møller, 2012). Most of the cancer-related deaths are caused by the lung, liver, stomach, colorectal, breast and oesophageal cancers (Table 1.1)

Table 1.1: Deaths caused by leading cancers (World Cancer Report, 2014)

Cancer	Deaths
Lung	1590 000
Liver	745 000
Stomach	723 000
Colorectal	694 000
Breast	521 000
Oesophageal	400 000

For many decades, scientists have been trying to understand the underlying cause(s) of cancer. Both internal and external factors contribute in the process of carcinogenesis and they often work in combinations to cause malignancy (M. T. Williams & Hord, 2005). For different cancers, risk factors can be different, though the same factor may contribute to several cancers.

1.3 Epithelial to Mesenchymal Transition (EMT) and its role in metastasis

Cancer progresses through successive genetic (and epigenetic) changes. Research outcomes of past several decades have improved the understanding of these steps and associated molecular, biochemical and cellular anomalies. In 2000, Hanahan and Weinberg proposed that the progression can be described as six ‘hallmarks’ of cancer such as: sustaining proliferative signalling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis (Hanahan & Weinberg, 2000). This list of cancer ‘hallmarks’ were extended in a follow up review (Hanahan et al., 2011). Among them, the invasion and metastasis of cancer cells is considered as the most detrimental step that causes the death of the cancer patients.

In the last decade, many researchers have worked to understand the mechanism of invasion of cancer cells and to identify the factors that can transform normal cells into malignant, motile invasive cells. Among other mechanisms, Epithelial to Mesenchymal Transition (EMT) has been considered to be a key contributor to invasion and metastasis (Brabletz et al., 2005).

Epithelial to Mesenchymal transition (EMT) is a well-studied developmental process that allows polarized and well-attached epithelial cells to go through a number of changes that enable them to acquire mesenchymal cell-like phenotype, such as- change in the shape, enhanced motility, invasive ability, resistance to apoptosis, and a greatly increased production of extra cellular matrix (ECM) components (Kalluri & Neilson, 2003). During vertebrate embryogenesis, EMT transforms epithelial cells into mesenchymal cells that invade and migrate through the ECM. In amniotic vertebrates, primary mesenchymal cells originated from epithelial cells in primitive streak migrate and differentiate into mesodermal and endodermal epithelia (Hay, 1995).

A number of transcription factors that induce EMT have been identified and studied quite extensively. Amongst them ZEB1, TWIST1, TWIST2, SNAI1, SNAI2 are regarded as the master regulators of EMT. All or some combination of these factors are needed in any given tissue to influence the EMT process. Cell lines data in the Bodmer Laboratory suggest that it may be the ZEB1 that regulates the EMT process in the colorectal epithelium. The transcription regulators of EMT negatively influence the production of cell-cell junction proteins such as E-Cadherin (CDH1), Claudine 3,7 (CLDN3, CLDN7), ZO-1 etc., whereas Vimentin (VIM) and N-Cadherin (CDH2) level increases with the progression of EMT.

The opposite process of EMT is called MET (Mesenchymal to Epithelial Transition) which restores the epithelial characteristics (Fig 1.1) This process is particularly important when the metastasised cells colonise in a distant organ. Amongst the factors that influence this MET process includes the microenvironment of the host tissues, growth factors, receptors, and signalling molecules (Yao, Dai, & Peng, 2011). Unlike the transcription factors of EMT, those of MET are less defined and characterised. In most of the cases, MET is associated with the down-regulation of the EMT regulators.

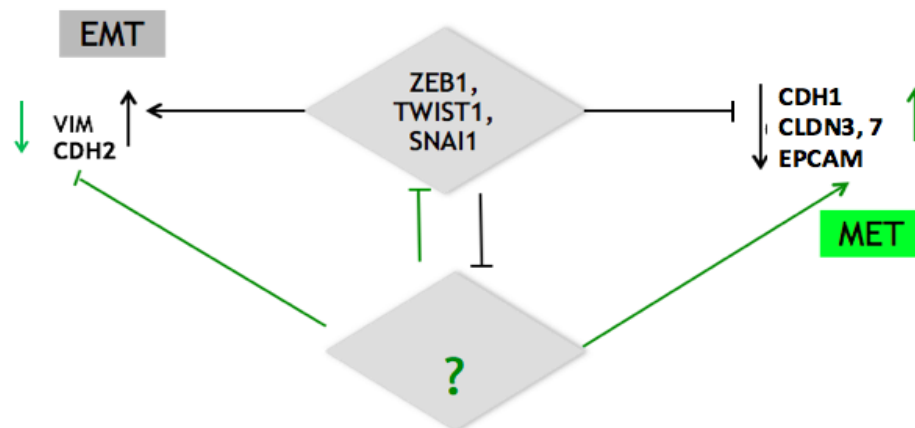


Fig 1.1 The regulators and markers of EMT (shown with black lines) and MET (shown with green lines). ZEB1, TWIST1, SNAI1 upregulate VIM and CDH2 whereas suppress the production of E-Cadherin (CDH2), Claudin 3 or 7 (CLDN3, CLDN7) and EPCAM. During the process of MET, the expression patterns become opposite.

1.3.1 Loss of E-Cadherin as the marker of EMT

E-Cadherin is a Ca^{2+} dependent cell adhesion glycoprotein, which forms homotypic junctions with neighbouring cells (van Roy & Berx, 2008).

Mature E-Cadherin contains 728 amino acids, forming a transmembrane protein of three domains: cytoplasmic, transmembrane and the ectodomain (van Roy & Berx, 2008). It is pivotal in the formation of cell-cell adherence junctions (Mohamet, Hawkins, & Ward, 2011; van Roy & Berx, 2008). They connect the extracellular environment to the intracellular signalling through interactions with the catenins, in particular p120-, α - and β -catenins (van Roy & Berx, 2008).

Down regulation of E-Cadherin has been correlated with aggressiveness and invasiveness of various cancers of epithelial origin (Wendt, Taylor, Schiemann, & Schiemann, 2011). In the opposite process, there is evidence that once cells reach their site of metastasis, they undergo mesenchymal to epithelial transition (MET) and revert back from a more mesenchymal to a more epithelial phenotype. Re-introduction of E-Cadherin can sometime restore the epithelial characteristics of cancer cells and reduce their growth (Jiao, Miyazaki, & Kitajima, 2001).

The consequence of E-Cadherin down regulation on the differentiation of cancer cells has been a matter of continuous interest as the differentiation status of cancer cells correlate with their invasiveness and malignancy. It has been showed that in colorectal cancer, lower level of E-Cadherin are associated with less differentiated tumours (Bendardaf et al., 2005; Karatzas et al., 1999; Khoursheed et al., 2003).

1.4 Heterogeneity in Cancer

It is clear from the above discussion that cancer cells can go through transitions with the expression or suppression of certain markers and their regulators. Moreover, cancer can occur almost in any tissue of the human body. All these matters can contribute to the heterogeneity in cancers which can be considered from three aspects: 1) heterogeneity based on the cell of origin of the cancer, i.e., cancers of different tissues, 2) heterogeneity between cancers from the same type of cellular origin and 3) heterogeneity within a given cancer. Both type 1 and type 2 heterogeneities are called inter-tumour heterogeneity, whereas type 3

is called intra-tumour heterogeneity. These heterogeneities are of intense research interest for the basic understanding and clinical management of cancer.

1.4.1 Inter-tumour heterogeneity

Between tumours of various tissues there are differences in gene expressions as different cell types express different combinations of genes. The heterogeneity between tumours is partly due to the fact that the cancer stem cells in different tumours vary in the extent to which they can still differentiate. Such inter-tumour heterogeneity is well recognised and often interpreted in the clinic as different treatment choice. Prognosis and response to therapeutic agents which are the results of varying interactions of the genetic and epigenetic changes with cell-intrinsic biological properties vary among cancers of different tissues (Zellmer et al., 2014). However, some signalling pathways, driver genes, genomic signatures etc. can be shared among different cancers which often are explored for the understanding of common evolution of different cancers and common druggable targets.

One way to deal with the inter-tumour heterogeneity is to identify the tissue specific signatures of the tumour and use them as the common treatment targets. One approach in breast cancers was to group them based on the expression of hormone receptors with the amplification of the gene *HER2* (Weigelt & Reis-Filho, 2009) and use different drugs (or combination of them) for different groups.

1.4.2 Intra-tumour heterogeneity

Intra-tumour heterogeneity refers to the presence of cancer cells with diverse geno- and pheno-typic attributes within the same tumour. The phenotypic heterogeneity within a tumour has been recognised for a long time. Investigations from as early as 1950s by Makino Sijiro and colleagues reported the presence of sub-populations with varying tumorigenic potentials. Makino coined the term ‘stem line(s)’ to indicate a cancer cell subpopulation that were capable of causing cancers on their own when inoculated in mice (Makino, 1956). Subsequent investigations by many researchers have revealed underlying genotypic cues that run parallel to the phenotypic differences of cancer cells in a tumour. In recent years, high throughput sequencing of single cancer cells has revealed the complex extensity of spatial and temporal intra-tumour heterogeneity.

1.4.2.1 Origin and evolution of intra-tumour heterogeneity

The intra-tumour heterogeneity can be examined in the genetic, epigenetic and phenotypic level (Gay, Baker, & Graham, 2016). Cancer cells continuously gather (epi)mutations that can contribute to clones with diverse (epi)genetic backgrounds. The cancer cells evolve and survive following the Darwinian evolutionary mechanism once any spontaneously change in the cells give them selective advantage over their peers (Nowell, 1976). Successive accumulation of changes can lead to malignancy and a malignant tumour can be heterogeneous with the co-existence of cells from different stages of the evolutionary process.

If, for the sake of simplicity, it is assumed that just one cell acquires advantages from such mutations, the progeny cells can be considered as the clones of the initial abnormal cell. This clonal expansion and selection will exist until another mutation with selective advantage occur in the cell(s) which will then follow the same expansion and selection cycle (Fig 1.2)

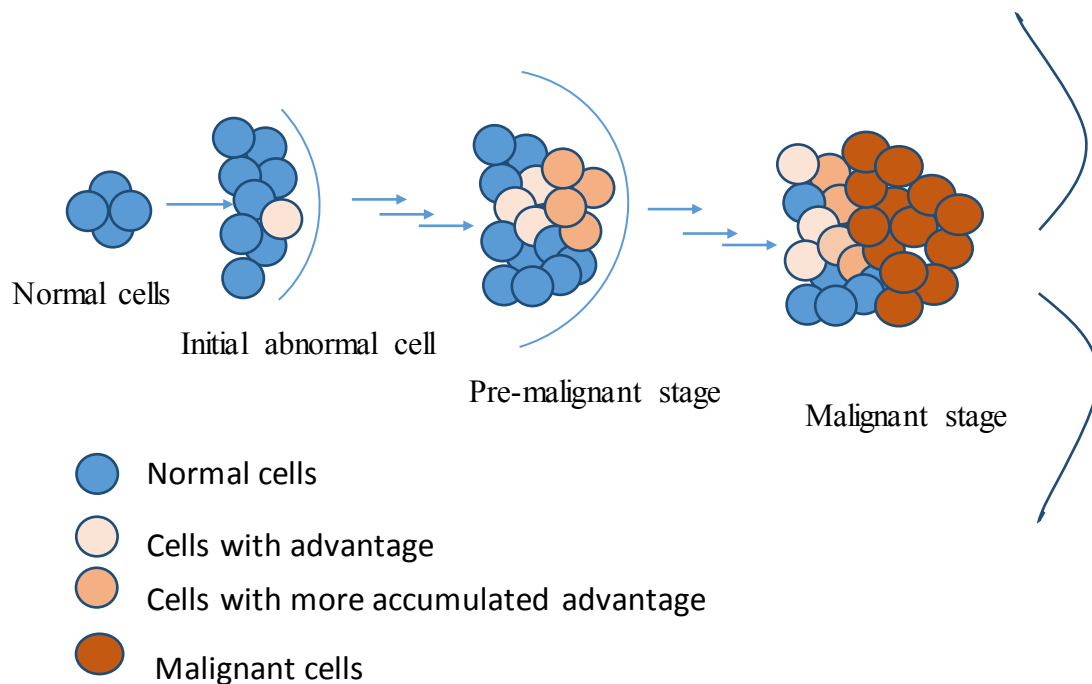


Fig 1.2 Successive accumulation of changes eventually leads to malignancy. The malignant stage can be a mixture of cells from previous stages which may remain as the minority in the population.

1.4.2.2 Causes of Intra-tumour heterogeneity: Genetic factors

Intra-tumour heterogeneity can arise from a complex interplay of i) genetic and ii) non-genetic factors.

Genetic instability (also called genomic instability) refers to the increased alteration in the genome of a cell (Shen, 2011). Genetic instability is a broad term that covers alterations by simple mutations, such as, base pair deletion/substitution which may affect one or a few nucleotides, to big mutational events such as chromosomal rearrangement where a segment of the genome is altered (Burrell, McGranahan, Bartek, & Swanton, 2013). Mutations can occur randomly in the active tumour suppressor and/or in the latent oncogenes causing their inactivation and activation respectively that can lead to the abnormal growth properties in cells, eventually making them cancerous. These mutations, if give any selective advantage to cancer cells, become established and will be present in all the progeny derived from that cell (Tomlinson, Novelli, & Bodmer, 1996).

Contributions of mutation in cancer development have been well studied regarding the driver mutations. Compared to non-selective passenger mutations, driver mutations provide selective advantage to cancer cells (Greaves & Maley, 2012). For example, in colorectal cancer, mutations in *KRAS*, *BRAF* and *APC* genes are considered driver events. Inherited or spontaneous mutations in these genes can make cells vulnerable to develop as colorectal cancer.

Compared to the single (or several) base pair deletion, insertion or substitution; a bigger change in the genome occurs when a part of a chromosome gets missing or rearrange with other chromosomes during cellular multiplication. One of the major consequences of chromosomal rearrangement is the copy number variation (CNV) where certain part of a genome is repeated in different numbers in different cells (Redon et al., 2006).

1.4.2.3 Non-genetic factors of heterogeneity: Epigenetic contributions

Besides genetic contributions, epigenetic mechanisms are key non-genetic contributor in tumour heterogeneity.

Like mutation, epigenetic changes can affect gene expression/function. The changes can be of two types: DNA methylation and histone acetylation. Methylation is a heritable change and once this happens to a cell or group of cells, methylation can be carried to the next generations.

It is already known that cancer cells have much higher rate of methylation changes than normal cells. Although most of these changes occur in the repeat sequences, a significant number of changes affect the regulatory regions of the genes. Hypomethylation of the promoter can express a silent oncogene, whereas, hypermethylation can stop the expression of a cancer suppressor.

Methylation of oncogenes and/or demethylation of oncosuppressor genes can cause knock-on effects by affecting the activation or suppression of other genes. Unlike mutation, evidence suggests that methylation is more dynamic and reversible (Ramchandani, Bhattacharya, Cervoni, & Szyf, 1999). This increases the complexity to a further level by creating a plastic state of heterogeneity.

1.4.2.4 Source of intratumour heterogeneity: Genotypic heterogeneity and Phenotypic heterogeneity

Cancer cells with different genetic and epigenetic makeup can manifest phenotypic differences such as morphological difference and/or expression differences in certain markers. The resulting intra-tumour heterogeneity from these differences can be examined from two aspects: i) the genotypic- (or clonal-) heterogeneity where the cancer cells are grouped according to their genetic and epigenetic variations and ii) the phenotypic heterogeneity with respect to the state of differentiation of cells within a tumour.

Genotypic heterogeneity in cancer arises from the genetic and epigenetic changes that occur in cancer cells. Any cell that sustains in the tumour milieu can create progeny that can be

considered clones of the original cell. When a mutation occurs that gives a growth advantage to the cancer, it will be selected for and give rise to a new subclone of the cancer carrying this mutation. Genotypic heterogeneity is an inevitable outcome of a somatic evolutionary process. That is because a new advantageous mutation does not immediately replace all the other cells in a tumour (Fig 1.2).

Heterogeneity at the genetic and epigenetic level has been documented in a variety of malignancies, including leukemia (Campbell et al., 2008), breast (Fujii, Marsh, Cairns, Sidransky, & Gabrielson, 1996; Shipitsin et al., 2007), prostate (Konishi et al., 1995), colon (González-García, Solé, & Costa, 2002), brain (Mora, Cheung, & Gerald, 2001), esophagus (Maley et al., 2006), bladder (Sauter, Moch, Gasser, Mihatsch, & Waldman, 1995), gynecological carcinomas (Fujii et al., 2000) and multiple myeloma (Pantou et al., 2005).

In addition to the above-described tumour heterogeneity that is due to genetic/epigenetic differences between subclones within a tumour, individual cells that are derived from the same tumour or clones may exhibit phenotypic heterogeneity that then contribute to the overall heterogeneity within tumours, and must be distinguished from genetically determined heterogeneity.

Within a tumour, the cancer cells exist with a range of phenotypic differences such as expression of certain sets of markers and/or resultant differentiation states. A small subset of cancer cells, called cancer stem cells (CSCs), are always maintained as the core which have the ability to produce the whole tumour. It was first shown by Makino Sijiro in 1950s that tumours have different subset of cells with varying tumorigenic potential and the stem-line subset is the most tumorigenic (Makino, 1956). In recent years, many lines of experimental evidence have shown that the cancer stem cells can maintain themselves in their stem-like state and also give rise to a range of differentiated cells which are less tumorigenic (Yeung et al., 2010; Humphries et al., 2013; Leedham et al., 2013). Whether the cancer stem cells will remain as stem cell or follow a route towards differentiation is determined by its niche which is defined as the anatomically different location in a tumour where cancer stem cells reside (Plaks et al., 2015). “The location rather than special patterns of gene expression may

characterize stem cells” (Spradling, Drummond-Barbosa, & Kai, 2001). Stem cell niches are structurally and functionally diverse in different tissues.

Recent advancements in experimental techniques and model systems have allowed to study the microenvironment that surrounds the cancer stem cells. The regulation of cancer stem cells depends on the interaction with the surrounding cells together with their internal mechanisms. In the intestine, Paneth cells play important role in the niche of LGR5⁺ stem cells in the crypt base. CD24⁺ Paneth cells express EGF, TGF α , Wnt3 and the Notch-ligand Dll4 which are essential signals for stem cell maintenance in culture. Another evidence came when sorted stem cells were co-cultured with Paneth cells resulting in significant improvement of organoid formation (Sato et al., 2011). The mesenchymal cells under the basal layer of the crypt also secrete factors to maintain the crypt. One such signaling is mediated through the Wnt signaling pathway.

A particular cancer can be examined simultaneously for these two aspects of heterogeneity as shown in Fig 1.3. Clonal heterogeneity can be seen as the genetic (and/or epigenetic) differences that contribute to different clones; and phenotypic heterogeneity can be examined from the expression-differences of certain genes that are associated with stemness or differentiation to these cell populations.

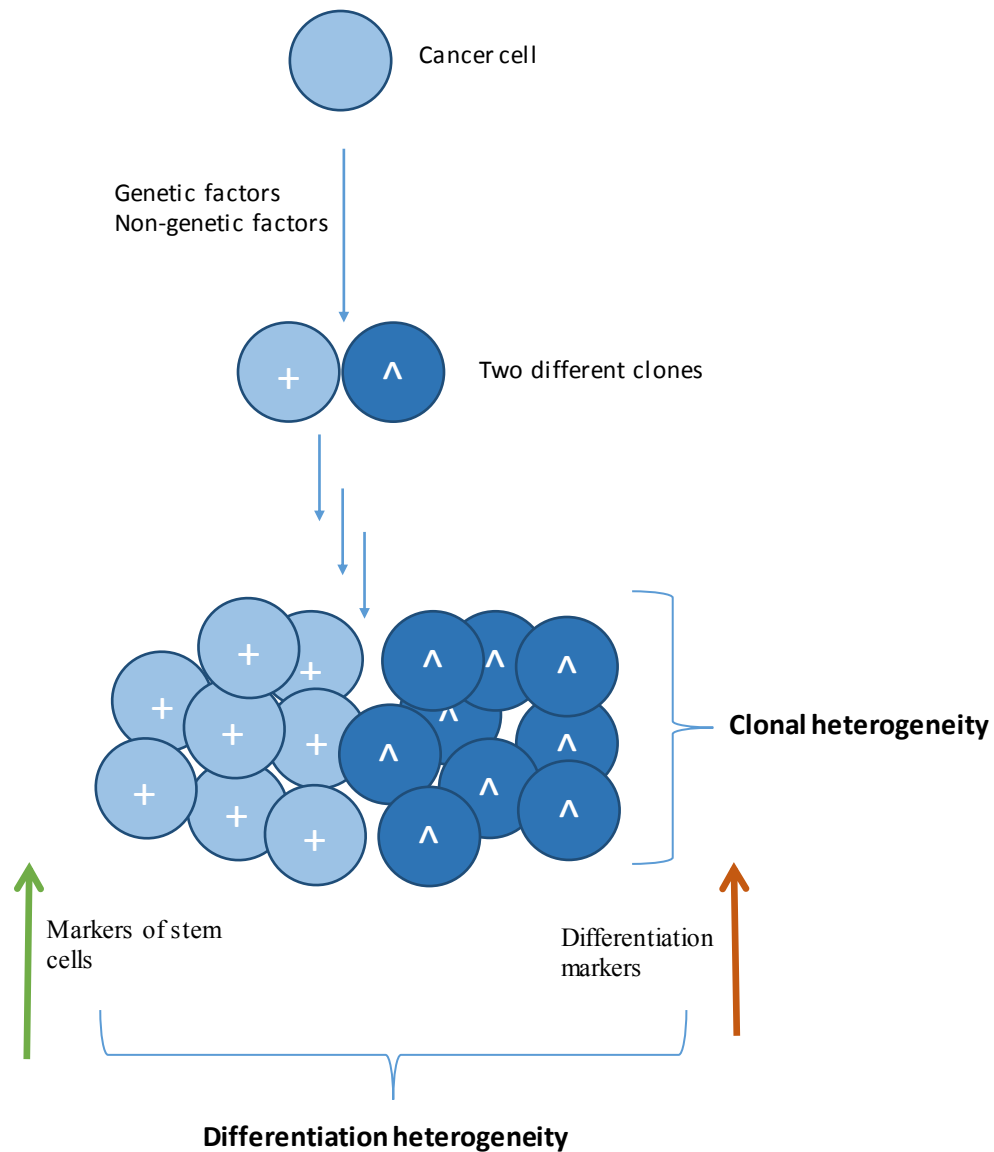


Fig 1.3: Two sources of intra-tumour heterogeneity. (Top to bottom) Clonal populations can arise from certain genetic/epigenetic changes in cancer cells. (Left to right) Those clones can vary in their differentiation status based on the expression of certain genes. If one clonal population expresses stem cell markers (such as, CD44, CD24, CD133), they are expected to have more stem-cell properties whereas the cells that will express markers of differentiation (such as, CDX1, KRT20), will be more differentiated.

1.4.3 Interaction among clones in a heterogeneous environment

The presence of multiple clonal populations in a tumour indicates that there might be positive interactions among the clones which support their co-existence. The co-existence of different clones in a tumour can be explained by non-autonomous models (Tomlinson & Bodmer, 1997). According to this model, a cooperative interaction among clones is essential for their stable coexistence. There are two important mechanisms which can support their coexistence, such as: synergism and complementation. In synergism, the interactions among clones will result in the expression of novel protein(s) in the mixture that will exert a beneficial effect. In complementation model, one clone expresses factors which are received and used by others and vice versa. A common example of this model of interaction involves ligands and receptors interactions between clones.

1.4.4 Consequence of tumour heterogeneity

One of the most important significances of tumour heterogeneity is its implications in resistance and relapse of cancer. Cancer stem cells and/or the plasticity in the stem cell populations may cause drug resistance which can be selected for just as are any of the changes that occur in the progression of a cancer. If all the differentiated cells are killed but not the cancer stem cells, the cancer will continue to grow. It is one of the main reasons of resistance against drugs. Understanding the nature and evolution of tumour heterogeneity is a paramount research interest of current time.

1.5 Colorectal Cancer as a model for studying cancer

Colorectal cancer is the cancer of colon and rectum. Colorectal cancer develops through a progression of stages from adenoma to carcinoma, which allows careful analyses of the genetic changes of the cancer at various stages (Bodmer, 2006).

The colon is the first part of the large intestine and extends upto 5 feet in length. The primary function of the colon is to absorb water from the passing digested foods and the waste material is defecated through the rectum, the end of the large intestine.

Cancer can develop in any part of the colon and rectum. It is now known that etiology of cancers in different parts of the colon is different, however, though the difference between left versus right colon is well recognised, it is not clear for rectum versus the rest.

Colorectal cancer has been- and is continuously being -studied as a model to understand the genetic and epigenetic changes that underlie the evolutionary process of cancer. There are clear inherited cases of colorectal cancer (discussed below) which give important clues to what happens in the sporadic non-inherited cancers with regards to disease susceptibility and somatic evolution. Colorectal cancer progresses through distinctive stages (Fig 1.4) and specimens from those stages are easily obtainable, making it easier to follow the disease from the view of point of successive progression (DeVita, Lawrence, & Rosenberg, 2011).

Colorectal cancers can arise either sporadically or in patients with predisposing germline mutations. There is a clear example of familial colorectal cancer syndrome known as Familial Adenomatous Polyposis (FAP) which provides good model of Mendelian genetics in cancer study. Germline mutations in *APC* gene on chromosome 5q21 (Miki et al., 1992; Nishisho et al., 1991) result in the development of hundreds to thousands of precancerous polyps in the colon and rectum which can progress to cancerous lesions by the age of 20 or 30. This rare and dominantly inherited condition comprises 1% of the total incidence of colorectal cancer. However, the progress from the precancerous polyps to full blown cancer means that FAP syndrome presents an easy to study model of various stages of cancer. It was later determined that 80% of all colorectal cancers start with the first mutation in the *APC* gene (Bodmer, 2006).

Another form of inherited CRC is Hereditary Non-inherited Polyposis Coli (HNPCC) which occurs because of a defective mismatch repair gene, mainly- *hMLH1* and *hMSH2* (discussed in the next section). These germline mutations result in the microsatellite instability (MSI) in HNPCC patients with many novel somatic changes in their genes.

For most sporadic CRCs, it is believed that the oncogenic process starts with somatic mutations in *APC* followed by the mutations in genes involved in apoptosis such as *TP53*;

RAS oncogenes such as *KRAS*, *NRAS* (Vogelstein et al., 1988) and cell cycle check point genes- *p14 (LAMTOR2)* and *p16 (CDKN2A)* etc. (Bodmer, 2006). Somatic mutations can lead to the loss of heterozygosity (LOH) when such mutations provide the ‘second’ hit to inactivate the functional allele of a gene whose one allele was inactive from germline mutations (Knudson & Jr., 1971; Ryland et al., 2015). LOH for *APC* probably occurs quite early and mostly before *TP53* mutations which are known to occur mainly in late adenomas.

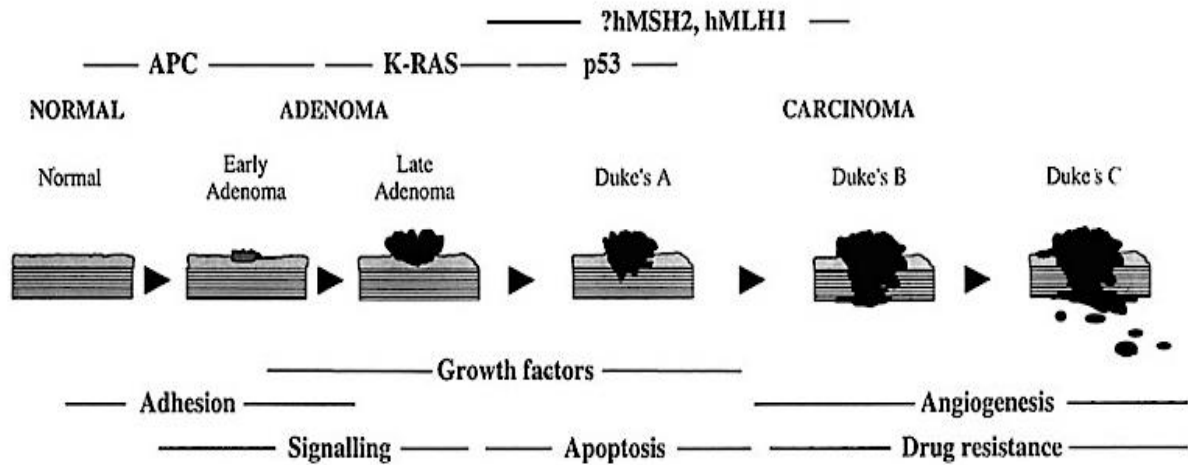


Fig 1.4: Stages from adenoma to colorectal carcinoma. Expression of genes in different stages are also shown (Bodmer, 2006).

It is believed that most, if not all cancers progress through such multistep process (Hanahan & Weinberg, 2000; Nowell, 1986). CRC is clearly a good model to understand and explore the multistep process of cancer progression.

1.5.1 Colorectal Cancer cell lines as the models of CRC

Cancer cell lines present convenient source of material to study cancers. Colorectal cancer (CRC) cell lines are very widely used models to study various aspect of colorectal cancer, such as- tumour biology, biomarker discovery and *in vitro* drug response tests.

Recently (Jennifer L. Wilding & Bodmer, 2014) has reviewed issues related to the use of cell lines as preclinical models and scrutinised the reasons of the failures often observed in phase III clinical trials. They have pointed out that it is fairly clear from the review of

publications which used cell lines for drug screening that, most of these studies were based on a very limited number of any particular tissue type of cancer cell lines. For example, the NCI-60 cell line panels (Shoemaker, 2006) has been used extensively for many studies, and although this includes 60 cell lines, there are only six, on average, of each tissue type included. This small number of cell lines in the NCI-60 panel has limitations regarding representing the genetic, molecular and cellular heterogeneity of cancers (Jennifer L. Wilding & Bodmer, 2014).

So, the problem was not with the use of cell lines, it was rather with the number of cell lines from each cancer category used for such studies. The comprehensive analysis of Cancer Cell Line Encyclopaedia (CCLE) project showed that large, annotated cell line panels can be valuable preclinical model for drug studies (Barretina et al., 2012). Compared to the NCI-60 panel, CCLE analysed 947 cell lines for gene expression, genome sequence, copy number variations and chromosomal instability and for 479 cell lines, the data were coupled with the pharmaceutical profiles of 24 anticancer drugs. This study has successfully identified the genetic, lineage, and gene-expression based predictors of drug sensitivity, which shows the power of a bigger panel of cell lines that can be representative of the diversity of primary cancer. It shows that there is a need for large enough number of cell lines of a given type of cancer to detect changes that matter in cancer.

Another study by Mouradov et al., (2014) which compared 70 colorectal cancer cell lines with the primary cancer data from The Cancer Genome Atlas (TCGA) database found notable resemblance in mutations, copy number variations and expression profiles (Fig 1.5). The similarities include the two hypermutation phenotype defined by the defective DNA mismatch repair and DNA polymerase ϵ (Pol E) proofreading deficiency. The cell lines were also concordant with the mutation profiles in the broadly altered *WNT*, *MAPK*, *PI3K*, *TGF β* , and *p53* pathways.

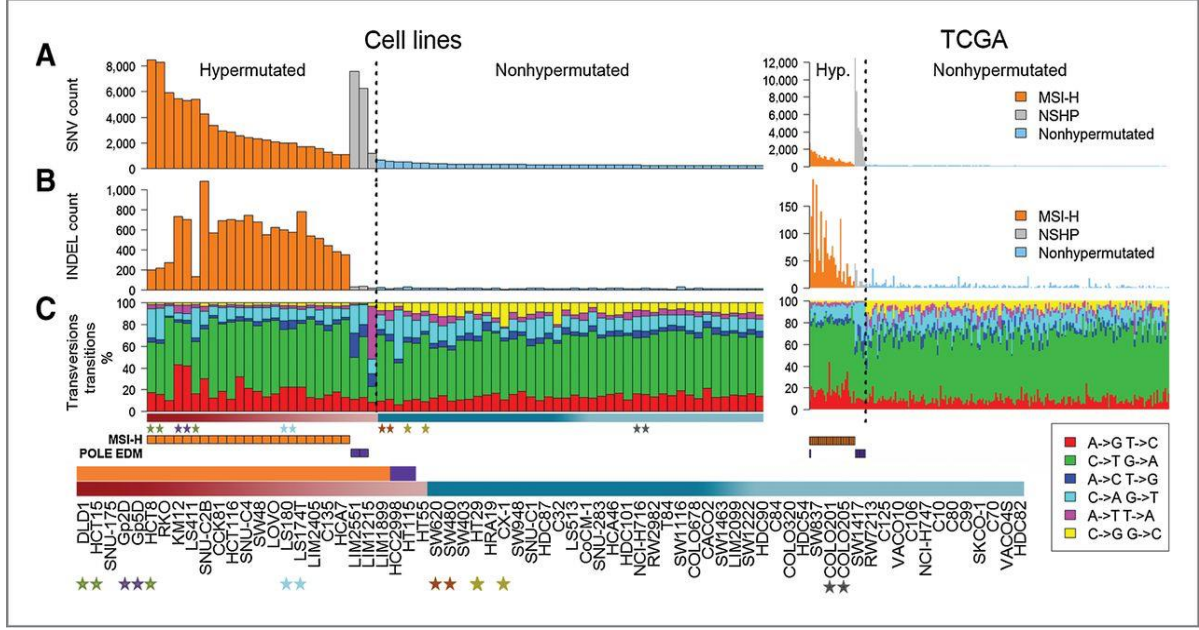


Fig 1.5: Types of mutations for 70 human colorectal cancer cell lines and 223 TCGA-analysed primary cancers. Counts of SNVs (A–C), InDels (A and B), and proportions of nucleotide transitions and transversions (C). Identical coloured stars indicate paired cell lines originating from the same tumour.

Legends: MSI-H, microsatellite instability-high; NSHP, nucleotide-substitution hypermutator phenotype; POLE EDM, missense mutation in the POLE exonuclease domain (Mouradov et al., 2014).

This study and the review by J Wilding and W Bodmer (Jennifer L. Wilding & Bodmer, 2014) highlight the importance of using the right kind of cell lines for a particular study. Like primary cancers, different cell lines can present different molecular signatures and this should be considered during designing a study.

One of the key considerations in this regard is the Mismatch Repair (MMR) status of the cell lines. Mismatch repair mechanism is one of a number of DNA repair pathways which in particular repairs differences that lead to DNA strand mismatches. Mutations in the genes *hMLH1*, *hMSH2* can lead to the deficiency of mismatch repair system in cells (Fishel et al., 1993; Wheeler et al., 1999). In case of HNPCC, MMR genes are mainly mutated in the germ

line, whereas in sporadic cancers, *hMLH1* is silenced mostly due to the methylation of its promoter (Wheeler et al., 1999). The regions of the genome which are enriched in mono-, di-, tri- or tetra- nucleotide repeat sequences (known as ‘microsatellites’), are particularly vulnerable to the errors caused by the defective MMR system which leads to the variable lengths of these repeat regions. This phenomenon is called microsatellite instability (MSI). Through the seminal work of Perucho et al., (Ionov, Peinado, Malkhosyan, Shibata, & Perucho, 1993) and Thibodeau et al., (Thibodeau, Bren, & Schaid, 1993), and subsequent studies, MSI has been detected in 15% colorectal cancer, 3% are of these are associated with Lynch syndrome and the other 12% are caused by sporadic, acquired hypermethylation of the promoter of the *MLH1* gene. MSI positive colorectal cancers have some distinctive features, such as their tendency to occur in the proximal colon and the lymphocytic infiltration in the cancer. They are poorly differentiated, mucinous, may have signet ring appearance and slightly better prognosis than colorectal tumors without MSI ((J. L. Wilding, McGowan, Liu, & Bodmer, 2010; Gryfe et al., 2000; Kim, Jen, Vogelstein, & Hamilton, 1994).

In the MMR deficient cells, somatic frameshift mutations have been documented in many genes which contain repetitive sequences. In colorectal cancer, *TGFB2R* is a frequently mutated gene in the MMR deficient cell lines, whereas in the MMR proficient cell lines *SMAD4* is generally mutated causing to the inactivation of TGF β pathway (J. L. Wilding, McGowan, Liu, & Bodmer, 2010; Woodford-Richens et al., 2001).

Besides MMR related genes, cell lines vary regarding the mutation in the driver genes such as *KRAS*, *BRAF*, *CTNBB1*, *APC* (Mouradov et al., 2014). It has been already mentioned that methylation mediated control of gene expressions also show variations among cell lines. So it is important that during any study, all these molecular characteristics of a cell line are considered.

1.5.1.1 3D culture: making the most from the study of colorectal cancer cell lines

Conventional 2D culture is very effective to study many conceptual aspects of molecular and cellular biology. However, when asking more complex biological questions such as differentiation of the cells, one of the major drawbacks of 2D *in vitro* cell cultures is the absence of extracellular matrix (ECM), vascular system and complex cellular interactions to resemble the actual 3D environment of *in vivo* system (Buono, Pignatelli, & Hall, 1991; Mueller-Klieser, n.d.; Richman & Bodmer, 1988); Yamada et al., 2007). Traditionally, animal models (such as nude mice) are often used to study such features. However, xenogenic animal models also present limitations of species difference and immunological and other incompatibilities. A major progress to find a culture system that will satisfyingly overcome the gap came from 3D cell culture- mainly as an extra cellular matrix isolated from Engelbreth-Holm-Swarm (EMS) sarcoma of mouse. The matrix, now commercially available as matrigel (by Corning and BD) was developed from decades long research by The Laboratory of Developmental Biology and Anomalies (LDBA) in the 1970s (Orkin et al., 1977; Smith, Martin, Miller, Dorfman, & Swarm, 1975). However, in the form of matrigel, it was first elaborately characterized by Kleinman *et.al.*, (H K Kleinman et al., 1982) . The matrix contains many components of the basement membrane of epithelial cells (Table 1.2) and gives a reasonable 3D culture environment for epithelial cell *in vitro*.

Table 1.2 Components of the basement membrane (Hynda K Kleinman & Martin, 2005)

Abundant components	Growth factors
Laminin	TGFβs
Type IV collagen	FGF
Perlecan (heparan sulfate proteoglycan)	EGF
Nidogen/entactin	PDGF
Proteases	IGF
72 kda MMP-2	Other proteins
92 kda MMP-9	Amylase
Urokinase	Transferrin
Tissue type plasminogen activator	Clusterin

In the Bodmer laboratory, colorectal cancer cell lines have been cultured in 3D medium such as in collagen gel and matrigel quite routinely to elucidate characteristics such as differentiation (Ashley, Yeung, & Bodmer, 2013), polarization and stemness (Yeung et al., 2010) and lineage specification (Yeung, Gandhi, & Bodmer, 2011) of colorectal cancer cells.

These 3D cultures have enabled the study to identify and characterise the markers and regulators of differentiation/stemness in colorectal cancer cells. The degree of differentiation of primary cancers is a key consideration for clinopathological interference. Based on the presence or absence of glandular structures in colorectal tumour tissue, they are graded as well/moderately differentiated (low grade cancer) or poorly differentiated (high grade cancer). The more differentiated the cancer is, the better is the prognosis.

The connection of glandular structures in primary cancer with their differentiation status was very convincingly replicated in the 3D culture of colorectal cancer cells. It has been shown that the lumen surface of the glandular structures are enriched in differentiation markers such as villin, ezrin and F-actin (Ashley et al., 2013).

Ashley et al., (2013) from the Bodmer lab have extensively tested F-actin staining with TRITC-Phalloidin in the CRC cell lines and have shown that the cell lines, like primary cancers, can be classified according to their differentiation ability. A number of cell lines when cultured in matrigel can grow into colonies with lumens (Fig 1.6). The luminal surface stains for differentiation markers such F-actin, Ezrin and Villin (like the lumens of glandular structures of the normal colon and low grade tumour). Cell lines such as SW1222, LS180 and C80 form recognizable lumens clearly marked with F-actin staining. However, there are other cell lines, such as, HCT116, DLD1, SW620 etc. that lack the ability to form lumens.

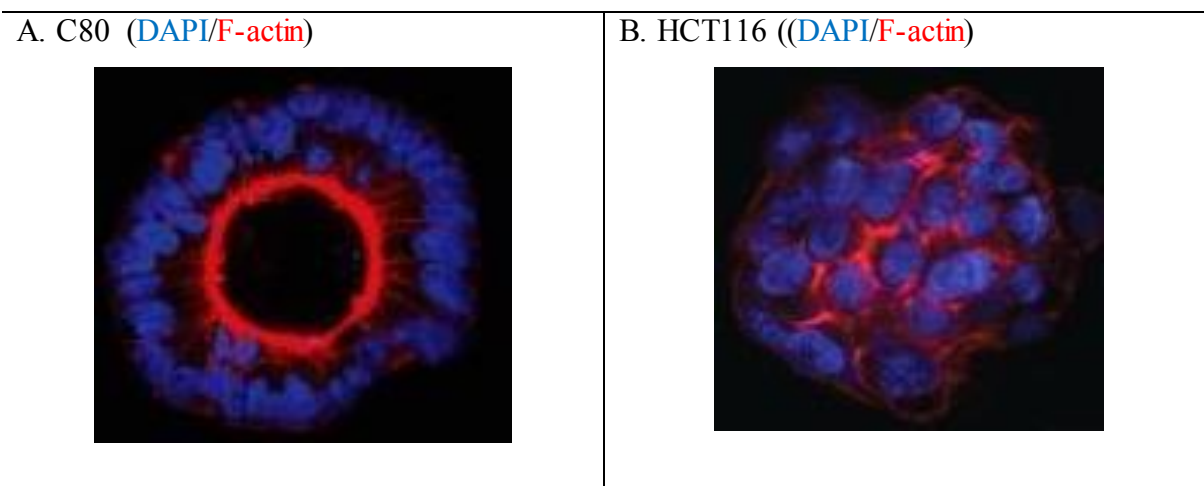


Fig 1.6: C80 and HCT116 cell lines in matrigel. A. C80 can differentiate in matrigel and forms well defined lumen. B. HCT116 cell does not polarise and differentiate in matrigel. Red: TRITC-Phalloidin stains indicating F-actin, Blue: DAPI stains indicating nucleus (Ashley et al., 2013)

A panel of 90 cell lines has been analysed in 3D culture in the Bodmer laboratory to understand this differentiation ability, mostly by Eric Liu, Shaan Ghandi and Neil Ashley (internal lab data). They described their observations of morphology in the 3D culture in four terms (LIDN- Lumen, Intermediate, Dense, Network) as shown in Table 1.3.

Table 1.3: Morphology of cell lines as observed in 3D culture in the Bodmer Laboratory

Morphology	Number of cell lines	Example of cell line
Lumen	25	SW1222, LS180, C80
Intermediate	30	Colo206, RW2982, NCIH716
Dense	32	HCT116, T80, SW620, CAR1
Network	3	RKO, HDC111, CC20

These characteristics of the cell lines together with their microarray data are routinely used to select the appropriate cell line for a study.

1.5.1.2 Relationship between lumen formation and cancer stem cell

Cell polarization in lumen structure in the 3D culture may not always be accompanied by cell differentiation. However, in the case of CRC cell lines, Yeung *et al.*, from the Bodmer Laboratory has showed that stem cells from lumen-forming-cell-lines can produce lumens containing all three types of cells: enterocytes, enteroendocrine cells and goblet cells (Yeung et al., 2010) in matrigel. The study also showed that it was the CD44⁺/CD24⁺ cells of the lumen forming cell lines which were able to form large colonies with lumens, whereas CD44⁻/CD24⁻ cells could not. The CD44⁺/CD24⁺ cells showed higher tumorigenicity in nude mice. This provided the key evidence that cancer stem cells are present in CRC cell lines, that they fulfil the criteria for stem cells as they can self-renew and give rise to all three lineages. These observations supported the use of CD44 and CD24 as colorectal cancer stem cell markers. However, clear understanding of the functional and regulation aspects of these markers remain elusive.

1.6 Colorectal cancer stem cell markers

‘Stem’ cells have the ability to generate more stem cell progeny or to differentiate in a number of other cells when induced. ‘Cancer stem’ cells are believed to have similar properties. Cancer stem cells can self-renew or follow a certain route of differentiation based on their niche and differentiation ability. One way to measure the stemness is to study the clonogenicity of cancer cells (together with their differentiation ability). If a cancer cell population consists of only stem cells, then a high proportion of the single cells from the population will be able to renew themselves and form individual colonies. In other cases, where only a subset of cancer cells has self-renewal ability, not all the seeded single cells will be able to form colonies, and so the number of individual colonies will be lower than the number of seeded cells.

Researchers have discovered a number of stem cell markers for colorectal cancer. Many of the new discoveries are made from the association with already known markers for normal stem cell. Colorectal cancer stem cell markers like CD44, CD24, CD133, BMI1 etc. have been associated with properties such as self-renewal, tumorigenicity, invasiveness or metastasis.

The expression of certain sets of cancer stem cell markers can be indicative of the status and fate of the stem cells they are expressed on. It can also be indicative of the signalling pathways that are active in the cells. It has been known for a long time that some signalling pathways active in normal stem cells may be disrupted in cancer stem cells, such as PI3K/Akt/mTOR pathway, Nuclear Factor KappaB (NF- κ B) pathway, mitogen activated protein kinase (MAPK) pathway, Myc and Polycomb group (PcG) proteins such as, B Lymphoma Mo-MLV insertion region 1 homolog (BMI-1) pathway, Wnt pathway etc (Garza-Treviño et al., 2015).

In the normal colon, the Wnt signalling pathway plays a key role in the development and maintenance of the intestinal epithelium (de Lau, Peng, Gros, & Clevers, 2014; Leedham et al., 2013) and stem cell populations (Barker et al., 2007). In the canonical Wnt signalling pathway WNT proteins bind their receptor Frizzled which leads to the formation of a cell surface complex with LRP5/6. R-spondins increase the sensitivity of Wnt ligands by inhibiting the activity of ZNRF3 and RNF3 that ubiquitinate and remove Frizzleds from the surface. Thus R-spondins work as an agonist of the Wnt pathway.

Canonical Wnt signalling pathway results in the LCF/LEF mediated transcription of target genes; among those, LGR5 is an important one which marks stem cells in the adult intestinal crypt and colon cancer. R-spondins, work as the ligands for the LGR5. Wnt pathway activation and deactivation can decide whether stem cell population will self-renew or they will follow certain route of differentiation (Zeuner, Todaro, Stassi, & De Maria, 2014).

Among others, CD44 and CD24 are two well studied markers in the breast, colorectal and other cancers (Jaggupilli, Elkord, Jaggupilli, & Elkord, 2012). In the Bodmer laboratory, CD44 and CD24 are routinely used as the markers for stem cell enrichments in colorectal cancer cell lines since the findings of Yeung *et al.*, (2010) had shown their importance as cancer stem cell markers.

Though the roles of *CD44* among various cancers are quite similar, understanding about *CD24* is still very inadequate. In breast cancer, CD44⁺/CD24⁻ cells are considered to be stem like (Al-Hajj, Wicha, Benito-Hernandez, Morrison, & Clarke, 2003), whereas in colorectal,

cancers, CD44⁺/CD24⁺ selections enrich for stem cells (Yeung et al., 2010). The regulation of *CD24* is still poorly understood. Its prognostic significance is also a matter of controversy- in some studies it was associated with poor prognosis (Agrawal et al., 2007), in other studies, no association was found (Ahmed et al., 2009).

1.6.1 CD24: Structure and function

CD24 is a small surface protein linked with the membrane with a glycosyl-phosphatidylinositol (GPI) anchor. It was first discovered in mouse as a heat stable antigen by Springer in 1970 (Springer, Galfrè, Secher, & Milstein, 1978). Mouse *Cd24* was first cloned in 1990 and mapped on Chromosome 10 (Kay, Takei, & Humphries, 1990). Human *CD24* was successfully cloned in the same year and mapped on Chromosome 6. Two other pseudogenes of *CD24* have been mapped on Chromosome Y and 15.

Mature CD24 protein is heavily glycosylated and this pattern of glycosylation is tissue specific (Tan, Zhao, Xiang, Chang, & Lu, 2016). A mouse mature *Cd24* consists of 27 amino acid residues whereas human mature CD24 consists of 32. Based on the level of glycosylation, the molecular weight can vary from 20 to 70 kDa in different tissues and cells.

CD24 can go through various pattern of glycosylation with both O- and N-glycosylation. Functional consequence of this extensive glycosylation has been detailed by Edwin *et al.* They analysed human CD24 using MALDI-TOF-MS and showed that the major O-glycans linked to the CD24 protein core are: Neu5Ac α -2,3/6Gal β -1,3GalNAc and NeuAc₂Gal β -1,3GalNAc₁; and the major N-glycans are: GalNAc₂GlcNAc₂Man₃Fuc, Gal₁GalNAc₂GlcNAc₂Man₃Fuc₁, and Gal₂GalNAc₂GlcNAc₂Man₃Fuc₁. Among these oligosaccharides, Neu5Ac α -2,3/6Gal β -1,3GalNAc is more commonly found than others.

CD24 does not have any cytoplasmic domain. The N-terminus domain is exposed to extracellular environment where it can bind to several ligands. Based on the nature of the

glycans, CD24 can bind different ligands which include- P-selectins, sialic-acid-binding immunoglobulin-like lectin-G (Siglec-G, mouse) or Siglec 10 (human) (Chen, Tang, Zheng, & Liu, 2009), neural recognition molecule L1 (Sammar, Aigner, & Altevogt, 1997), TAG-1, Contactin (Kleene, Yang, Kutsche, & Schachner, 2001), and some damage-associated molecular patterns (DAMPs) such as HMGB1 (Chen et al., 2009).

CXCR4 has been reported to be a binding partner of CD24 (Schabath, 2006). This interaction affects CXCR4-triggered cell migration and ERK1/2 phosphorylation.

CD24 is expressed in a variety of cells. Most hematopoietic cells such as immature T cells and B cells (Hubbe & Altevogt, 1994; Li, Zheng, & Liu, 2004; L. A. Williams, Hock, & Hart, 1996), developing neuronal cells and regenerating muscle cells express CD24. It is also expressed on granulocytes, epidermal Langerhans cells, monocytes, granulocytes, dendritic cells (DCs), macrophages, keratinocytes and epithelial cells in various tissues.

CD24 expression is elevated in many cancers compared to their normal tissue counterpart. In a study during this thesis work (where three tumour tissue were compared with the normal counterparts from the same patient), it has been found that CD24 level is increased in colorectal cancer. In cancer tissue, CD24 expression was considerably higher than normal tissue. However, it could be an artefact due to the fact that normal tissue is mostly differentiated whereas cancer tissue is not.

1.6.2 Plasticity in the CD24 expression

CD24 has variable expression across various tissues and cancer. However, within the same tissue the expression levels can fluctuate over time. It was observed in breast cancer by (Meyer et al., 2009) that CD44⁺/CD24⁺ cells populations give rise to CD44⁺/CD24⁻ population. Bidirectional cell plasticity and dedifferentiation go hand in hand and the plastic changes can allow cells to enter into tumour initiating stem cell state. However, they can revert back to differentiated state based on their interactions with the stroma (Schwitalla et al., 2013).

In the study in Colorectal cancer cell lines SW1222, Yeung *et al.*, had similar observation. It was found that CD44⁺/CD24⁺ cells when seeded in matrigel can give rise to CD44⁺/CD24⁻ population among other combinations (Yeung *et al.*, 2010).

The mechanism of this dynamic expression of CD24 has not been revealed, although it was speculated that methylation mediated regulation might be a reason for such dynamic change in CD24 expression (Meyer *et al.*, 2009). However, Meyer *et al.*, could not show direct association between the CpG methylation and CD24 expression in their study.

The most likely mechanism of such plastic change in the expression of any gene is methylation, as reversible mutational changes are unlikely to occur in a short period of time. The standard experimental technique to find methylation mediated regulation is bisulfite sequencing. Bisulfite sequencing is performed to see the methylation of CpG positions in target region(s). It gives information about direct involvement of methylation in regulating gene expression. Its principle is based on the fact that if the cytosine of a CpG dinucleotide is methylated, then upon treatment with Sodium bisulfite, the unmethylated Cytosine becomes Uracil but methylated cytosine (5-Methylcytosine, 5-mC) remains unaffected (Fig 1.7).

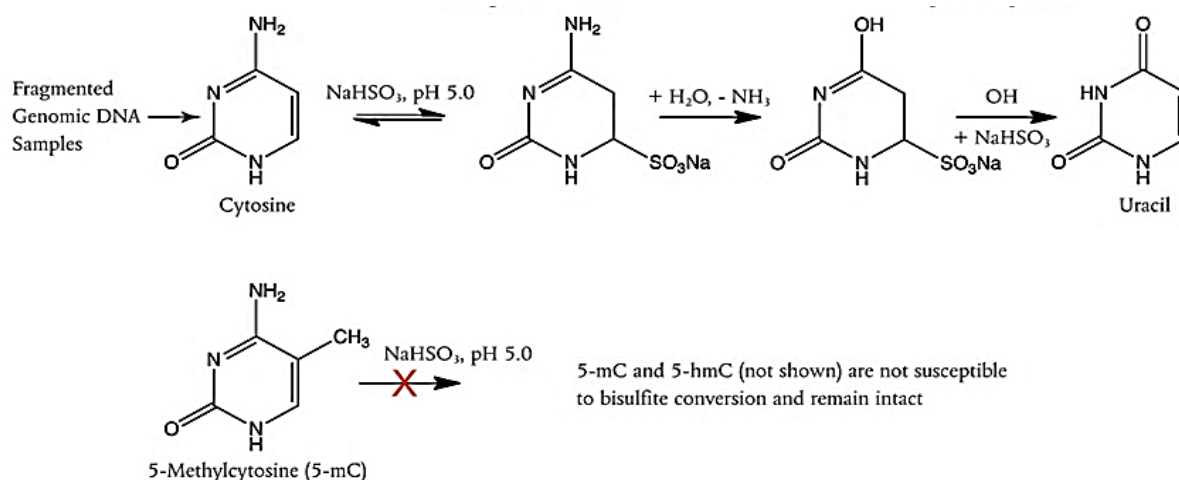


Fig 1.7: Sodium bisulfite treatment converts an unmethylated Cytosine (C) into a Uracil (U). On the other hand, a methylated Cytosine is not susceptible to such conversion. (image acquired from NEB website).

This difference in CpG positions can be detected by PCR amplification and sequencing. During PCR amplifications, all Uracil will change into Thymine (T) whereas the unchanged Cytosine remains as Cytosine (C) (Fig 1.8)

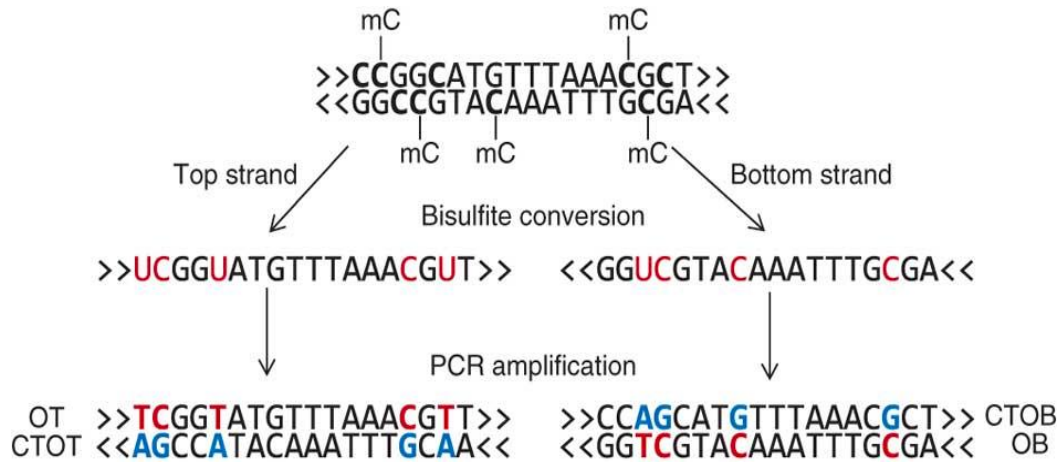


Fig 1.8: Bisulfite conversion results in 2 different strands from where 2 different amplicons can be obtained after PCR amplification. OT= Original Top strand, CTOT = Strand Complimentary to Original Top strand; OB= Original Bottom Strand, CTOB= = Strand Complimentary to Original Bottom strand (Acquired from (Krueger, Kreck, Franke, & Andrews, 2012)).

1.7 Scopes of this thesis work

Dynamic nature of CD24 expression creates dynamic heterogeneity in cell population. From a homogeneous CD44⁺/CD24⁺ population, when a sub-population with CD44⁺/CD24⁻ profile is generated, it creates a heterogeneous environment where both CD24⁺ and CD24⁻ cells coexist. This gives a system that can be studied as a microcosm of tumour heterogeneity regarding the regulation of such heterogeneity, factors that contribute and maintain such evolution and the resultant clinical significance.

The core of the work presented in this thesis was based on an observation that the LS174T, a widely used cell line in colorectal cancer research contains a stable mixture of CD24⁺ and CD24⁻ cells. It started from a cloning experiment where it was found that both the CD24⁻ and CD24⁺ clones of the LS174T cell line can survive and propagate independently and stably. With the availability of such clones, an opportunity was created to use the cell lines to study heterogeneity regarding the expression of CD24, an important cancer stem cell marker. It was later found that a number of other cell lines (including SW480) may also have distinctive populations regarding CD24 expression.

A key observation during the study was the dynamic regulation of *CD24*, a previously described feature of this gene.

These observations were considered important for two reasons; firstly, it allowed the study of CD24 based heterogeneity using cell line models, which makes the study of the evolution and regulation of such heterogeneity possible within a very controlled and measurable setting of cell culture; and secondly, it allowed to explore how the CD24 is dynamically regulated, a question has not been answered yet.

This work has explored and successfully answered those questions.

Chapter 2

Materials and Methods

Chapter 2

2.1 Chapter introduction

Colorectal cancer cell line models were used for all studies during this thesis work. Standard cell culture techniques, cell transfections, single cell cloning; DNA, RNA and protein techniques such as PCR, agarose gel electrophoresis, (Bisulfite) sequencing, qRT-PCR, western blot, poly acrylamide gel electrophoresis, flow cytometry; microscopy (light and fluorescence) were frequently used. Microarray data analysis, accessing online databases and software from NCBI and UCSC genome browsers were among computational techniques and tools used. The Graphpad Prism 6.0 software suite was used for statistical analysis.

2.2 Cell culture techniques

2.2.1 Cell lines

For the different parts of this project, cell line models were selected based on their gene expression profiles across a panel of more than 100 human colorectal cancer cell lines in the Bodmer laboratory. For the E-Cadherin (*CDH1*) related work (Chapter 3), LS180 and LS174T cell lines were studied. Other cell lines which are negative or low for CDH1 expression such as, Colo320DM, Colo320HSR, SW620, SW480 and RKO were used for studying the effect of CDH1 expression in 3D culture.

For the work related to CD24 expression regulation analysis, LS174T cell line and its clones (LS174T_clone 1 and LS174T_clone 2) and subclones (Sub 3, Sub 5), SW480 and its clones (SW480_clone C1 and SW480_clone D2), CC20, RKO, DLD1, NCIH716 and HCT116 were used.

During this project, cell lines were routinely checked for mycoplasma contamination using the MycoAlert™ Mycoplasma Detection Kit (Lonza, Maine, USA) according to the manufacturer's protocol.

2.2.2 Cell culture

All cell lines (except NCIH716) were cultured in Dulbecco's modified Eagle medium (DMEM, catalogue number: 11965084) supplemented with 10% FCS (catalogue number: 26140079) and 1% Penicillin/Streptomycin (catalogue number: 17-602E). For NCIH716 culture, RPMI medium (catalogue number: 11875093) was used instead of DMEM. Cells were incubated in 10% CO₂ at 37°C in a humidified incubator. Cultures were passaged when they reached ~70-80% confluence. For this, cells were washed in PBSA twice and then trypsinised using Trypsin/EDTA Solution (TE) (Gibco® catalogue number: 15400-054). After 3-5 minutes, Trypsin was neutralised by adding 3-5 ml complete media. Cells were then transferred in appropriate dilution in fresh media.

2.2.3 Transient cell transfection

2.2.3.1 siRNA transfection of LS180 cells

Cells were washed in PBSA twice and trypsinised in a culture flask when they were approximately 70% confluent. Trypsinised cells were transferred in a 15 ml falcon tube and counted using a Cellometer™ Auto T4 (Nexcelom Bioscience). For this, 20 µl of the cell suspension was transferred in the chamber of the cell counting slide which was then counted in the Cellometer™ Auto T4. The software captures eight fields of the counting slide and gives directly the number of cells per millilitre.

Approximately 1×10^5 cells were 'reverse' transfected in a 24 well plate. For this, a transfection mixture containing 20 picomole of siRNA for E-Cadherin (Origene, USA)

and 3 μ l of RNAiMAX (Life Technologies) diluted in a total volume of 100 μ l of OptiMEM® was added to each well. Approximately 1×10^5 cells were then added in those wells and incubated for 24 hours before washing and replacing with fresh complete medium and incubated for a further 24, 48 or 72 hours.

2.2.3.2 E-Cadherin plasmid transfection in LS174T cells

Cells for transfection were prepared and counted as described above. After plating 1×10^5 cells into each well of a 24 well plate, a transfection mix of 500 ng of plasmid and 3 μ l of LTX-Plus (Invitrogen) diluted in a total volume of 100 μ l of OptiMEM® was added in each well. After 24 hours, the medium was changed and the cells were incubated for periods of 48, 72 and 96 hours.

2.2.4 Cell culture in Matrigel

Matrigel (Corning) was used for 3D culture of cells (reviewed in section 1.5.1.1). For this, 3-4 hours before the experiment, the required volume of matrigel aliquots were transferred from -20°C to 4°C to allow the matrigel to liquefy.

Appropriate numbers of wells in a 96 well plate were coated with 30 μ l of 50% matrigel diluted in DMEM (1% P/S, without serum).

The plate was then kept at 37°C for 20-30 minutes to allow the solidification of matrigel. Cells were trypsinised and counted and approximately 500 cells in a volume of 20 μ l complete medium (with 10% serum, 1% P/S) were mixed with 20 μ l of the ice-cold liquid matrigel and added in the matrigel-coated wells of the plate.

The plate was incubated at 37°C for 30 minutes. Finally, 100 μ l complete medium was added in each well and the plate was maintained at 10% CO_2 and 37°C in an incubator for up to 4 weeks, based on the experimental design. The medium was changed every three days.

2.2.4.1 F-actin staining assay

In order to measure the differentiation ability of the cell lines, F-actin staining assay was used as described by Ashley *et al*, (Ashley, Yeung, & Bodmer, 2013). Generally, after two weeks, cells that have differentiation ability can arrange themselves around a central focus. Such foci have been termed as 'lumen' in the literatures from the Bodmer laboratory. In bigger colonies, there were often multiple lumens, whereas small colonies were generally observed with just one central lumen. Cells which are unable to differentiate, do not form lumen structures in matrigel.

To stain the lumens, cells in matrigel were allowed to grow at least for two weeks. In some experiments, the lumens were stained at the end of week four.

For this, colonies growing in matrigel were first fixed by adding 100 μ l of 4% paraformaldehyde for 10 minutes. Colonies were then washed once with 150 μ l of PBSA, and permeabilised with 0.5% Triton-X for 10 minutes. Following 3 washes with 150 μ l of PBSA, those colonies were incubated for 3 hours at room temperature or overnight at 4 °C in 100 μ l TRITC-phalloidin diluted at 1:1000 in PBSA. After the incubation period, they were washed twice with 50mM glycine (in PBSA) and 100 μ l of 10 μ g/ml DAPI was added to each well. The colonies were observed under fluorescent microscope (Zeiss cell observer).

2.3 DNA methods

DNA isolation, polymerase chain reactions (PCR) and sequencing of the PCR amplicons were an important part of this work. In general, the extracted DNA was used for: i) cell line identification and mutation analysis, and ii) revelation of methylation status which involved: isolation of DNA and bisulfite modification of DNA, PCR amplification, gel electrophoresis, purification of PCR products and sequencing.

2.3.1 DNA isolation

DNA was isolated from cultured cells using a DNA isolation kit from Qiagen (catalogue: 69504) following the manufacturer's protocol. Briefly, $\sim 1 \times 10^7$ trypsinised cells were collected in a 1.5 ml Eppendorf tube in 1 ml PBSA buffer. Cells were spinned down for 3 minutes at $2320 \times g$ using a table top centrifuge (Eppendorf, 5415 D). The supernatant was removed using micropipette after centrifugation and cells were dissolved in 200 μ l of PBSA by gentle tapping followed by addition of 20 μ l of proteinase. After that, 200 μ l of lysis buffer AL was added in the mix and vortexed for 15 seconds. When solution became clouded, 200 μ l ethanol (95-100%) was added and vortexed again for 10-15 seconds. The solution (usually) appeared clear after this step.

Total content from the Eppendorf tube was then transferred to the pre-labelled DNAeasy spin columns (provided with the kit). The columns were spun at $14499 \times g$ for 1 minute. The flow-through was discarded and the bound DNA on the membrane was successively washed with 500 μ l of wash buffer AW1 and AW2 by centrifuging at $14499 \times g$. After these two washing steps, the DNA was eluted in 30-50 μ l of nuclease free water. For this, the water was added directly in the middle of the membrane and then left for 1 minute at room temperature. After that, the column was transferred in a properly labelled Eppendorf tube and centrifuged at $14499 \times g$. Those tubes were stored on ice and DNA was quantified using NanodropTM spectrophotometer (Thermofisher). The quantified DNA was used instantly or stored at 4 °C for short term or at -20 °C for long term preservation.

2.3.2 Validation of the mutation status of the cell lines

2.3.2.1 Primer design and PCR

LS174T and LS180 cell lines have been developed from the same primary cancer. Interestingly, they differ regarding two mutations in *CDH1* gene (E-Cadherin). LS174T contains two heterozygous mutations but LS180 does not. These mutations were used for confirming the identity of the LS174T cell line and the clones (Clone 1 and Clone 2)

developed from it. In order to detect those two mutations, two pairs of sequencing primers were designed using NCBI PCR primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) with a consideration of keeping the putative mutations in the middle region of the amplified sequence. The primers were then validated using the UCSC *in silico* PCR tool (<http://genome.csdb.cn/cgi-bin/hgPcr>). The primer information is summarized below:

Table 2.1: Primer information for E-Cadherin mutation detection

Primer Orientation	Sequence	T _m (°C)	Target Mutation	Amplicon Size
Forward	5'-TACGTATACCCTGGTGGTTCA-3'	61.7	Exon 8 boundary mutation	477 bp
Reverse	5'-TGTGTGATGGTCTCTCCAGTG-3'	64		
Forward	5'-TTTCCCTCGACACCCGATTC-3'	68.3	Exon 3 single base pair deletion	506 bp
Reverse	5'-TACACTACCCGATCCCCACA-3'	65.1		

PCR was performed using AmpliTaq Gold® (Applied Biosystems®) master mix on a G-STORM GS4 PCR platform for 35 cycles at the conditions of denaturing at 95°C for 30s, extension at 72°C for 120s and annealing at 56°C and 60°C for 60s for the 477 bp and 506 bp amplicons respectively.

2.3.2.2 Gel Electrophoresis

To determine the success of the PCR reactions, PCR products were run in 1% agarose gel. For this, 1 gram of agarose was dissolved in 100 ml 1X TBA buffer. It was then heated at 900W in a microwave oven for 1 minute which was followed by a swirling and another 30 seconds of heating. At this point, agarose was completely dissolved in the TBA buffer. Then, 1 µl of SYBR® Safe DNA gel stain was added in 100 ml of the

dissolved gel which was casted on a tray designed for the Biorad gel electrophoresis system. A comb was set at one end of the gel and the gel was allowed to solidify.

After 30 minutes, the comb was gently pulled out of the gel. Then 10 µl (or 5 µl) of the PCR sample, mixed with DNA loading buffer (Biolabs, catalogue number: B7025S) at a 6:1 ratio was loaded in the wells of the gel. After completion of loading all the samples, the gel was submerged in TBA running buffer in the electrophoresis tank, which was filled to its maximum level. The system was then connected to power and run for 1 hour at 80 V. When the dye front ran $\frac{3}{4}$ of the gel length, the power was switched off and DNA bands were visualised using Multi-image II (Alpha Innotech) gel documentation system.

2.3.2.3 Purification of PCR products and sequencing

After the confirmation of successful PCR reaction using gel electrophoresis, the remaining PCR products (~40 µl for each sample) were cleaned using Qiagen PCR purification kit (catalogue number: 28104) according to the manufacturer's instructions. Briefly, PCR products were mixed with buffer PB in 1:4 ratio and gently vortexed for 15 seconds. The mixture was then transferred into the labelled PCR purification columns and centrifuged in the Eppendorf table top centrifuge at 14499xg for 1 minute. After discarding the eluate, 700 µl of buffer PE was added in the columns and centrifuged as mentioned before. The eluate was discarded and 20-30 µl of DNase free water was added directly in the middle of the membrane of the columns. The columns were left in the room temperature for 1 minute and then centrifuged for 14499xg for 1 minute. The elutes were stored on ice and quantified using Nanospectrometer.

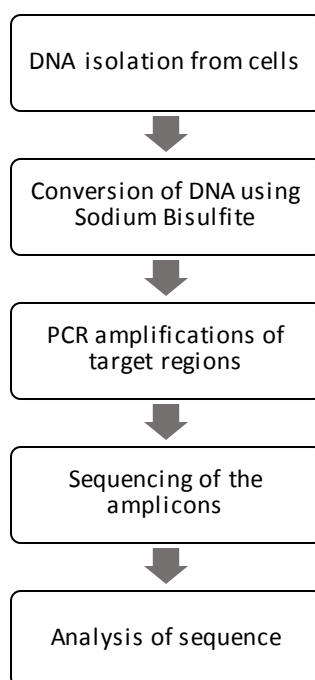
Volumes of 5 µL of the purified PCR products were sent for sequencing in the local facility. Forward and/or reverse primers (10 µL of a 3.3 picomolar solution) were provided for the sequencing.

Sequence results were visualised and analysed using 'Sequencher 5.0' software.

2.3.3 Bisulfite sequencing

Bisulfite sequencing was performed to identify the methylation of CpG positions in target region(s) (reviewed in section 1.6.2). Sequence differences observed between samples can indicate differences in CpG methylation at target sites.

A schematic of the bisulfite Sequencing process:



Workflow 2.1: Steps in bisulfite sequencing of DNA.

2.3.3.1 DNA isolation and bisulfite conversion

DNA was isolated using the procedure described in section 2.3.1. After measuring the DNA concentration, samples can be directly used for bisulfite conversion or kept at 4°C for short term storage and -20°C for a long term storage. For consistency purposes (as it was given for sequencing), the DNA was bisulfite converted on the same day using Qiagen Epitect Fast Bisulfite kit (catalogue number: 59824) according to the manufacturer's instruction. The table below was followed to prepare the sample for conversion.

Table 2.2: Recipe for bisulfite conversion reaction mix

	For high concentration (100 ng-2 μ g/ μ l) DNA	For low concentration (<100 ng/ μ l) DNA)
DNA	Volume ($\leq$ 20 μ l) to get 500 ng DNA	Volume ($\leq$ 40 μ l) to get maximum ($\leq$ 500ng) DNA
Nuclease free Water	Variable	Variable
Bisulfite Solution	85 μ l	85 μ l
DNA protection Buffer	35 μ l	15 μ l
Total	140 μ l	140 μ l

In a typical case, for high concentrations of DNA, the following recipe was used.

Table 2.3: Composition of a typical bisulfite conversion reaction

	For high concentration (100 ng-2 μ g/ μ l) DNA
DNA	5 μ l
Nuclease free Water	15 μ l
Bisulfite Solution	85 μ l
DNA protection Buffer	35 μ l
Total	140 μ l

After mixing the ingredients, the bisulfite conversion reaction was carried out in a thermal cycler (G-Storm). The following conditions were used for the conversion based on manufacturer's suggestion:

Table 2.4: Conditions for bisulfite conversion reaction

Step	Time (min)	Temperature ($^{\circ}$ C)
Denaturation	5	95
Incubation	20**	60
Denaturation	5	95
Incubation	20**	60
Hold	Indefinite	12*

** To increase the efficiency of the process, the maximum incubation time was used according

* The suggested temperature is 20 (°C). But in G-Storm 12 is the maximum temperature that can be set for storage.

Converted DNA can be stored after the reaction finishes. But during this work, the purification step was completed immediately after the conversion step.

2.3.3.1.1 Purification of the bisulfite converted DNA

30 minutes before starting this step, buffer BD and spin columns were transferred in room temperature from their 4°C storage.

To purify the converted DNA, protocol of Qiagen Epitect Fast Bisulfite Kit was followed.

The process started with adding 310 µl of Buffer BL to each sample in an Eppendorf tube. If starting total DNA amount was high (500 ng - 2 µg) then no carrier RNA (provided with the kit) was added in buffer BL. However, for low DNA amount (<100 ng), carrier RNA was added to the buffer BL according to the following table:

Table 2.5 Carrier RNA addition during the purification of bisulfite converted DNA

Number of samples	Volume of buffer BL (ml)	Volume of Carrier RNA added (µl)
1	0.35	3.50
5	1.75	17.5
10	3.50	35.0
15	5.25	52.5
25	8.75	85.5

Then 310 µl of this mixed buffer was added to each sample. The tubes were briefly vortexed after adding BL buffer and then 250 µl of Ethanol (95-100%) was added in each tube. All these tubes were centrifuged for 5-7 seconds (not exceeding 4547 x g) in a table top centrifuge.

The mixture was transferred in the appropriately marked spin columns, which were then centrifuged at 14499xg for 1 minute. Flow-through was discarded and the columns were

washed with 500 µl of buffer BW. After this, 500 µl of buffer BD was added in each column and kept at room temperature for 15 minutes.

After discarding the flow-through from the columns, they were washed twice with buffer BW as before. This step was followed by washing with 250 µl of ethanol. The columns were dried at 60 °C for 5 minutes.

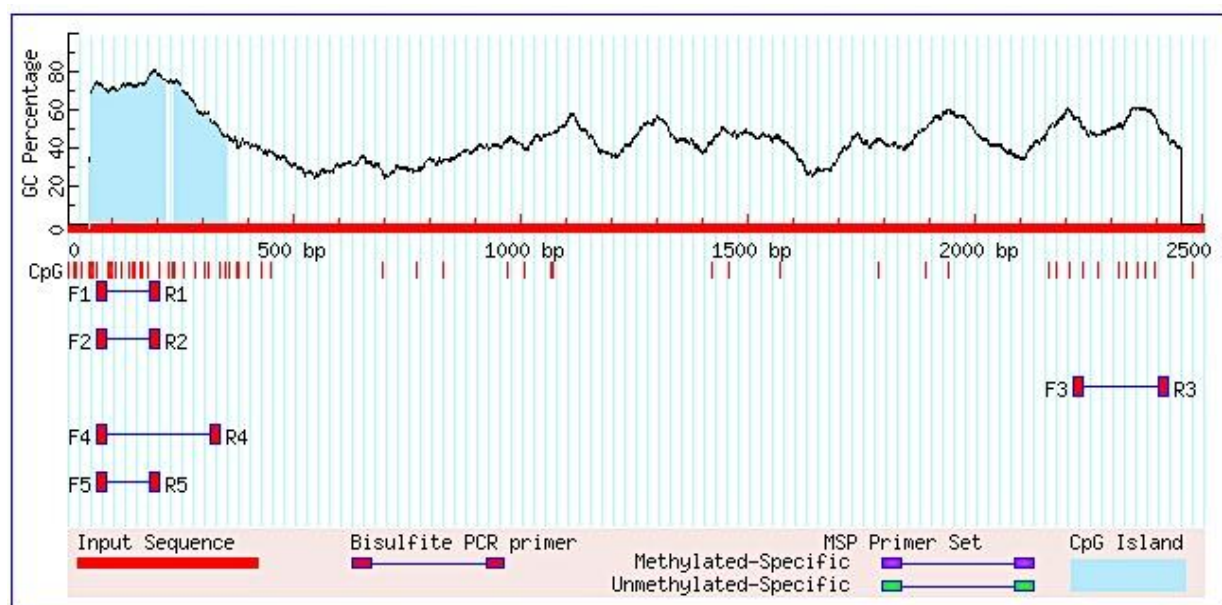
While eluting DNA, the starting amount of DNA (before conversion) was taken into consideration again. For high starting DNA amount, the elution was done in 25-30 µl of water. In case of low starting DNA amount, converted DNA was eluted in 12-15 µl of water. The elution step was completed by centrifuging at 14499xg for 1 minute in the table top centrifuge.

2.3.3.1.2 PCR amplification of purified bisulfite converted DNA

PCR amplification of bisulfite converted DNA was performed using primers designed by Methprimer platform (<http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi>).

In this particular case, chr6:106974606-106977143, flanking region of the 5'UTR of the *CD24* gene was used as the input sequence to design primers (sequence in the appendix: Ax.1).

Functionally significant methylation is expected to occur in the promoter region of *CD24* gene. Notably, a large part of the predicated CpG island spanned the first and second exons and thus, although the whole 2538 bp of the predicted CpG island was considered for analysis, a preference was given for the selection of primers that spanned the 5' flanking region of the 5' UTR. For this, a 2501 bp sequence spanning part of exon 1 and 5' UTR of *CD24* was examined on Methprimer which gave 4 sets of primers (Fig 2.1).



Sequence Length: 2501

CpG island prediction results

(Criteria used: Island size > 100, GC Percent > 50.0, Obs/Exp > 0.6)

2 CpG island(s) were found in your sequence

	Size	(Start - End)
Island 1	166 bp	(48 - 213)
Island 2	115 bp	(236 - 350)

Primer picking results for bisulfite sequencing (or restriction) PCR

	Primer	Start	Size	Tm	GC%	'C's	Sequence
1	Left primer	65	24	58.18	66.67	7	GTATTTTTTTTAGGAGTTGGGGAGT
	Right primer	201	23	58.76	78.26	11	AAAAAATCCCCTCAAATAAACC
	Product size: 137, Tm: 73.5, CpGs in product: 12						
2	Left primer	65	24	58.18	66.67	7	GTATTTTTTTTAGGAGTTGGGGAGT
	Right primer	202	24	59.86	79.17	12	AAAAAATCCCCTCAAATAAACC
	Product size: 138, Tm: 73.7, CpGs in product: 12						
3	Left primer	2214	25	51.98	52.00	5	TAGGTAGATTATATGGTTAGGAGAT
	Right primer	2422	25	54.15	64.00	9	CATCCAAACTAAAATACAATAACAC
	Product size: 209, Tm: 75.9, CpGs in product: 7						
4	Left primer	65	24	58.18	66.67	7	GTATTTTTTTTAGGAGTTGGGGAGT
	Right primer	335	23	54.84	39.13	4	AAAAATATATCCCCAAAAATTC
	Product size: 271, Tm: 75.4, CpGs in product: 21						

Fig 2.1. Primer pairs suggested by Methprimer program (Li & Dahiya, 2002): Primer pair 4 covered the longest region and the largest number (21) of CpGs (shown in red box).

Mainly primer pair 4 (shown in the box above) was used in the study as it covers the regions of primer pair 1, 2 and beyond. PCR was performed using NEB EpiMark® Hot Start *Taq* DNA Polymerase (Catalogue number: M0490)

The following recipe was used for mixing PCR components in sterile thin-walled PCR tubes:

Table 2.6: Recipe for PCR with bisulfite converted DNA

Component	25 μ l Reaction	50 μ l Reaction	Final Concentration
5X EpiMark® Hot Start <i>Taq</i> Reaction Buffer	5 μ l	10 μ l	1X
10 mM dNTPs	0.5 μ l	1 μ l	200 μ M
10 μ M Forward Primer	0.5 μ l	1 μ l	0.2 μ M (0.05-1 μ M)
10 μ M Reverse Primer	0.5 μ l	1 μ l	0.2 μ M (0.05-1 μ M)
EpiMark® Hot Start <i>Taq</i> DNA Polymerase	0.125 μ l	0.25 μ l	1.25 units/ 50 μ l PCR
Template DNA	Variable	variable	< 1,000 ng
Nuclease-free water	to 25 μ l	to 50 μ l	

PCR products were visualised using Agarose gel electrophoresis, purified and sent for sequencing in local sequencing facilities following procedures described in sections 2.3.2.2 to 2.3.2.3.

2.3.3.1.3 Analysing bisulfite converted sequence for CpG methylation

To quantify the methylation percentage in a particular CpG position, the relative height of the C and T representing the methylated and unmethylated cytosine respectively in the chromatograms was examined. All unmethylated Cs were expected to be converted to Ts

when they were not part of CpG dinucleotides. The efficiency of the bisulfite conversion reaction was measured from the presence of C when it was expected to be T.

2.4 RNA methods

qRT-PCR was performed to measure the relative expression of the gene of interest in different cell lines and clones. The process started with RNA isolation.

2.4.1 RNA Isolation

RNA was isolated from cells using Qiagen miRNAeasy kit (Catalogue number: 74104) following manufacturer's protocol. Briefly, cells were trypsinised and a suspension of approximately 0.5-1 million cells was transferred in the Eppendorf tubes. Cells were centrifuged at 2320xg for 3 minutes on a table top Eppendorf centrifuge.

Supernatant was then discarded and cell pellet was dispersed by gentle tapping. Cell lysis was carried out by adding 350 µl of lysis buffer RLT and vortexing for 1 minute with intermittent pulse.

The lysate was then transferred to labelled RNAeasy spin column. The columns were centrifuged for 1 min at 14499xg and the flow-through was discarded. The membrane from the column now binds the RNA ready to be washed and eluted.

For this, 700 µl of wash buffer RW1 was added in each column. The columns were then centrifuged at 14499xg for 1 minute. Flow-through was discarded. The columns were then washed with 500 µl of buffer RPE twice by centrifuging at 14499xg for 1 minute each time.

To remove residual buffer from the membrane and column, they were centrifuged for another two minutes at 14499xg.

The columns were then dried in room temperature for 3-5 minutes. RNA was eluted in water. The volume of water varied based on the number of initial cells. If cell number was low such as 2000 (or lower) to 10000, then 10-15 µl water was added directly in the

middle of the column. If cell numbers were higher, such as 100,000 to 1 million, then 35-50 μl water was added to elute RNA.

To collect the elute, the columns were transferred into Eppendorf tubes labelled appropriately for each samples. One minute after adding water into columns, they were centrifuged at 14499 $\times g$ for 1 minute in the Eppendorf table top centrifuge.

The elute now collected in the tubes was kept on ice before further use. Concentration of RNA was measured using (Thermo) nanodrop spectrophotometer. Blank was set using water and 1-2 μl of RNA was measured against the blank.

RNA was diluted as 10 ng/ μl in nuclease free water for the qRT-PCR.

2.4.2 Quantitative RT-PCR (qRT-PCR)

qRT PCR was performed in two steps: cDNA synthesis from RNA and real time PCR with the cDNA.

For cDNA synthesis, Applied Biosystem's High-Capacity cDNA Reverse Transcription Kit (Catalogue number: 4368814) was used. Briefly, diluted RNA (10 ng/ μl) was mixed with 10 μl of master mix prepared using the following recipe:

Table 2.7: Master mix for cDNA synthesis

Components	Volume/Reaction (μl)	
	Kit with RNase inhibitor	Kit without RNase inhibitor
10x RT buffer	2.0	2.0
25x dNTP Mix (100 mM)	0.8	0.8
10x RT random primers	2.0	2.0
MultiScribe Reverse Transcriptase	1.0	1.0
RNase Inhibitor	1.0	-
Nuclease free water	3.2	4.2
Total	10.0	10.0

Total 20 μ l reaction volume (10 μ l RNA + 10 μ l master mix) was taken for each sample in PCR microtubes which was placed in a G-storm G4S PCR system for cDNA synthesis in the following conditions:

Table 2.8: PCR conditions for cDNA synthesis

	Step 1	Step 2	Step 3	Step 4
Temperature ($^{\circ}$ C)	25	37	85	4
Time	10 min	120 min	5 min	∞

After cDNA synthesis, quantitative PCR was performed using Taqman[®] probes and Taqman[®] Fast Universal PCR Master Mix (Applied Biosystem) on an Applied Biosystems AB7500 fast machine. For this, cDNA was diluted at a concentration of 25 ng/ μ l. Then, 10 μ l universal master mix was added with 1 μ l Taqman[®] probe for each target gene. For qRT PCR, 16 μ l of this mixture was added to 4 μ l cDNA (25 ng/ μ l). The resultant 20 μ l volume was added in triplicate wells of a MicroAmp[®] Optical 96-Well Reaction Plate. Ubiquitin C (UBC) was used as an internal control in all of these experiments. The plate was spun to collect the reaction mixture at the bottom of the well.

The PCR was run in a quick synthesis program (~45 minutes) using an Applied Biosystems AB7500 Fast Real-Time PCR System. The $\Delta\Delta C_T$ method was used for calculating the relative changes in gene expression.

2.5 Protein methods

2.5.1 Protein Isolation

For western blot analysis, protein was isolated from cells following standard protocol. Briefly, cells were washed with ice cold PBSA three times. Remaining PBSA was drained and then 30 μ l of ice cold lysis RIPA buffer (150 mM sodium chloride, 1.0% NP-40 or Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS (sodium dodecyl sulfate), 50 mM Tris, pH 8.0) was added in each well. Cells were scraped from these wells and the lysates transferred to the 1.5 mL microcentrifuge tubes were kept on ice for 40 minutes with intermittent vortexing.

Lysate was then centrifuged at 4°C for 20 minutes at 18000xg and the protein containing supernatant was transferred into fresh tubes and stored at -20°C for short term storage and -80°C for long term storage until use.

2.5.2 Western Blot

Western blot was performed using a mini PROTEAN 3 apparatus (BioRad). Protein concentration of the lysates was quantified using Pierce™ BCA Protein Assay Kit (Thermo Scientific) according to manufacturer's instruction. For western blot, 30 mg of each protein sample was mixed in 2x Laemmli buffer (4% SDS, 10% 2-mercaptoethanol, 20% glycerol, 0.004% bromophenol blue, 0.125 M Tris HCl, pH 6.8) to obtain a total volume of 40 µl, then heated at 95°C for 10 minutes. Denatured protein was then loaded into the wells of a polyacrylamide gel which were prepared as 3.2% stacking- and 8% resolving-layer respectively. The electrophoresis was performed at 100 volts for 2 hours.

Once the gel run completed, proteins were transferred on an Amersham™ Hybond™ ECL Nitrocellulose Membrane using semi dry transfer system. For this, the gel was rinsed in ice-cold TBST buffer (20 mM Tris-Cl, pH 7.5, 150 mM NaCl, 0.1% TWEEN 20). The membrane was cut and rinsed into ice-cold methanol for 2 minutes and then a sandwich of the gel and the membrane was made according to the protocol. Transfer was performed at 20 volts for one hour. The transfer efficiency was visualised using Ponceau red stain. Then the membrane was washed in water to remove the dye.

The membrane was blocked in 5% skimmed milk in TBST buffer for 30 minutes and then incubated at 4°C overnight with a primary mouse monoclonal anti-E-Cadherin antibody (BioLegend, catalogue number: 324102) at 1:1000 dilution in 2.5% skimmed milk blocking buffer. This was followed by 3 washing steps of 10 minutes and an incubation with a Rabbit polyclonal anti-mouse antibody (DAKO) at 1:2000 dilution in 2.5% skimmed milk blocking buffer for one hour at room temperature.

After washing 3 times in TBST buffer, the membrane was developed using Amersham™ ECL™ Prime Western Blotting Detection Reagent (GE Healthcare Life Sciences)

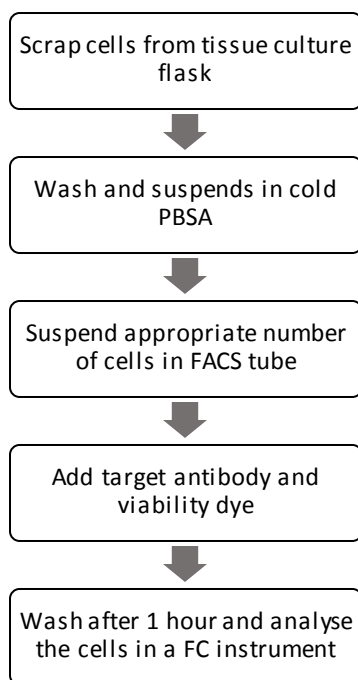
according to manufacturer's instructions using chemiluminescent gel doc system. Alternatively, the membranes were exposed to x-ray films and developed in a dark room using a film processor.

2.5.3 Flow cytometry

Flow Cytometry can measure cellular protein expression (both cell surface and cytoplasmic). It gives quantitative data and has the advantage over western blot analysis as it can show the distribution of negative and positive cells within a bulk population.

During this study, flow cytometry was used extensively for checking the expression of CD24 and CD44. Only surface expression of these genes were measured and that's why no permeabilisation step was needed.

The following flow diagram shows the basic steps of the process:



Workflow 2.2: A simple schematics of flow cytometry

2.5.3.1 Preparation of cells (scrapping and washing cells)

Cells are collected by trypsinisation. For this, cells were washed 3 times with PBSA and 700 μ l trypsin was added to the flask and incubated at 37°C for 5 mins. A gentle shake detached the cells from the flask surface. Trypsin was neutralized by adding 5 ml of complete medium.

The cell suspension was transferred into a 15 ml falcon tube and washed twice with PBSA. Those tubes were centrifuged at 2571xg for 5 mins at 5°C in a Beckman Coulter centrifuge machine (Allegra [™] 6R centrifuge)

2.5.3.2 Antibody staining

Mouse anti-CD24-PE (catalogue number: MHCD2404) and mouse anti-CD44-FITC (catalogue number: MHCD4401) were obtained from Life Technologies and used for the detection of surface expression of CD24 and CD44, respectively.

For this, cells were suspended in 300-500 μ l of PBSA. A volume of 100 μ l of this cell suspension (0.5-1 million cells) was transferred into FACS tubes and added with 5 μ l of antibody. DRAQ7 (Biolegend, Catalogue number: 424001) was added as viability testing dye (1:100). The tubes were kept on ice for 45-60 minutes with occasional shaking.

After this incubation, 3 ml PBSA was added in each FACS tube and then centrifuged at 2571xg for 5 mins (at 5 °C) in the Beckman Coulter (Allegra [™] 6R centrifuge). The supernatant was discarded and cells were dissolved in 300-500 μ l PBSA.

2.5.3.3 Flow cytometry analysis

Samples were analyzed on a CyAn ADP analyzer using the Summit 4.3 software. Cells were first gated based on their size (using FS Lin vs SS lin) followed by the Pulse width vs FS lin to select for the single cell population. The final gating was based on DRAQ7 staining where only the live cells (DRAQ7 negative cells) were selected. The representation of CD24 and CD44 expression was displayed using both a histogram and a dot plot representation.

2.5.4 SRB assay

The sulforhodamine B (SRB) assay (Vichai & Kirtikara, 2006), optimised by Djamila Quaret for CRC cell lines in the Bodmer laboratory, was used for cell density determination by measuring total cellular protein content. Briefly, in the 96-well plates, the cells were fixed by adding 15uL of TCA (50%). The plates were left at 4°C for at least 4 hours or overnight. After this, the plates were washed with water (3x) and dried in the air for at least 1 hour and 30 minutes. At this stage, SRB (0.04%) was added in the plate and left for 1 hour. After this incubation period, the SRB was removed from the plates by flicking and the plates were washed 3 times or more with 1% acetic acid to get rid of the non-specific staining. The plates were dried for at least 1 hour and 30 minutes in room temperature. Plates were stored at this stage in room temperature or suspended with 200uL of Tris solution to be read in a plate reader.

2.6 Microarray Data Analysis using Partek Genome Suite

Partek® Genomics Suite™ was used to analyse the microarray data. Data were first normalized and the batch effect was removed before performing analysis of differentially expressed genes.

Cell lines divided in group 1 and group 2 (based on their level of gene expression) were compared in order to identify the most significantly differentially expressed genes between these two groups.

All microarray data referred in this report are non-log-transformed, absolute fluorescence intensities obtained using GCRMA normalised data.

2.7 Statistical test

In this thesis, correlation between the expressed genes across cell lines was assessed on several occasion. The GraphPad Prism 6.0 software package was mostly used for such analysis. The Chi-squared test and Fisher's exact test were used to find statistical significant correlation from a 2x2 table with two variables. GraphPad Prism's online service was used in those cases.

Chapter-3

LS174T- A Model for Studying Clonal Heterogeneity and E-Cadherin

Chapter-3

3.1 Chapter introduction

Two LS cell lines –LS174T and LS180– were established in 1974 from a colon adenocarcinoma removed from a 58-year-old woman (Tom et al., 1976). LS174T was developed after ten months of primary culture as a trypsinised variety. In this case, cells were detached from the flask during sub culturing using 0.1% trypsin in 0.02% versane (PH 7.4), whereas LS180 was sub-cultured using only mechanical scraping (with a rubber policeman) without trypsinisation.

Initial characterisation of LS174T by Tom *et al.*, revealed a doubling time of 96 hours till passages 6-11, but in later passages (passage 20), the doubling time changed to 26 hours. LS174T formed loosely connected island-like colonies in culture that became densely packed if left in culture for over a week. LS174T secreted mucin and produced high level of CEA (Tom et al., 1976) .

Since its establishment, LS174T has been widely used in the studies of colorectal cancer (CRC). In the Bodmer Laboratory, LS174T has been characterised well over many years of work on genetic mutations, epigenetic modifications, whole genome mRNA expression levels and drug responses. It grows well in DMEM medium with 10% serum. It is highly clonogenic (~85%) and can be reliably cloned using FACS of single cells seeded into a 96 well plate.

LS174T is a mismatch repair (MMR) deficient cell line with a mutation in *hMLH1* and, as such, displays microsatellite instability with a pseudodiploid karyotype that was initially characterised by Tom *et al.* as 45X, with loss of one of the X chromosomes. Later characterisation of LS174T using SKY spectral karyotyping revealed a karyotype between chromosomes 46 and 47 (Fig 3.1), that included trisomies for both chromosomes 7 and 15, together with loss of one of the X chromosomes (Abdel-Rahman *et al.*, 2001).

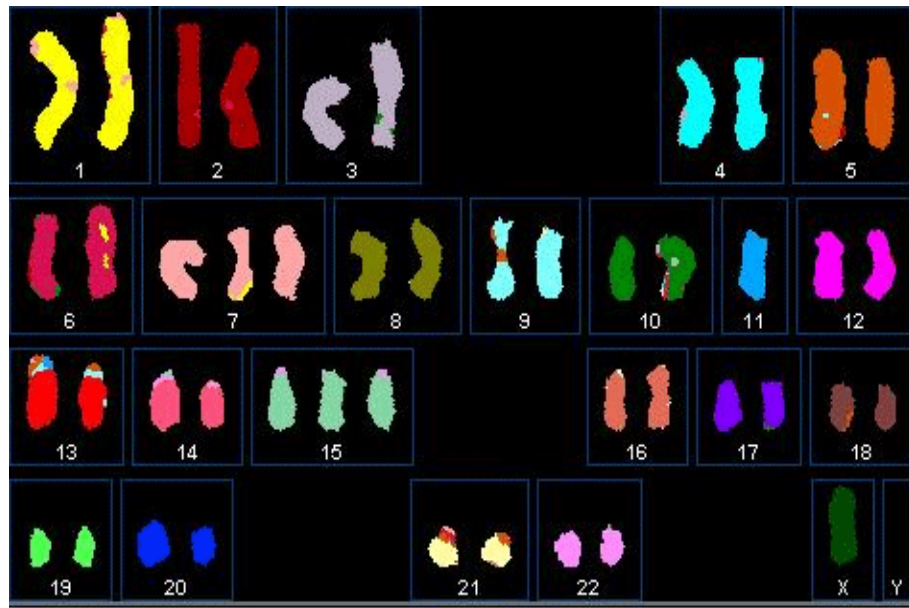


Fig 3.1: LS174T Karyotype by multicolour chromosome painting- spectral karyotyping (SKY) showing trisomies of chromosomes 7 and 15, and loss of an X chromosome. (Image retrieved from: Abdel-Rahman *et al.*, 2001).

3.2 Molecular characterisation of LS174T

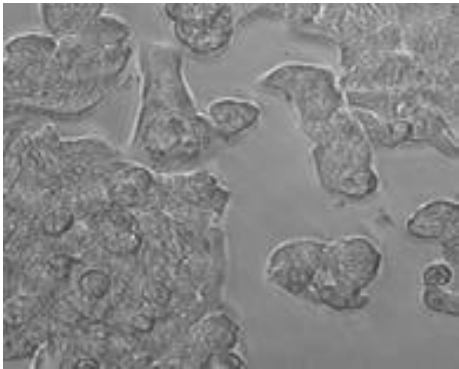
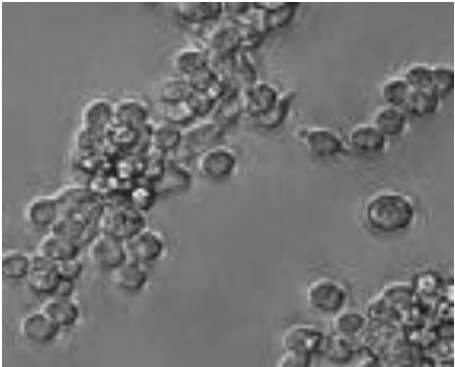
LS174T is a replication error (RER) positive cell line which is indicative of its relatively higher rate of mutation. LS174T contains mutations in several driver genes including *KRAS* (in codon 12: Gly → Asp), *CTNNB1* (codon 45), and *PIC3CA* (3140 A→G H1047R; het ex20). According to the public database and internal lab data, it is wild type with regards to *TP53*, *APC* and *BRAF*. However, during this research work, a new mutation in *BRAF* gene (chapter 4) has been identified in the LS174T.

3.3 LS174T and LS180: A duo with differences

3.3.1 Morphological differences

Despite being derived from the same primary cancer, LS174T and LS180 showed some visible differences. Tom et al., (1976) made the observation that “most of the scraped cells of the LS174T grow as multicellular clusters or spheroids that float in the medium until they become heavy enough to adhere to the flask surface”. LS180 was growing as relatively more compact colonies.

In this current study, the colonies of LS174T have been characterised as grape-like clusters of loosely attached cells and are easier to trypsinise, whereas LS180 grows firmly attached to the flask surface and requires longer treatment with trypsin (Fig 3.2 A). The most important difference between the cell lines was observed when they were grown in matrigel for 3D culture; LS180 forms colonies with well-defined central lumens (Fig 3.2 B)— a feature previously used as a measure for the capacity of a cell line to differentiate (Ashley, Yeung, & Bodmer, 2013). LS180 grown in matrigel gives rise to two types of lumen-containing colonies: smaller colonies with a single central lumen and bigger colonies with multiple lumens. On the contrary, LS174T does not form well-defined lumens in matrigel, rather its colonies may contain some micro-lumens which are visible as spots of actin foci (Fig 3.2 B). By the end of week 3 in matrigel, the LS180 colonies remain mostly compact with defined edges, however, the LS174T colonies spread out in matrigel (Fig 3.2 C).

LS180	LS174T
A. 	

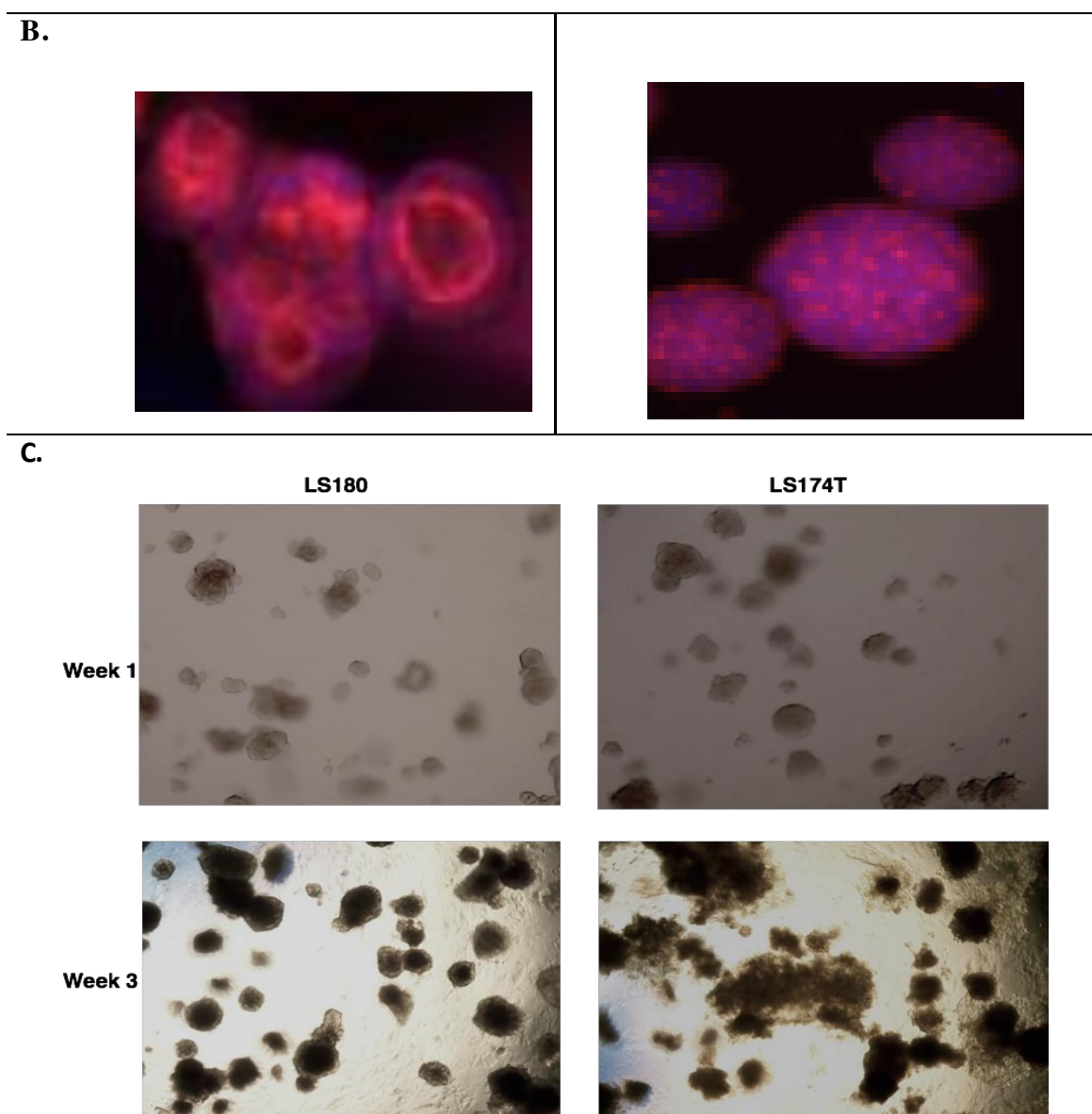


Fig 3.2: Difference between LS180 and LS174T cell lines. A. Light microscopy image (20x) of LS180 (left) and LS174T (right). B. Colonies of LS180 (left) and LS174T (right) stained in matrigel after 14 days. F-actin stained with TRITC-Phalloidin (red), nucleus stained with DAPI (blue). C. Comparison between the colonies of LS180 (left) and LS174T in matrigel after week 3. Top panel: colonies from both cell lines at week 1, bottom panel: colonies from both cell lines at week 3.

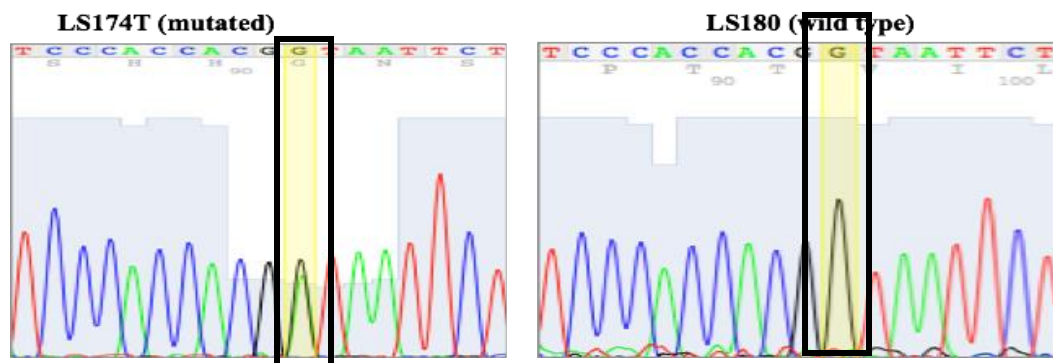
3.3.2. E-Cadherin expression difference between LS174T and LS180

A key observation from the Bodmer laboratory identified two heterozygous *CDH1* mutations in the LS174T cell line: i) exon 8 splice donor site (ACGgtaa ->ACGataa) mutation, and ii) a frame shift deletion mutation in exon 3 that results in a downstream premature stop (TGA) codon at position 216 (Efsthathiou et al., 1999). As the LS174T does not express E-Cadherin but LS180 does, it seems highly likely that these mutations cause the loss of E-Cadherin expression in LS174T and also provide a convenient cell line identification marker for LS174T.

At the start of this study, and periodically throughout the progression of this body of work, the presence of these two mutations was confirmed using sequencing oligos that encompass the regions containing those two mutations, in order to validate the LS174T cell line.

Sanger sequencing of the PCR amplified products confirmed the presence of both mutations in LS174T but not in LS180 (Fig 3.3 A and 3.3 B).

A. Exon 8 splice donor site: **ACGgtaa -> ACGataa**



B. Exon 3 frame shift deletion mutation: **GCCCCCG -> GCCCCG**

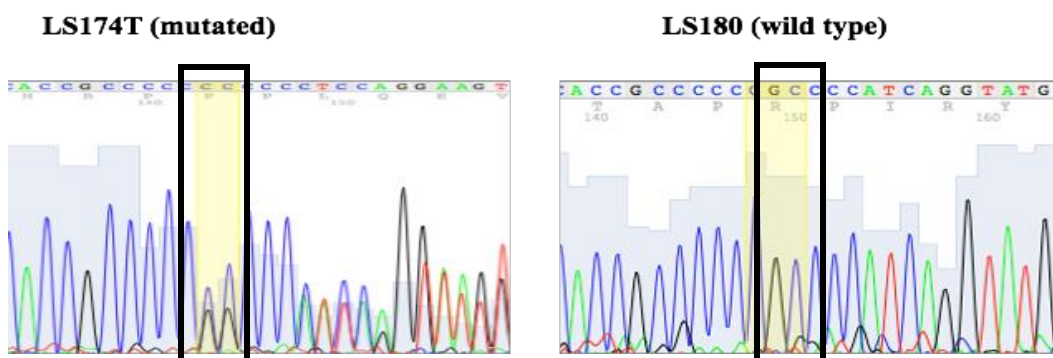


Fig-3.3: E-Cadherin mutation difference between LS174T and LS180, A: Heterozygous mutation at Exon 8 splice donor site (highlighted in boxes): ACGgtaa ->ACGataa. Here the presence of **A** (green peak) with **G** (black peak) indicates the heterozygous mutation. B: Heterozygous Exon 3 frame shift deletion mutation (highlighted in boxes): GCCCCCCG -> GCCCCCCG. The single base pair deletion results in the characteristic shift in the sequence trace of one allele after the point of the deletion (to the right of the highlighted box in LS174T sequence).

Despite this difference in E-Cadherin mutations and expression, confirmation that the two cell lines, LS174T and LS180, are indeed derived from the same patient comes from SNP genotyping using a panel of 34 unlinked SNPs with high heterogeneity scores that is routinely used in the Bodmer laboratory to authenticate cell line identity. The mutation signature of key genes such as *KRAS*, *PIK3CA* and *CTNNB1* provide further validation of the shared origin of the two cell lines.

3.4 LS174T: A model cell line to study the role of E-Cadherin in the regulation of differentiation

Down regulation of E-Cadherin is a key marker of epithelial to mesenchymal transition (EMT). EMT can be described as a cellular process where a polarised epithelial cell undergoes multiple biochemical changes, enabling it to assume a more mesenchymal phenotype with enhanced migratory capacity, invasiveness, resistance to apoptosis and increased production of extra cellular matrix (ECM) components. The key molecular features include (but are not limited to) downregulation of expression of the epithelial cell marker CDH1 and upregulation of mesenchymal markers such as Vimentin (VIM), together with the expression of known inducers of EMT such as the transcription factors ZEB1, TWIST1 or SNAI1. De-differentiation has often assumed to be a part of the EMT process (Medema, 2013).

The above evidence of i) the common origin of the LS174T and LS180, ii) their differences in E-Cadherin mutations and expression and iii) their differences in the capacity to differentiate when grown in matrigel, facilitated further study to understand the exclusive role of CDH1 in differentiation, as the lack of CDH1 expression in LS174T is not brought about by the more usual inducer of EMT process (such as the ZEB1, TWIST1 or SNAI), but rather more directly through the mutation of *CDH1*. These two

cell lines do not differ in the expression of other genes that have been implicated with EMT (Table 3.1).

Table 3.1: EMT associated gene mRNA expression levels in LS174T and LS180. Values are normalised absolute linear fluorescent intensities obtained from Affymetrix U133+2 microarray data.

Gene	Level in LS180	Level in LS174T	Status in EMT
E-cadherin (<i>CDH1</i>)	18115	677	Not expressed
Claudin7 (<i>CLDN7</i>)	4493	4737	Not expressed
Claudin3 (<i>CLDN3</i>)	6857	3670	Not expressed
<i>ZEB1</i>	12	14	Expressed
<i>TWIST1</i>	6	6	Expressed
<i>SNAI1</i>	8	8	Expressed
Vimentin (<i>VIM</i>)	27	25	Expressed

The loss of E-Cadherin (CDH1) expression in the LS174T cell line without changing any EMT makers and regulators allows for separating the key aspect of EMT, namely the downregulation of CDH1 from the expression of EMT inducing transcription factors—ZEB1, TWIST1, SNAI1 and markers *VIM*, *CDH2* etc. For this, the experimental design involved exogenous expression of E-Cadherin in LS174T and knocking down of E-Cadherin in LS180 to see if these interventions could change the two cell lines' original growth features in matrigel. To that end, F-actin staining was used to detect lumen formation.

3.5. Knockdown of E-Cadherin in LS180

3.5.1: E-Cadherin knockdown efficiency in LS180

To investigate whether E-Cadherin is necessary for the differentiation of LS180 into lumen structures, the cell line was transfected with siRNA of E-Cadherin to knock down its expression.

First, the knockdown efficiency of the siRNA was measured. For this, 500,000 cells from two E-Cadherin expressing lines, LS180 and SW1222 were transfected with siRNA of E-Cadherin using RNAimax transfection reagents and incubated for periods of 48, 72 and 96 hours.

It was found that siRNA knockdown resulted in at least 79% silencing of the gene in LS180 and 84% silencing in SW1222 at 48 hours (Fig 3.4 A). The knockdown was observed up to 60% at 96 hours for LS180 and 78% for SW1222. The knockdown was confirmed at the protein level in LS180 (Fig 3.4 B).

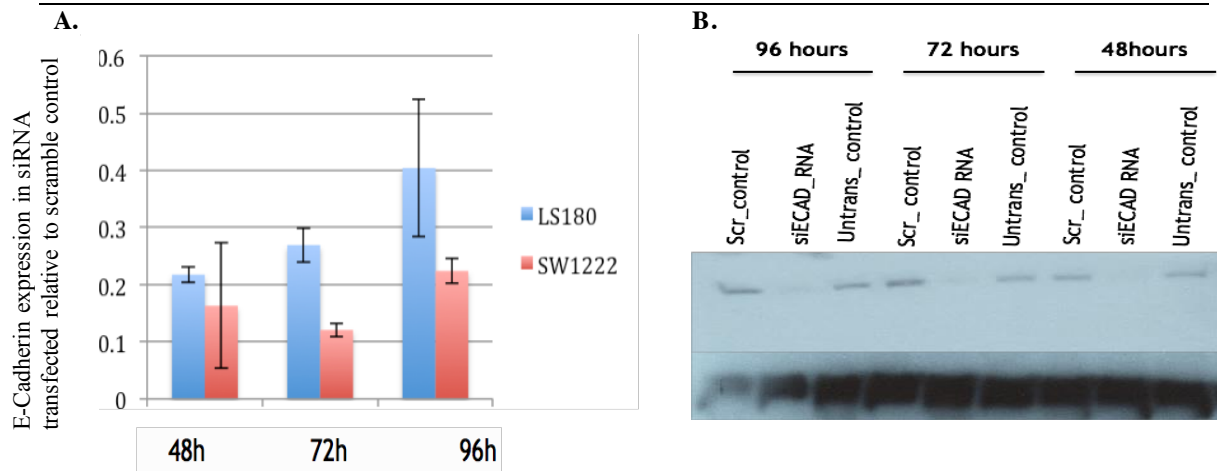


Fig 3.4: E-cadherin siRNA knockdown efficiency: A. Ratio of E-Cadherin expression in siRNA^{E-Cadherin} to scrambled transfection control in LS180 (blue bars) and in SW1222 (red bars) at 48, 72 and 96 hours. The error bars indicate the standard error of the mean calculated from three replicate wells in the same experiment. B. Western blot of E-cadherin siRNA transfected LS180. Scrambled control (Scr_control), siRNA^{E-Cadherin} (siECAD) transfected and untransfected control (Untrans_control) have been shown at three time points. The bottom lane shows Beta-tubulin as loading control.

The knockdown efficiency suggested that the effect should be there beyond these time points (4 days). Based on these observations, LS180 transfected with siRNA of E-Cadherin was observed in matrigel after 7 days to investigate the effect E-Cadherin knock down on colony morphology and lumen formation.

3.5.2. Effect of E-Cadherin knockdown on differentiation

Cells transfected with siRNA were seeded in triplicate wells of matrigel in a 96 well plate. 500 cells per well were seeded 24 hours after transfection. The plate was incubated at 37°C for 7 days and then assessed for differentiation by F-actin staining to detect lumens.

Fluorescent microscopy of the transfected colonies in matrigel showed that E-cadherin knockdown could not abolish the lumen forming ability of LS180 (Fig 3.5). Like scrambled control LS180, siRNA^{E-Cadherin} transfected LS180 formed similar lumens.

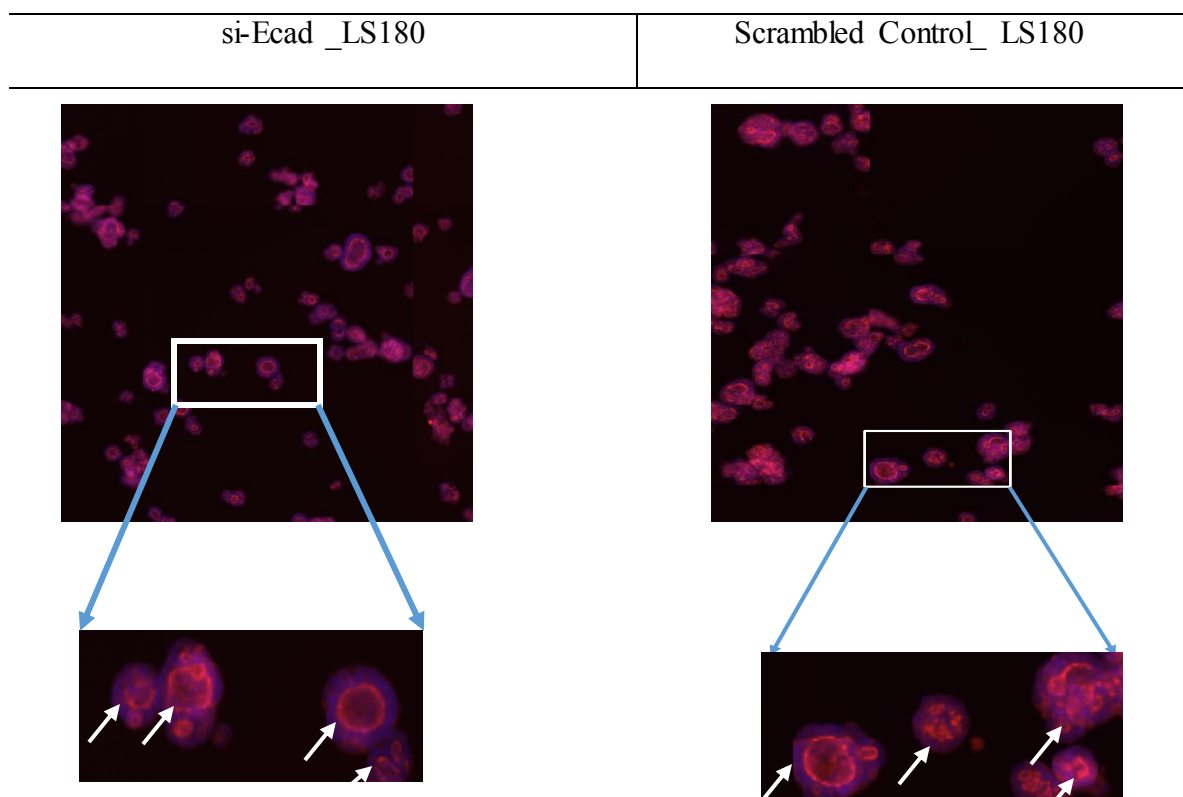


Fig 3.5: siRNA^{E-cadherin} transfected (left) and scramble transfected (right) LS180 cell line. Lumens were observed in E-Cadherin knocked down cells as well as in the scramble control cells. The bottom panel shows the enlarged image.

3.6. E-Cadherin knock-in in LS174T and its effects

3.6.1. E-Cadherin knock-in using plasmid

LS174T cell line was transfected with a GFP tagged expression vector containing E-Cadherin cDNA (Addgene, pCDNA3.1). E-Cadherin expression up to 96 hours after transfection was investigated to see the efficiency of this transfection. Western blot was performed to check the level of E-Cadherin (using E-Cadherin antibody from Biolegend, catalogue No. 324101) in untransfected, empty vector transfected and E-Cadherin transfected cells. As expected, neither one of the untransfected or empty vector transfected cells showed expression of E-Cadherin, whereas cells transfected with E-Cadherin cDNA showed a significant expression even after 4 days (Fig 3.6).

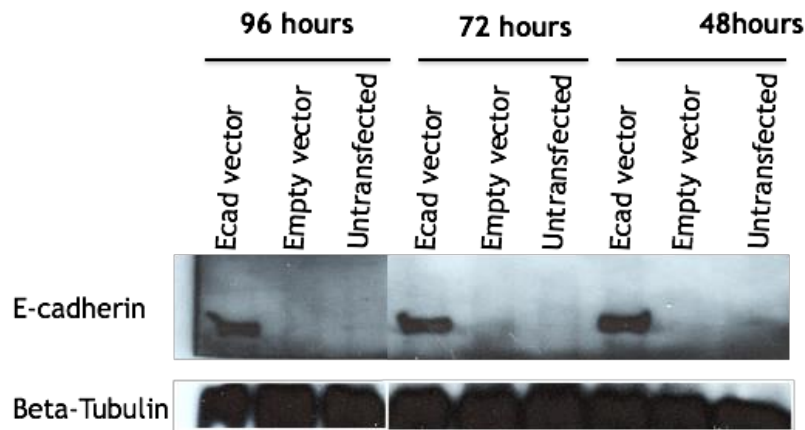


Fig 3.6: Western blot of E-Cadherin in LS174T cells transfected with a plasmid containing the gene. Top panel: E-Cadherin expression shown at three time points for E-Cadherin cDNA transfected (Ecad vector), empty vector transfected and untransfected LS174T. Bottom panel: Beta-tubulin expression in all cells shown as the loading control.

3.6.2. Effects of E-Cadherin expression on LS174T differentiation

After confirming the transfection efficiency and stability of E-Cadherin protein expression in LS174T, the effect of E-Cadherin expression on lumen formation was assessed as a measure of differentiation. Both the control and transfected cells were seeded in matrigel and F-actin staining assays were performed after seven days as described previously.

Fluorescent microscopy of colonies in matrigel showed that despite exogenous expression of E-Cadherin in LS174T, they did not show any difference regarding lumen

formation (Fig 3.7 A, B). The colonies from the E-Cadherin-transfected LS174T have similar morphology and structure to the untransfected LS174T and are not comparable with the morphology of the colonies of E-Cadherin expressing LS180 (Fig 3.7 C). It is expected that, upon its transfection with the cDNA of E-Cadherin, LS174T would show some differentiation like the LS180 if E-Cadherin plays any causal role in differentiation.

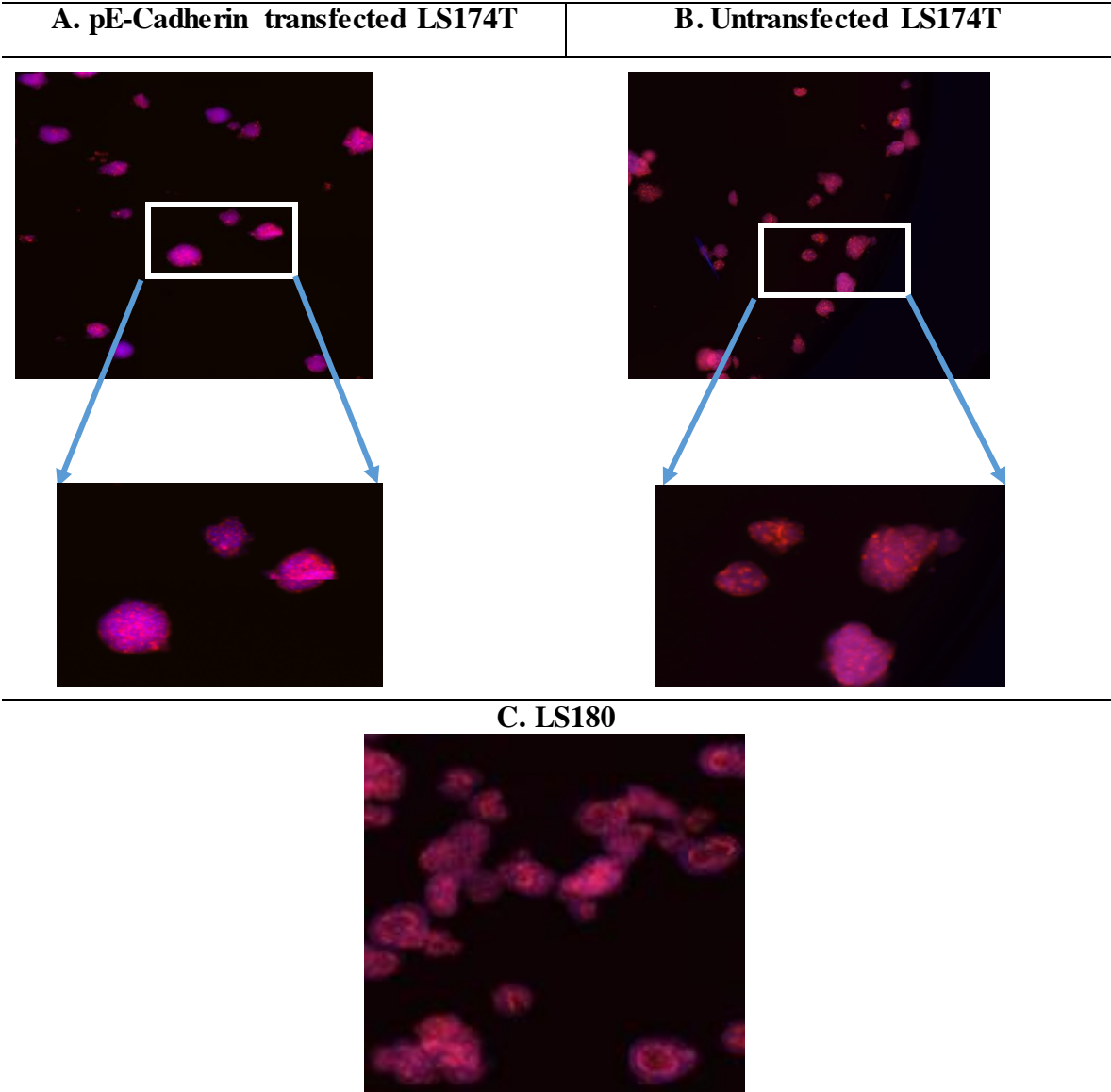


Fig 3.7: Effect of E-Cadherin expression on lumen formation: A. Morphology and structure of the colonies of LS174T expressing E-Cadherin from the transfected E-Cadherin cDNA. B. Morphology and structure of the colonies from untransfected LS174T. In both A and B, the bottom panel shows enlarged sections. C. LS180, the E-Cadherin expressing cell line forms lumens in matrigel. In all cases, cells are stained for F-actin using TRITC-Phalloidin (red).

3.7 Loss of E-Cadherin due to mutation versus E-cadherin silencing due to EMT have different consequences

In the panel of 95 CRC cell lines in the Bodmer laboratory, 11 cell lines (COLO320HSR, CC20, OXCO1, NCH716, RKO, COLO320DM, CCO7, SW620, HDC9, LS174T, SW480) express very low levels of E-Cadherin (CDH1). Interestingly, except HDC9 and LS174T, in all other cell lines, loss of E-Cadherin was accompanied by the expression of transcription factors ZEB1 and another well characterised marker of EMT, Vimentin (VIM) (Fig 3.8).

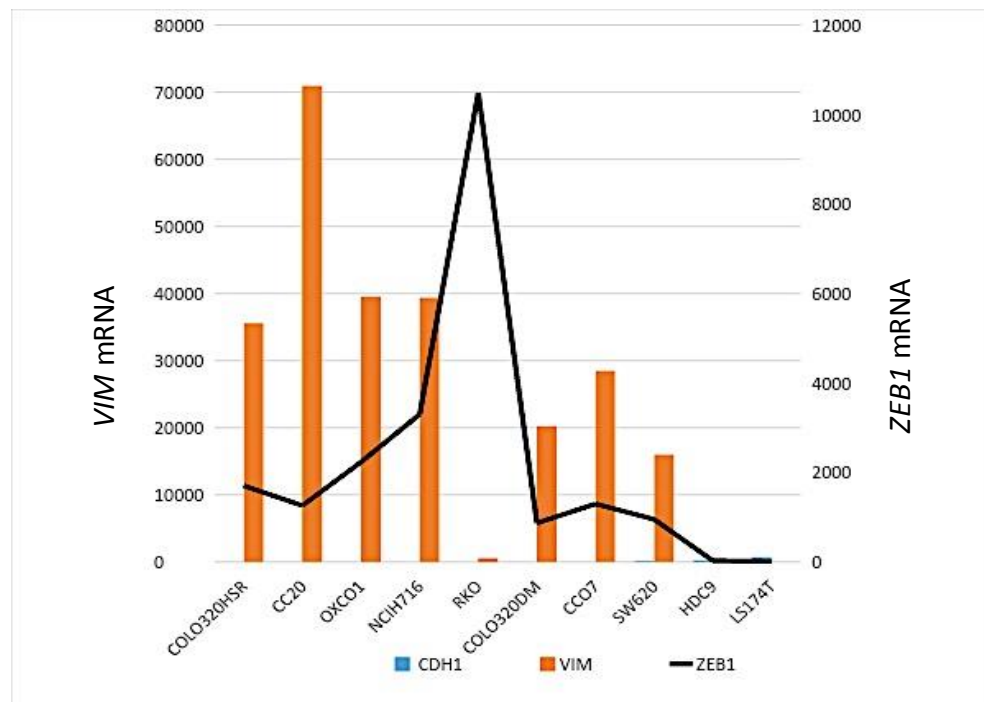


Fig 3.8: mRNA expression of *CDH1*, *VIM* and *ZEB1* in E-Cadherin-low CRC cell lines. In this 2 dimensional graph linear expression values of *VIM* and *ZEB1* have been shown based on microarray analysis.

Expression of ZEB1 and VIM is indicative of the activation of EMT which resulted in the down regulation of E-cadherin in those cell lines (Kalluri & Weinberg, 2009; Ocaña et al., 2012; Romero-Pérez et al., 2013; Sánchez-Tilló et al., 2010; Thiery et al., 2009). In the LS174T, however, the suppression of E-Cadherin was ascribed to two mutations (section 3.3.2) and was not associated with any evidence of activation of the EMT process (LS174T does not express ZEB1 or any other key marker of EMT). A previous PhD student, Olybmia Dikomitou, from the Bodmer Laboratory has shown that the cell

line HDC9 also has two mutations in the E-Cadherin gene resulting its silencing in this cell line (Fig 3.9, unpublished data).

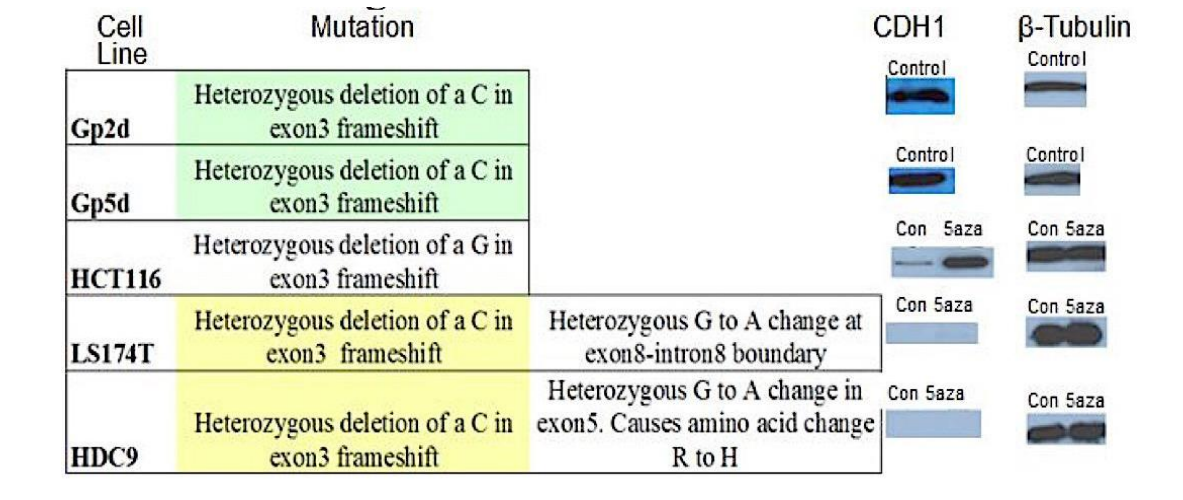


Fig 3.9: Effects of mutations on E-Cadherin expression. The presence of the heterozygous deletion of ‘C’ in exon 3 frameshift did not impact the expression of E-Cadherin in Gp2d and Gp5d. In HCT116, the low expression of E-Cadherin was probably due to methylation mediated suppression rather than the heterozygous deletion of ‘G’ in exon 3 frameshift, as here the 5-aza-2'-deoxycytidine treatment rescued the expression. Only LS174T and HDC9 have additional mutations that cause the total loss of E-Cadherin in them. Treatment with 5-aza-2'-deoxycytidine does not have any impact on CDH1 expression in these two cell lines which rules out the methylation mediated suppression in them compared to HCT116 where CDH1 expression clearly increases with 5-aza-2'-deoxycytidine treatment and that explains its low level of CDH1 in spite of having only one mutation (unpublished data, Olymbia Dikomitou).

The observation of the E-Cadherin silencing due to mutation versus due to ZEB1 expression helps to infer the lack of a causal role of E-Cadherin in EMT.

The absence of E-Cadherin in the LS174T or HDC9 did not have any impact on the expression of ZEB1 or any other regulators or markers of mesenchymal cell. However, in all the other 9 cell lines, the presence of ZEB1 has most likely repressed the expression of E-Cadherin (Fig 3.8). So it can be inferred that the loss of E-Cadherin cannot cause obvious induction of EMT associated markers or regulators, however the

expression of EMT regulators (ZEB1) can most probably cause down regulation of E-Cadherin.

Another important observation was that none of the E-Cadherin^{low} cell lines (9 EMT like cell lines plus LS174T and HDC9) can form lumens when grown in matrigel (see data in Appendix: Ax.9). So, a question was asked whether the absence of E-Cadherin could be the reason for their inability to differentiate in matrigel. Notably, in the Bodmer Laboratory, out of 83 cell lines that have been characterised in 3D culture to check their differentiation, a chi-square test for a 2x2 table showed there is no association between E-Cadherin expression and lumen formation. In that analysis it has been shown that though 20 E-Cadherin^{high} cell lines can differentiate (form lumens), 52 (out of 83) E-Cadherin^{high} cell lines cannot. This indicates that E-Cadherin expression cannot dictate the differentiating ability of the cell line. In other words, it is not that a cell line will form lumen because it is an E-Cadherin high line; but rather because it is a non-EMT line, that correlates with its capacity to differentiate.

A likely postulate based on these observations is that when E-Cadherin^{low} cell lines do not express E-cadherin because of the induction of EMT, they will not differentiate, and it is related to their EMT induction, not to their E-Cadherin loss. It supports the evidence that the lack of differentiation ability is an important attribute of EMT like cells (Romero-Pérez et al., 2013).

E-Cadherin knock down in a E-cadherin^{high} /lumen forming cell line (LS180) or knock in in the E-Cadherin^{low} /non-lumen cell line (LS174T) did not change their ability to differentiate in the matrigel, which supports the lack of a causal role for E-Cadherin in differentiation.

In order to find other factors that may contribute to the control of differentiation, seven E-Cadherin^{low/-ve} mesenchymal-like cell lines were compared with seven E-Cadherin^{high} epithelial cell lines (Table 3.2).

Table 3.2: List of cell lines selected for the analysis of E-Cadherin

'Mesenchymal-like' E-Cadherin low/ Vimentin high/non-lumen	Epithelial E-Cadherin high/ Vimentin low/ lumen
COLO320HSR	SNUC1

CC20	HCA46
OXCO1	HT55
COLO320DM	CACO2
RKO	C106
SW620	C80
SW480	SW1222

The microarray analysis showed a number of genes to be significantly higher in the mesenchymal like cell lines (Table 3.3), including known markers of EMT such as VIM and ZEB1.

Table 3.3 : List of genes that are upregulated in the mesenchymal like cell lines (E-Cadherin^{neg})

Gene symbol	Gene Name	Bonferroni corrected p value	Fold increase in mesenchymal cell lines
<i>VIM</i>	Vimentin	6.9E-07	149
<i>ZEB1</i>	zinc finger E-box binding homeobox 1	7.0E-08	28
<i>LIX1L</i>	Lix1 homolog (mouse)-like	4.4E-08	179
<i>TTL</i>	tubulin tyrosine ligase	2.8E-06	4
<i>VMA21</i>	VMA21 vacuolar H ⁺ -ATPase homolog (S. cerevisiae)	5.7E-06	3
<i>MAPK12</i>	Mitogen-activated protein kinase 12	1.8E-06	2
<i>CCDC111</i>	Coiled-coil domain containing 111	1.8E-06	2
<i>CAPRIN2</i>	Caprin family member 2	3.3E-05	2

This list of genes was then compared in LS174T and LS180 in order to identify any key EMT associated genes that might explain the different capacities of LS174T and LS180 to form lumens in matrigel. Among all these genes, only *TTL* (Tubulin Tyrosine Ligase) was differentially expressed in LS174T and LS180 where LS174T expressed almost 2 fold higher *TTL* than LS180. *TTL* is a ligase that tyrosinates tubulin, which is required for the morphological differentiation from non-polarized cells into an epithelial monolayer in 2D culture (Zink et al., 2013). The expression of *TTL* is almost 4 fold higher (significant after Bonferroni correction) in the mesenchymal like cell lines. Using

a breast cancer cell line, (Whipple et al., 2010) reported *TTL* down regulation with the exogenous expression of TWIST and SNAI1. However, in the CRC cell line panel, SNAI1 and TWIST1 expression was not conclusive (from microarray data) and so it was not possible to check such a relationship. It might be worth silencing this gene in LS174T to see if that makes any change in its ability to differentiate in matrigel. (mRNA expression profile of *TTL* in CRC cell line panel has been provided in the appendix: Ax.9).

3.7 Chapter Discussion

Down regulation of E-Cadherin is considered a key marker of EMT (Guarino, Rubino, & Ballabio, 2007). In the literature, there are reports on E-Cadherin gene transcriptional repression (Batlle et al. 2000; Cano et al. 2000; Hajra et al. 2002), promoter methylation (Graff et al. 1995; Kanai et al. 1997; Saito et al. 1998) and protein phosphorylation and degradation (Zhou et al. 2004; Bachelder et al. 2005; Lester et al. 2007) during the EMT process. It is, however, still not clear whether reduced levels of E-Cadherin expression are just a consequence of this process or whether they play a causal role during the EMT.

The paired cell lines, LS174T and LS180 gave an opportunity to discern the role of E-Cadherin in the differentiation of cells. It has been mentioned before that LS174T has two mutations in the E-Cadherin gene, whereas LS180 is wild type (Section 3.3.2). Because of the mutations, LS174T does not express E-Cadherin. It is noteworthy that the E-Cadherin expressing LS180 forms lumens in matrigel, while the E-Cadherin negative LS174T does not (Section 3.3.1). This creates the scope to study the effect of the loss of E-Cadherin due to mutation (without any influence of the transcription factors ZEB1, ZEB2, TWIST1, TWIST2, SNAI1, SNAI2 etc.) and uncouple the exclusive role of E-Cadherin in the process of EMT when other regulators of the process remain constant.

The current experiments with LS180, a cell line that forms well defined lumens in matrigel (which was used as a phenotype of differentiation), found that E-Cadherin knockdown could not abolish its ability to form lumens. Both the siRNA^{E-Cadherin} transfected- and untransfected- LS180 showed similar morphology and differentiation in matrigel. This implies that the loss of E-Cadherin may not *cause* the de-differentiation of cells going through the EMT process.

The conclusion about the ‘causality’ of E-Cadherin was corroborated with an opposite experiment with LS174T, an E-Cadherin negative counterpart of LS180. In this case, gene knock-in experiment was done on LS174, where plasmid that contains cDNA of E-Cadherin was transfected in the cells. It showed that in spite of this exogenous expression of E-Cadherin, LS174T could not differentiate like LS180 when cultured in matrigel. Thus, it was confirmed from both directions, by both knockdown and knock-in of E-Cadherin, that E-Cadherin cannot be the cause of differentiation of cells.

This conclusion can be further supported by comparing the capacity for lumen formation with E-Cadherin expression levels across several cell lines. When lumen forming versus non-lumen forming cell lines were compared in a panel of 83 CRC cell lines (Table 3.4), it was found that E-Cadherin expression was high in both lumen and non-lumen groups, suggesting a lack of association between E-Cadherin expression and differentiation.

Table 3.4: Chi-square test for E-Cadherin expression association with lumen formation ability of CRC cell lines. E-cadherin mRNA levels in 83 CRC cell lines from Affymetrix U133+2 assays were divided into low and high, and cell lines were grouped according to colony morphology as described in section 1.5.1.1.

E-Cadherin status	Lumen forming cell lines	Non-lumen cell lines	Total
E-Cad Low	0	11	11
E-Cad High	20	52	72
Total	20	63	83

The results of a chi-square analysis of the data in Table 3.2 revealed a chi-square statistic of 2.650, with 1 degrees of freedom. The two-tailed p value equals 0.1035, which indicates that the association between E-Cadherin expression and lumen formation (as a measure of differentiation ability) is not significant at $p < 0.05$ level.

It is important to note that during this work, as mRNA and protein levels of E-Cadherin were measured on day 4 and not measured at the end time-point (day 7) of the lumen formation assay, it is possible that recovery of expression levels between days 4 and 7 was sufficient to obscure any phenotypic effect of knockdown at the earlier time points. One focus for future studies can be to derive stable transfectants of E-Cadherin

knockdown or knock-in to study the effects on differentiation at longer time points of growth in matrigel.

These preliminary results, however, show that during EMT, not the loss of E-Cadherin but the transcriptional regulators and other factors are the likely causes of the suggested EMT-induced changes in cells, as when mutations cause the loss of E-Cadherin, no impact on the differentiation or other EMT related phenotypes was observed. Therefore, low E-Cadherin on its own does not seem to lead to whatever is usually described as EMT.

During the culture and maintenance of LS174T for the above E-Cadherin study, an important feature of LS174T was noticed: its morphological heterogeneity. Unlike LS180, which has very homogeneous appearance in 2D culture, LS174T grew in two layers: an attached layer of mostly flat cells and a more loosely connected population of grape-like bunches on top of the attached layer (Fig 3.10). If shaken, the top layers were easily detached and floated in the culture medium. These cells were confirmed by trypan blue to have high viability and looked healthy under microscope.

This raised the question of whether the morphological heterogeneity observed in the LS174T cell line could suggest the co-existence of subclones within LS174T and prompted much of the work presented in the remainder of this thesis.

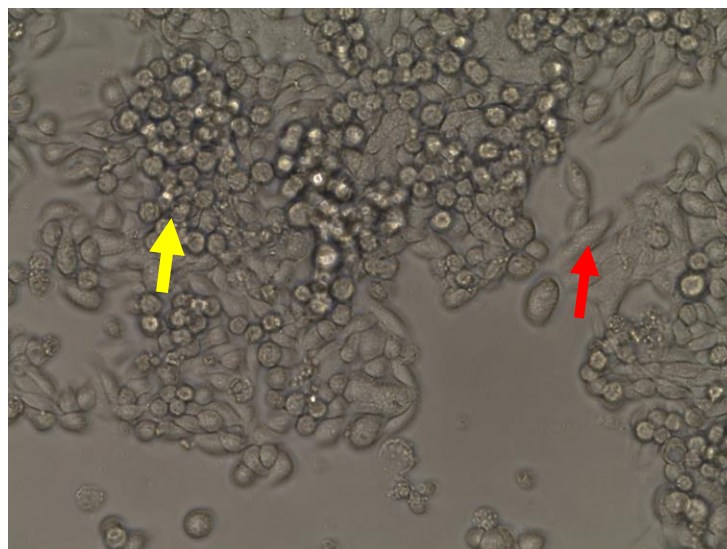


Fig 3.10: Heterogeneous appearance of LS174T in culture flask. Some cells were firmly attached (red arrow) and others are loosely attached (yellow arrow).

Chapter 4

CD24 Based Clonal Heterogeneity in LS174T

Chapter 4

4.1 Chapter introduction

One of the major criticisms against using cell lines as pre-clinical models relies on the notion that they do not retain the heterogeneity that is present in the primary cancer. (Auman & McLeod, 2010). Survival and stable maintenance of genetically distinct stable clones in a single cell line is considered rare, if not completely unprecedented (Baguley & Leung, 2011; Hastings & Franks, 1983). The traditional cell-autonomous model of tumour initiation and progression proposes continuous selection of dominant clones over their less fit counterparts (Cairns, 1975; Loeb, 1991), where cells with advantageous genetic or epigenetic variations compete with other cells. Contrary to this model, a non-autonomous model of interactions proposes that cancer cells may gain some changes that not only benefit themselves but also affect other cells in the tumour either directly or as a side effect (Tomlinson & Bodmer, 1997). According to this model, if an interaction becomes beneficial for both cell populations, they would co-exist in a stable manner and thus, clonal heterogeneity will arise and sustain. This heterogeneity may be accompanied by observable morphological differences.

On this ground, during the study of the effects of E-Cadherin on the differentiation ability of the LS174T cell line (Chapter 3), the observation of its heterogeneous appearance (Fig 3.8) was deemed significant, especially considering the implications of the above-mentioned non-autonomous model proposed by Tomlinson and Bodmer (1997). However, to support this model, the observed morphological heterogeneity needs to be associated with the co-existence of stable clones in LS174T. This requires studying

the nature of the heterogeneity, i.e., clonality, stability of clones, and genetic and epigenetic variations in the subclones of LS174T.

In this chapter, the presence of stably maintained subclones in the LS174T cell line has been explored. Differences between the subclones with respect to gene expression, genetic sequence and responses to chemotherapeutic drugs have been investigated.

4.2 Morphological heterogeneity of LS174T is separable as stable clones

Firstly, a series of sub-cloning experiments were conducted to pursue the preliminary indication of stable clonal heterogeneity in the LS174T cell line. In order to obtain distinct clones from LS174T, single cells were sorted in a 96 well plate by FACS (BD Fusion). After one week, the plate was checked for the presence of single colonies. LS174T is a highly clonogenic cell line. On average, it shows ~85% clonogenicity in 2D culture and so colonies appeared in most of the wells of the 96 well plate. Any well that contained more than one colony was marked to be excluded from further analyses in order to increase the probability of establishing clones that were derived from single cells.

At week 4, twelve single colonies were randomly picked and transferred in a 12 well plate for further analyses. Bright field photographs were taken for each clone in respective wells using Zeiss Z1 Inverted Fluorescence Microscope (Fig 4.1)

Based on the morphology, those clones could be roughly grouped into 2 sub-categories: i) Clone 1 like morphology (which includes Clones 1, 3, and 5-12), where cells are mostly flat and firmly attached to the surface; and ii) Clone 2 like morphology (which includes Clones 2 and 4), where cells are mostly globular and appear loosely attached to the surface.

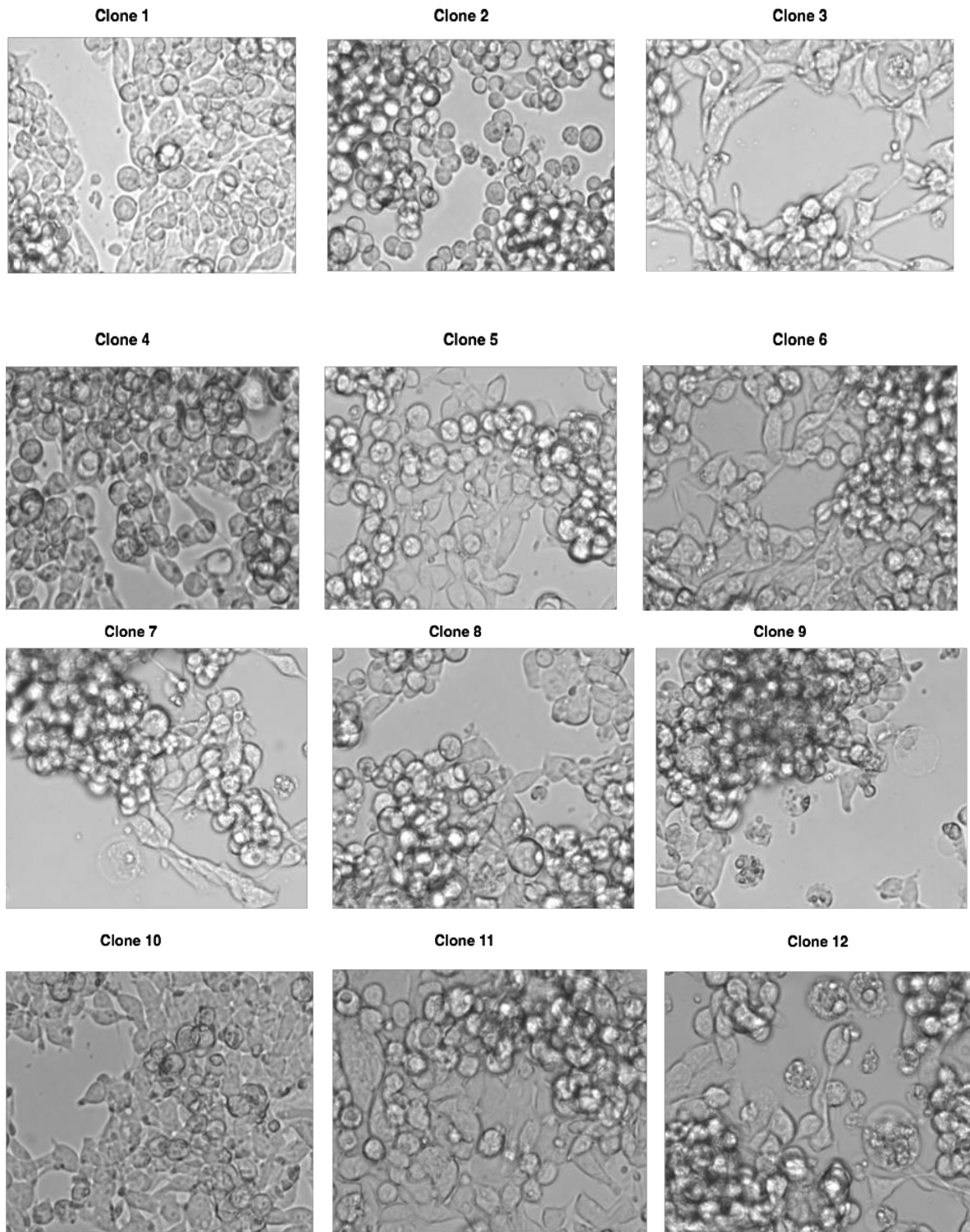


Fig 4.1: Bright field images (20×) of 12 randomly selected clones from LS174T. Out of these 12 clones, two (Clone 1 and Clone 2) showed most distinctive morphology (flat versus round) and were therefore selected for further analysis.

Clone 1 and Clone 2 were selected for further analyses as they displayed the most distinct morphologies. Clone 1 cells were mostly flat in early passages and required longer exposure to trypsin during sub-culturing. On the contrary, Clone 2 cells showed mostly globular morphology and their colonies had the appearance of the grape-bunches and were easier to trypsinise. Retrospectively, both of these morphological categories closely matched the two subpopulations in the parental LS174T (Fig. 4.2) that characterised the heterogeneity initially observed within that cell line (Fig 3.8).

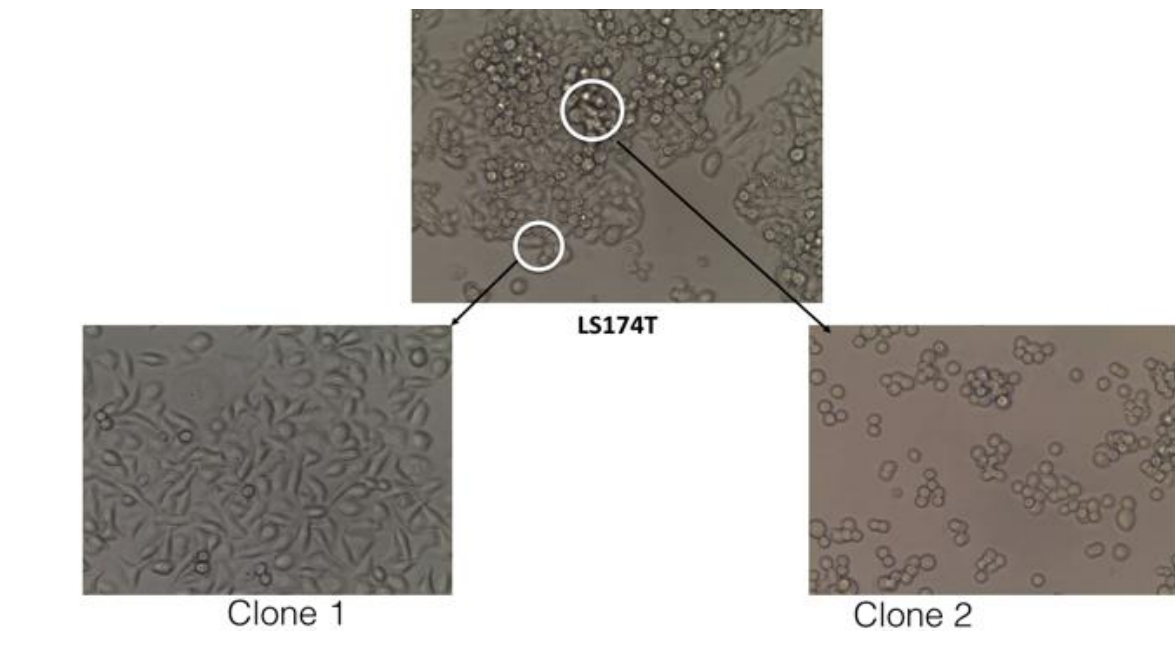


Fig 4.2: Morphological differences between Clone 1 and Clone 2. Bright field images (5×) of Clone 1 (left) and Clone 2 (right) show distinct morphologies, which can be retrospectively observed in parental LS174T (top).

4.3 Confirmation of the cellular origins of Clone 1 and Clone 2 in LS174T

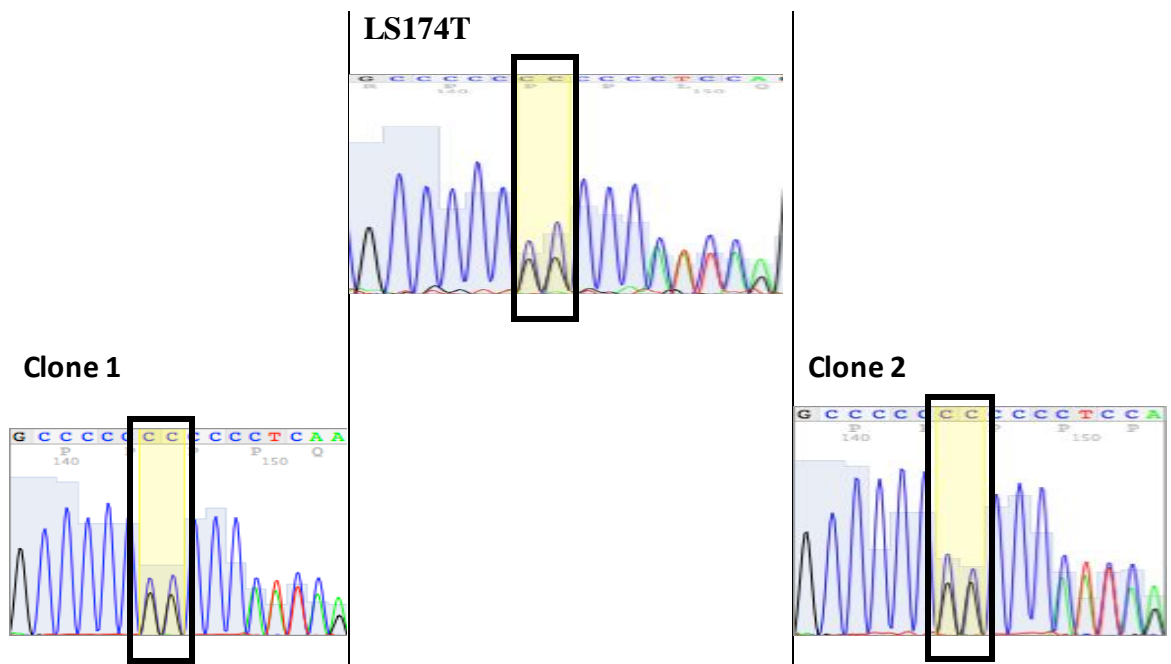
Before any further study, analyses were done to confirmed that Clone 1 and Clone 2 were indeed isolated from LS174T, and not from a possible mixture of a contaminating cell line. The unique E-Cadherin mutations known to be present in the LS174T, but not in any other of the colorectal cancer cell lines for which the Bodmer laboratory holds whole exome sequence data, was used for this confirmation. As described in section 3.3.2, LS174T contains two mutations in the E-Cadherin gene: i) Exon 3 frame shift

mutation and ii) Intron 8 splice donor mutation. Thus, if both of these mutations can be observed in Clone 1 and Clone 2, this would provide conclusive confirmation that LS174T was indeed the source of the clones.

In order to perform this validation analysis, primers used for checking two E-Cadherin mutations in LS174T were used to amplify the same regions of Clone 1 and Clone 2. After PCR amplification and purification, the amplicons were sequenced, which showed that both Clone 1 and Clone 2 contained exon 3 frame shift and intron 8 splice donor mutations in the same positions. These results confirmed that the clones came from LS174T rather than from any contaminating cell line (Fig 4.3 A, B).

Further validation of the origins of Clone 1 and Clone 2 was provided from analysis of 34 SNPs, revealing identical SNP profiles for LS174T (see appendix: Ax.2)

A. Exon 3 frameshift deletion mutation: GCCCCCG>GCCCCCG



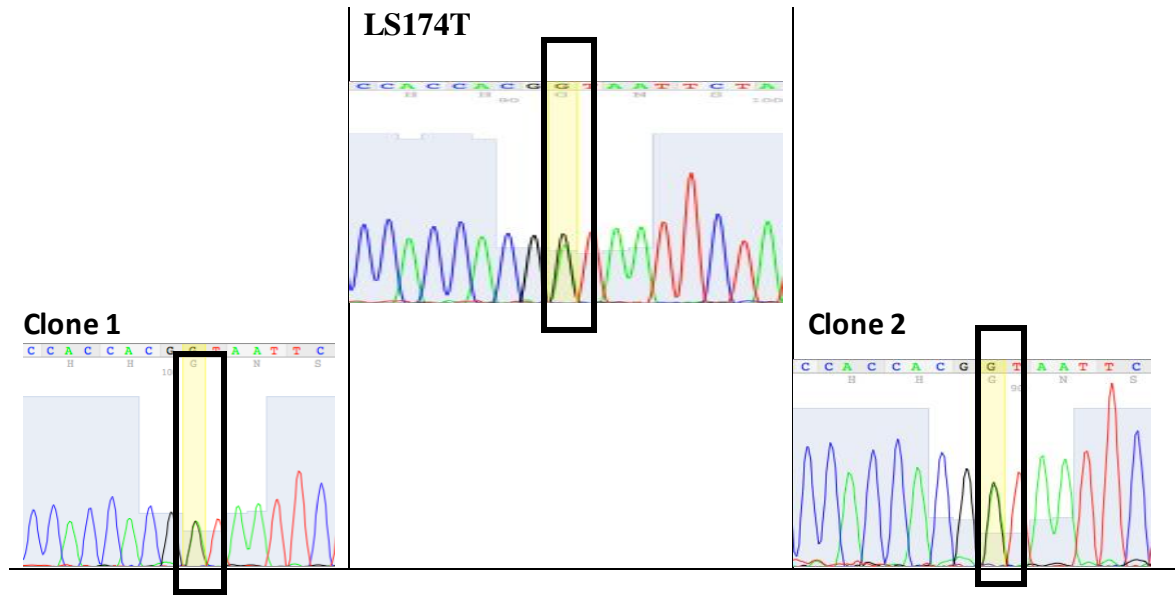
B. Intron 8 splice donor site mutation: ACGGTAA > ACGATAA

Fig 4.3: Confirmation of the origins of clones by Sanger sequencing of the E-Cadherin mutations. A. Heterozygous exon 3 frameshift deletion mutation is represented as jumbled peaks to the right of the mutation position (highlighted) in LS174T, Clone 1, and Clone 2. B. Heterozygous Intron 8 splice donor mutation (G>A) is represented by the peaks for both G (black) and A (green) in LS174T, Clone 1, and Clone 2 at the same position (in the black boxes).

4.4 Differences between Clone 1 and Clone 2 of LS174T

After observing the morphological differences between Clone 1 and Clone 2 of LS174T (Fig 4.2), more differences were revealed between them. Experiments were conducted to uncover differences regarding their growth rates, clonogenicity and differentiation in matrigel, gene expression, mutation and response to certain chemotherapeutic drugs.

4.4.1 Clones have different growth rates

To reveal the growth rates of Clone 1 and Clone 2, their doubling times were calculated across different passages from cells stored in liquid nitrogen. To get the growth curve, an SRB assay was carried out after culturing cells over a period of 7 days. Six wells were seeded with 3500 cells from each passage of the Clone 1, Clone 2 and parental LS174T in a 96-well plate. The number of cells in each well from the seven plates were

estimated by SRB assay and growth curves for each passage of each cell lines were obtained by analysing data with Prism software.

From the growth curve the following doubling times were obtained for Clone 1, Clone 2 and their parent cell line LS174T. (Here 'p' indicates passage numbers for each cell line/clone)

Cell lines	Clone 1				Clone 2				LS174T			
	p12	p14	p33	p50	p10	p15	p33	p43	p74	p107	p123	p182
Doubling time (DT) (in hours)	23	18	17	20	29	37	24	23	21	18	23	26
Mean DT(+/- SD)	19.5 (+/-2.6)				28 (+/-5.94)				22 (+/- 3.36)			

A one-tailed t-test shows that the mean doubling time of Clone 1 (19.5 hours) is significantly ($p= 0.0017$) lower than the mean doubling time of Clone 2 (28 hours). A two-tailed t- test also shows the significant difference ($P= 0.0035$). This gives conclusive evidence that Clone 2 grows slower than Clone 1.

4.4.2 Clonogenicity and colony morphology differences between Clone 1 and Clone 2 of LS174T

In order to explore whether the doubling time difference between the clones affects clonogenicity or other growth properties in 3D culture, 500 cells from each clones were seeded in matrigel in triplicate wells. After 3 weeks, colony numbers were counted in each well and colony morphology was observed. Differentiation status of the colonies were measured by observing lumen structures in colonies by F-actin staining with TRITC-Phalloidin.

The counts of the colonies from triplicate plates showed that Clone 1 had higher clonogenicity than Clone 2 (Fig 4.4).

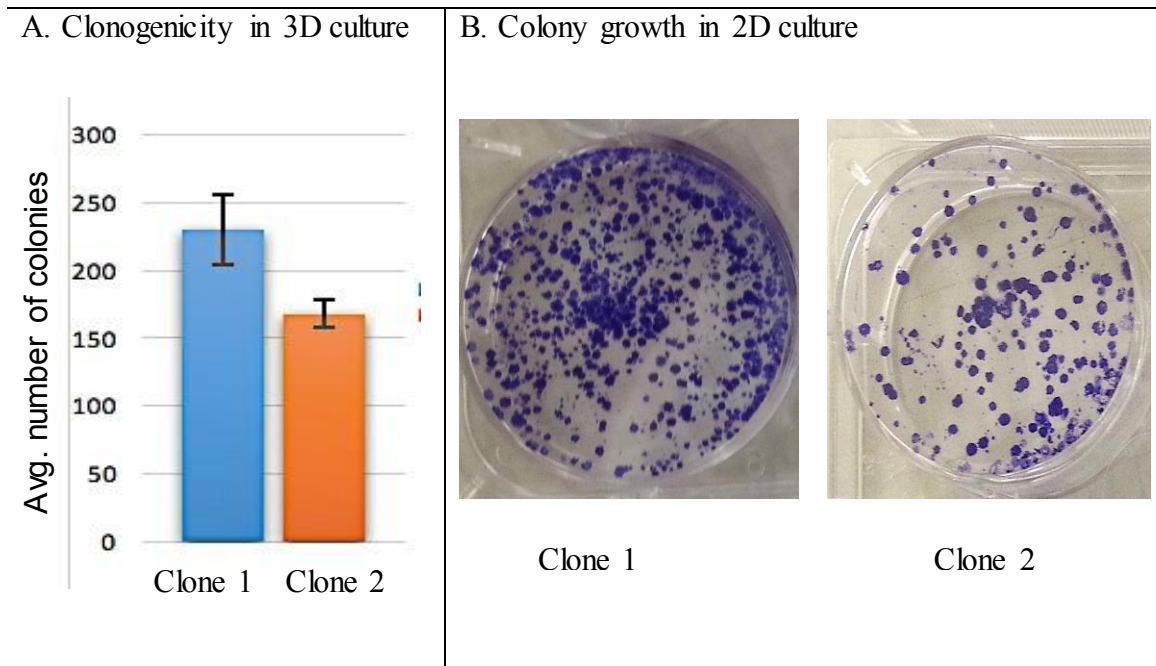


Fig 4.4: Clonogenicity difference between Clone 1 and Clone 2. A. In matrigel, Clone 1 shows higher clonogenicity than Clone 2. Error bars indicate standard deviation calculated for colony numbers in triplicate wells. B. Colony growth of Clone 1 and Clone 2 in 2D culture condition. Clone 1 shows higher growth than Clone 2.

Colonies of Clone 1 were mostly compact and of relatively smaller size, whereas Clone 2 formed bigger colonies with diffused borders (Fig 4.5, Table 4.1).

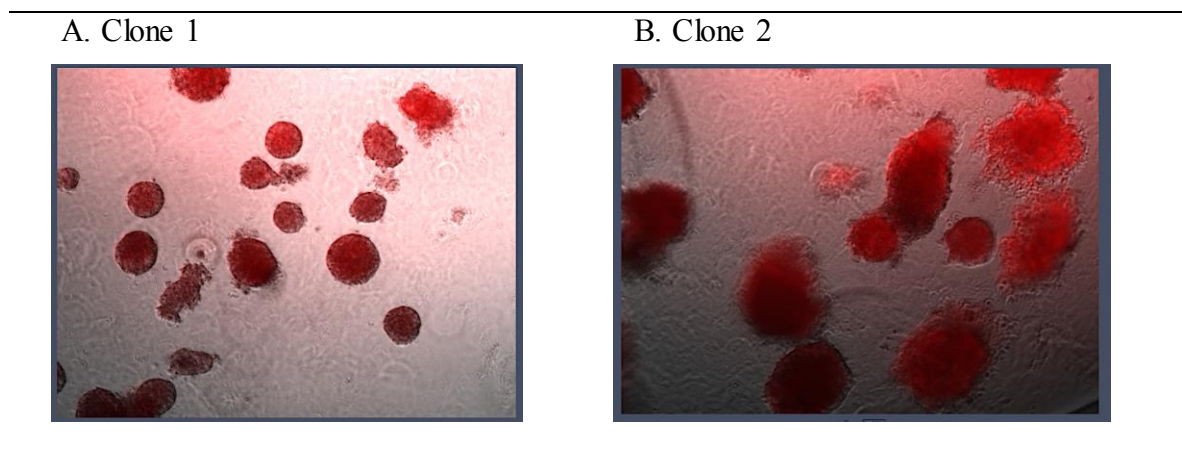


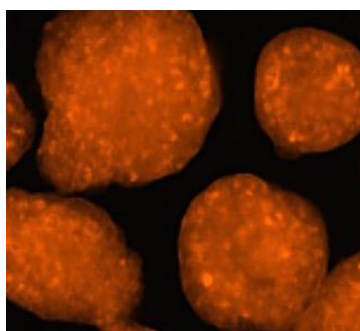
Fig 4.5: Colony number and morphology of Clone 1 (left) and Clone 2 (right). Images of colonies (5x) in matrigel stained with Phalloidin (Red) at week 3. Clone 1 formed more compact colonies than Clone 2.

Table 4.1: Clone 1 and Clone 2 showed differences regarding colony size and colony morphology in matrigel.

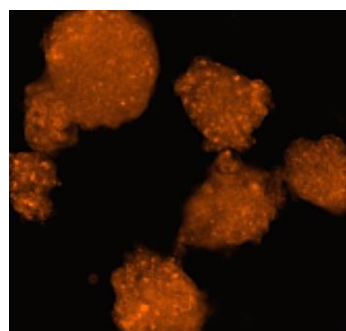
Characteristics	Clone 1	Clone 2
Size of colonies	Smaller	Bigger
Colony morphology	Mostly compact with sharp edge	Mostly Diffused

Neither Clone 1 nor Clone 2 showed obvious signs of well-defined lumen formation in matrigel (as observed in LS180 or other differentiating cell lines) (Fig 4.6).

A. Clone 1



B. Clone 2



C. LS180

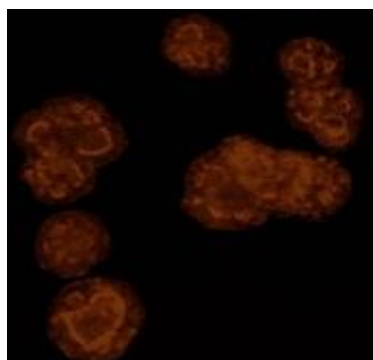


Fig 4.6: Colonies of Clone 1 and Clone 2 and LS180 in matrigel. Fluorescent microscopy image (5x) of F-actin staining with TRITC-Phalloidin (red) shows that neither (A) Clone 1 nor (B) Clone 2 has lumen forming characteristics like differentiating cell lines (C) LS180.

However, analysis at higher magnification revealed the presence of ~2-5% of colonies from both Clone 1 and Clone 2 that had multiple focal concentrations of F-actin staining that were termed “micro-lumens” which are also observed in parental LS174T.

4.5 Clonal heterogeneity in LS174T can be linked to differential expression of CD24

Yeung *et al.* had shown that CD44^{high} and CD24^{high} sub-populations of cells isolated from cell lines by FACS were more clonogenic *in vitro* and made many more tumours per cell in mice than CD44^{low} and CD24^{low} cells (Fig 4.7). These observations confirmed that CD44 and CD24 markers may be used for the enrichment of stem cells within individual colorectal cancer cell lines (Yeung, Gandhi, Wilding, Muschel, & Bodmer, 2010).

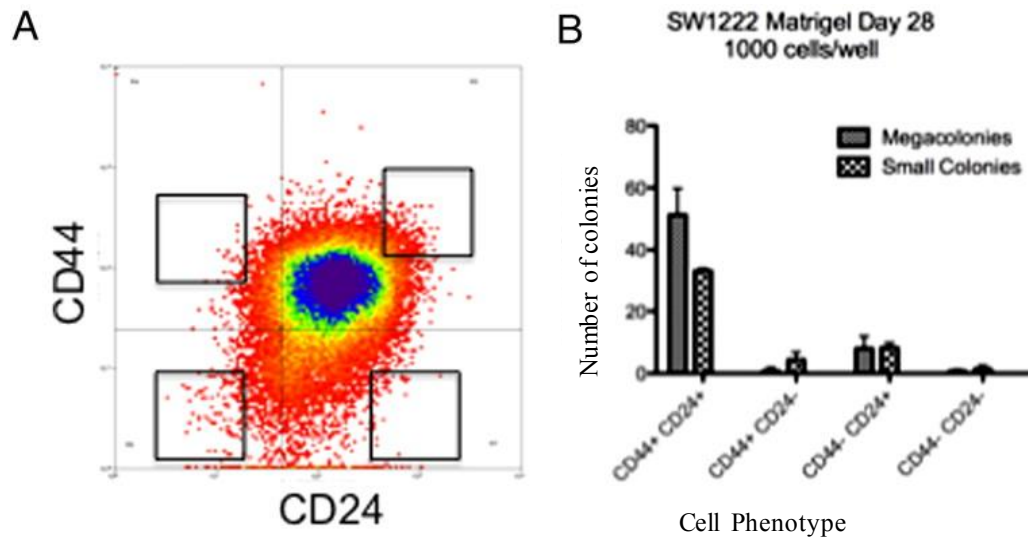


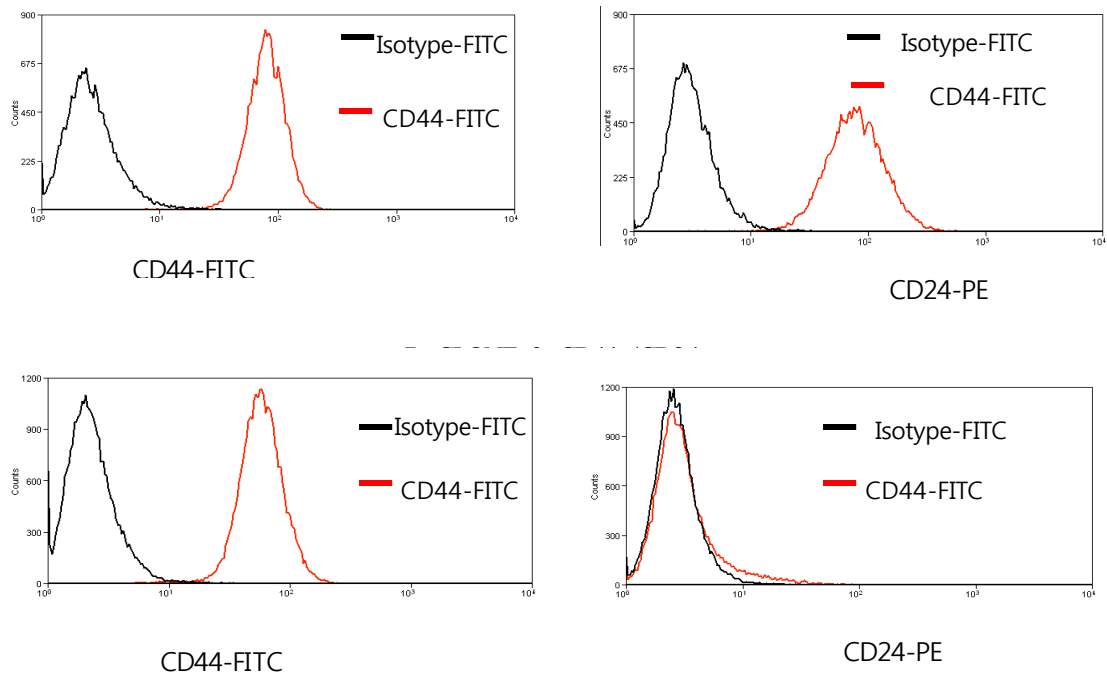
Fig 4.7: CD44/CD24 levels are related to clonogenicity and colony morphology in matrigel (Retrieved from Yeung *et al.*). A. Cells sorted based on their CD44/CD24 expression level. B. Colony numbers in matrigel for cells with all four combinations of CD44/CD24 expression. Here, mega colonies refer to those that make lumens, while small colonies refer to non-lumen ones.

Based on this prior knowledge, further investigations were conducted on the observed differential clonogenicity and colony morphology in Clone 1 and Clone 2 in matrigel (described in section 4.4.2) to check for their expression of CD44 and CD24 in the clones.

The investigations started by assessing the surface expression of CD24 and CD44 in Clone 1 and Clone 2 using Flow Cytometry. CD44-FITC and CD24-PE antibodies were used in these experiments.

When analysed, both clones showed similar levels of CD44 expression. Strikingly, however, CD24 expression was quite different between them. Clone 1 was clearly CD24⁺ whereas Clone 2 was CD24⁻. Most importantly, parental LS174T showed the presence of both CD24⁺ and CD24⁻ sub-populations. Almost 20% cells were CD24⁻ in LS174T. However, LS174T had no significant heterogeneity regarding CD44 (Fig 4.8).

A. CLONE 1: CD44⁺/CD24⁺



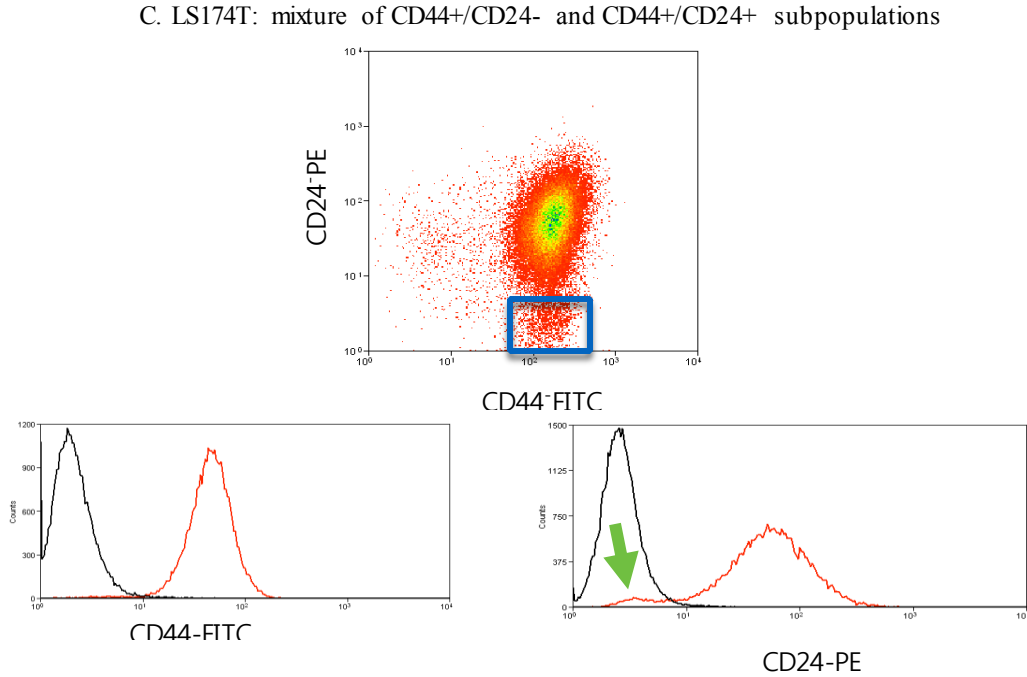


Fig 4.8: Assessment of CD44 and CD24 expression in clones using flow cytometry. A. Clone 1 expresses both CD24 and CD44. B. Clone 2 does not express CD24 but expresses CD44 in similar levels to Clone 1. C. LS174T, the parent of Clone 1 and Clone 2, expresses CD44, but it has a significant proportion of CD24⁻ sub-population (shown with green arrow), which is consistent with the coexistence of Clone 1^{CD24⁺} and Clone 2^{CD24⁻} in LS174T. In all cases, black histograms indicate isotype control.

Differential expression of CD24 between Clone 1 and Clone 2 was further confirmed using live cell imaging with Fluorescent microscope. It showed abundant expression of CD24 in Clone 1 and complete absence in Clone 2 (Fig 4.9).

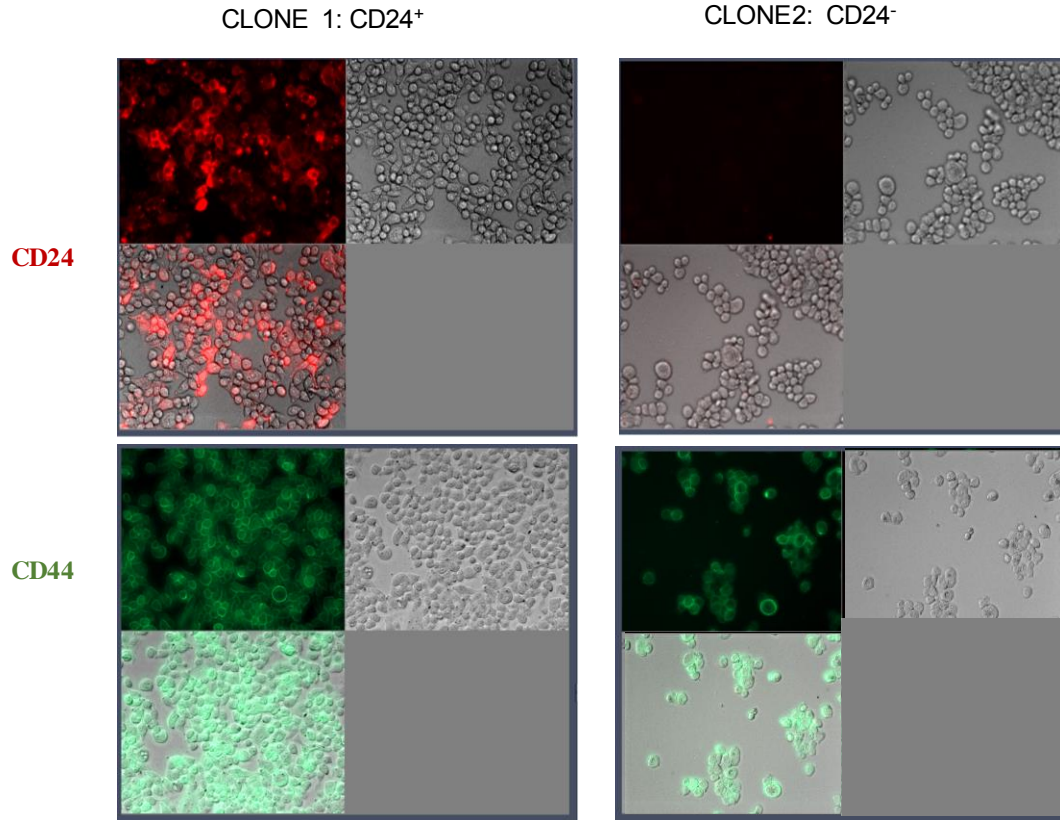


Fig 4.9: Live cell imaging of Clone 1 and Clone 2 for CD44 and CD24. A. Clone 1 stains for both CD44 (green) and CD24 (red). B. Clone 2 does not stain for CD24 (absence of red) though it stains quite well for CD44 (green).

Further, qRT-PCR was conducted to determine the level of CD24 message; the results confirmed that Clone 1 and Clone 2 differed in the levels of CD24 mRNA. After normalising against the housekeeping gene UBC, Clone 2 had a $\Delta\Delta CT$ (cycle time) of ~ 8 , whereas Clone 1 had ~ 4 , which indicated ($2^4 =$) 16 times higher mRNA levels in Clone 1 than in Clone 2 (Fig 4.10).

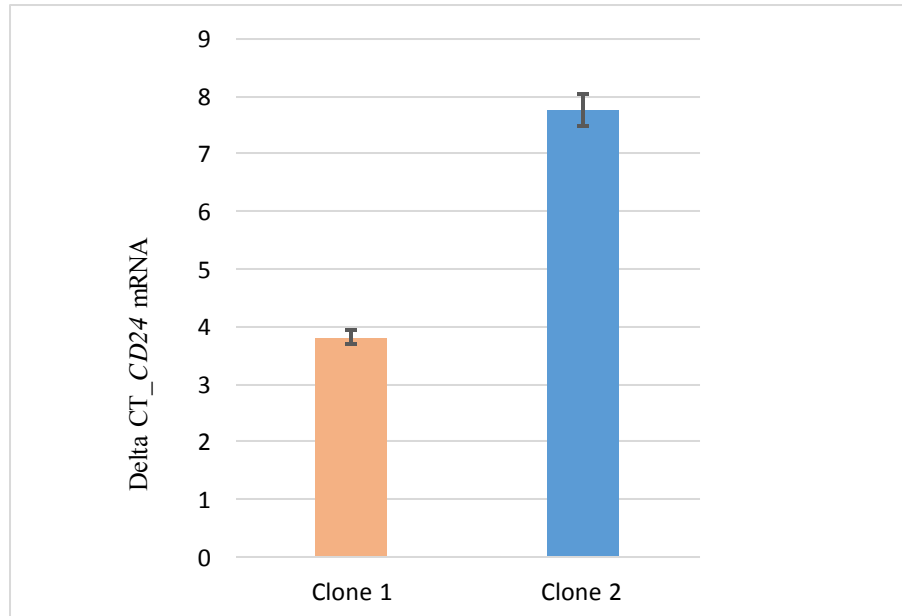


Fig 4.10: Delta CT values for *CD24* in Clone 1 and Clone 2. *CD24* mRNA level in Clone 1 is $2^{(8-4)}=16$ times higher compared to Clone 2. The error bars indicate the standard deviations among triplicate values of cycle time (CT) in each clone.

4.6 Clone 1 and Clone 2 did not differ in other stem cell markers

After confirming the difference of *CD24* expression between Clone 1 and Clone 2, the clones were checked for expression levels of other stem cell markers that are important in colorectal cancer.

A panel of 11 markers (*CD44*, *CD24*, *LGR5*, *EphB2R*, *CEACAM6*, *CD10*, *CD58*, *CD26*, *CD55*, *CD166* and *c-Kit*) reported in the literature for their association with stemness and/or differentiation in CRC were analysed using flow cytometry (see appendix: Ax.3). When the current data were analysed for these markers, 8 out of the 11 markers showed no difference between Clones 1 and 2. *CEACAM6* (*CD66c*) and *CD26* showed very little difference. However, in concurrence with the previous observation, *CD24* showed significantly higher expression in Clone 1 than in Clone 2.

4.7 Proof of stability: When subcloned, Clone 1 and 2 reproduced their CD24 expression profiles

In order to see if the clones were stable with respect to their CD24 expression, they were sub-cloned by single cell sorting of the bulk population using FACS. Clone 1 and Clone 2 gave rise to 11 and 7 subclones, respectively.

All of the 11 subclones of Clone 1 were CD24⁺ and all of the 7 subclones of Clone 2 were CD24^{-/low} when tested for CD24 expression at passage 5. Fig 4.11 shows three representative subclones from Clone 1 and Clone 2. These results implied that the clones were stable enough to reproduce themselves.

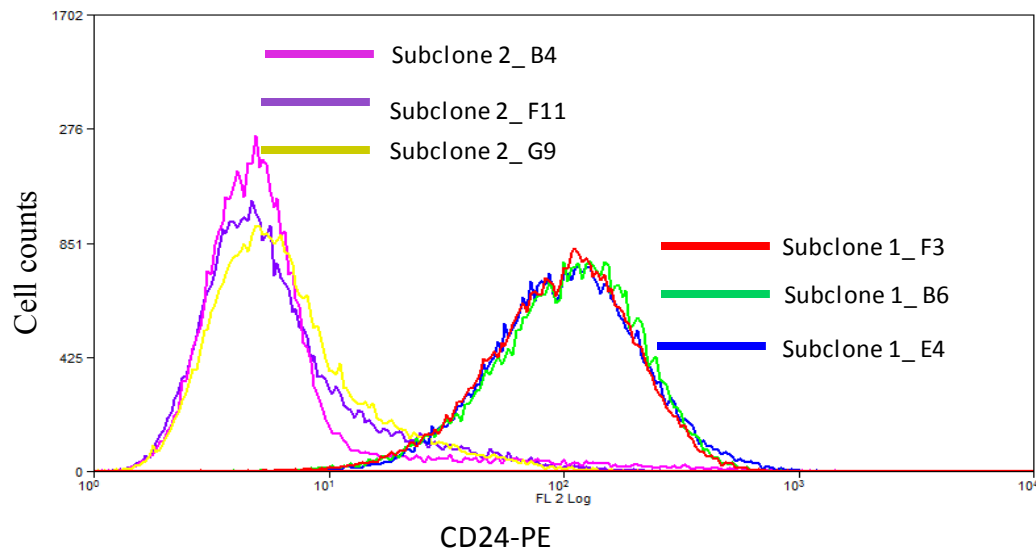


Fig 4.11: Reproducibility of Clones. Subclones generated from Clone 1 were CD24⁺, whereas, subclones from Clone 2 were CD24⁻ indicating the reproducibility and stability of Clone 1 and Clone 2.

4.8. Microarray analysis revealed expression differences of key genes between the clones of LS174T

Establishing that Clone 1 and Clone 2 differentially express CD24 and confirming that the two clones are stable created an opportunity to use these clones as a model to study the regulation and function of CD24, and to study the nature of the clonality in LS174T.

In order to get a more comprehensive understanding about the expression difference between Clone 1 and Clone 2, their global gene expression was revealed using Affymatrix U133+2 Microarray platform. After necessary normalisation, expression data were analysed using Partek Genome Suit 6.0 as has been described in chapter 2, section 2.6.

Microarray analysis between Clone 1 and Clone 2 showed that they differed in the expression of more than 100 genes. CD24 expression showed 6 times higher expression in Clone 1 compared to Clone 2 in this analysis. As CD24 difference has already been confirmed using flow cytometry and immunofluorescence, a 6-fold difference in expression was considered as the cut-off value for significant difference, which gave a list of 36 genes (see appendix: Ax.4) expressed in higher amounts in Clone 1 than in Clone 2. Table 4.2 shows a list of some selected genes which have been implicated in important mechanisms such as drug resistance and EMT etc.

Table 4.2: Microarray analysis shows differential expression of several important genes between Clone 1 and Clone 2 of LS174T. Differences between Clone 1 and Clone 2 in the expression of a selected a group of genes are shown here. The left two columns show genes that are expressed more highly in Clone 1, whereas the right two columns show genes that are more highly expressed in Clone 2.

Gene name	Ratio of expression Clone 1:Clone 2	Gene name	Ratio of expression Clone 1:Clone 2
<i>LAPTM4B</i>	181:1	<i>ARMCX6</i>	1: 48
<i>CXCR4</i>	39:1	<i>PAGE1</i>	1: 43
<i>PPIC</i>	32:1	<i>CSA</i>	1: 37
<i>ELL2</i>	20:1	<i>HBE1</i>	1: 26
<i>NEBL</i>	17:1	<i>ERVH-3</i>	1: 20
<i>IL8</i>	13:1	<i>PCP4</i>	1: 17
<i>TGFBI</i>	6:1	<i>HIST1H2AB</i>	1: 17
<i>FOXQ1</i>	6:1	<i>DMXL2</i>	1: 13
<i>SI00A4</i>	6:1	<i>SHOX2</i>	1: 13
		<i>SLC10A2</i>	1: 11
		<i>TUSC3</i>	1: 10

The most differentially expressed gene between Clone 1 and Clone 2 was Lysosomal Protein Transmembrane 4 Beta (*LAPTM4B*) which has been associated with drug resistance and poor prognosis in cancer (Y. Li, Iglehart, Richardson, & Wang, 2012; Meng, Chen, Song, & Lou, 2015).

CXCR4 has been shown to influence the progression of Epithelial to Mesenchymal Transition (EMT) by interacting with *CXCL12* (SDF-1) (Hu et al., 2014). *FOXQ1* has also been implicated with EMT promotion (Qiao et al., 2011) and S100A4 is a strong marker of EMT (Lo et al., 2011; Schneider, Hansen, & Sheikh, 2008). As can be seen in the microarray analysis results reported in Table 4.2, these genes were expressed at significantly higher levels in Clone 1. In Clone 2, *SHOX2* and *TUSC3* were much highly expressed than in Clone 1. Both of these genes have been reported to be suppressed by promoter methylation in several cancers (Duppel, Woenckhaus, Schulz, Merk, & Dietmaier, 2016; Kneip et al., 2011; Pils et al., 2013).

From microarray analysis, the 43 genes (Appendix 4) which are expressed >6 fold higher in Clone 1 to Clone 2 were analysed using PANTHER tool (<http://www.pantherdb.org/>) for their Gene Ontology (GO) and pathway information.

Table 4.3, 4.4 and 4.5 respectively contain information about the number of genes which are associated with GO terms Molecular function, Biological process and Cellular component. It shows 34 hits across 6 broadly categorised molecular functions (Table 4.3), whereas 65 hits were found across 12 biological process (Table 4.4) and 27 across 6 cellular components (Table 4.5). Regarding pathway involvement, 12 hits were found (Table 4.6).

Table 4.3 Molecular Function

GO terms	Number of hits
translation regulator activity (GO:0045182)	1
binding (GO:0005488)	12
receptor activity (GO:0004872)	3
structural molecule activity (GO:0005198)	6
signal transducer activity (GO:0004871)	1
catalytic activity (GO:0003824)	11

Table 4.4 Biological Process

Go terms	Number of hits
cellular component organization or biogenesis (GO:0071840)	5
cellular process (GO:0009987)	17
localization (GO:0051179)	2
reproduction (GO:0000003)	1
biological regulation (GO:0065007)	6
response to stimulus (GO:0050896)	6
developmental process (GO:0032502)	6
multicellular organismal process (GO:0032501)	3
biological adhesion (GO:0022610)	3
locomotion (GO:0040011)	1
metabolic process (GO:0008152)	13
immune system process (GO:0002376)	2

Table 4.5 Cellular component

Go terms	Number of hits
membrane (GO:0016020)	2
macromolecular complex (GO:0032991)	1
extracellular matrix (GO:0031012)	2
cell part (GO:0044464)	12
organelle (GO:0043226)	3
extracellular region (GO:0005576)	7

Table 4.6 Pathway

Pathway terms	Number of hits
Pyrimidine Metabolism (P02771)	1
Purine metabolism (P02769)	1
Blood coagulation (P00011)	1
Axon guidance mediated by Slit/Robo (P00008)	1
Interleukin signaling pathway (P00036)	1
Integrin signaling pathway (P00034)	2
Inflammation mediated by chemokine and cytokine signaling pathway (P00031)	2
p38 MAPK pathway (P05918)	1
CCKR signaling map (P06959)	2
Oxidative stress response (P00046)	1

PANTHER analysis with 39 genes (Appendix 5) which are ≥ 4 fold higher in Clone 2 than Clone 1 gave 30 hits in molecular functions (Table 4.7), 52 in biological process, 16 in cellular components (Table 4.8) and 12 across 11 pathways (Table 4.9).

Table 4.7 Molecular function

Go terms	Number of hits
binding (GO:0005488)	11
receptor activity (GO:0004872)	5
signal transducer activity (GO:0004871)	3
catalytic activity (GO:0003824)	10
transporter activity (GO:0005215)	1

Table 4.8 Biological process

Go terms	Number of hits
cellular component organization or biogenesis (GO:0071840)	3
cellular process (GO:0009987)	12
localization (GO:0051179)	4
reproduction (GO:0000003)	2
biological regulation (GO:0065007)	6
response to stimulus (GO:0050896)	5
developmental process (GO:0032502)	2
multicellular organismal process (GO:0032501)	4
metabolic process (GO:0008152)	13
immune system process (GO:0002376)	1

Table 4.8 Cellular component

Go terms	Number of hits
membrane (GO:0016020)	3
macromolecular complex (GO:0032991)	1
cell part (GO:0044464)	7
organelle (GO:0043226)	4
extracellular region (GO:0005576)	1

4.9 Pathway

Pathway terms	Number of hits
Gonadotropin-releasing hormone receptor pathway (P06664)	1
JAK/STAT signaling pathway (P00038)	1
B cell activation (P00010)	1
Insulin/IGF pathway-protein kinase B signaling cascade (P00033)	1
p53 pathway (P00059)	1
Alpha adrenergic receptor signaling pathway (P00002)	1

Huntington disease (P00029)	2
Hypoxia response via HIF activation (P00030)	1
Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway (P00026)	1
p53 pathway feedback loops 2 (P04398)	1
T cell activation (P00053)	1

The lists obtained from this analysis give indication about the functional roles of the differentially expressed genes between the clones. These genes can be candidates for future experiments to reveal their specific roles.

4.9 The clones of LS174T have genetic differences

To reveal any discernible genetic difference between Clone 1 and Clone 2, whole genome of these clones and parent LS174T was sequenced using Illumina technology. The sequences were aligned to the latest reference human genome assembly (GRCh38/hg38) by Dr Kamila Koprowska and analysis of sequence differences performed with support from the Computational Biology Research Group at the WIMM (Dr. Simon McGowan).

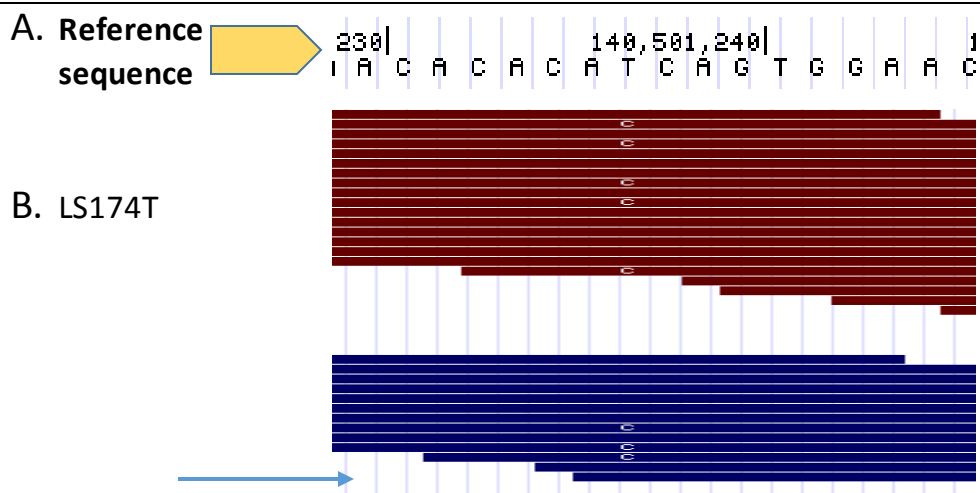
Initially, the data from Clone 1, Clone 2 and LS174T were compared for sequence differences in 40 selected genes (see list in appendix: Ax.5), which included driver genes for colorectal cancer such as *APC*, *KRAS*, *BRAF*, *CTNBB1*, *TP53* etc. and other genes of interest such as E-Cadherin (*CDH1*), *EPCAM*, *VIM*, *CD24*, *CD44*, *MLH1*, *MSH2* etc.

When analysed, a total of 9696 positions in those 40 genes showed variations in the sequences of Clone 1, Clone 2 and LS174T compared to the reference genome. Of these 9696 variations, only 76 were present in the exons of 30 of the 40 unique genes analysed. Among these 76 exonic variations, Clone 1 and Clone 2 differed in 12 positions in 9 genes: *BRAF*, *TP53*, *MSH6*, *PIK3C2B*, *BCL9L*, *MLH3*, *HLA-A*, *HLA-B* and *HLA-C*. As the *HLA* gene family naturally has a high frequency of error because of closely homologous pseudogenes, their variations were not considered significant in this particular analysis. As a RER+ cell line, the LS174T and its clones are expected to have mutations in the MMR genes and that is why the variations in *MLH3*, *MSH6* were also ignored. This left *BRAF*, *TP53*, *PIK3C2B* and *BCL9L* genes to look for true genetic differences between Clone 1 and Clone 2 (Table 4.10).

Table 4.10: Genetic differences between Clone 1 and Clone 2. Out of 40 genes selected to compare between Clone 1 and Clone 2, variations were found in 4 exonic positions of 4 genes. WT= Wild Type; Het/Mt= Heterozygous mutant; Het/Del= Heterozygous Deletion

Gene name	Chromosome start site	Chromosome end site	Reference base	Observed base	Clone 1	Clone 2	LS174T
TP53	7675236	7675236	A	G	WT	Het/Mt	WT (?)
BCL9L	118902396	118902396	A	Del	Het/Del	Het/Del	Het/Del
PIK3C2B	204446002	204446002	A	G	WT	Het/Mt	WT(?)
BRAF	140801437	140801437	T	C	Het/Mt	WT	Het/Mt

In order to confirm the sequence changes observed from the high throughput sequencing, the variations in *BRAF* were verified in the parental LS174T, Clone 1 and Clone 2 by Sanger sequencing. No *BRAF* mutation has previously been reported in the LS174T, which is known to contain a well characterised, key oncogenic Gly to Asp mutation at codon 12 of the *KRAS* gene. Mutations in *KRAS* and *BRAF* are generally mutually exclusive, as they occur in the same signalling pathway, making the presence of an additional, heterogeneous T>C change in a codon (p. M279>V) of exon 6 in *BRAF* in Clone 1 (Fig 4.12) somewhat unusual. The relatively low frequency with which the mutation occurs in the parental LS174T cell line (8/30 sequences in Fig 4.12) may account for the fact that this mutation has not previously been recorded in this cell line. However, the whole genome sequence data, confirmed by Sanger sequencing, is consistent with the hypothesis that the parental cell line consists of co-existing stable subpopulations of Clone 1 (heterozygous mutant) and Clone 2 (homozygous wild type) cells.



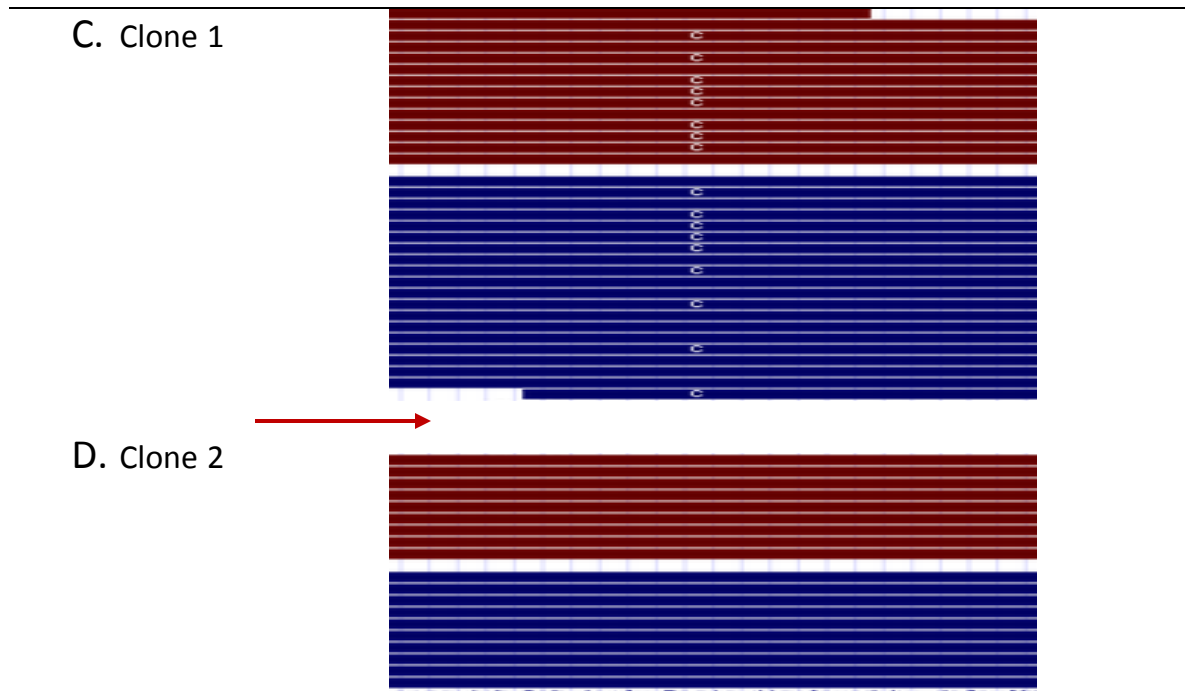


Fig 4.12: Heterozygous (T>C) change in *BRAF*. In place of 'T' in reference genome (A, pointed with yellow arrow) Clone 1 has heterozygous: T>C change (C, bordered with a blue arrow) but Clone 2 is homozygous wild type (D, bordered with a red arrow). As the mixture of Clone 1 and Clone 2, LS174T appears heterozygous for the T>C change in *BRAF* (B). Here, some strands contain wild type 'T' but others contain 'C', indicating the heterozygosity for this variation. Here, red bars represent sequence reads for forward strands and blue bars represent reads for reverse strands. Number of bars indicate sequence depth. Base positions different from reference genome are visible in LS174T and Clone 1, where the presence of 'C' differs from reference sequence 'T'.

To confirm this variation, Sanger sequencing was performed in this region of *BRAF*. The sequence results revealed that Clone 1 and LS174T had T>C heterozygous change, whereas Clone 2 was wild type (Fig 4.13).

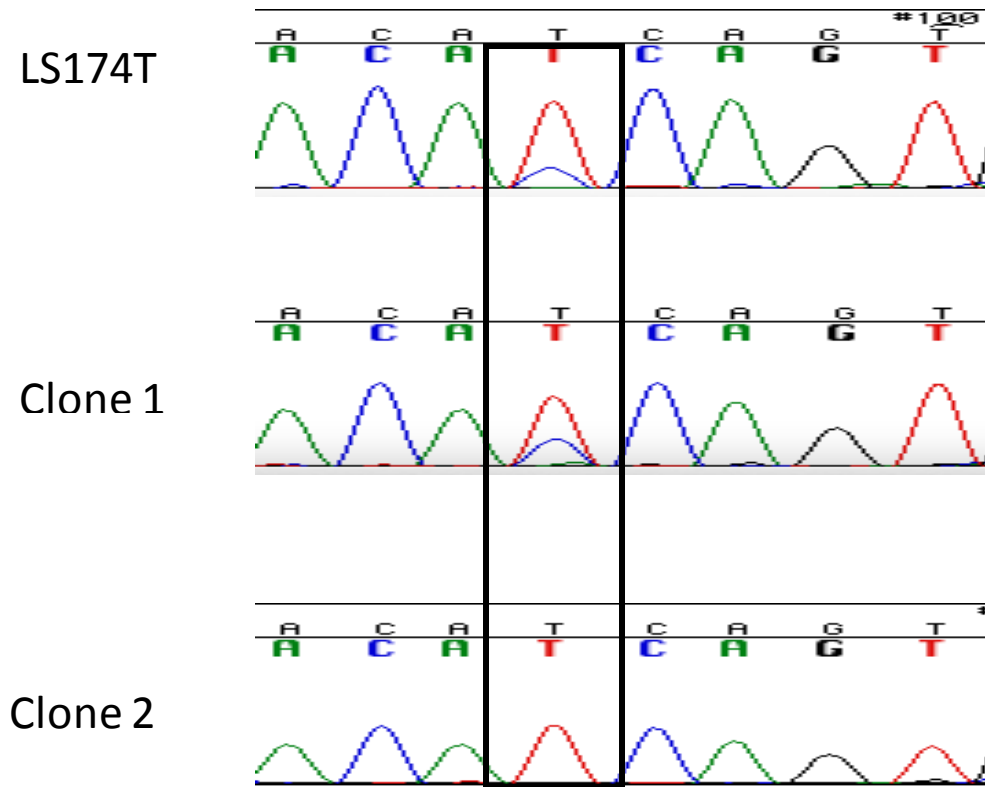


Fig 4.13: Confirmation of T>C change in *BRAF*. LS174T and Clone 1 show heterozygous T>C change in exon 6 of *BRAF* gene, indicated by the presence of peaks of both T (red) and C (blue); whereas Clone 2 is wild type, indicated by the single peak of 'T' (red).

Among these 4 genes, for *TP53* and *PIK3C2B*, there were discrepancies regarding the relation between the sequences of the clones and LS174T. It is expected that as the LS174T is the parent of Clone 1 and Clone 2, if any of these clones are heterozygous for a particular locus, the parental sequence should also be heterozygous in that locus. However, from the visualisation of the sequence tracks on the UCSC genome browser, the LS174T appeared wild type for both *PIK3C2B* and *TP53*, whereas in both cases, Clone 2 was heterozygous mutant (Fig 4.14 A, B).

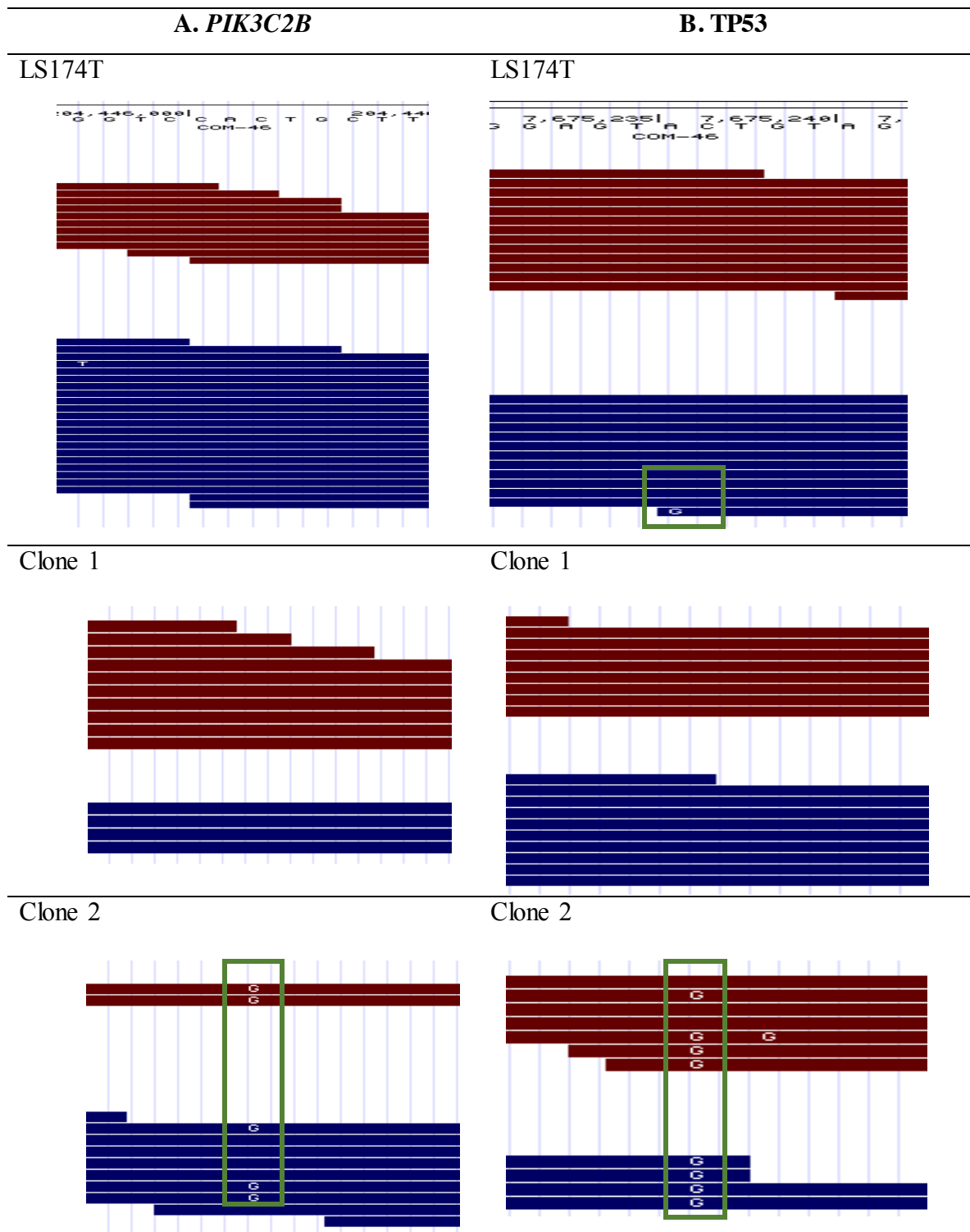


Fig 4.14: A. In the LS174T bulk sequence, *PIK32C2B* gene appears wild type. However, in Clone 2, out 10 reads, 5 showed A>G variation (shown in green box); Clone 1 was wild type. B. For *TP53*, in the case of parental LS174T, only one read appeared to have A>G variation. Clone 2 had heterozygous A>G change. Clone 1 was wild type. Here, red bars represent sequence reads for forward strands and blue bars represent reads for reverse strands. Number of bars indicate sequence depth.

The sequence change for *TP53* was selected for confirmation by Sanger sequencing to find out whether it could pick heterozygous A>G variation in the parental LS174T. However, Sanger sequencing concurred with the observation of the high throughput sequencing, where Clone 2 is heterozygous, Clone 1 is homozygous wild type and parental LS174T is homozygous wild type (Fig 4.15).

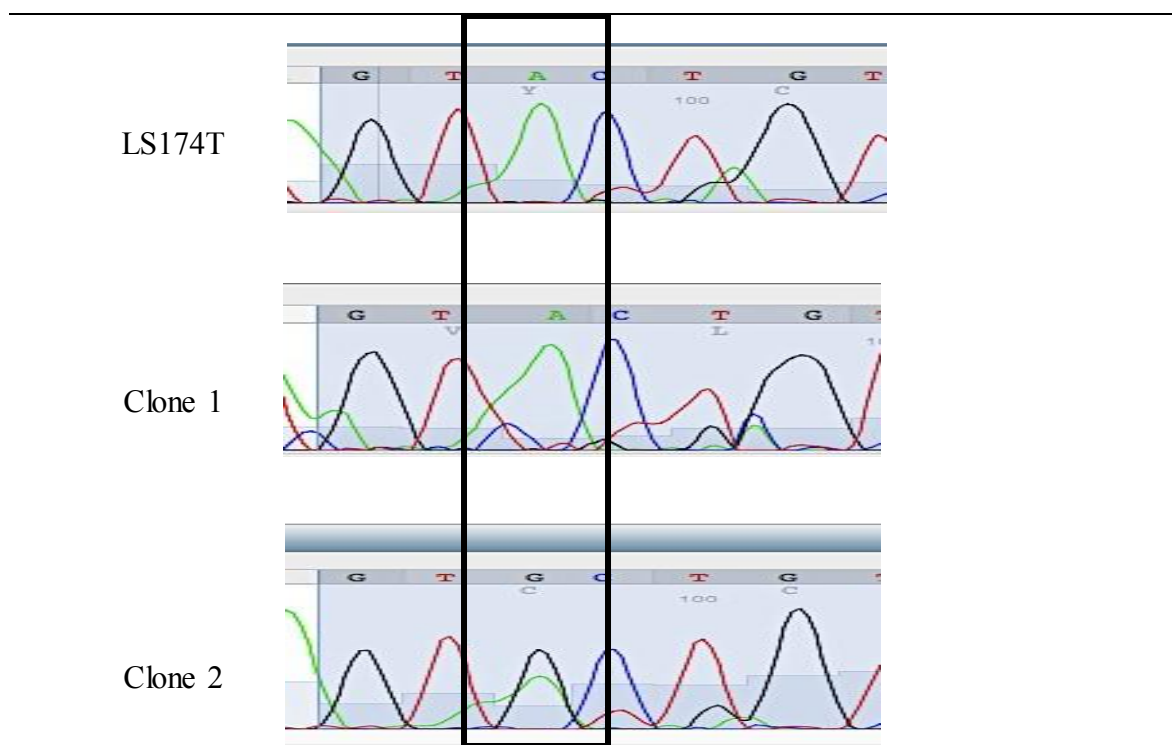


Fig 4.15: Sanger sequence results for *TP53* variation. Clone 1 (middle) shows only base 'A' (as green peak) meaning the homozygosity of this position, whereas in Clone 2 (bottom), both 'G' (black peak) and 'A' (green peak) are present, indicating A>G heterozygosity. The parental LS174T (top) was expected to be heterozygous, but it appeared homozygous.

One explanation for the Sanger sequencing results can be drawn from the fact that Clone 2 is the minor population (~17-20%) in the parental LS174T. If a small portion of Clone 2 (as it is heterozygous) contains the mutant versions of gene *TP53* and *PIK3C2B*, it is not unlikely that those variations will not be picked during sequencing. Another explanation would be if the mutation had only been generated and selected for subsequent to the clones being isolated from the parental LS174T (which is unlikely to occur in a short period of time in the culture).

4.10 Drug response in clones

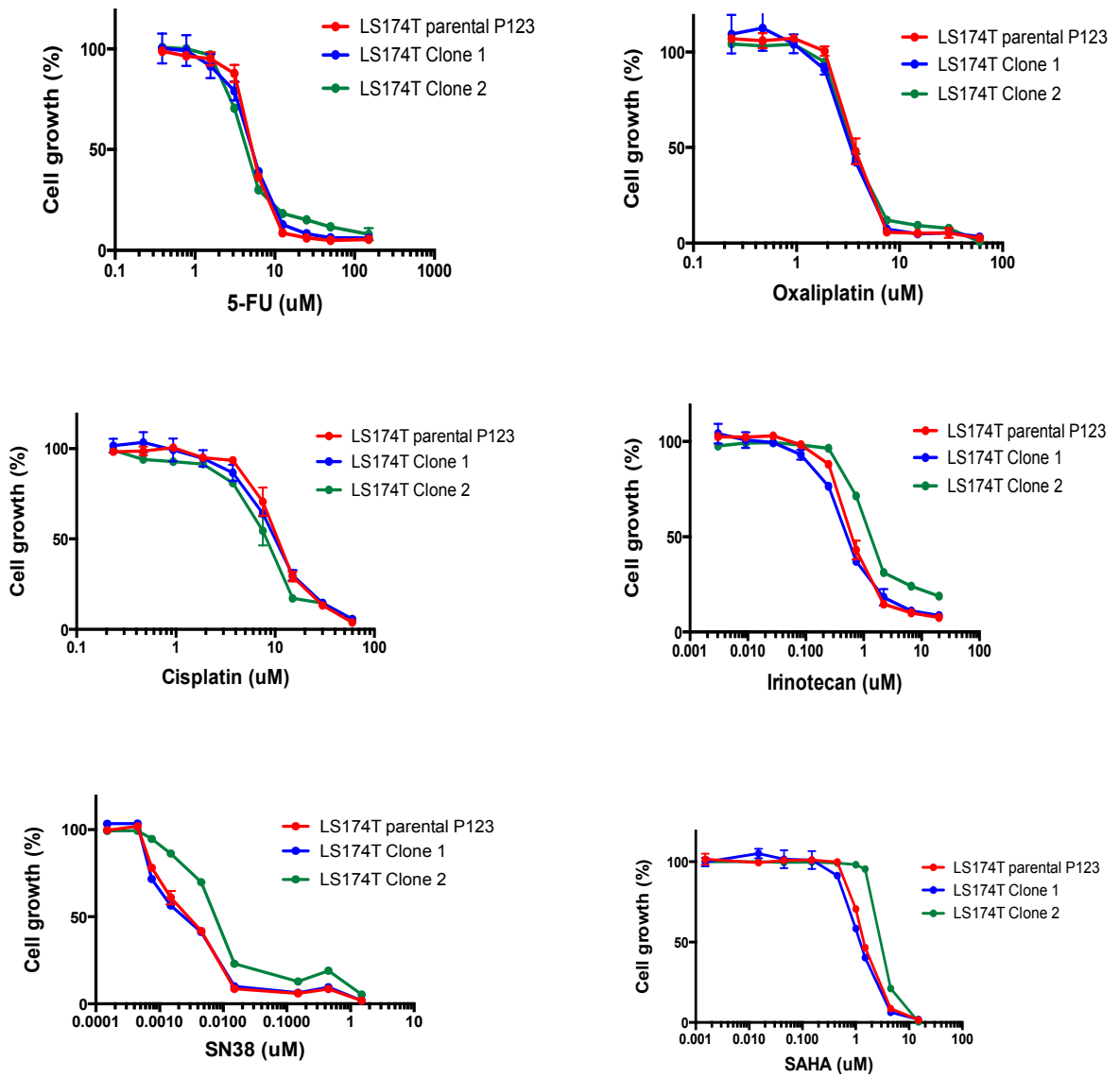
Clone 1 and Clone 2 showed expression differences in a number of genes that play potential roles in drug resistance and EMT (Table 4.11). Consequently, an obvious question was whether they would respond differently when treated with chemotherapeutic drugs used commonly in colorectal Cancer. To explore this question, a panel of 10 drugs (Table 4.11) were tested on Clone 1 and Clone 2 and parental LS174T (courtesy, Djmaila Ouaret).

Table 4.11: List of 10 drugs used in CRC treatment.

No.	Name of the drug	Mode of action
1.	5-Fluorouracil	Inhibits thymidylate synthase, anti-metabolite agent
2.	Oxaliplatin	Alkylating agent
3.	Cisplatin	Alkylating agent
4.	Irinotecan	Topoisomerase Inhibitor
5.	SN38	Active metabolite of Irinotecan
6.	OSI906	IGF1R tyrosine kinase inhibitor
7.	Figitumumab	anti-IGF1R mAb
8.	Lapatinib	EGFR/HER2 tyrosine kinase inhibitor
9.	Cetuximab	anti-EGFR mAb
10.	SAHA	HDAC inhibitor

Clone 1, Clone 2 and parental LS174T were treated with different concentrations of drugs on two time points. Cell killing was measured by SRB assay (describe in section 2.5.4)

When analysed, no striking difference across the 10 tested agents was observed (Fig 4.16). However, minor differences were observed in Clone 2 for the following agents: SAHA, Lapatinib, Irinotecan and SN38.



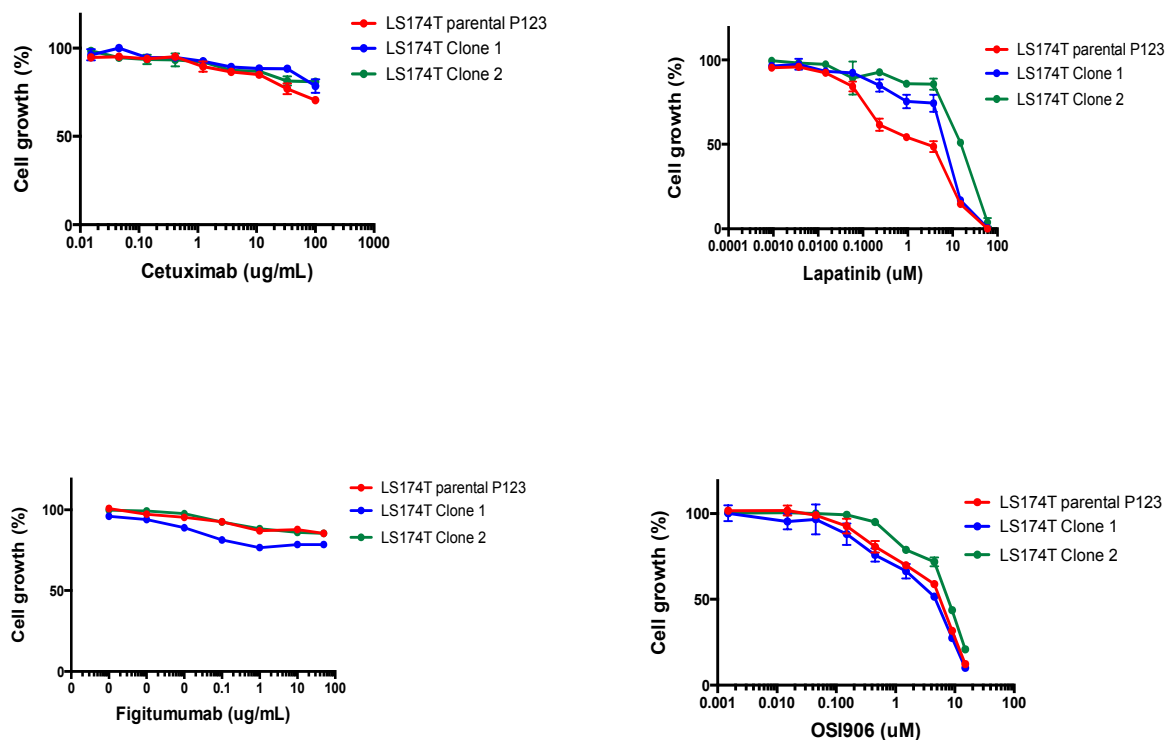


Fig 4.16: Responses to drug treatments: Clone 2 showed slightly higher resistance to drug SAHA, Lapatinib, Irinotecan and SN38. For all other drugs, there were no differences in responses among Clone 1, Clone 2 and parental LS174T. **Red lines:** Parental LS174T, **Blue lines:** Clone 1, **Green lines:** Clone 2.

The above data showed that Clone 2 might have shown slightly higher resistance against SAHA, which is an HDAC inhibitor. Microarray data revealed (Table 4.2) that Clone 2 expresses a higher level of *HIST1H2AB*, an important histone for chromatin organisation. Therefore, a study was designed to repeat the experiment with SAHA and other HDAC inhibiting drugs (CAHA and Trichostatin) to find out if there was any correlation between *HIST1H2AB* expression and resistance to HDAC inhibitors in general. The assay results showed that Clone 1 and Clone 2 did not show any notable difference to treatment with HDAC inhibitors, despite their apparent difference in *HIST1H2AB* expression.

4.11 Chapter discussion

The LS174T cell line was developed from a Duke B stage adenocarcinoma (Tom et al., 1976). This cell line has been used in many studies and over the years characterised with regards to its growth, mutation status and tumorigenicity etc. Most conveniently, a pair of easily detectable mutation in the E-Cadherin gene (Efsthathiou et al., 1999) gives a handy tool to check the authenticity of this cell line at any stage of a study. This is why when the LS174T cell line appeared heterogeneous under the microscope, it was considered a significant phenomenon to follow up, especially, as many commonly used cell lines –such as LS180, DLD1, HCT11T, HT29, SW1222 etc.– do not present such obvious heterogeneity.

In this chapter, it has been shown that the morphological heterogeneity observed in LS174T is separable as distinct clones. By using FACS based single cell cloning, morphologically different clones, Clone 1 and Clone 2 were isolated from LS174T. Importantly, these morphological differences remained stable for the entire length of the study. Obtaining such morphologically distinct clones is uncommon in many other cell lines. When cell lines like HCT116, LS411, DLD1 and RKO were cloned using FACS, distinguishable morphological differences were not found among the successful clones.

However, heterogeneity in several cell lines have been reported previously on various contexts. Walter & Kissel (1995) reported heterogeneity in Caco 2, a widely used colorectal cancer cell line. In a similar effort, Hastings & Franks (1983) showed that a cell line generated from bladder cancer was heterogeneous with distinct clones which have differences in morphological appearance, growth in agar and tumorigenicity. Among breast cancer cell lines, MCF7 is well-known for its heterogeneity regarding progesterone receptor (PR) expression and other characteristics (Baguley & Leung, 2011).

The observation of heterogeneity in the LS174T and successful isolation of Clone 1 and Clone 2 was investigated further for more characterisation. In addition to their obvious morphological differences, the two clones showed different ability of attachment on plastic surface during regular subculturing using trypsin. Clone 1 had faster growth rate than Clone 2 and when seeded in matrigel, and they showed differences in colony

morphology and clonogenicity. Previously, Yeung *et al* had shown that differential clonogenicity and colony size of cells from SW1222, HCT116 and HT29 were correlated with the enrichments of stem cell marker CD44 and CD24 (Yeung et al., 2010). Several other studies have reached similar conclusion regarding CD44 and CD24 (Battula et al., 2010; Jaggupilli, Elkord, Jaggupilli, & Elkord, 2012).

When the expressions of CD44 and CD24 were checked in Clone 1 and Clone 2 of LS174T, it appeared that both of the clones expressed the same level of CD44, but regarding CD24 expression, they were very different. Clone 2 did not express CD24 (designated as Clone2^{CD24-}), but Clone 1 was highly expressing CD24 (designated as Clone 1^{CD24+}). This finding is important, as it created the opportunity to uncouple the roles of CD24 from CD44. In various cancers, although the presence and significance of CD44 has always been consistent, the role of CD24 was variable. In breast cancer, for example, stem cells are designated as CD44⁺/CD24⁻ (Beça et al., 2013; Kai Yoshimi Arima AE Toshio Kamiya AE Hideyuki Saya, 2010). However, in colorectal cancer (Yeung et al., 2010), pancreatic cancer (C. Li et al., 2007) and renal cancer (Lee et al., 2008), stem cells are enriched with CD44⁺/CD24⁺. Some studies have identified CD24 as an important prognostic marker in several cancers (Agrawal et al., 2007; Kristiansen et al., 2003; Lee et al., 2008), whereas others have dismissed its importance for such use (Ahmed et al., 2009).

The fact that the Clone 1^{CD24+} and Clone 2^{CD24-} (i) have originated from the same cell line and (ii) had naturally occurring differences in CD24 expression which is stable in the culture for a long time, make them powerful models to investigate the relative importance of CD24, especially, as these two clones do not differ in other known stem cell markers.

Beyond this CD24 expression difference, these two clones were also investigated for genetic differences. Their whole genomes, together with their parent (LS174T) genome, were sequenced and compared. Comparisons in 40 selected genes showed that Clone 1 and Clone 2 differed in *BRAF*, *TP53*, *PIK3C2B* and *BCL9L* genes. A 'T>C' variation in *BRAF* (resulting in p. M279>V) and an 'A>G' variation in *TP53* gene (resulting in p. Y87>C) were confirmed by Sanger sequencing. To date, the LS174T cell line was considered wild type for *BRAF* and *TP53*. However, by sequencing the genomes of the

clones, now two heterozygous variations have been revealed in those genes. The presence of these mutations in the cell lines or clones may not have any functional significance if they are just passenger mutations. However, they can still be useful as the genetic markers to differentiate between the clones.

Microarray expression analysis revealed differences in other important genes between the clones, such as *CXCR4*, *LAPTM4B*, *TGFBI* (highly expressed in Clone 1), *SHOX2*, *CST1*, *SPINK1* (highly expressed in Clone 2) (Table 4.2). These differential expressions can be further explored using these two clones to study the functional and regulatory aspects of these genes.

The above observations of (i) morphologically distinct clones in the LS174T, (ii) differential expression of CD24 in Clone 1^{CD24+} and Clone 2^{CD24-}, (iii) the presence of sequence variations between Clone 1 and Clone 2, and (iv) expression differences in functionally relevant genes between the clones (Table 4.2) led to the interpretation that the LS174T cell line was essentially a stable mixture of Clone 1 and Clone 2. In future research, genetic mutations and the expression profile of CD24 can be used as the distinguishing markers between these clones.

Cell lines have been used as models to study various aspects of primary cancers for a long time. This clonal heterogeneity in the LS174T cell line can be exploited to understand the nature and evolution of tumour heterogeneity using an *in vitro* system. However, there was an argument to search for more cell lines with such heterogeneity in the cell line panel to further validate the notion of heterogeneity in CRC cell lines and their intended use. This has been explored in the next chapter.

Chapter- 5

CD24 Based Heterogeneity in CRC Cell Lines

Chapter- 5

5.1 Chapter introduction

The presence of stable clones in LS174T, their genetic differences and differential expression in CD24 suggested more investigation for heterogeneity in other CRC cell lines. One aspect of the study was to use prior knowledge about cell lines with heterogeneous morphology in culture. Among prospective cell lines, SW480 was a prime candidate because of its known heterogeneous appearance in culture (Fig: 5.1).



Fig: 5.1: Heterogeneous appearance of CRC cell line SW480 in culture. A bright field (5x image after 48 hours in culture. Red and Yellow arrow indicates cells with different morphology.

SW480 is a paired cell line with SW620. However, while the paired cell lines LS174T and LS180 were both derived from the same primary tumour using different cell culture techniques, SW480 and SW620 cell lines were derived from spatially separated tumours. In this case, SW480 and SW620 cell line was developed respectively from the primary and metastasised tumours resected from the same patient (Hewitt et al., 2000).

SW480 is a commonly used cell line in the Bodmer laboratory. In order to investigate the nature of the observed morphological heterogeneity, the cell line was cloned by sorting single cells using FACS and analysed following the same procedures described in Chapter 4 for LS174T. Ultimately, the expression of CD24 was assessed in the clones isolated from SW480 to see if CD24 could also be used as a marker to reveal stable heterogeneity in this cell line.

5.2 Isolation and characterization of morphologically distinct clones with differential expression of CD24 in SW480

5.2.1 SW480 is a cell line with morphologically distinct subpopulations

SW480 and SW620 exhibit some striking differences despite being derived from the same patient (Luo et al., 2016). In the Bodmer laboratory, it has been known from microarray analysis that these cell lines differ in CD24 and CD44 expression (Table 5.1).

Table 5.1: Linear value of CD24 and CD44 message in SW480 and SW620

Gene	SW480	SW620
CD24	520	13440
CD44	3011	10

This observation has been confirmed at the protein level using flow cytometry (courtesy: Kamila Koprowska) where SW480 is CD24⁻/CD44⁺, but SW620 is CD24⁺/CD44⁻ (Fig 5.2).

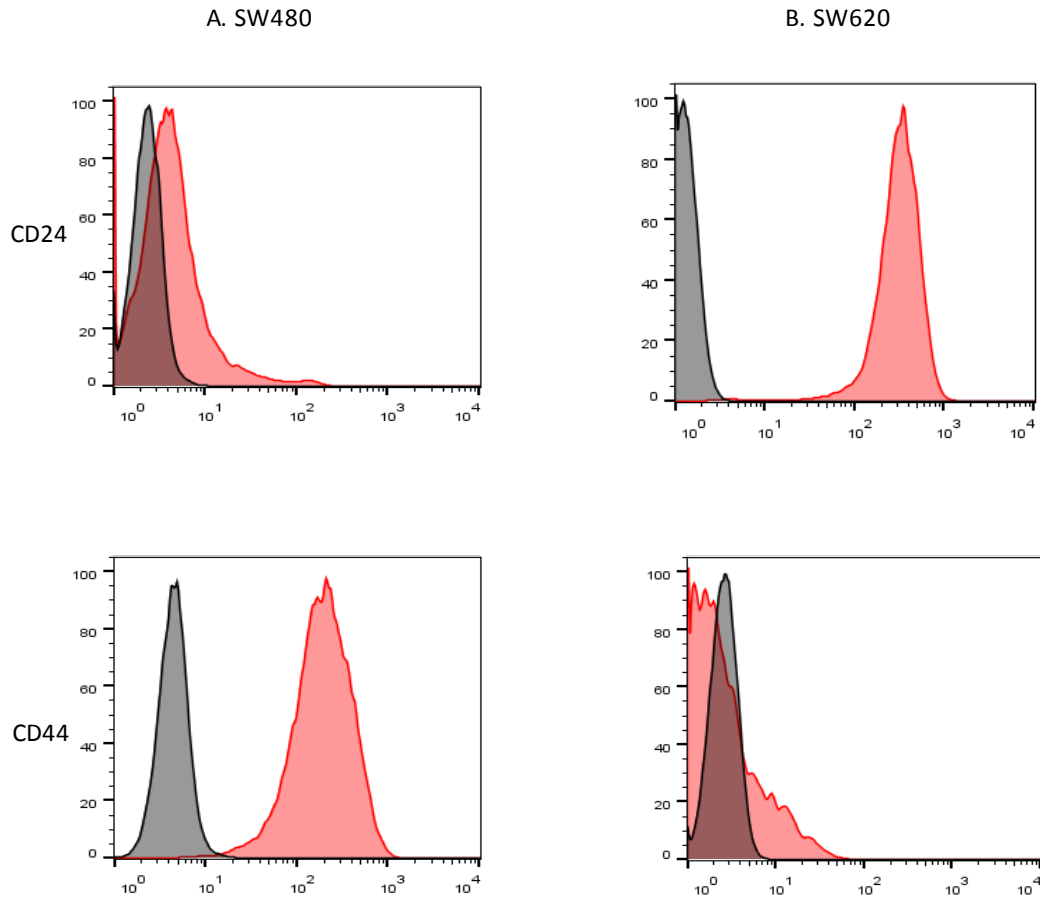


Fig 5.2: CD44 and CD24 expression assessments in A. SW480 and B. SW620 using Flow cytometry. Grey histogram: Isotype control, Red histogram: CD24 and CD44 staining with specific antibody. X-axis: CD24-PE and CD44-FITC; Y-axis: Cell counts

When cultured in standard plastic cell culture flask, SW480 and SW620 showed differences in their morphology: SW620 appeared homogeneous with mostly elongated cells, whereas SW480 looked heterogeneous with two distinct cell populations, i) slender elongated cells and ii) blob-like cells (Fig. 5.3).

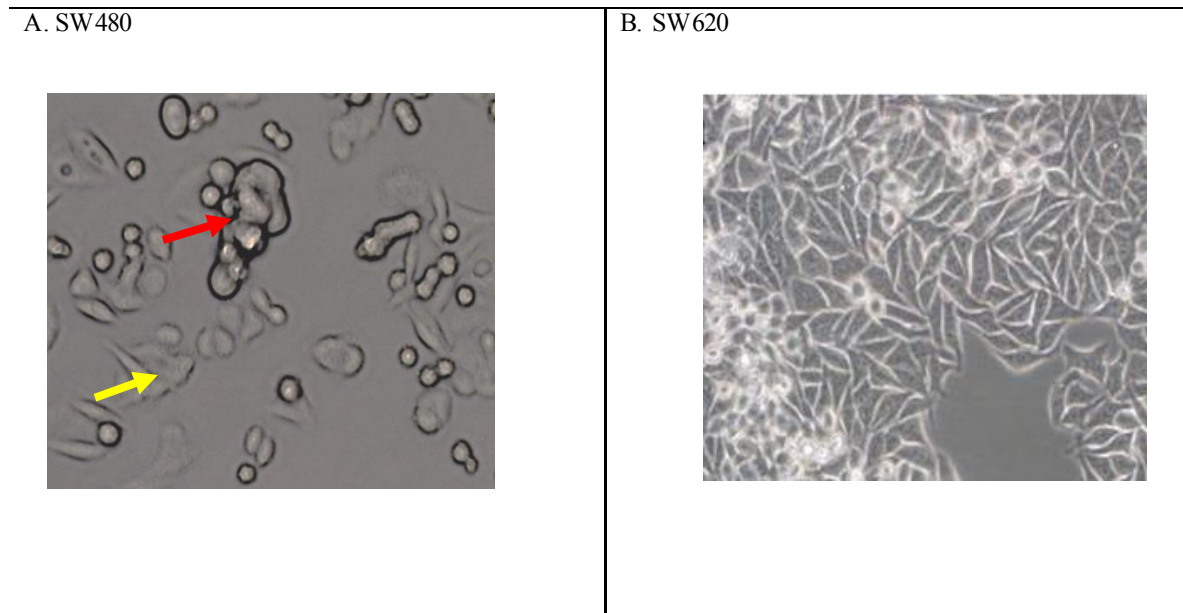
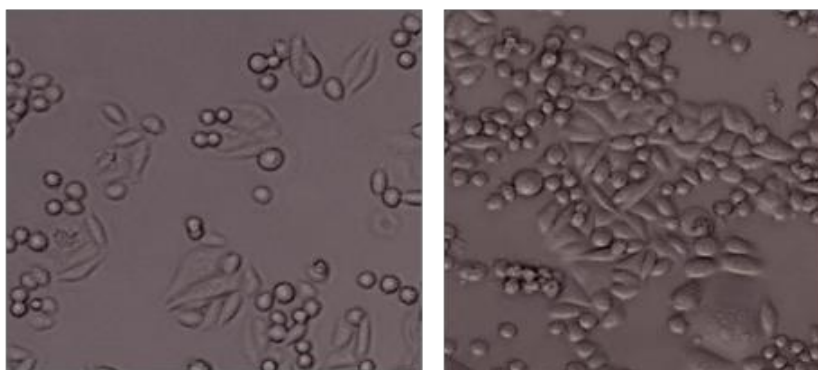


Fig 5.3: Bright field images of SW40 and SW620. A. SW480 appears to have two types of cells: one with slender elongated morphology (shown with yellow arrow) and another with a more rounded, blob-like morphology (shown with red arrow). B. SW620 has a mostly homogeneous appearance with slender cells in 2D culture.

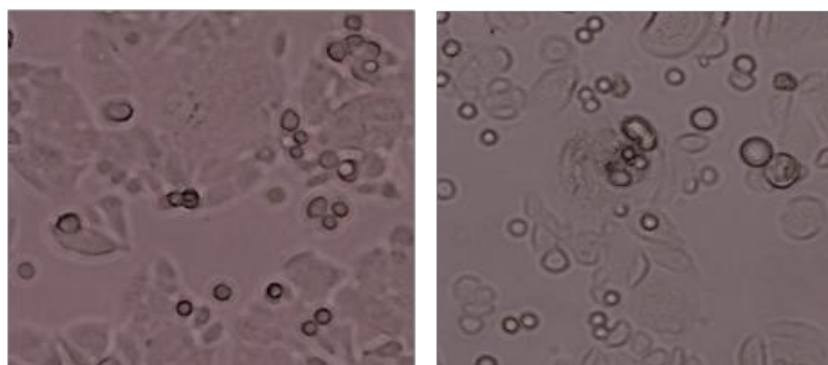
Following this observation of morphological heterogeneity, SW480 was cloned by FACS sorting. Single cells were sorted in a 96-well plate. After one week, wells were checked for appearance of colonies. Wells that showed more than one colony were excluded from further studies. After four weeks, 24 single colonies were picked from 24 different wells and transferred in two 12 well plates to obtain enough cells for further analysis.

Cells from these 24 clones with different morphology could be grouped into three different categories. Group 1 consisted mostly of elongated cells with some round (which might be mitotic) cells on top of them. Group 3 clones formed blob-like colonies and Group 2 colonies were of ‘intermediate’ morphology where the difference was rather less stark compared to other two groups (Fig: 5.4).

Group 1



Group 2



Group 3

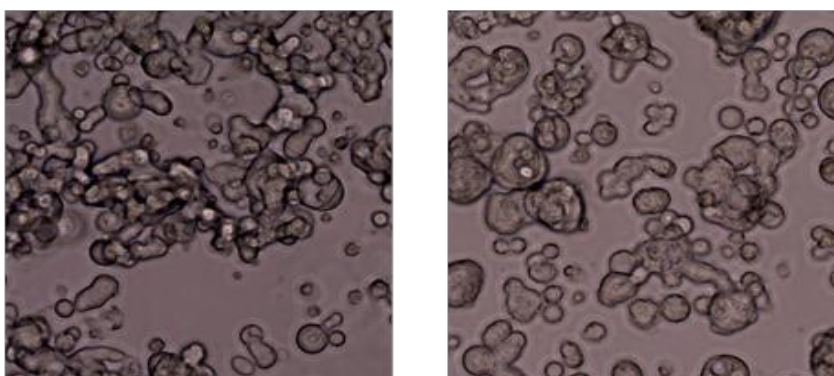
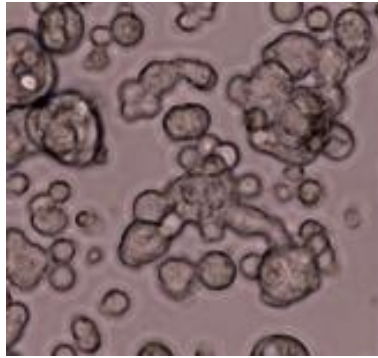


Fig 5.4: Bright field image (5x) to show morphology of clones from SW480. Group 1 clones are mostly elongated cells. Cells in Group 3 form a blob-like morphology. Group 2 clones showed intermediate morphology

Clones with the two most distinct morphologies (from Group 1 and Group 3) were chosen for further analysis (Fig: 5.5): they were named as “SW480_Clone D1” and “SW480_Clone C2” (the letter indicates the well the clones were picked from).

A. SW480_Clone C2



B. SW480_Clone D1

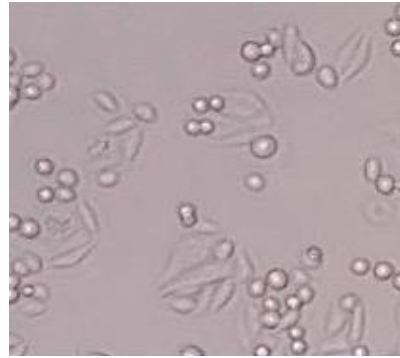


Fig 5.5: Morphology of selected clones. A. SW480_Clone C2, representative of blob-like morphology and B. SW480_Clone D1, representative of extended/slender morphology.

5.2.2 CD24 expression difference between SW480_Clone C2 and SW480_Clone D1

As seen in chapter 4, morphologically distinct Clone 1 and Clone 2 of LS174T differed in CD24 (and not in CD44) expression. It suggested to check whether SW480_Clone C2 and SW480_Clone D1 would show similar pattern.

In these clones CD24 and CD44 expressions were assessed using flow cytometry (courtesy: Kamila Koprowska). For this, SW480_Clone C2 and SW480_Clone D1 at passage 5 (p5) were incubated with CD24-PE and CD44-FITC antibodies and compared against an unstained control (no antibodies).

When analysed, SW480_Clone C2 showed a broad expression profile for CD24. The majority of the cells in this clone are CD24⁺ according to the FACS profile, but there is a clear subpopulation of CD24⁻ cells. In contrast, Clone D1 is mostly CD24⁻. Both of these clones were CD44⁺, though SW480_Clone D1 were more homogeneous than SW480_Clone C2 (Fig 5.6).

A. SW480_Clone C2: CD44+/CD24+

B. SW480_Clone D1: CD44+/CD24-

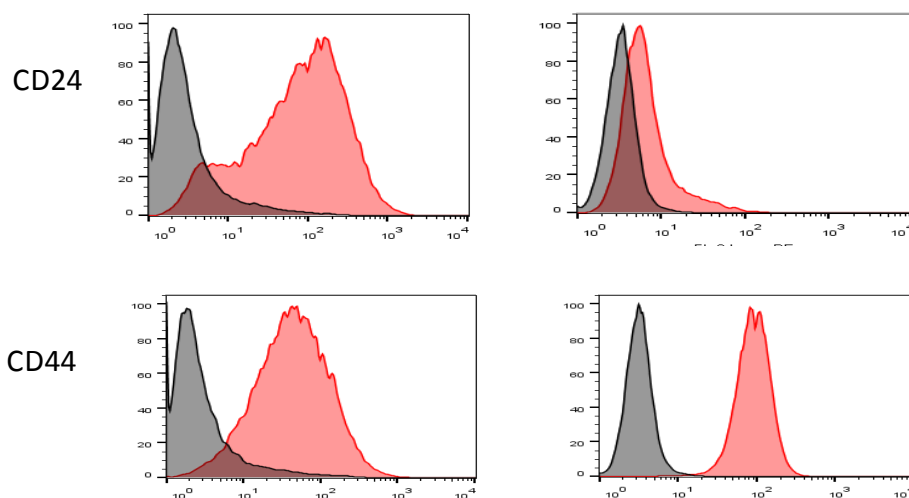
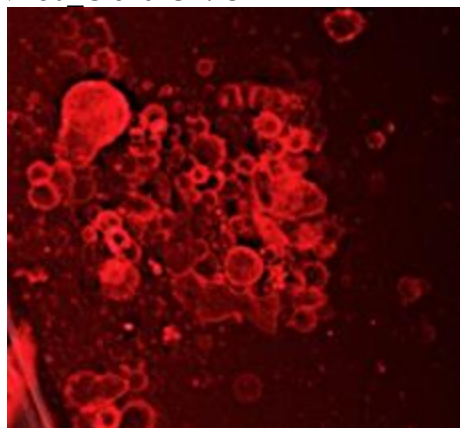


Fig 5.6: Flow Cytometry profile of CD24 and CD44 expression in SW480 clones. A. SW480_Clone C2 expresses both CD24 and CD44 (red histogram). B. SW480_Clone D1 is CD24- and expresses high level of CD44 (shown as red histogram). Grey histograms represent unstained control in all samples.

Live cell imaging with fluorescence microscopy confirmed this observation (Fig 5.7).

A. SW480_Clone C2: CD24+



B. SW480_Clone D1: CD24-

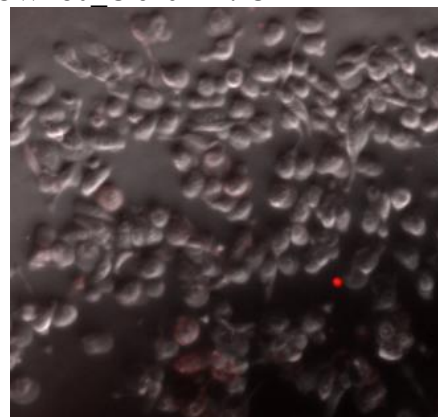


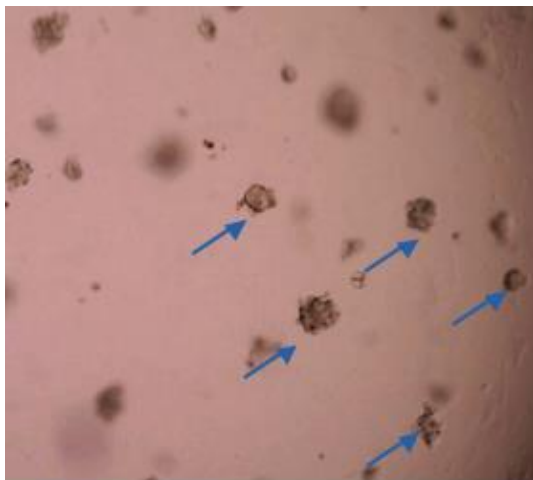
Fig 5.7: Fluorescence microscopy of SW480_Clone C2 and SW480_Clone D1 for CD24. A. Clone C2 shows expression of CD24 (red staining on the cell surface). B. Clone D1 does not stain at all indicating no expression of CD24.

5.2.3. SW480_Clone C2 and SW480_Clone D1 differ in clonogenicity and differentiation in matrigel

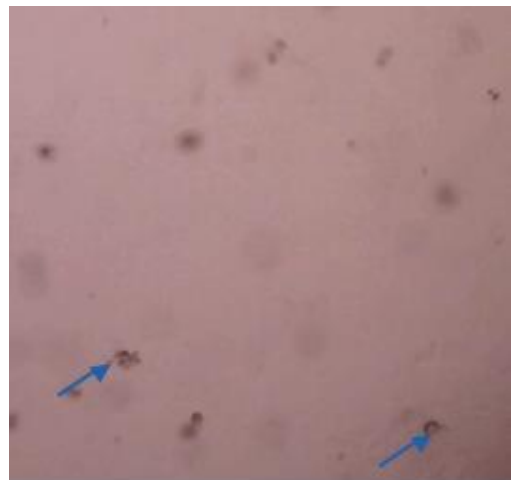
SW480_Clone C2 and SW480_Clone D1 were grown in matrigel to test their clonogenicity, morphology and differentiation ability. From each clone, 500 cells were seeded in matrigel in duplicate wells of a 96 well plate. They were imaged in every seven days from week 1 through week 3.

SW480_Clone C2 started to show signs of cell-cell aggregation from week 1. At that time, well defined colonies appeared in the wells of SW480_Clone C2 whereas just sparsely aggregated cells were found in the Clone D1 wells (Fig 5.8). By week 2, colonies of SW480_Clone C2 formed organized structures. Compared to them, SW480_Clone D1 colonies were much smaller and appeared as aggregates of loosely connected cells.

A. SW480_Clone C2



B. SW480_Clone D1



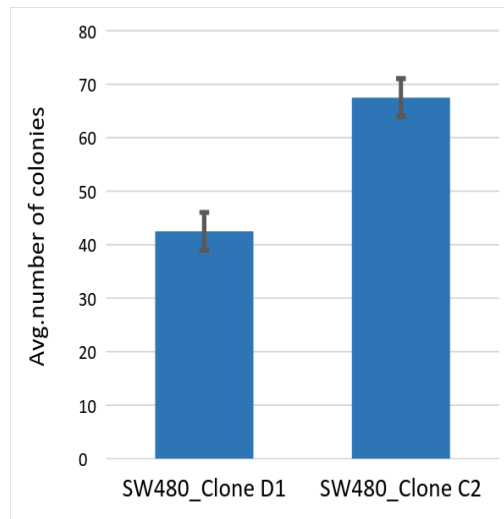
C. Differences in clonogenicity between clones in matrigel after two weeks


Fig 5.8: Bright field image of colonies of SW480_Clone C2 and SW480_Clone D1 in matrigel and clonogenicity. A. Clone C2^{CD24+} forms bigger and more colonies than B. Clone D1^{CD24-} in matrigel after week 1. C. After 14 days, counting of colonies (number out of 500 cells plated) in matrigel shows that SW80_Clone C2 has higher clonogenicity than SW80_Clone D1. The error bars indicate the standard deviation calculated from colony numbers of triplicate wells.

In order to assess any evidence of lumen formation in colonies of SW480_Clone C2 and SW480_Clone D1, colonies were stained with TRITC-Phalloidin. The observations revealed significant differences between the clones. SW480_Clone C2 formed two types of colonies: i) colonies with well-defined central lumens, and ii) colonies with no central lumens but with multiple foci. On the contrary, ~80% colonies of SW480_Clone D1 appeared without any organised structure (Fig: 5.9).

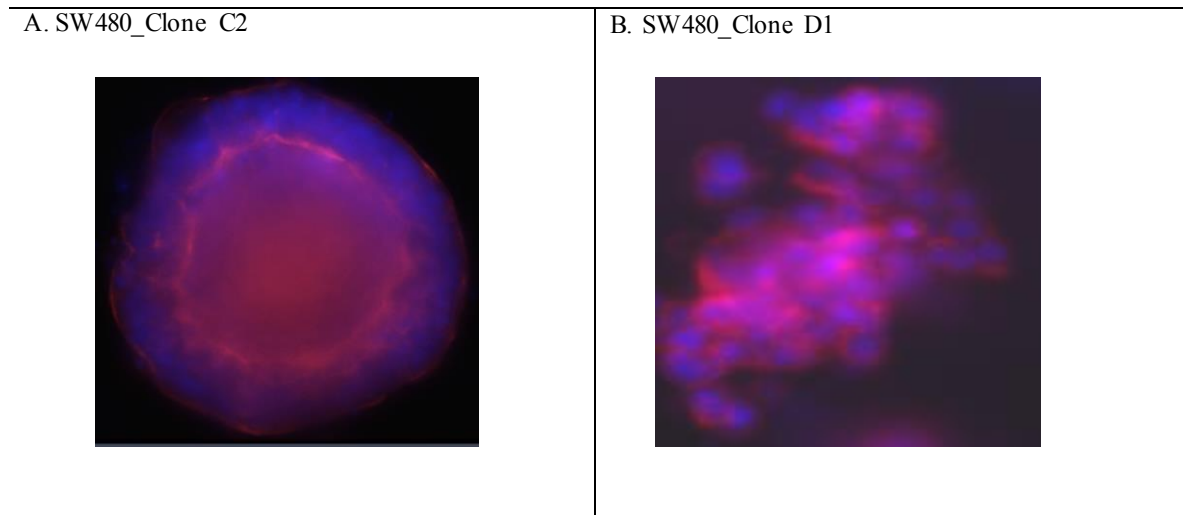


Fig 5.9: Image (20x) of TRITC-Phalloidin stained colony of SW480_Clone C2 and SW480_Clone D1 in matrigel after week 2. A. A representative SW480_Clone C2 colony with well-structured lumen. B. A representative SW480_Clone D1 colony without any organized structure. Red: TRITC-Phalloidin staining of F-actin. Blue: Nucleus staining with DAPI.

At week 3, all SW480_Clone D1 colonies started to spread in matrigel, but the lumen-forming SW480_Clone C2 colonies remained compact (Fig 5.10).

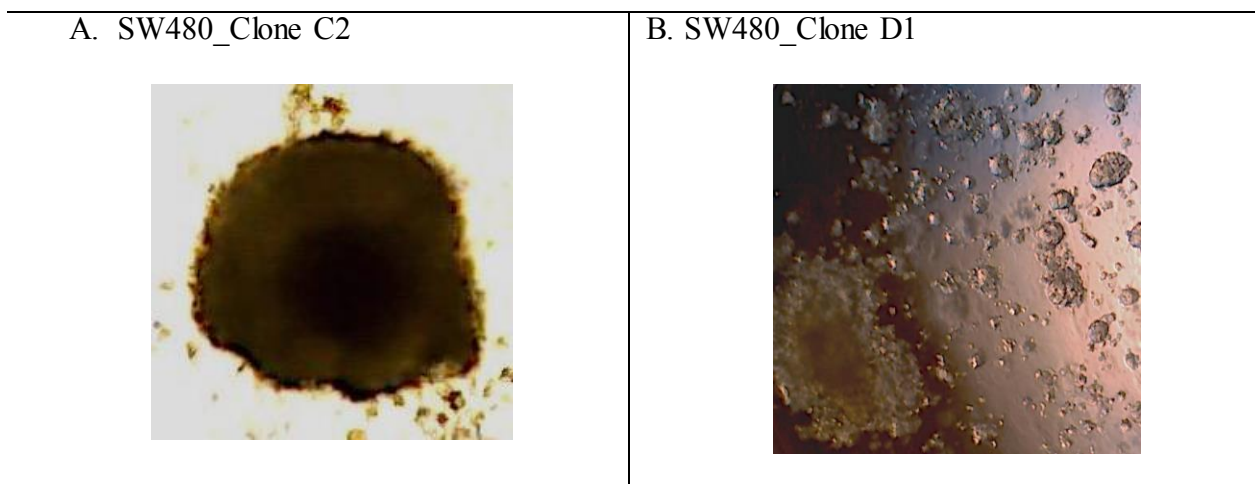
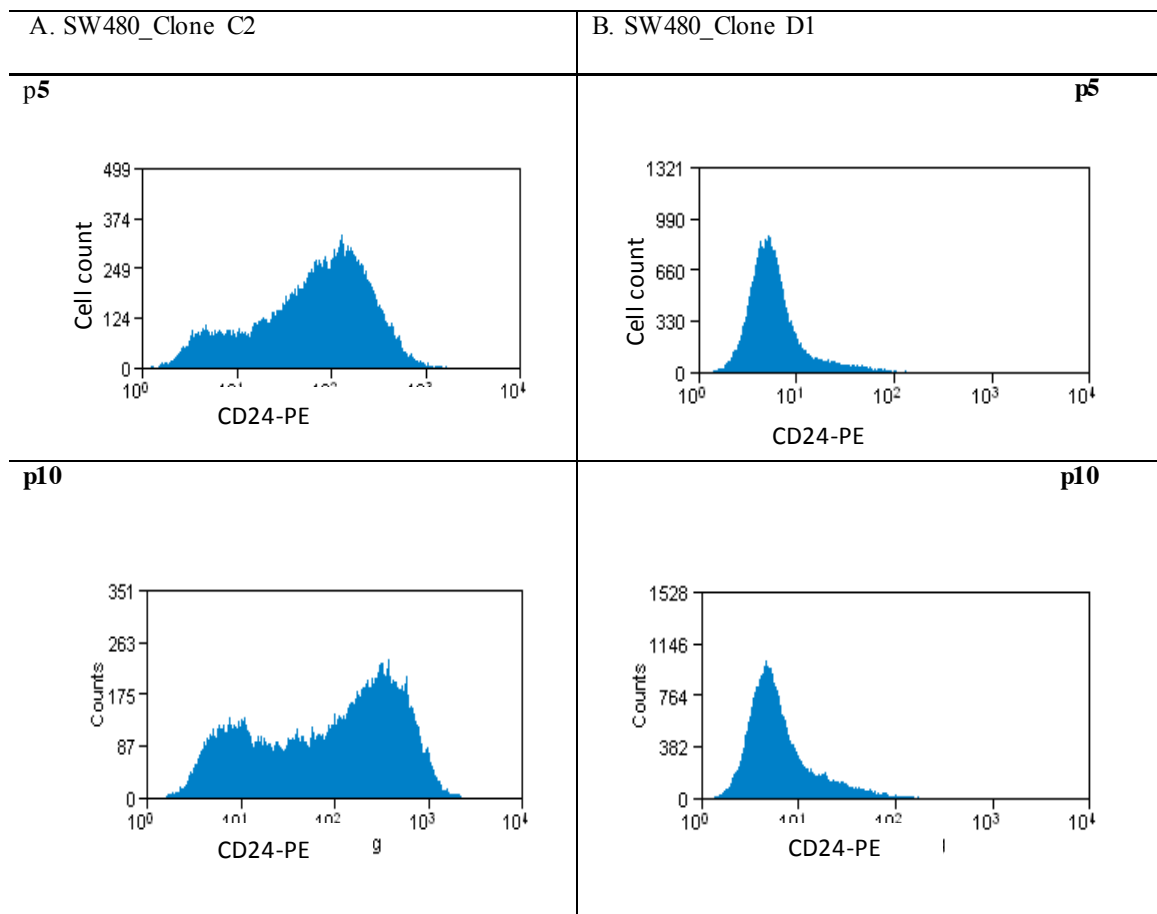


Fig 5.10: Colonies in matrigel after week 3. A. A representative colony of SW480_Clone C2, which remained compact. B. Colony of SW480_Clone D1 spread out into the matrigel. Here, the smaller colonies have grown after cells broke away from early bigger colonies.

5.2.4. SW480_Clone C2 changes the expression of CD24 over time but SW480_Clone D1 remains stable

Early passage SW480_Clone C2 (up to passage 8 - p8) is mostly CD24⁺. However, when subcultured, later passages of SW480_Clone C2 starts to lose CD24 expression with time (Fig 5.11 A). By passage 17 (p17), SW480_Clone C2 lost most of its CD24 expression. On the contrary, SW480_Clone D1 did not change (Fig 5.11 B) at all during this whole time.



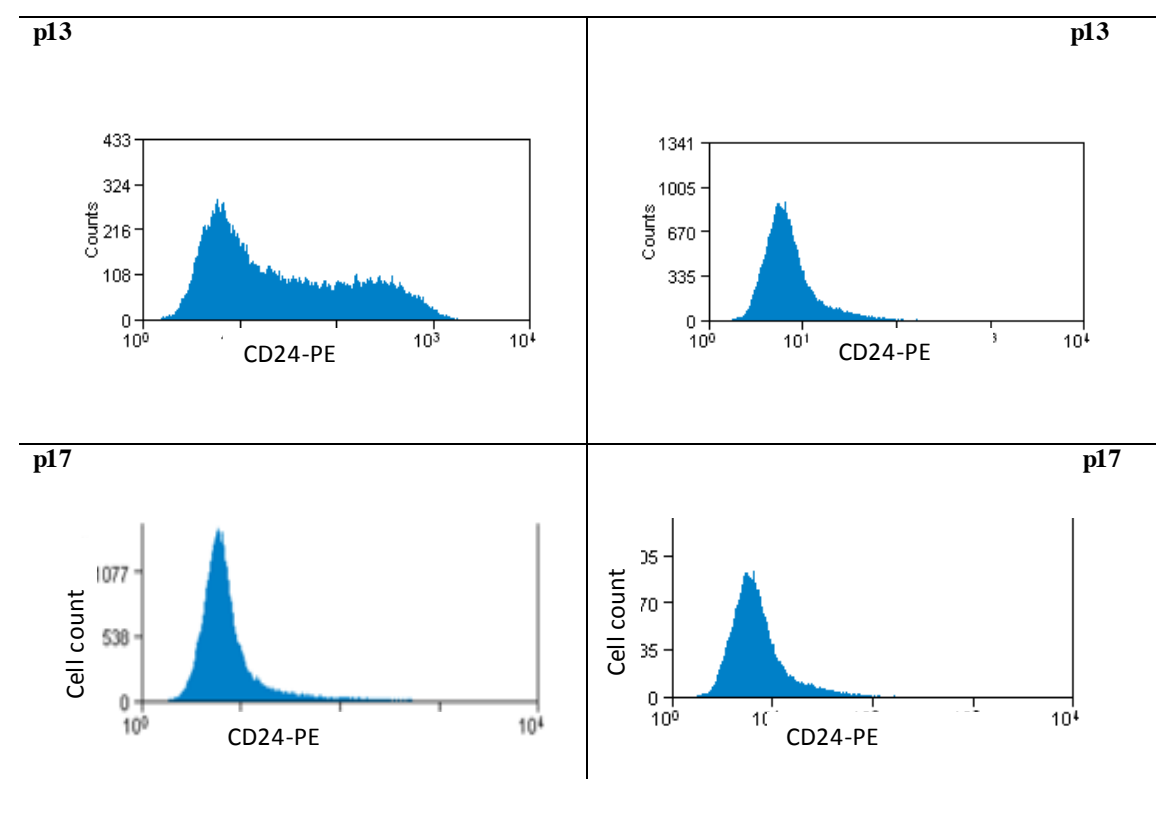


Fig 5.11 Change of CD24 expression in SW480_Clone C2. A. panel of SW480_Clone C2 at different passages. B. Panel of SW480_Clone D1 in the same passage as SW480_Clone C2. X-axis: CD24-PE, Y axis: Cell counts. 'p' indicates passage number.

5.3 Chapter discussion

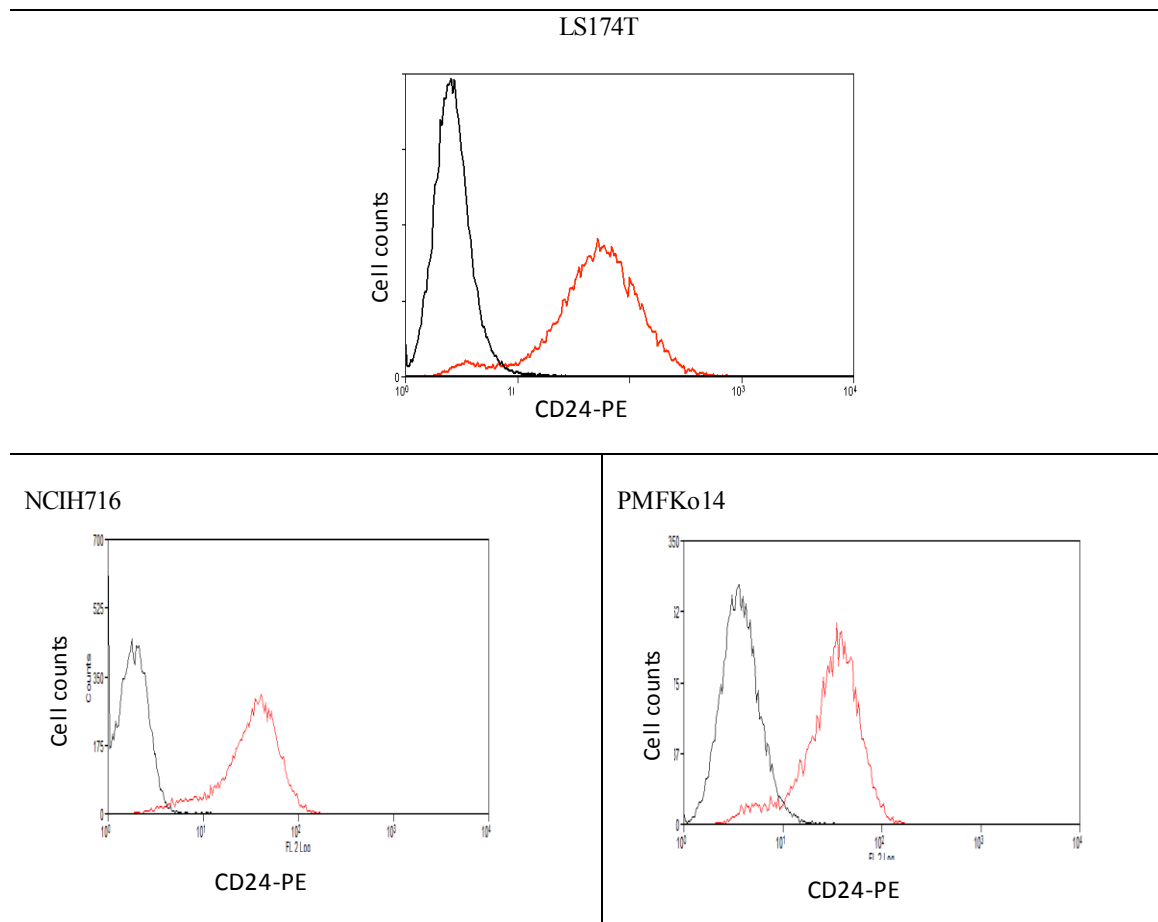
5.3.1 CD24 based heterogeneity in CRC cell lines

The observations from chapter 4 and this chapter provided convincing evidence that both LS174T and SW480 cell lines are heterogeneous and contain morphologically distinct clones which differ in the expression of CD24.

In LS174T, Clone 2 was representative of cells that appeared in the CD24⁻ tail, whereas Clone 1 represented CD24⁺ area of the FACS histogram (Fig 4.8). Clone 1 and Clone 2 were stably co-existing in LS174T and showed genetic differences in genes like *BRAF* and *TP53*. In the case of SW480, Clone C2 and Clone D1 were morphologically different and

early passages of SW480_Clone C2 were CD24⁺ (which changed in later passage) but SW480_Clone D1 was CD24⁻ throughout the time of the investigation. These observations prompted the hypothesis that more CRC cell lines might be heterogeneous regarding CD24 expression.

A supporting observation for this hypothesis came from data generated independently to this thesis work. Seven CD24⁺ CRC cell lines chosen for a drug study (courtesy: Jens Puschhof) from the panel of 96 cell lines and were analysed using flow cytometry (Fig 5.12). Out of them, cell lines NCIH716, PMFKo14, VACO10MS and LS411 showed bimodality in CD24 expression, similar to the LS174T. All these cell lines showed a CD24⁻ subpopulations in their FACS profile, though most of the cells were predominantly CD24⁺.



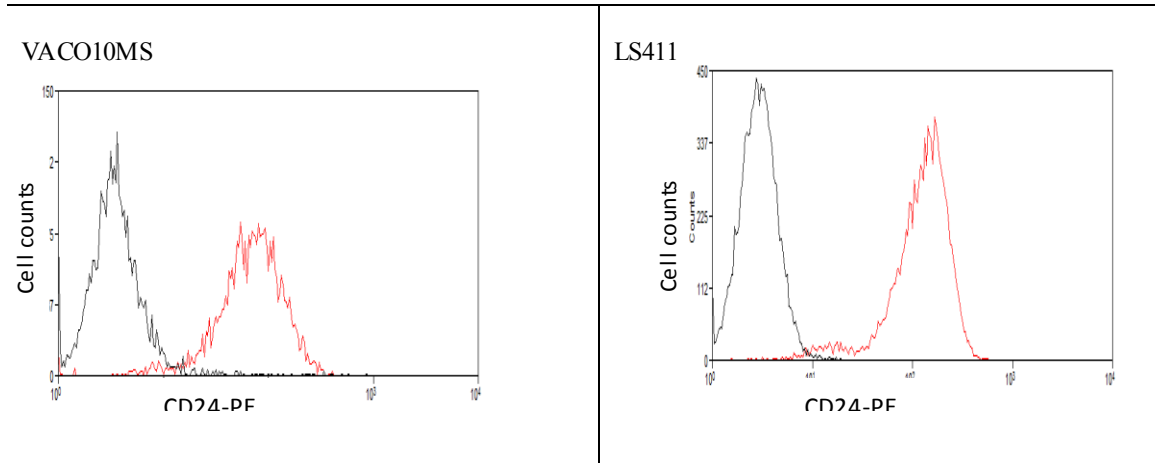


Fig 5.12: Bimodality of CD24 expression in cell lines. Like LS1754T, four other cell lines showed CD24⁺ subpopulation, suggesting heterogeneity of the cell line. Black histogram: Isotype control, Red histogram: CD24 level in respective cell lines.

Three other CD24⁺ cell lines- SW1222, C80 and SW2972 did not show such obvious bimodality. (Fig 5.13).

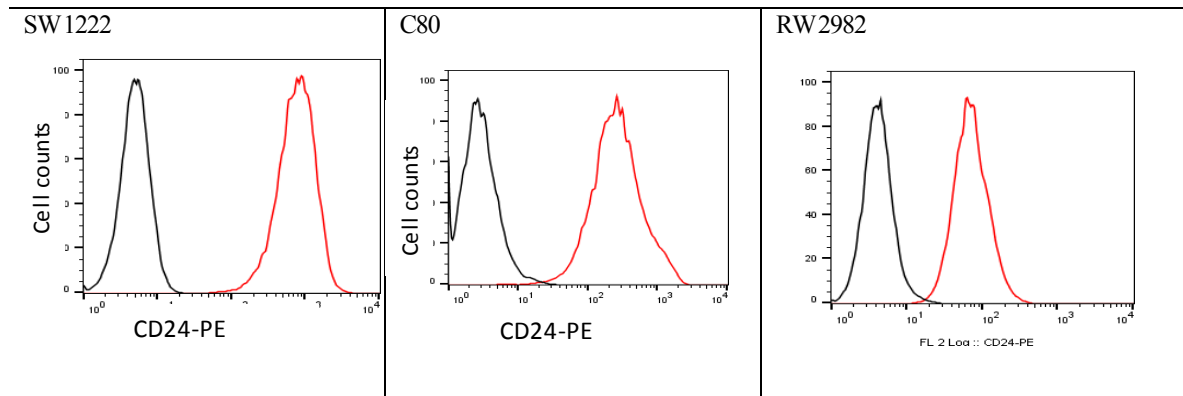


Fig: 5.13: Homogenous expression of CD24 in cell lines. (From left to right) SW1222, C80 and SW2972 did not show bimodal expression of CD24 as seen in other cell lines in Fig 5.12.

Red histograms: CD24 expression, Black histograms: isotype control.

In the cohort of five CD24⁺ cell lines, four (SW480, C3BBel, SW837, SNUC2B) showed presence of CD24⁺ subpopulations in the form of a tail on the right hand side of their FACS profile (Fig 5.14). As expected, the majority of cells were CD24⁺ in these cell lines.

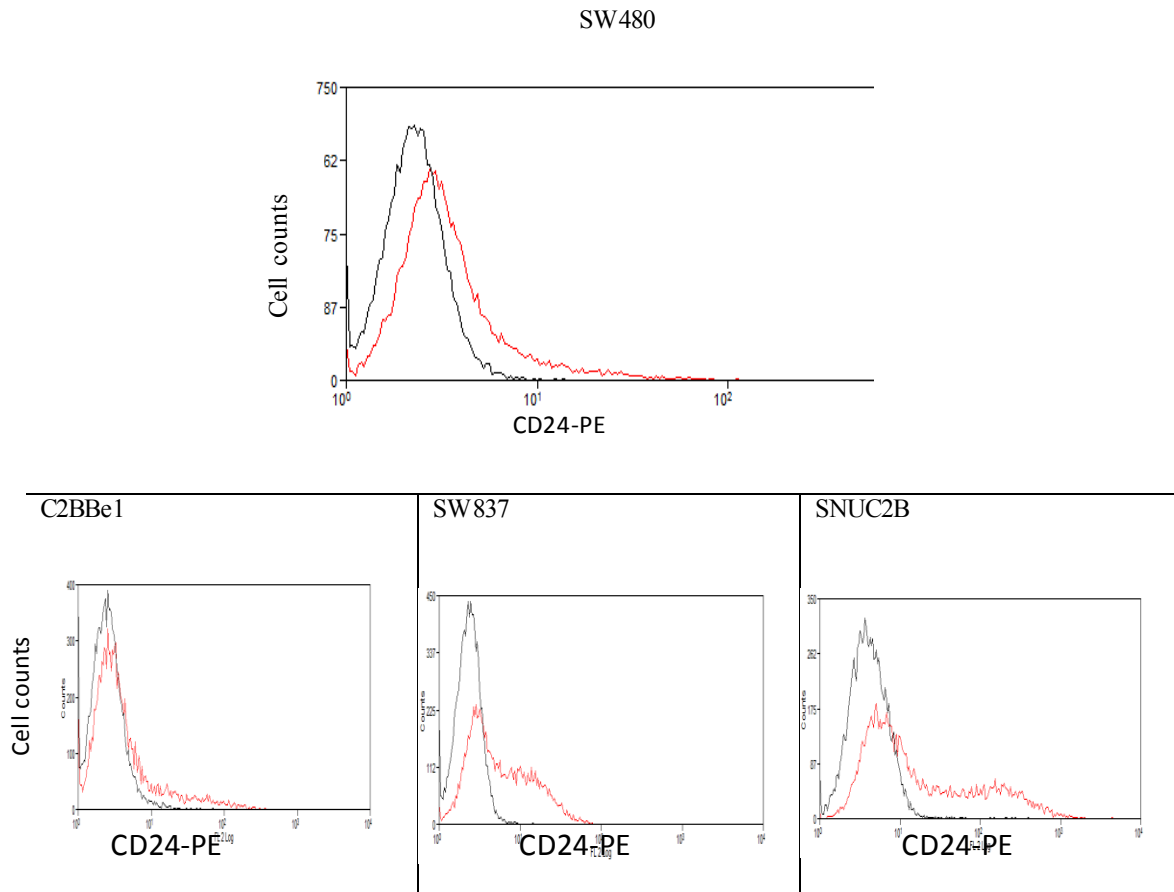


Fig 5.14: Heterogeneity in CD24⁻ cell lines. Like SW480, CRC cell lines C2BBel, SW837 and SNUC2B are mostly CD24⁻ with a minor subpopulation of CD24⁺ cells.

One CD24⁻ cell line, OXCO1 did not show any obvious CD24⁺ tail in its FACS profile (Fig 5.15) indicating presence of homogeneous population regarding CD24 expression.

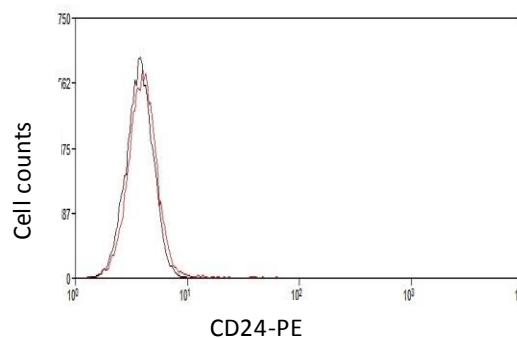


Fig 5.15: OXCO1 was homogeneously CD24⁻, unlike other heterogeneous CD24⁻ cell lines (Fig 5.14).

5.3.2 Using FACS profile as the suggestion for heterogeneous cell lines

The presence of CD24⁺ and CD24⁻ clonal populations in LS174T and SW480 cell lines was clearly evident in the bimodality of their FACS profiles. This CD24 based heterogeneity cannot be predicted from microarray, qPCR or western blot data as they will indicate total expression. However, flow cytometry revealed the presence of CD24⁻ subpopulations in the CD24⁺ cell lines and vice versa. Cell lines like NCIH716^{CD24+}, PMFKo14^{CD24+}, VACO10MS^{CD24+}, and LS411^{CD24+}, C3BBel^{CD24-}, SW837^{CD24-} and SNUC2B^{CD24-}, where FACS profiles suggest the presence of both CD24⁺ and CD24⁻ cell populations, can be considered as additional examples of heterogeneous cell lines. However, these cell lines need to be further investigated for the confirmation of stable presence of CD24⁺ and CD24⁻ cell populations. If confirmed, they can be additional models for *in vitro* study of tumour heterogeneity, and can provide important information to understand and explain heterogeneity in primary cancer.

Chapter 6

Methylation Mediated Regulation of CD24 in CRC Cell Lines

Chapter 6

6.1 Chapter introduction

The data in chapters 4 and 5 convincingly show the presence of CD24 differentially expressing subpopulations in the CRC cell lines, LS174T and SW480. In the LS174T cell line, two clones: Clone 1^{CD24+} and Clone 2^{CD24-} were representatives of the CD24⁺ and CD24⁻ subpopulations of the cell line respectively. In the SW480 cell line, Clone C2^{CD24+} and Clone D1^{CD24-} were two similarly representative clones. The co-existence of CD24⁺ and CD24⁻ clones in those cell lines raises the question as to the mechanism underlying the differential CD24 expression. The obvious possibilities are gene mutation, or promoter methylation affecting transcription controlled expression either directly or indirectly.

Mutation difference between CD24⁺ and CD24⁻ clones can be revealed by direct sequencing. Any mutation that results in a functional amino acid change, out of frame insertions or deletions leading to truncated proteins can impair the function of CD24. Alternatively, mutations in regulatory regions can lead to alterations in transcriptional control. Either one or both of these mechanisms could result in the loss of CD24 expression in CD24⁻ clones (such as in the Clone 2^{CD24-} of LS174T).

However, the observation that during the continuous culture, individual clones go through a marked change in the expression of CD24 over time, as measured by FACS, put the mutation-centric hypothesis into question. This suggests a more dynamic control of CD24 expression in these clones. Genetic mutation would be expected to lead to stable changes in

expression. Reversible mutations resulting in fluctuations in gene expression are unlikely to occur and be selected for over a short period of time in the culture. Thus, the observation of dynamic changes in *CD24* expression led to the hypothesis of promoter methylation-mediated regulation of *CD24* expression in the clones of LS174T and SW480, as methylation changes occur at a relatively higher frequency compared to mutations and so can be expected to modify the expression of *CD24* in a relatively short period time.

Investigation of this hypothesis required studying the methylation status of the regulatory region of *CD24* using the standard bisulfite sequencing technique. The results of these experiments revealed that *CD24* expression was strongly correlated with the methylation status of certain CpG positions in the regulatory region of this gene. Further extensive study confirmed the regulation of *CD24* expression in CRC cell lines by direct promoter methylation. All these results are described from here on.

6.2 Mutation is not responsible for down/up regulation of *CD24* in the clones of LS174T

In the Bodmer Laboratory, exome sequence data of all major cell lines including LS174T and SW480 are available. The analysis of these data showed no mutation in the *CD24* exomes in these two cell lines (internal lab data). However, such analysis may miss mutations which are present in a 'minor' clone of a cell line (as shown in section 4.9 for *TP53*). Therefore, if there is any mutation in the *CD24* gene of Clone 2^{CD24-}, which is a minor sub population in the LS174T, the mutation could go undetected from the high throughput exome sequence analysis. This limitation could be overcome by sequencing Clone 2^{CD24-} and Clone 1^{CD24+} separately.

During this study, whole genomes of the LS174T cell line and clones from this cell line (Clone 1^{CD24+} and Clone 2^{CD24-}) were sequenced and were accessible to extract sequence information of any gene. If mutations were responsible for the *CD24* down regulation in Clone 2^{CD24-} or upregulation in Clone 1^{CD24+} of LS174T, then that should be detectable by examining genome sequence differences between the clones. Such comparative analysis was done with the help of CBRG Bioinformatics Group of the Weatherall Institute of Molecular

Medicine (WIMM) and Dr. Kamila Koprowska of the Bodmer Laboratory as described earlier in section 4.9. The whole *CD24* gene, plus 1000 bp from both 3' and 5' regions, were analysed. Results of this comparison showed that there was no variation/change in the structural parts (exons, introns) of *CD24* gene. One heterozygous deletion of an Adenine (A) was found in the 5' upstream region of the longest of the seven splice variants of *CD24*. However, this change was not expected to affect transcription of any of the major splice variants as it was hundreds of base pairs away from their 5' UTRs. The sequence analysis further showed that there were five deletions in the 3' downstream region of *CD24* that varied between LS174T and Clone 1^{CD24+} or Clone 2^{CD24-} (Table 6.1). These deletions had no predicted impact on the expression or function of CD24 according to the algorithm developed by CBRG (in the WIMM). The algorithm predicts the severity of the variation, reported as a value between 0 and 6, by considering the site of the change and whether it would substitute amino acids. On this scale, a score of 6 indicates a functionally significant change and 0 indicates no functional significance.

Table 6.1: Changes in the up and downstream region of *CD24*. A total of 5 variations were found in the 3' UTR and 1 change was found in the 5' UTR of the *CD24* gene. The significance of all these changes was predicted as zero (according to the CBRG algorithm) *i.e.* not having any effect on the expression/function of the gene. Here, the first two columns show the position in the genome where the variations could be found. The reference allele sequence and the type of changes in LS174T and Clone 1 and Clone 2 are also listed.

Variant position from	Variant position to	Position in gene	Significance score	Reference allele	Observed allele	LS174T	Clone 1	Clone 2
106969510	106969512	3' UTR	0	AAA	Deletion	hom_del	hom_del	hom_del
106969693	106969706	3' UTR	0	CACAAAAAAAAAAAA	Deletion	het_del	het_del	wild type
106969694	106969707	3' UTR	0	ACAAAAAAAAAAAAA	Deletion	het_del	het_del	wild type
106969710	106969725	3' UTR	0	AAAAAAAAAAAAAAAAA	Deletion	het_del	wild type	hom_del
106969712	106969725	3' UTR	0	AAAAAAAAAAAAAAAAA	Deletion	het_del	hom_del	wild type
106976877	106976877	5' UTR	0	TCAAAAAAAAAAAAA	Deletion	het_del	het_del	het_del

It is noteworthy that all of these deletions were in the mononucleotides repeat, which is not unusual for an MMR deficient (Johannsdottir et al., 2000) cell line like LS174T. None of these sequence variations could, however, provide a likely explanation for the CD24 expression difference between Clone 1^{CD24+} and Clone 2^{CD24-}.

6.3 Dynamic change in CD24 expression in continuous culture

Since no genetic changes were identified that could provide a possible explanation for the CD24 differential expression in the clones of LS174T, the alternative hypothesis of methylation mediated regulation was investigated. This hypothesis was further supported by the observation of persistent gradual change of CD24 expression in the clones in continuous culture.

6.3.1 If grown separately for a long time, clones tend to approach their parental profiles

If CD24 differentially expressing clones were continuously maintained in culture, they tended to shift from their initial homogeneity towards a more heterogeneous expression profile over time.

This plastic change in CD24 expression was bidirectional (described below). The clones would change the expression in both directions to approach their parental profile. However, the changes can occur at different rates among clones from different cell lines or different clones of the same cell line.

It has been shown in chapter 4 that early passage (p5-p25) Clone 1 of LS174T is CD24⁺ and early passage Clone 2 (p5-p20) is CD24⁻. Parental LS174T contains both CD24⁻ and CD24⁺ populations (Section 4.5).

However, when Clone 1^{CD24+} was maintained in the culture for a long time, it regained a CD24⁻ subpopulation after being passaged for more than 50 times (p50 equivalent to ~ 7 months in culture). The FACS profile of this late passage Clone 1 (Fig 6.1 B) looked very

similar to parental LS174T (Fig 6.1 A). The early Clone 1^{CD24+} (p7-p30) lacked the pronounced presence of the CD24⁻ subpopulation (Fig 6.1 C).

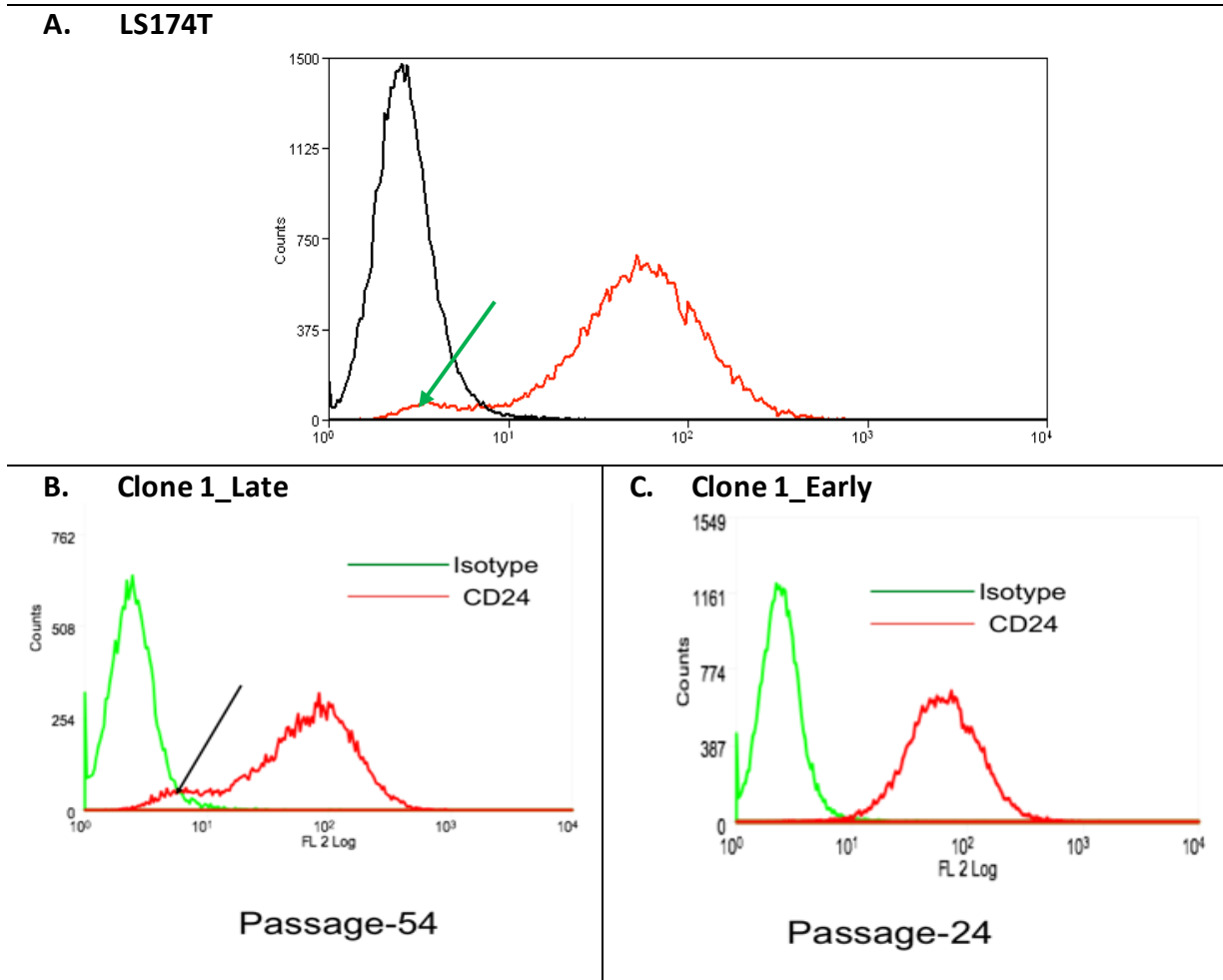


Fig 6.1: Change in CD24 expression in the late passages of LS174T Clone 1: A. Parental LS174T has a bimodal expression profile for CD24 (red histogram). The green arrow indicates a CD24⁻ subpopulation which can be interpreted to be due to the presence of a Clone 2^{CD24-} population, whilst the CD24⁺ population of the histogram represents Clone 1^{CD24+}. B. In late passage Clone 1, a CD24⁻ subpopulation appeared (indicated by the black arrow) and its CD24 profile looks very similar to the parental LS174T. C. In early passage Clone 1^{CD24+}, all cells are CD24⁺ (red histogram). Black histogram (in A) and Green histograms (in B and C) indicates isotype control.

A shift in the pattern of CD24 expression in Clone 2 was also observed. By passage 40, it increased its CD24⁺ proportion considerably (Fig 6.2).

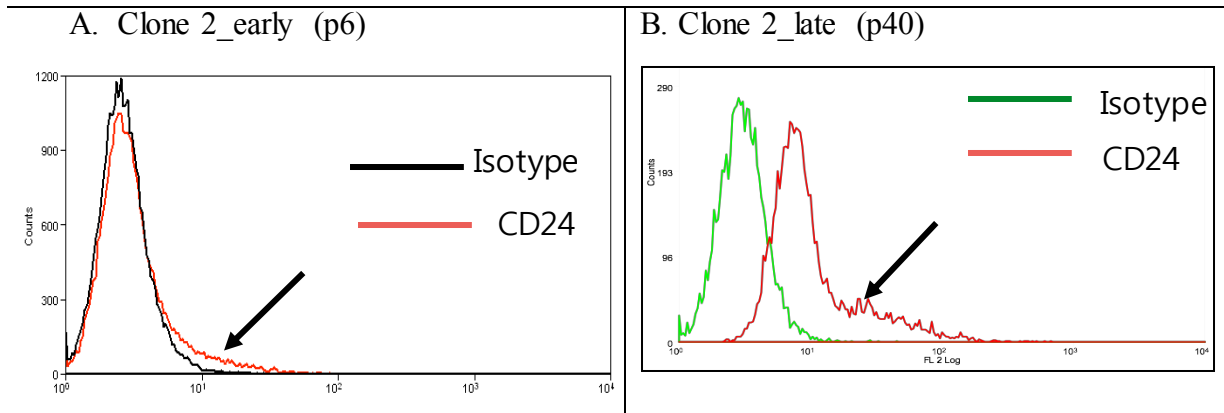


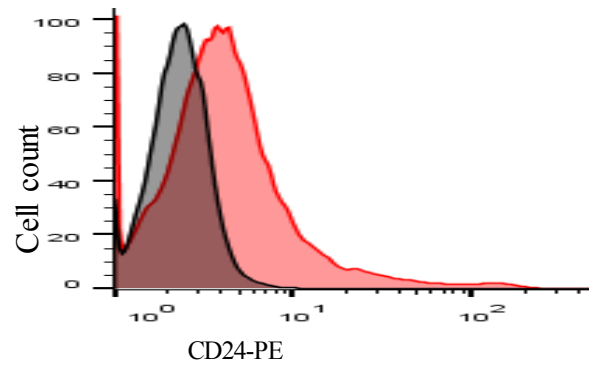
Fig 6.2: Increase of CD24⁺ cell populations in Clone 2. A. Early Clone 2 is CD24⁻ (only a very small CD24⁺ tail, indicated by an arrow on red histogram). B. Late passage Clone 2 produced a significant CD24⁺ population (indicated by the shoulder on red histogram).

A similar change towards the parental expression profile of CD24 occurred in the clones of SW480. Here, parental SW480 was mostly CD24⁻ (Fig 6.3 A). Between two clones analysed from SW480, Clone C2 was mostly CD24⁺ and Clone D1 was CD24⁻ (Fig 6.3 B)

However, in the subculture, later passages of SW480 Clone C2 started to lose CD24 expression with time (Fig 6.3 C). On the contrary, Clone D1 which had a CD24 profile similar to SW480 from the beginning, did not change.

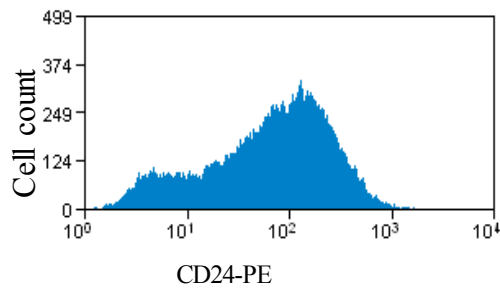
By passage 17 (p17), Clone C2 lost most of its CD24 expression and its FACS profile looked very much like the parental SW480 (Fig 6.3 C).

A. Parental SW480_ p20

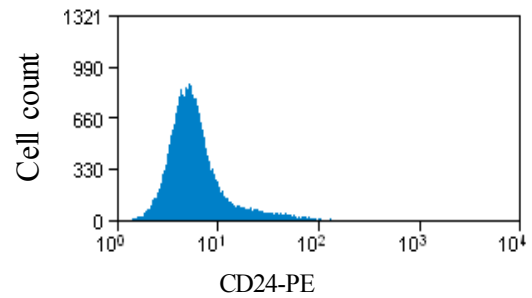


B. Early passage Clone C2^{CD24+} and Clone D1^{CD24-}

Clone C2_ p5

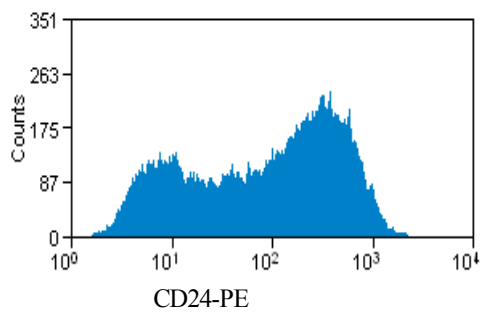


Clone D1_ p5

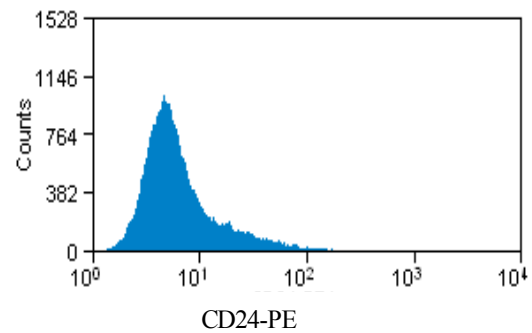


C. Changes in CD24 expression in Clone C2

Clone C2_ p10



Clone D1_ p10



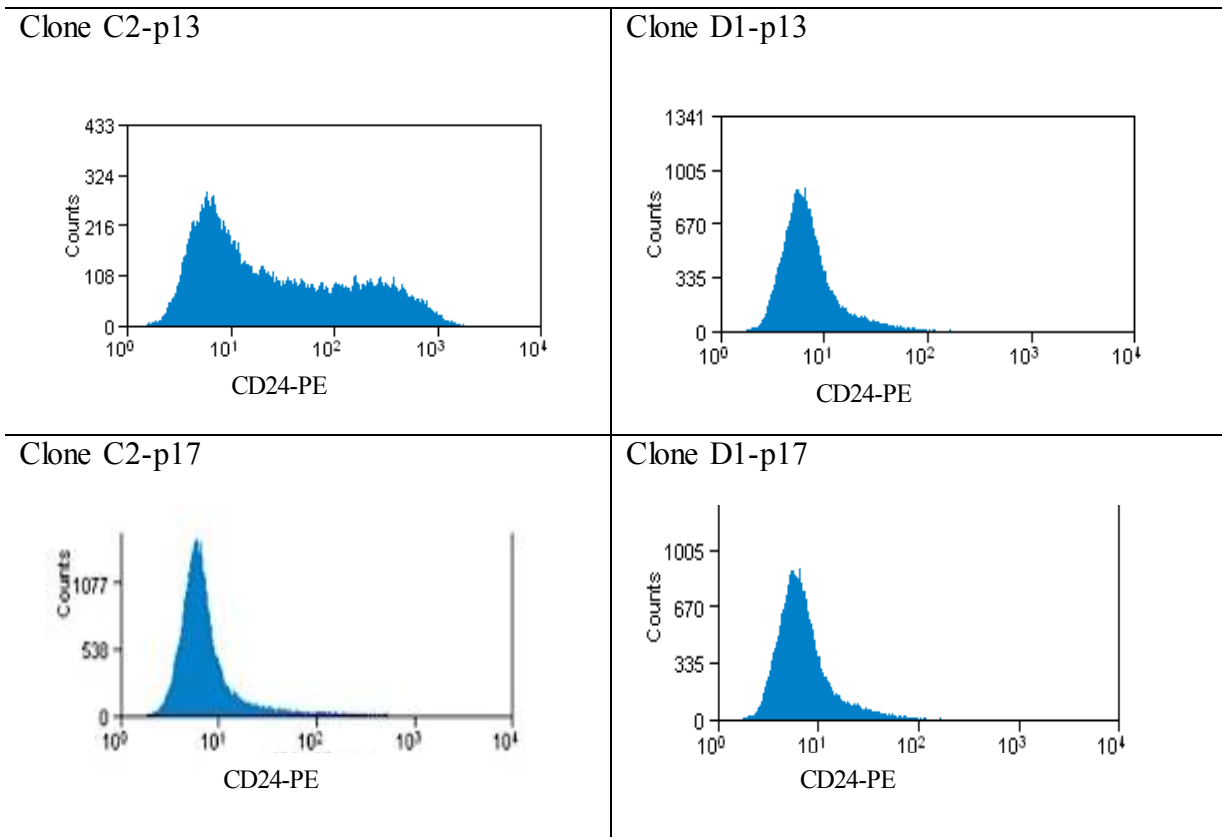


Fig 6.3 The change of CD24 in SW480 Clone C2 towards parental expression profile. A. Parental SW480 shows mostly CD24⁻ profile with a small CD24⁺ tail (red histogram). B. Early Clone C2 (left) is mostly CD24⁺ whereas early Clone D1 is CD24⁻. C. Clone C2 gradually changes its expression of CD24 but Clone D1 remains CD24⁻ throughout the time. By passage 17, Clone C2 is mostly CD24⁻ which is very similar to the parental SW480 profile. X-axis: CD24-PE, Y-axis: Cell counts.

It is noteworthy that compared to Clone 1^{CD24⁺} and Clone 2^{CD24⁻} of LS174T, where the change of CD24 expression took place at quite late passages, Clone C2 of SW480 changed the expression profiles of CD24 relatively earlier.

However, both examples of clones from LS174T and SW480 showed that the direction of change in CD24 expression was towards an increasing resemblance of their parental expression profiles. Similar trend was found in clones derived from another CRC cell lines: DLD1 (see in appendix, Ax.6).

6.3.2. CD24 expression change is bidirectional

Data from the clones of LS174T and SW480 showed that CD24 expression could change from positive to negative and vice versa. Such bidirectional change in CD24 expression was observed more clearly in further subcloning experiments.

The early passages of Clone 2^{CD24-} (p5-p20) of LS174T are mostly CD24⁻, but they contain a small fraction (around 7% - 9%) of CD24⁺ cells that are detectable as a positive tail in a FACS profile (Fig 6.4).

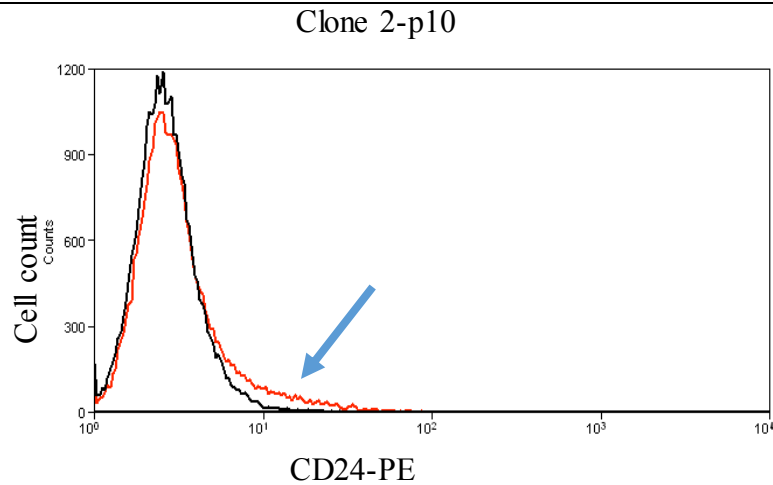


Fig 6.4: CD24⁺ cells in LS174T Clone 2. A small portion (~7-9%) of Clone 2^{CD24-} cells are CD24⁺ which appears as a tail on the right hand side of the FACS profile (indicated by blue arrow). Red histogram: CD24-PE. Black histogram: Isotype control. 'p' indicates passage number.

In order to further study the stability of CD24 expression, five subclones were established from single cells sorted from the top 3% of CD24⁺-expressing cells in the LS174T Clone 2^{CD24-} (at passage 11). These five subclones were then followed for changes in CD24 expression profiles over several passages. When analysed by FACS (Fig 6.5), subclone 1 (Sub1) and subclone 4 (Sub4) were predominantly CD24⁻ with FACS profiles indistinguishable from the original Clone 2^{CD24-}, from which they were subcloned and were not therefore further analysed at later passages. Subclone 2 (Sub2) initially had a bimodal distribution for CD24 expression but very quickly became predominantly CD24⁻ by passage 4. Interestingly, subclone 3 (Sub3) showed enrichment for CD24⁺ cells at early passage

which it maintained at passage 8 and subclone 5 (Sub5) (at passage 8) showed bimodality in CD24 expression.

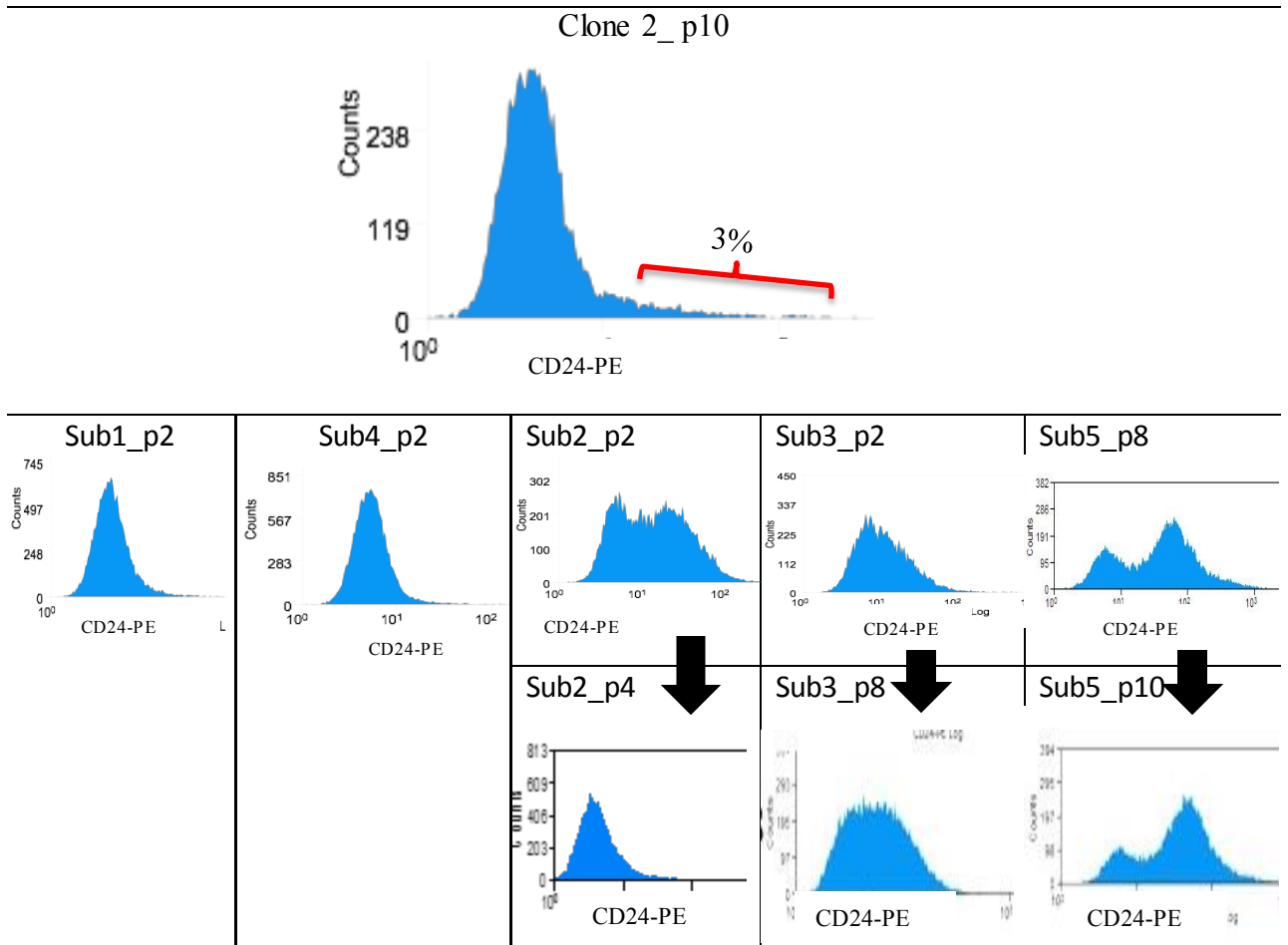


Fig 6.5: Flow cytometry histograms of CD24 expression in subclones established from the top 3% of the CD24⁺ cells of LS174T clone 2^{CD24-} (indicated by red bar). Among five subclones, subclone 5 (Sub 5) and 3 (Sub 3) showed expression of CD24. Subclone 2 initially showed bimodal expression but then became CD24⁻ very quickly. Subclone 1 and 4 (left) remained CD24⁻ from the beginning. In all images 'p' indicates passage number.

The CD24 expression profiles of both subclones 3 and 5 were then further followed and periodically analysed until passage 23. Subsequent passages of subclone 3 in continuous culture showed significant changes in CD24 expression. Up to passage 11, it showed CD24⁺ enrichment, but at passage 18, enrichment for CD24⁻ cells became prominent. By passage 23, subclone 3 settled as mostly CD24⁻, with a CD24⁺ tail (Fig 6.6 A).

In contrast, the trend in Subclone 5 was opposite. Here, the initial expression profile of CD24 was bimodal. In later passages, however, the proportion of CD24⁺ cells increased and Subclone 5 became stable (at passage 18) as mostly CD24⁺ with a minor subpopulation of CD24⁻ cells (Fig 6.6 B).

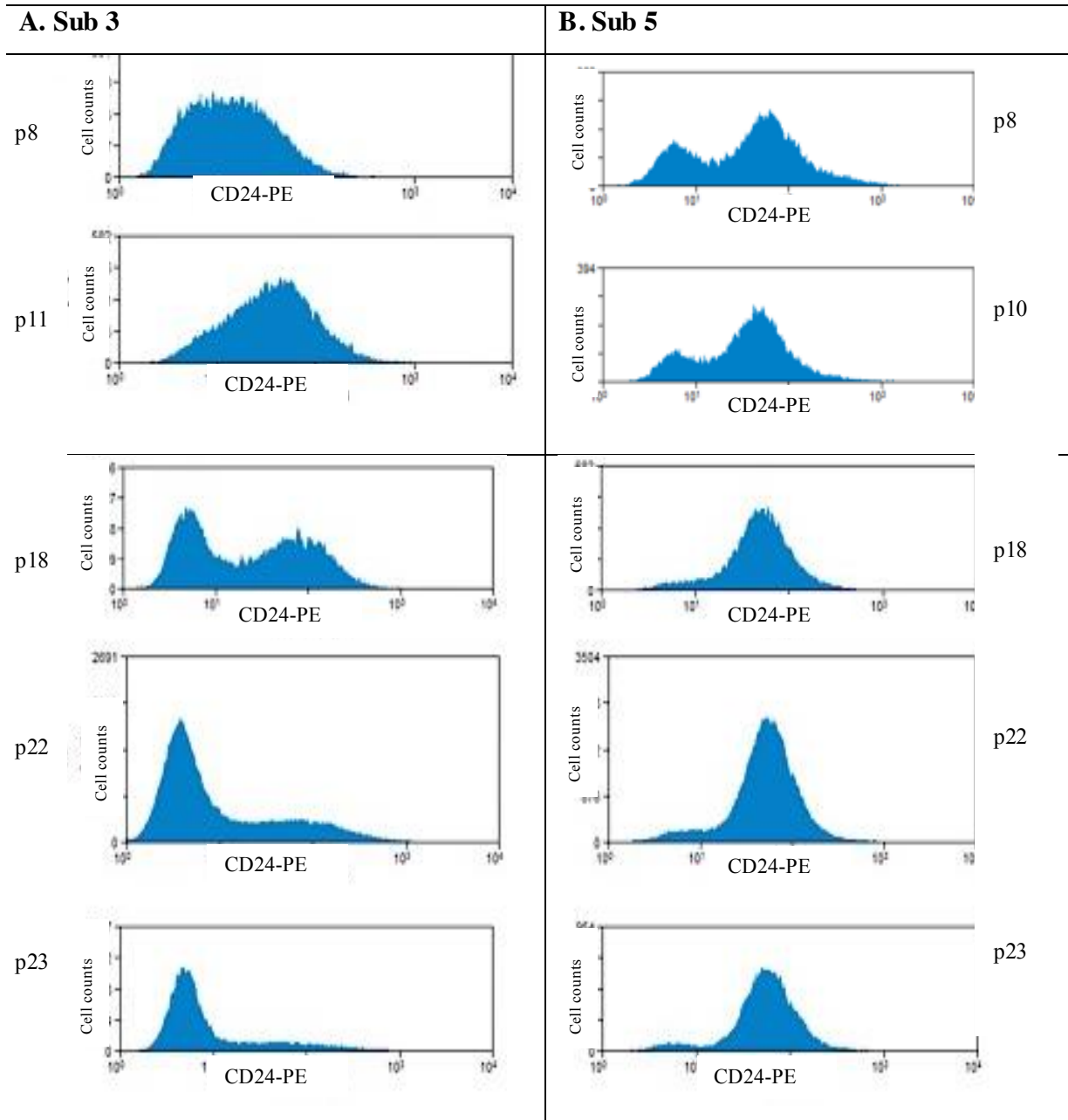


Fig 6.6: Subclone 3 (Sub 3) and Subclone 5 (Sub 5) of LS174T clone 2^{CD24⁻} showed opposite behaviour in CD24 expression over time. A. In Sub 3, expression of CD24 increased until passage 11. However, by passage 18, the CD24⁻ population started to become prominent and by passage 23, Sub 3 stabilised with a predominantly CD24⁻ profile that was

very similar to the profile for LS174T clone 2^{CD24⁻} (Fig 6.4), from which it was derived. B. Sub 5 shows a bimodal expression of CD24 at early passage but with progression in culture it became mostly CD24⁺ with an expression profile almost similar to parental LS174T (Figure 6.1A).

These data from subclones of Clone 2 of LS174T showed that expression changes of CD24 were bidirectional. In particular, one subclone, Sub 3, moved towards a more negative direction approaching its parental (Clone 2) profile (CD24⁻), and another subclone, Sub 5, moved in a positive direction, approaching the LS174T (parent of Clone 2) profile.

The key conclusion from all these observations is the plasticity in CD24 expression among clones and subclones of CRC cell lines. This dynamic change further strengthens the assumption made earlier that mutations might not be responsible for the differential expression of CD24 in Clones of LS174T (or in SW480). It is also consistent with the hypothesis of epigenetic regulation of *CD24*, because with this mechanism, it is much easier to explain the observed plastic change in CD24 expression.

6.4 Response to 5-aza-2'-deoxycytidine treatment indicates methylation mediated regulation of CD24 expression

6.4.1 Treatment with 5-aza-2'-deoxycytidine increases CD24 expression in CD24⁻ clones

Two key conclusions from sections 6.2 and 6.3, respectively, are that (i) there is no mutational difference between clones to explain their differential expression of CD24 and (ii) CD24 expression is a plastic and dynamic process.

These observations provided initial support for the hypothesis that *CD24* is under epigenetic control in those clones (Clone 1 and Clone 2 of LS174T and Clone C2 and Clone D1 of SW480) and subclones (Sub 3, Sub 5 of Clone 2). This is explainable by the notion that when individual clones are maintained in the culture, their divisions might allow for the

possibility of selection for variation in CD24 expression in either direction after being affected by methylation, and once they reach the parental composition, a favourable condition might be achieved by both CD24⁺ and CD24⁻ populations in the culture environment. During their 50 passages over 7 months in the culture, the clones of LS174T may have undergone around 150 divisions (on average 3 divisions per passage). Even if one cell with a different CD24 expression profile arises by chance and is selected for, sufficient numbers of cells will be generated from successive divisions within the course of 7 months to render them detectable by FACS analysis. In the absence of mutation-driven effects, the most plausible mechanism for such changes is DNA methylation which is a more frequent way for changing expression than mutation, due to its higher frequency of occurrence.

A test to gather evidence for the methylation mediated regulation of CD24 expression was initially carried out with 5-aza-2'-deoxycytidine treatment of Clone 2^{CD24⁻} of LS174T.

Methylation of CpG positions in the promoter region of genes is a key epigenetic mechanism of gene regulation (Deaton & Bird, 2011). To see if any gene is under epigenetic regulation, cells are often treated with 5-aza-2'-deoxycytidine (Christman, 2002). This is a major epigenetic drug, which prevents the function of methylase enzymes responsible for adding methyl group on target CpGs during replication (Lavelle, Sauntharajah, & DeSimone, 2008). 5-aza-2'-deoxycytidine can covalently bind to methylase enzyme which eventually leads to the loss of methylation from these CpG positions and makes the promoter accessible for transcription factors. Arguably, if in Clone 2^{CD24⁻}, methylation was responsible for suppression of *CD24*, then 5-aza-2'-deoxycytidine treatment would rescue its expression by removing methyl group from CpG positions. To test if 5-aza-2'-deoxycytidine had any effect on CD24 expression, Clone 2^{CD24⁻} was treated with this drug.

For these experiments, 5-aza-2'-deoxycytidine dissolved in acetic acid was used according to manufacturer's suggestion. Generally, the recommended dose of 5-aza-2'-deoxycytidine varies between 1 μ M and 10 μ M based on target cells. However, 5-aza-2'-deoxycytidine is toxic on cells at high dose. An initial treatment of Clone 2 with 5 μ M of 5-aza-2'-

deoxycytidine showed evidence of toxic effects (Fig 6.7). Therefore, the effects of different concentrations of 5-aza-2'-deoxycytidine were first checked on Clone 2.

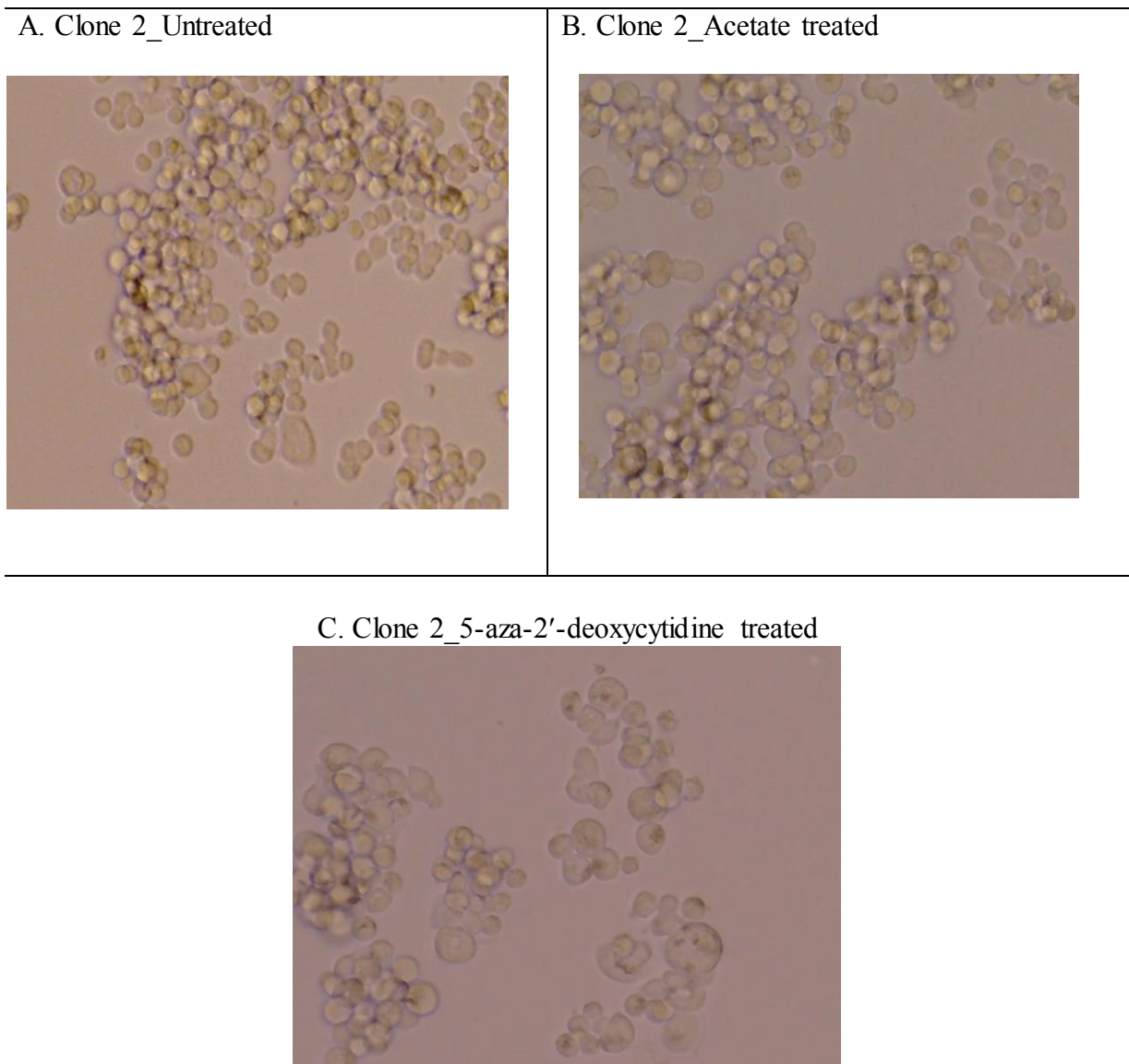
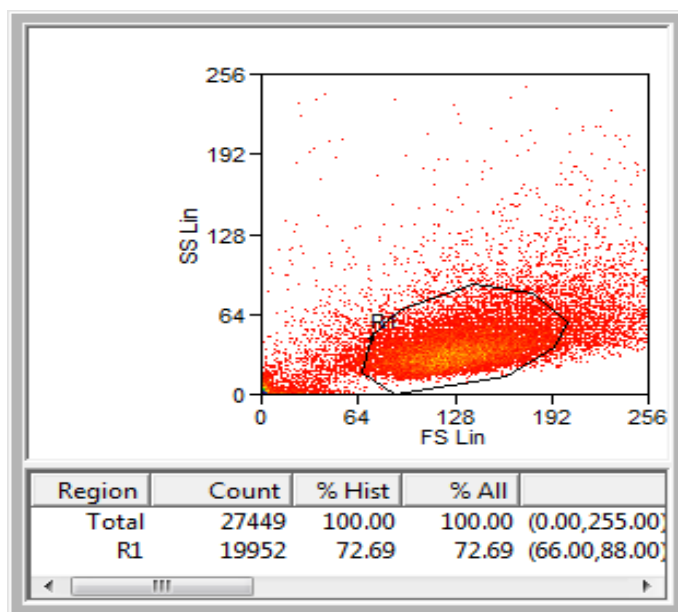


Fig 6.7: Effect of 5-aza-2'-deoxycytidine on Clone 2. A. Untreated Clone 2 and B. Acetate treated Clone 2 was morphologically very similar and their numbers were comparable. C. 5-aza-2'-deoxycytidine treated Clone 2 cells were bloated and fewer in number.

In order to identify a suitable concentration with minimal toxicity while maintaining efficient demethylation, the cells were treated with a range of concentrations of 5-aza-2'-deoxycytidine and then assayed by FACS for CD24 expression.

Doses of 0.5 μ M, 1.0 μ M, 3.0 μ M and 5.0 μ M were used for up to 5 days and the 5-aza-2'-deoxycytidine was changed every 24 hours (as it is unstable in culture medium). Cells were then analysed for CD24 expression using FACS. The scatter plots during this FACS analysis also provided a measure of fitness of the treated cells (Fig 6.8). The healthier cells were more homogeneous and concentrated in the FACS scatter plots. In case of the acetate control (Fig 6.8 A) almost 73% cells were within the central area plot, indicating good health and viability of the cells. Cells outside of the central cluster are considered non-viable or bloated. The analysis showed that among the range of concentrations of 5-aza-2'-deoxycytidine used, cells were most vigorous at 0.5 μ M concentration and least viable at 5.0 μ M concentration of 5-aza-2'-deoxycytidine (Fig 6.8 B). The same area in the scatter plots contained 46% cells at 0.5 μ M concentration of 5-aza-2'-deoxycytidine compared to the 24% at 5 μ M (Fig 6.8).

A. 5.0 uM Acetate control



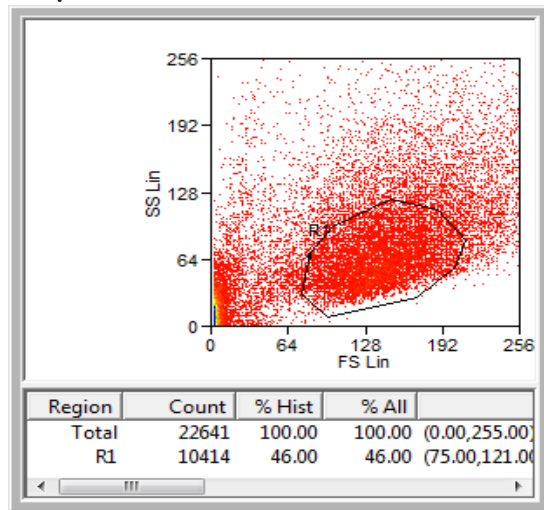
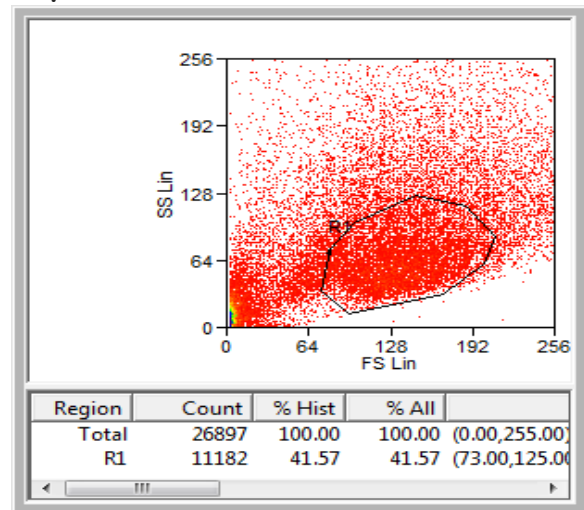
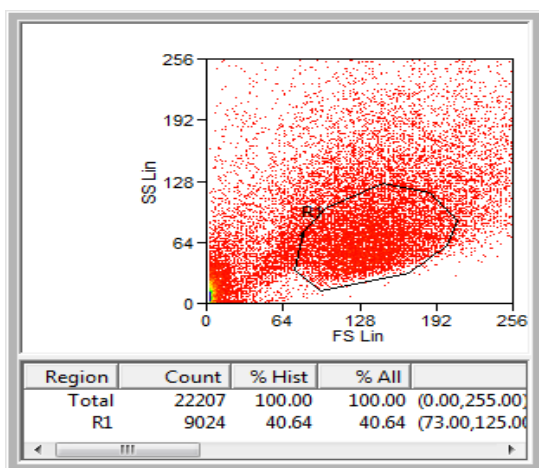
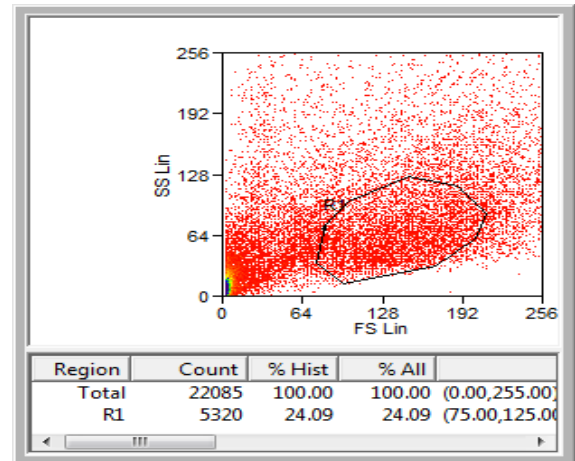
B. Different concentrations of 5-aza-2'-deoxycytidine treatment**0.5 μ M****1.0 μ M****3.0 μ M****5.0 μ M**

Fig 6.8: Effect of 5-aza-2'-deoxycytidine on the viability of Clone 2 cells. A. Control test: Clone 2 treated with 5 μ M Actetate showed very viable homogenous concentration in the middle in the scatter plot. B. Treatment with 5-aza-2'-deoxycytidine shows effects on viability of cells as % of cells in the same 'R1' area decreases with increased concentration of the drug. In all plots, the X axis shows forward scatter and the Y axis shows side scatter.

When the CD24 expression levels in the viable cell populations were compared in those concentrations, it was found that 5-aza-2'-deoxycytidine treatment increased the expression of CD24 in Clone 2 of LS174T in all cases (Fig 6.9). Compared to the acetate treated control, 5-aza-2'-deoxycytidine treated Clone 2 showed shifts of the histograms to the right in the FACS profile, indicating a higher expression of CD24. Treatment with 5.0 μM 5-aza-2'-deoxycytidine showed highest re-expression of CD24. However, even the 0.5 μM treatment was able to induce a measurable increase in CD24 expression, and given the considerable toxicity at higher doses, further experiments were conducted using 0.5 μM 5-aza-2'-deoxycytidine.

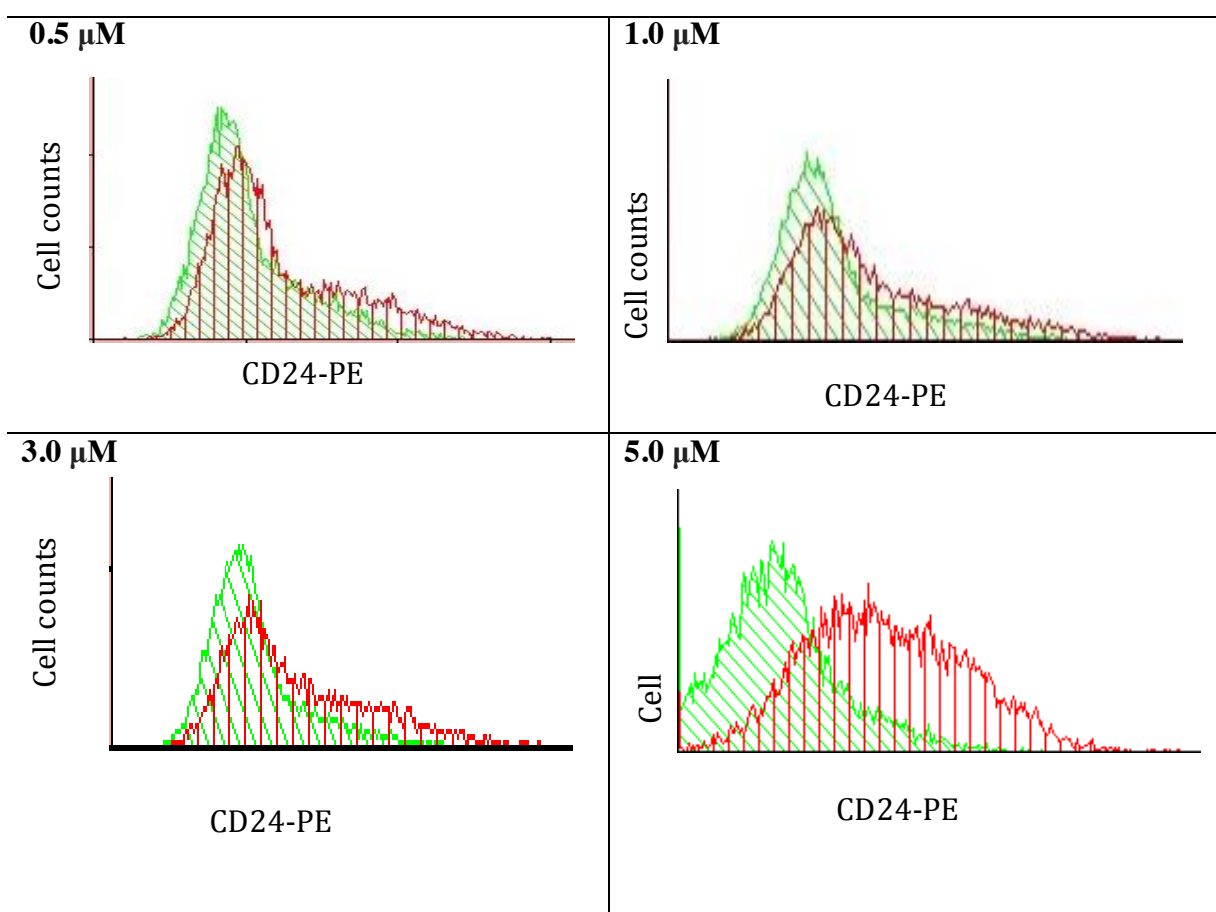


Fig 6.9: Effect of 5-aza-2'-deoxycytidine on CD24 expression. Treatment of LS174T Clone 2 with different concentration of 5-aza-2'-deoxycytidine resulted in enhanced expression of CD24. The positive effect of expression was found at as low as 0.5 μM concentration, although it was highest at 5.0 μM . Red: CD24 expression after 5-aza-2'-deoxycytidine treatment, Green: Acetate control.

The treatment with 0.5 μ M concentration was repeated again and 5-aza-2'-deoxycytidine treated Clone 2^{CD24-} cells were compared with untreated and Acetate treated controls. The results confirmed that CD24 level had increased in Clone 2 after 0.5 μ M 5-aza-2'-deoxycytidine treatment (Fig 6.10).

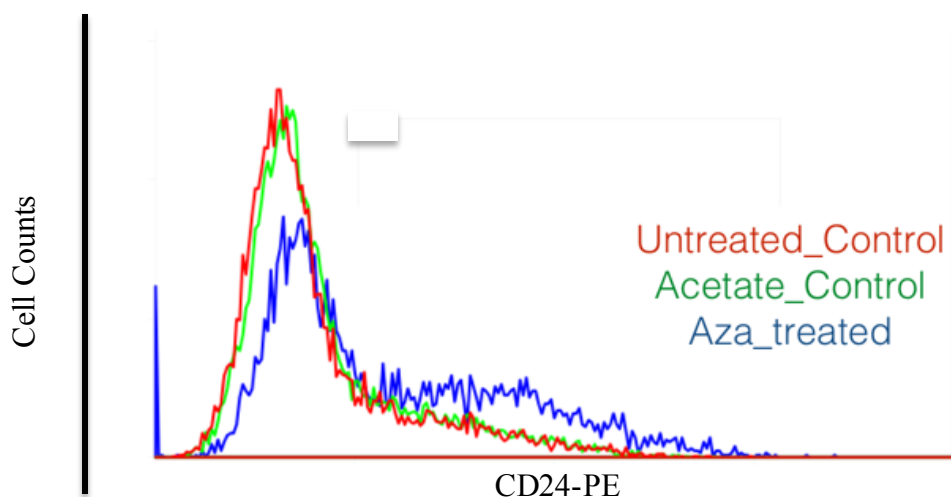


Fig 6:10: Enhanced expression of CD24 in Clone 2 after 0.5 μ M 5-aza-2'-deoxycytidine treatment. Untreated and Acetate treated controls showed no change in CD24 expression, but 5-aza-2'-deoxycytidine treatment showed increase expression with a positive shift.

qRT-PCR was performed to check the message level after 5-aza-2'-deoxycytidine treatment in the LS174T Clone 2^{CD24-}. A time course study showed that CD24 message level increased in 5-aza-2'-deoxycytidine treated Clone 2^{CD24-} compared to acetate control by up to 8-fold at day 3 (Fig 6.11 A). After 24 hours, the fold increase was 2.5, indicating that longer treatment increased the expression of CD24 in Clone 2. In all these time points, CD44 mRNA expression levels did not change (ratios of expression in treated versus untreated cells are close to 1 (Fig 6.11 B), indicating that CD44 expression is not directly or indirectly regulated by promoter methylation in this cell line. It is noteworthy that the basal level of CD44 in Clone 2 of LS174T is high and comparable to that of Clone 1 (Fig 4.8) and therefore, it can be inferred that *CD44* is already unmethylated at its promoter and other control regions.

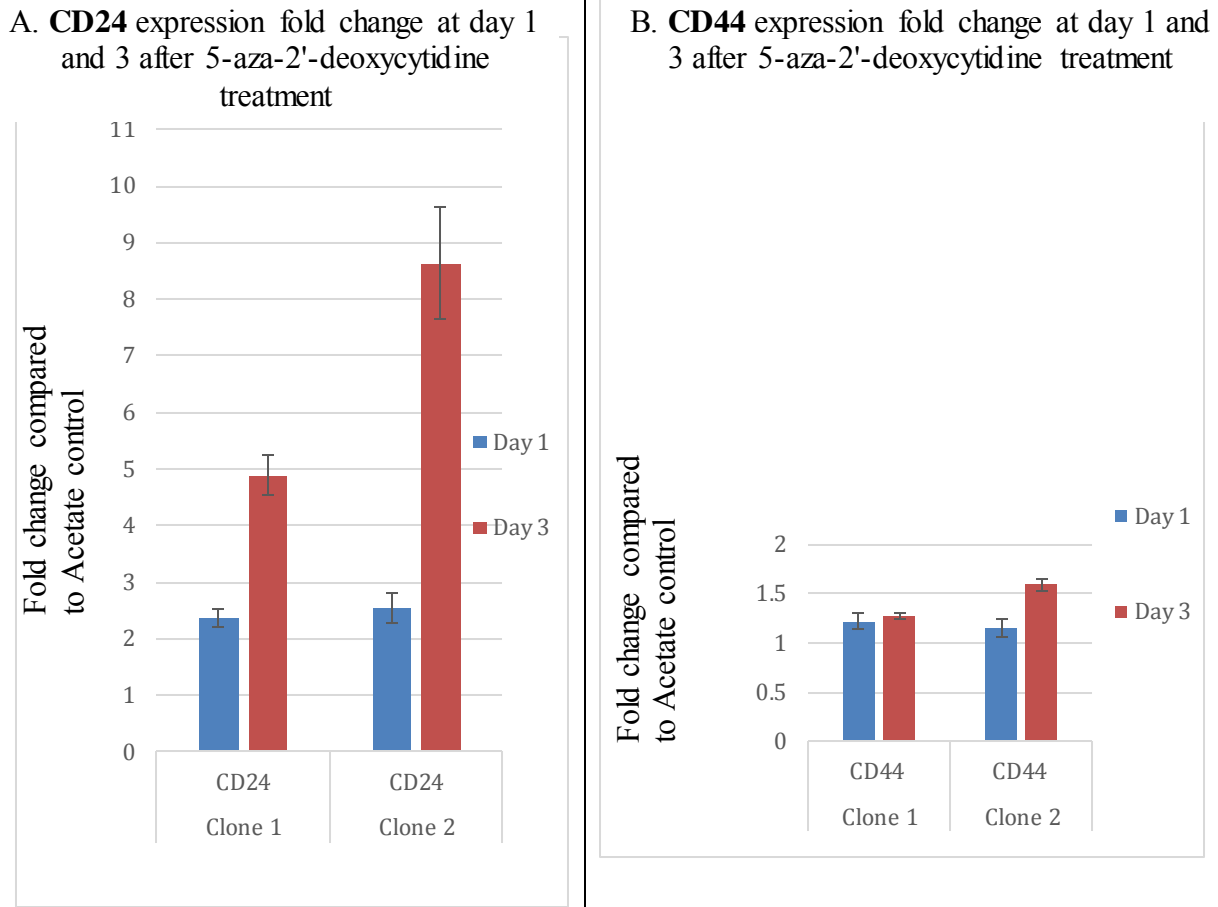


Fig 6.11: qRT-PCR of 5-aza-2'-deoxycytidine (0.5 μ M) treated Clone2^{CD24-}. A. CD24 level increases after the treatment. At day 3 (red bars) the increase is higher than day 1 (blue bars). In Clone 2^{CD24-} the CD24 expression is more than 8-fold higher in 5-aza-2'-deoxycytidine treated cells compared to Acetate control. In Clone 1^{CD24+} there is a modest increase after the treatment. B. There is no notable change in CD44 level in clone 2 (or in Clone 1) after 5-aza-2'-deoxycytidine treatment even at day 3. Error bars indicate the standard error of mean (S.E.M) of the individual fold changes calculated by $\Delta\Delta$ CT method from every experimental repeats.

6.4.2 5-aza-2'-deoxycytidine treatment in CD24⁻ cell lines increases CD24 expression

The observations of increase in the CD24 expression in Clone 2^{CD24-} of LS174T after 5-aza-2'-deoxycytidine treatment provided strong evidence for methylation mediated regulation of CD24.

In order to identify whether the above methylation effect in Clone 2^{CD24-} could be found in other cell lines, microarray expression data of 10 cell lines that were treated with 5-aza-2'-

deoxycytidine were analysed (Table 6.2). This expression data was available from previous work in the Bodmer laboratory where these 10 cell lines, treated with 5-aza-2'-deoxycytidine and also untreated counterparts, were analysed for gene expression using the Affymetrix microarray platform (internal lab data).

Of the 10 cell lines, the baseline CD24 expression is absent/low in three (CC20, RKO, Vaco5), moderate in two (HCT116, SW48) and high in six (HCA7, NCIH-716, DLD1, LS174T, LS411 and PC-JW) cell lines as can be seen in Table 6.2.

When compared with untreated counterparts, three 5-aza-2'-deoxycytidine treated CD24⁺ cell lines (CC20, RKO, Vaco5) showed the highest fold increase in CD24 expression. Moderately expressing lines (HCT116, SW48) showed a modest (~2 fold) increase and the CD24⁺ cell lines did not show any response at all (Table 6.2).

Table 6.2: Basal level of CD24 mRNA in 10 cell lines and fold increase in expression after treatment with 5-aza-2'-deoxycytidine. The basal level indicates the normalized linear value obtained from microarray analysis. Several of the cell lines with high basal levels of CD24 showed slight decreases in expression after 5-aza-2'-deoxycytidine, which could be due to toxic effects on gene transcription, or could reflect re-expression of epigenetically controlled negative regulators of CD24 expression in those cell lines.

Cell line	CD24 mRNA (normalized value)	Fold change after treatment
CC20	97	4.47
RKO	217	3.26
Vaco5	534	4.2
HCT116	1780	1.94
SW48	5457	1.39
HCA7	18976	0.26
NCIH-716	19896	0.21
DLD1	21104	0.58
LS174T	29992	0.48
LS411	64652	0.24
PC-JW	83611	0

This analysis suggests that when 5-aza-2'-deoxycytidine treatment causes demethylation of DNA, the cell lines with low basal level of *CD24* re-express *CD24* again. This indicates that methylation is the cause of *CD24* suppression in those cell lines.

Above observations, however, does not indicate whether methylation regulates *CD24* directly (promoter methylation of *CD24*) or indirectly (promoter methylation of the regulators of *CD24*), as 5-aza-2'-deoxycytidine removes methylation globally in the cell. The following section discusses experimental evidence for the direct methylation of *CD24* promoter.

6.5 Evidence of direct regulation of *CD24* by promoter methylation

The above protein and mRNA expression data for *CD24* in Clone 2^{CD24-} of LS174T and 3 other *CD24*⁻ cell lines showed that *CD24* could be re-expressed by preventing the methylation of the DNA in the cells. However, those data did not confirm whether *CD24* was directly methylated or if its suppression in Clone 2^{CD24-} and in other cell lines was a result of a secondary effect, as further explained below in Fig 6.12. Increased expression of a target gene after treatment with 5-aza-2'-deoxycytidine may occur through indirect mechanisms if this treatment removes methylation marks from a regulatory gene, such as 'Gene A', that will eventually positively affect *CD24* expression. In contrast, direct action of 5-aza-2'-deoxycytidine on *CD24* will remove methylation marks from the regulatory region of *CD24* and make it accessible for transcription.

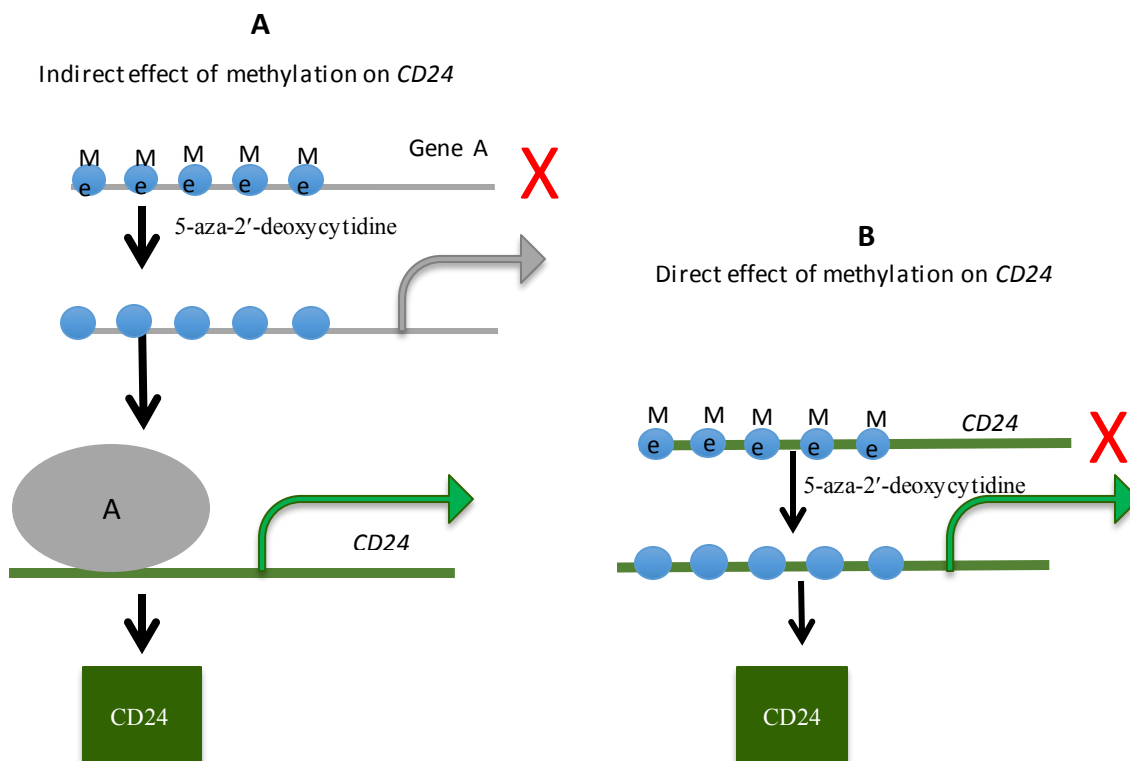


Fig 6.12: Two possible modes of 5-aza-2'-deoxycytidine action to increase CD24 expression. A. Indirect effect of 5-aza-2'-deoxycytidine removes methylation marks from a regulatory 'Gene A' resulting in expression of A (grey ellipse), which may ultimately bind to CD24 regulatory region and cause its expression (green square). B. Direct effect of 5-aza-2'-deoxycytidine removes methylation from the regulatory region of *CD24*, allowing re-expression by making it accessible to positive regulators of expression (green square).

In order to check whether the effect of 5-aza-2'-deoxycytidine on CD24 expression was direct or indirect, a standard technique of sequencing bisulfite modified DNA was used. This involved identification of candidate CpG islands in the promoter of *CD24*, followed by sequencing of bisulfite modified PCR fragments that span the target CpG islands. Bisulfite modification of DNA results in conversion of unmethylated cytosines ('C's) to thymidine (T), whilst any 'C's that are methylated are protected and remain as 'C's in the modified DNA. All 'C's that are not followed by a G (and therefore do not constitute a 'CpG') and those which are part of CpGs but unmethylated should be converted to 'T's in the bisulfite modified sequence. Thus, by comparing the unconverted cytosines ('C's) within the sequence obtained after PCR of the bisulfite modified DNA, it is possible to determine whether individual CpGs are methylated or not in the target DNA sequence.

6.5.1 UCSC genome browser shows large CpG islands in the 5' region of CD24

In order to identify the candidate CpG island sites in the promoter of *CD24*, the CpG island prediction track together with the percent GC track and the conservation between species tracks on the UCSC genome browser were examined for the DNA sequence spanning and flanking *CD24* gene.

The analysis showed a predicted 2500 bp long CpG island in the *CD24* gene which extended from the 5' UTR towards the first and second exons (and associated introns). A part of the island, flanking the 5' UTR, encompasses a sequence which shows full conservation between Human and Rhesus monkey and considerable conservation with mouse, dog and elephant genome (Fig 6.13) suggesting the importance of this region for the regulation of *CD24*. This part of the CpG island was therefore considered most probable target for CpG methylation

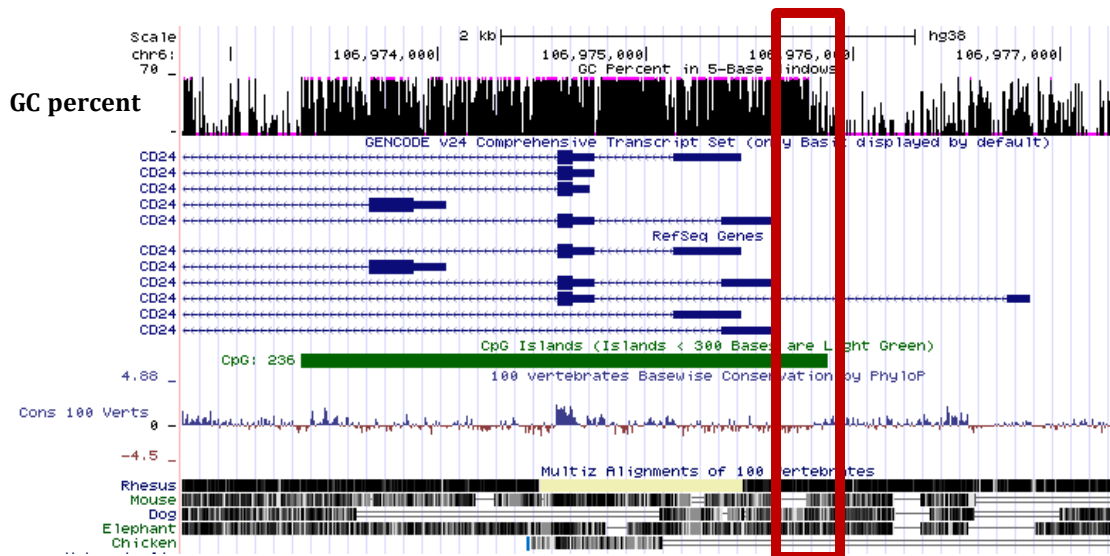


Fig 6.13: *CD24* is predicted to have a large CpG island (Green bar). A big part of the island is downstream of the first exon. A significant part is in the 5'-UTR and flanking region which is also highly GC rich (top panel). Conservations among species is shown on the bottom panel. The area highlighted by the red box was considered to be a prospective target for promoter methylation.

6.5.2 Bisulfite (BS) sequencing showed evidence of direct methylation of *CD24* promoter region

The CpGs in the promoter region of the LS174T Clone 2^{CD24-} were expected to be methylated if methylation was directly responsible for *CD24* suppression in this clone. In contrast, as LS174T Clone 1^{CD24+} expresses *CD24*, the prediction was that there would be little or no methylation of CpG positions in the promoter region of *CD24* in Clone 1^{CD24+}.

In order to test this idea, primers suitable for amplification of bisulfite (BS) converted DNA were designed using two platforms: EpiDesigner from Sequenom and Methprimer from the Li lab (Li & Dahiya, 2002) as described in section 2.3.3.1.2.

6.5.2.1 Bisulfite (BS) modification and sequencing of Clone 1^{CD24+} and Clone 2^{CD24-} of LS174T confirmed direct methylation of *CD24*

DNA from early passage (p7-p15) Clone 1^{CD24+} and Clone 2^{CD24-} of LS174T was bisulfite (BS) modified using EpiTect® Fast Bisulfite Conversion kit (Catalogue: 59824) according to the manufacturer's instructions. For each of the clones, an aliquot of cells was checked for *CD24* expression by FACS in parallel with the DNA preparation for methylation analysis.

BS modified and purified DNA from clones was amplified with the primers designed using Methprimer. Cleaned PCR products were sequenced using the sequencing facility in the WIMM.

BS sequence results of the CpG rich regions downstream to the first exon of *CD24* did not detect any methylation difference between Clone 1^{CD24+} and Clone 2^{CD24-} (see data in appendix: Ax.7).

However, a significant difference in methylation was observed between Clone 1^{CD24+} and Clone 2^{CD24-} in the amplicon from the primer pair 4 (designated as 'primer 4'), which

encompasses 5' region of exon 1 and the immediate 5' UTR of *CD24* (Fig 6.14). This amplicon was within the expected promoter region of *CD24*.

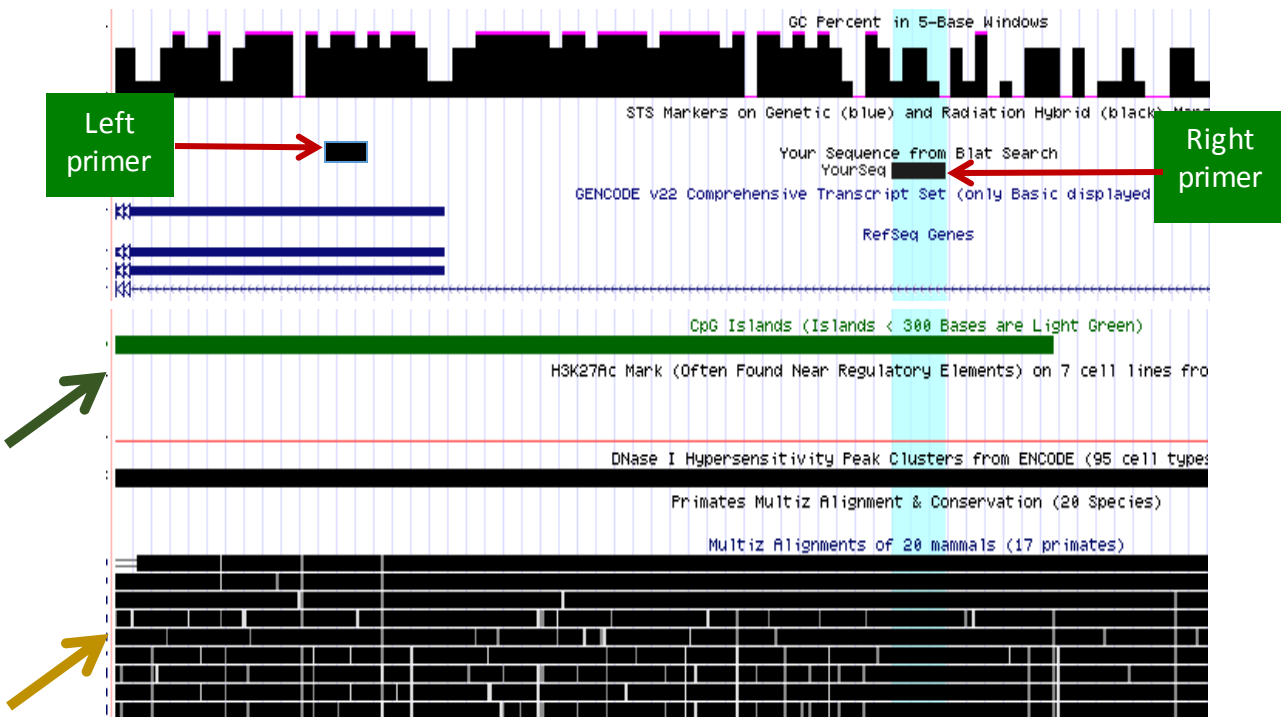


Fig 6.14: Amplicon with primer pair 4. A region between exon 1 and 5' flanking region of *CD24* gene which is part of the CpG island (indicated with a green arrow) is amplified by primer pair 4 (indicated with red arrows). The black horizontal bars at the bottom (indicated with a yellow arrow) shows high conservation of sequences in this region among different species. The topmost panel indicates the GC percent. The blue bars in the middle indicate exon 1 of different transcript variants.

Amplicon 4 included 21 CpGs in total. Sequence results allowed direct comparison of 7 CpG positions (indicated in boxes and with arrows in Fig 6.15 A and B) between the clones. Of these 7 positions, 4 positions (CpG-1, 2, 3, 5) in Clone 2^{CD24-} were significantly methylated, as indicated by the presence of blue peaks representing methyl protected 'C's in the Sanger sequence reads (Fig 6.15 A). CpG-2 showed 75% methylation which is approximated from the relative height of the 'C' and 'T' peaks in the same position (magnified in Fig 6.15 C). In other positions, Clone 2^{CD24-} had some low level methylation

as visible from blue peaks. Most importantly, in all these 7 positions, Clone 1^{CD24+} was completely unmethylated (Fig 6.15 B).

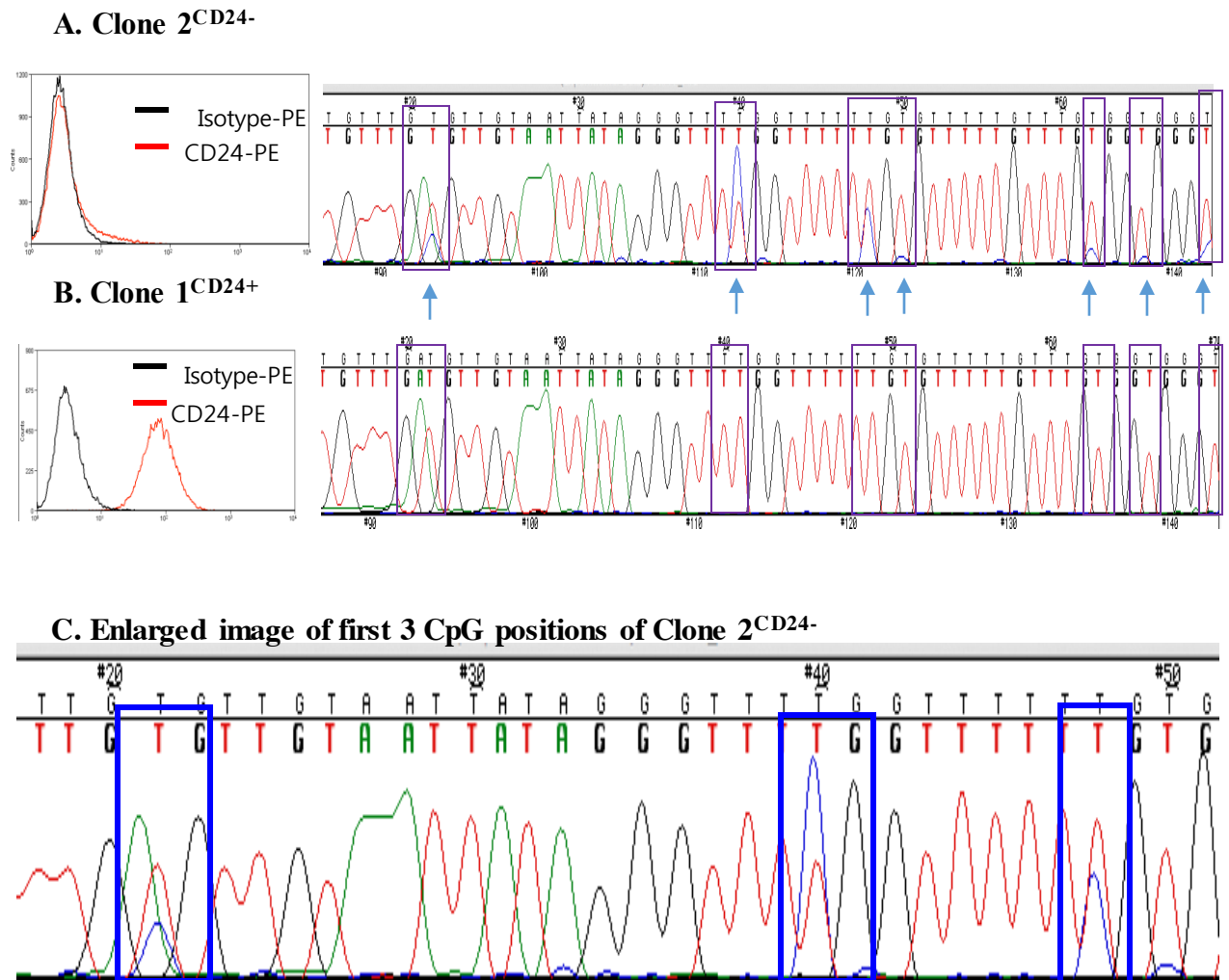


Fig 6.15: Clone 2 shows significant level of methylation compared to Clone 1. All 7 CpG positions which are directly comparable between Clone 2 and Clone 1 are given. On the left side of A and B, the CD24 FACS expression profiles are shown again for reference. Blue arrows indicate the CpG positions 1-7 from left to right. A. Clone 2^{CD24-} shows blue peaks of 'C's in all 7 positions (although positions 1, 2 and 3 shows more prominent blue peaks) indicating methylation. B. In Clone 1^{CD24+} there is no trace of 'C' represented by a blue peak, indicating the lack of methylation at these positions. C. Magnified image of the CpG positions 1, 2 and 3 where Clone 2^{CD24-} is approximately 25%, 75% and 40% methylated respectively.

These results mean that in Clone1^{CD24+}, the 'C's in the all seven CpG positions had been fully converted to 'T's. This would happen only when all the 'C's in those CpG positions were unmethylated. However, in Clone2^{CD24-}, CpG-1, CpG-2 and CpG-3 had 25%, 75% and 40% methylation respectively. Another position which showed lower but notable methylation is CpG-5 (~10%). The remaining two CpG positions showed very low levels of methylation. It is important to note that 'C' nucleotides that occur outside of CpGs can act as an internal control for the efficiency of bisulfite conversion. This facilitates the distinction between low levels of residual 'C's due to inefficient bisulfite modification, as compared to low levels of residual 'C's which reflect the true methylated state of a minor proportion of 'C's at a particular CpG. So the observation of no (or very low) blue peaks outside CpG would mean that the BS conversion was very efficient and at the same time, the conspicuous blue peak representing 'C' in CpG positions would be the evidence of methylation.

The presence of methylated 'C's in the LS174T Clone 2^{CD24-} as mentioned above suggested that the suppression of CD24 might be linked to the methylation of CpGs around the promoter region of the gene.

6.5.2.2 BS modification and sequencing of other CD24⁺ and CD24⁻ cell lines confirmed CpG position methylation observed in Clone 1^{CD24+} and Clone 2^{CD24-} of LS174T

Having detected methylation differences between Clone 1^{CD24+} and Clone 2^{CD24-} of LS174T that correlate with their CD24 expression status, the next goal was to find out if these CpG positions would show similar patterns of methylation or demethylation in other CD24⁻ and CD24⁺ cell lines, respectively.

To test this, two CD24⁻ cell lines: CC20 and RKO and two CD24⁺ cell lines: DLD1 and NCIH716 (Table 6.2) were BS modified and sequenced using the same set of primers (primer 4). Sequencing results of amplicon 4 revealed that the CD24⁻ cell lines, CC20 and RKO, showed similar patterns of CpG methylation in the 7 positions as was observed for LS174T Clone 2^{CD24-}, whereas the CD24⁺ cell lines, DLD1 and NCH716, showed no methylation at all which are similar to the observation in Clone 1^{CD24+} (Fig 6.16).

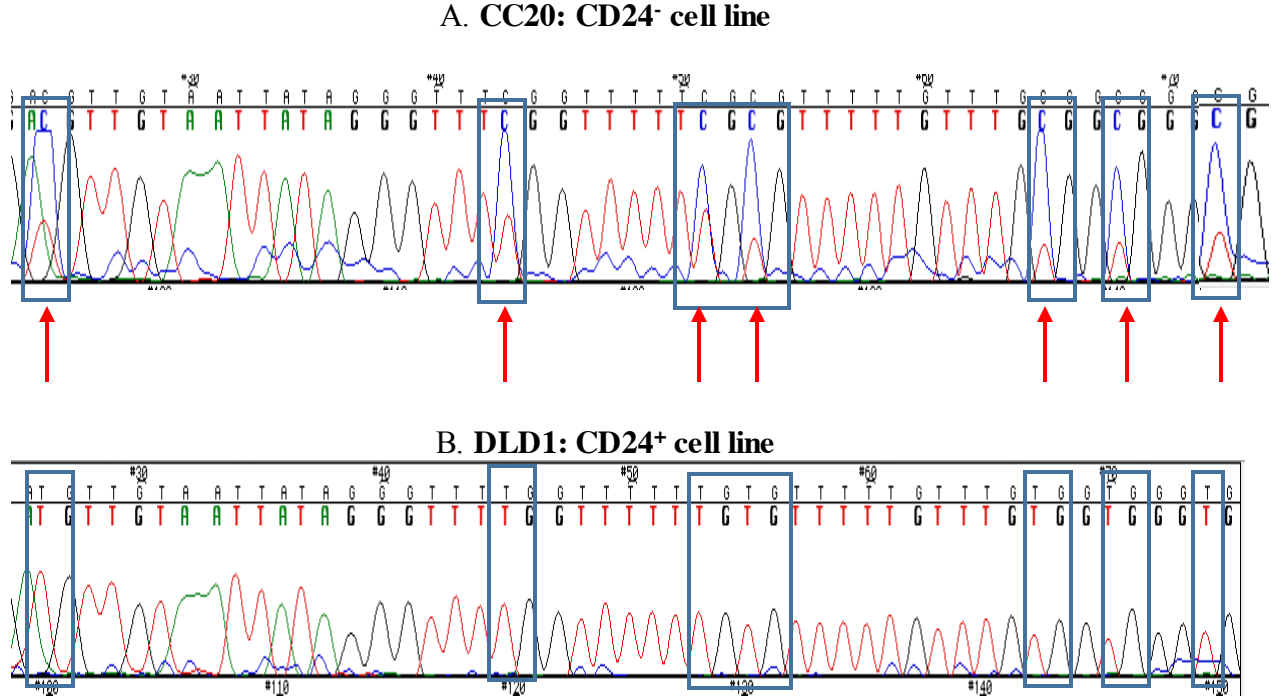


Fig 6.16: Methylation status in CD24⁻ and CD24⁺ cell lines. A. CC20 is a CD24⁻ cell line and has high level of methylation (indicated by blue peaks) in the CpG positions (indicated by red arrows) as similar to Clone 2^{CD24-} of the LS174T. B. DLD1 is a CD24⁺ cell line and lacks methylation in those CpG positions as like Clone1^{CD24+} of LS174T. Boxes and arrows indicate 7 comparable CpG positions.

6.5.2.3 BS modification and sequencing of subclones developed from 5-aza-2'-deoxycytidine treated Clone 2^{CD24-} confirms direct link of CpG positions with CD24 expression

It has been shown in section 6.4 that 5-aza-2'-deoxycytidine treatment resulted in enhanced expression of CD24 in the LS174T Clone 2^{CD24-}. The treatment should remove methyl marks from the DNA of Clone 2^{CD24-}, including those 7 CpG positions mentioned above. A direct association of those 7 positions to the regulation of CD24 can be drawn by determining whether, after 5-aza-2'-deoxycytidine treatment, those positions regain methylation with concomitant loss of expression of CD24, or whether there would be selection for CD24 expression in some clones where all those 7 positions would maintain the demethylated state.

To test this hypothesis, the top 5% of CD24⁺-expressing cells from 5-aza-2'-deoxycytidine treated Clone 2 of LS174T were sorted using FACS (Fig 6.17) to develop subclones.

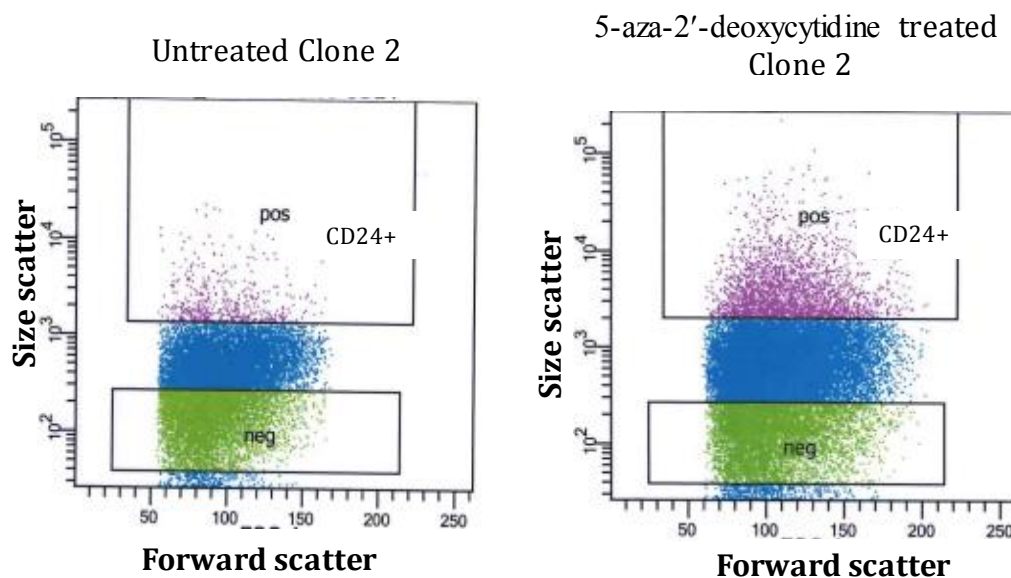
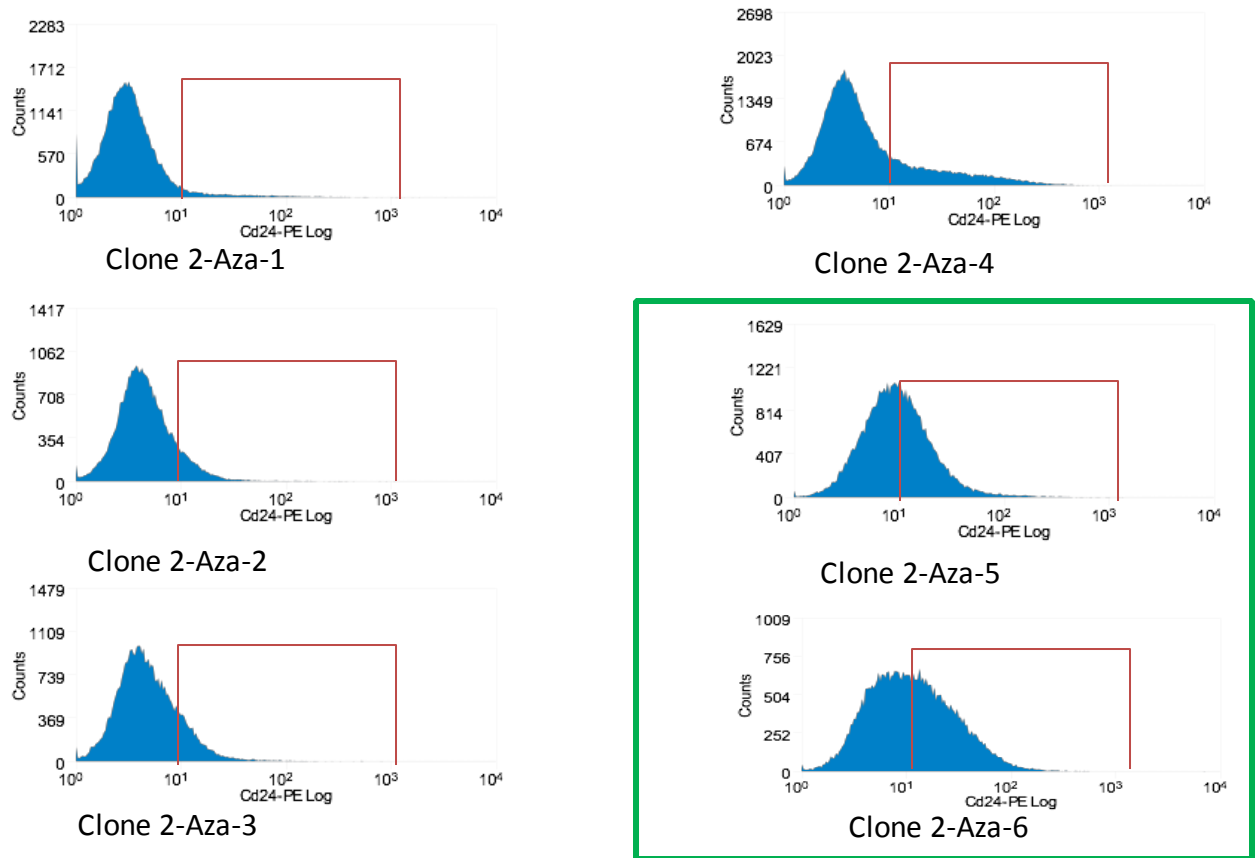


Fig 6.17: Effect of 5-aza-2'-deoxycytidine treatment and correlation with CD24 expression. Compared to untreated Clone 2 (left), the proportion of CD24⁺-expressing cells increased after 5-aza-2'-deoxycytidine treatment (right).

The objective of this subcloning experiment was two-fold: i) to find out if the 5-aza-2'-deoxycytidine induced increased expression of CD24⁺ would be selected and maintained. and ii) whether the methylation of these seven CpG positions will change consistently with regards to the expression (or suppression) of CD24 in the 5-aza-2'-deoxycytidine treated subclones.

Six subclones were successfully obtained from the experiment and propagated for further studies. They were in continuous culture for 2 months before enough cells were obtained for FACS analysis for CD24 expression. When checked, it appeared that 2 out of the 6 subclones (labelled Clone 2_Aza-5 and Clone 2_Aza-6) showed significant levels of CD24 expression (Fig 6.18 A).

A. Subclones of Clone 2 after treatment with 5-aza-2'-deoxycytidine



B. Clone 2_parental

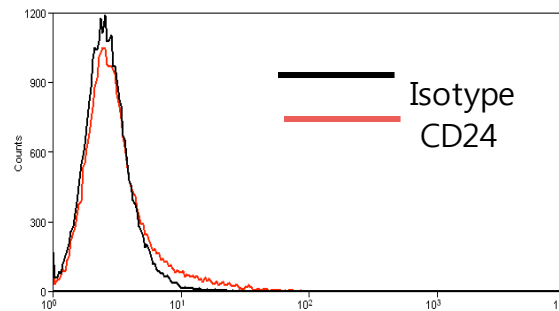
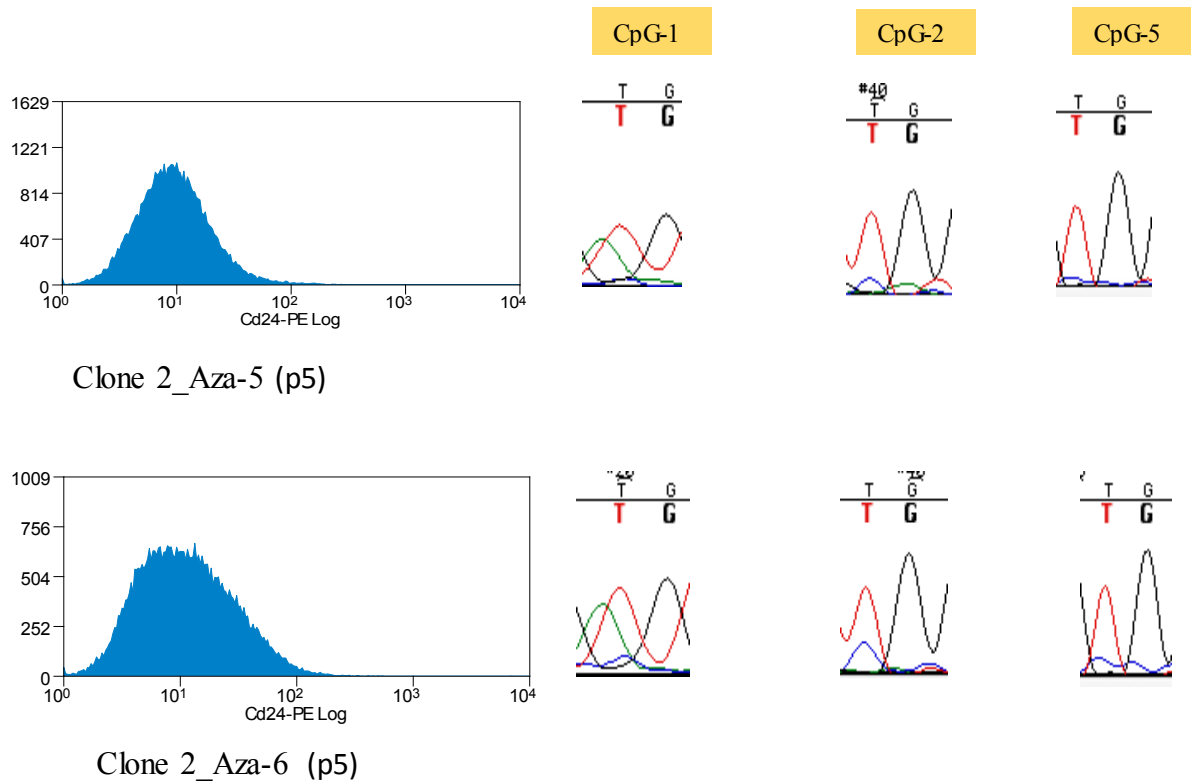


Fig 6.18: FACS profile of the subclones developed from 5-aza-2'-deoxycytidine treated Clone 2. A. Out of the 6 sub clones developed after 5-aza-2'-deoxycytidine treatment, 2 subclones (Clone 2_Aza-5 and Clone 2_Aza-6 in green box) showed considerably higher level of CD24 expression compared to other subclones and their untreated parent Clone 2^{CD24-} which is CD24⁻ (B). Red bars indicate the area where CD24⁺ cells should appear considering the isotype control.

All 6 subclones of the same passage (p5) were BS modified, amplified with primer pair 4 (as previously used for clones and cell lines) and sequenced to see methylation status of the previously detected 7 CpG positions. The 2 subclones, Clone 2_Aza-5 and Clone 2_Aza-6, with elevated level of CD24 expression, showed almost complete loss of methylation in all 7 CpG positions (Fig 6.20). In particular, attention was placed on 3 CpG positions (CpG-1, CpG-2 and CpG-5) (Fig 6.19A), as these positions were found to be highly methylated in untreated parental Clone 2^{CD24} (Fig 6.15). Importantly, all of the 4 CD24⁺ subclones retained methylation in these positions. (Fig 6.19B).

A. Methylation status of CD24⁺ subclones after 5-aza-2'-deoxycytidine treatment



B. Methylation status of CD24⁻ subclones after 5-aza-2'-deoxycytidine treatment

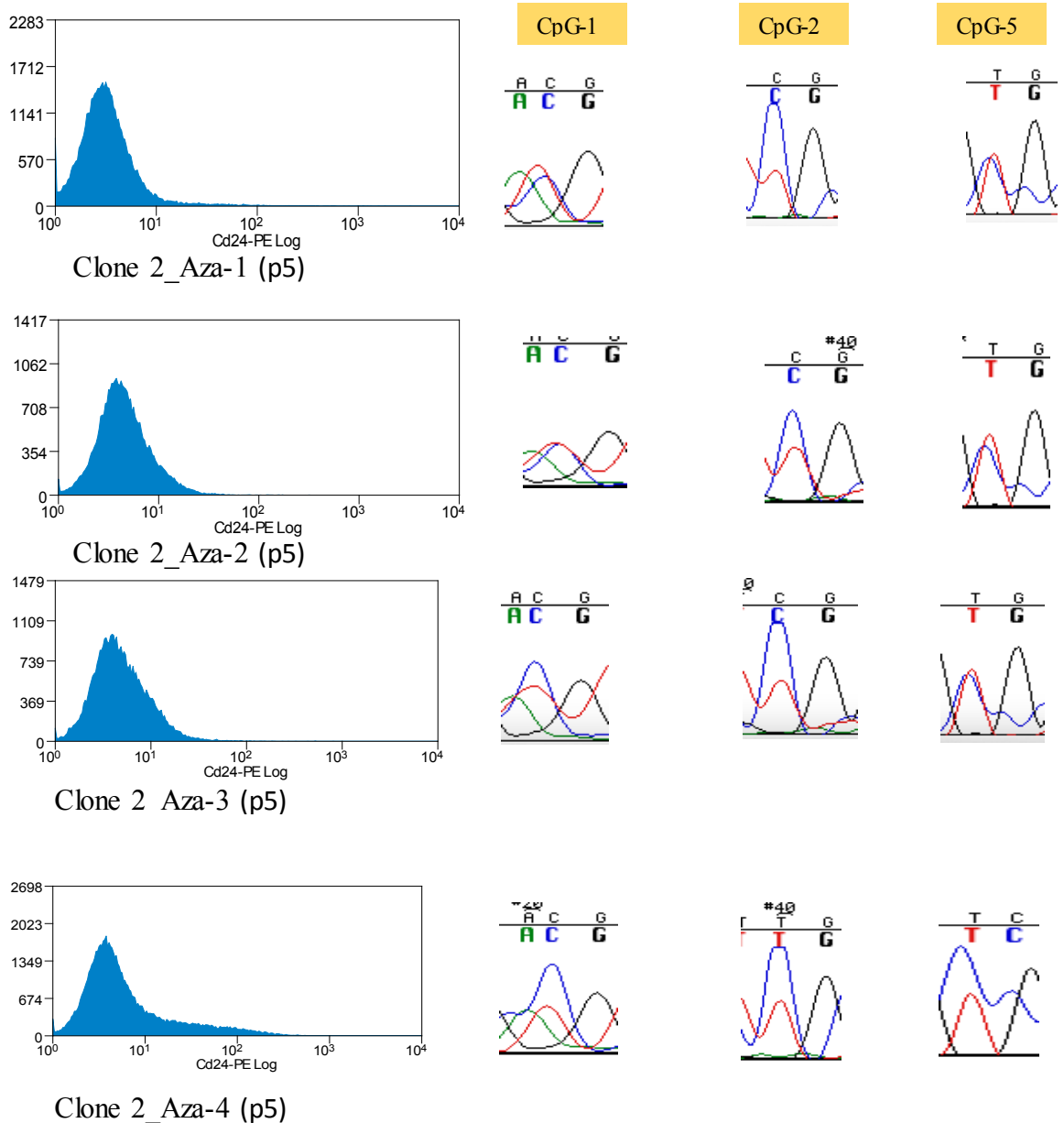


Fig 6.19: Selection of the 'unmethylated state' after 5-aza-2'-deoxycytidine treatment. A. Out of 6 sub clones derived from 5-aza-2'-deoxycytidine treatment, Clone 2_Aza-5 and Clone 2_Aza-6 were CD24⁺. Both of them have lost methylation in CpG positions 1, 2 and 5. B. The remaining 4 subclones that are largely CD24⁻, are highly methylated in these CpG positions. Here, on the left panel: CD24 expression profile of the subclones. In the sequence track, blue peaks indicate methylated 'Cs' and red peaks indicate unmethylated 'Cs' that have been converted to 'T's after bisulfite treatment.

It is not possible to be certain that the lack of methylation in the subclones, Clone 2_Aza-5 and Clone 2_Aza-6 was directly due to 5-aza-2'-deoxycytidine demethylation, rather than selection for a pre-existing population of cells that already have decreased methylation and thereby have higher levels of CD24 expression. However, there are differences in methylation profiles between the untreated CD24⁺ population compared to the treated CD24⁺ population. Naturally occurring CD24⁺ cells in Clone 2 do not lose the methylation marks from all CpG positions to the same degree that CD24⁺ cells of the 5-aza-2'-deoxycytidine treated Clone 2 do. In the case of 5-aza-2'-deoxycytidine treated CD24⁺ cells, all CpG positions are expected to lose methylation, either completely or to a high degree.

One compelling evidence for the difference between 5-aza-2'-deoxycytidine treated CD24⁺ subclones and naturally occurring CD24⁺ subclones came from a comparison of the BS-modified sequences of the CD24⁺ Subclone 2_Aza-5 and Subclone 2_Aza-6, developed from 5-aza-2'-deoxycytidine treated Clone 2, with the BS-modified sequence of Sub 5 (passage 19), developed as a (mostly) CD24⁺ subclone from untreated Clone 2 (Section 6.3.2). Sub 5 is a subclone of Clone 2 that was generated by isolating single cells from the top 3% CD24⁺ expressing cells of Clone 2. At passage 19, its FACS profile is bimodal where ~90% of cells are CD24⁺ but there is a small CD24⁻ tail. On the contrary, the two subclones: Sub-Clone 2_Aza-5 and Sub-Clone 2_Aza-6 were derived from the CD24⁺ end of 5-aza-2'-deoxycytidine treated Clone 2.

Comparison among their corresponding BS-modified sequences showed that for Sub-Clone 2_Aza-5 and Sub-Clone 2_Aza-6, the methylation marks are almost completely removed from all seven CpG positions. In Sub 5, however, although six CpG positions are demethylated, CpG-2 retains ~50% of methylation (Fig 6.20).

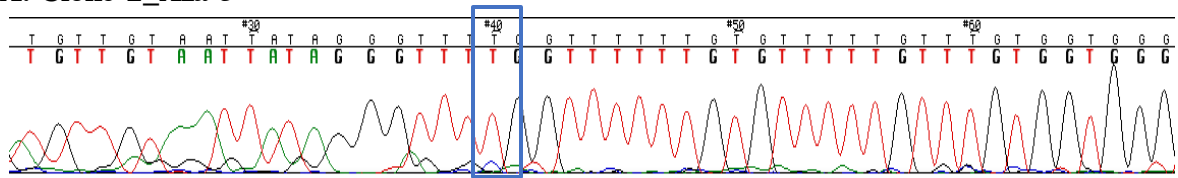
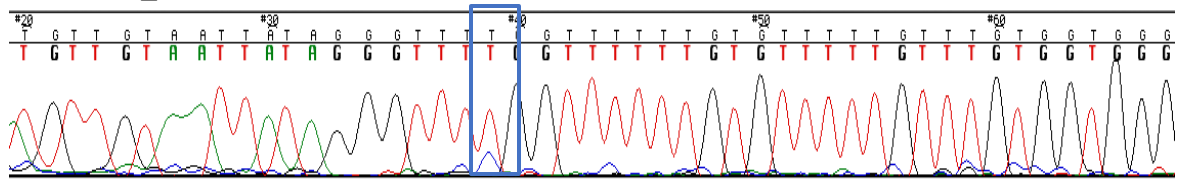
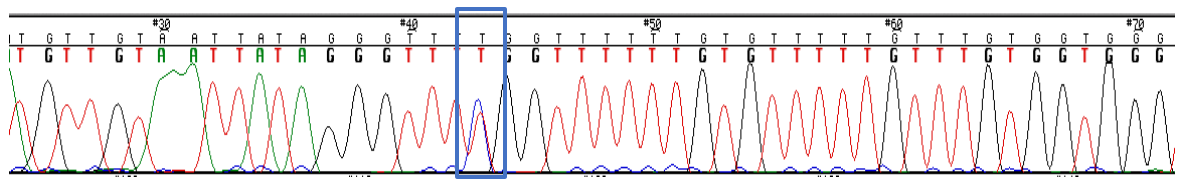
A. Clone 2_Aza-5**B. Clone 2_Aza-6****C. Sub 5**

Fig 6.20: Difference between 5-aza-2'-deoxycytidine treated CD24⁺ clones and untreated CD24⁺ clone. Blue boxes show difference between Sub-Clone 2_Aza-5, Sub-Clone 2_Aza-6 (derived from 5-aza-2'-deoxycytidine treated Clone 2) and Subclone 5 (derived from untreated Clone 2). Absence of prominent blue peaks in first two sub clones (A and B) confirms the effect of 5-aza-2'-deoxycytidine. In Sub 5 (C), there is ~50% methylation in CpG-2 (blue box), showing a difference between 5-aza-2'-deoxycytidine treated (A and B) versus untreated CD24⁺ (sub)clones (C).

This observation shows that naturally developed CD24⁺ (sub)clones may arise even though some partial methylation will remain in some CpG positions. This will be different from 5-aza-2'-deoxycytidine CD24⁺ (sub)clones where their demethylation states are achieved after complete removal of methylation marks by the treatment.

It is important to remember that the effect of 5-aza-2'-deoxycytidine to remove methylation of DNA in cells is transient. Once 5-aza-2'-deoxycytidine is removed from the culture environment, cells tend to return to their methylated state again. Out of 6 subclones, developed after 5-aza-2'-deoxycytidine treatment, 4 returned to the methylated state (Fig

6.19B). However, the other two (Sub-Clone 2_Aza-5 and Sub-Clone 2_Aza-6) retained their demethylated status (with resultant CD24 expression). This must in some way be due to selection, as the subclones were maintained in culture for 5 months after the treatment. It would be worthy to compare this observation with a control that might not have had any selection advantage, such as methylation of *hMLH1* in MMR deficient cell lines (like LS174T). In this case, one would expect to see loss of methylation immediately after 5-aza-2'-deoxycytidine treatment but then recovery of the methylation on subsequent propagation, since there is no presumed advantage to retaining *hMLH1* demethylation.

6.6 Methylations of CpG-2 and CpG-5 show strong association with CD24 suppression indicating 'seeding' effect of these CpG positions

Of the seven CpG positions clearly comparable between CD24⁺ and CD24⁻ cell lines or clones, it appeared that the methylation status of two positions (CpG-2 and CpG-5) is strongly related to the expression profile of CD24. When the level of methylation in the 7 CpG positions of CD24-positive Clone 1^{CD24+} of LS174T, NCIH716^{CD24+} and DLD1^{CD24+} were compared with CD24-negative Clone 2^{CD24-} of LS174T, CC20^{CD24-} and RKO^{CD24-}, the methylation of CpG-2 and CpG-5 showed a consistent association with the expression status of CD24. In all of the CD24-negative cases (Clone 2^{CD24-} of LS174T, CC20^{CD24-} and RKO^{CD24-}), CpG-2 was (75-85)% methylated and CpG-5 was (15-85)% methylated (Fig 6.21).

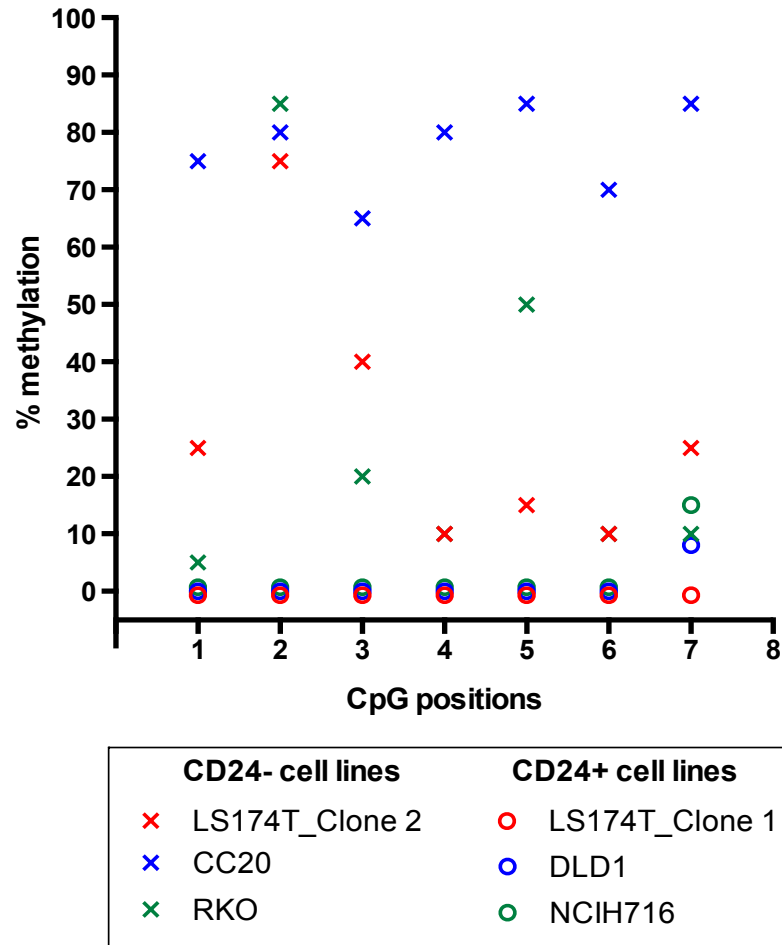


Fig 6.21: Role of a few CpG positions for methylation target in CD24 gene. In seven CpG positions, CD24⁻ clones and cell lines have methylation, whereas CD24⁺ cell lines and clones show no (or very little) methylation in these positions. Cell lines and clones which are CD24⁻ and CD24⁺ are indicated by a cross and circle respectively. Different colour represents different cell lines.

At the CC20 sequence (in Fig 6.16 and Fig 6.21), it is striking that there is much more consistent methylation of all the 7 CpG positions. A more stable lack of expression may perhaps require a higher proportion of the relevant CpG positions to be clearly methylated and as such the probability of reversion of expression is lower because it would require loss of methylation in more of the CpG sites.

To further investigate the relative contribution of individual CpG positions, 2 subclones Sub 3 and Sub 5 of Clone 2 of LS174T were BS-modified and sequenced. These two subclones change CD24 expression profiles in opposite directions over time (Section 6.6). From the sequence results it was found that in CD24-negative Sub 3 (at passage 28), CpG-2 and CpG-5 positions were 75% and 50% methylated respectively, whereas in CD24-positive Sub 5 (at passage 19), CpG-2 was ~55% methylated and there was no methylation at CpG-5. All other positions were unmethylated in these two subclones indicating that methylation change in these two positions contributed more than other positions to impact CD24 expression (Fig 6.22). This observation reinforces the importance of methylation of more/all sites for stable suppression as mentioned in the previous paragraph.

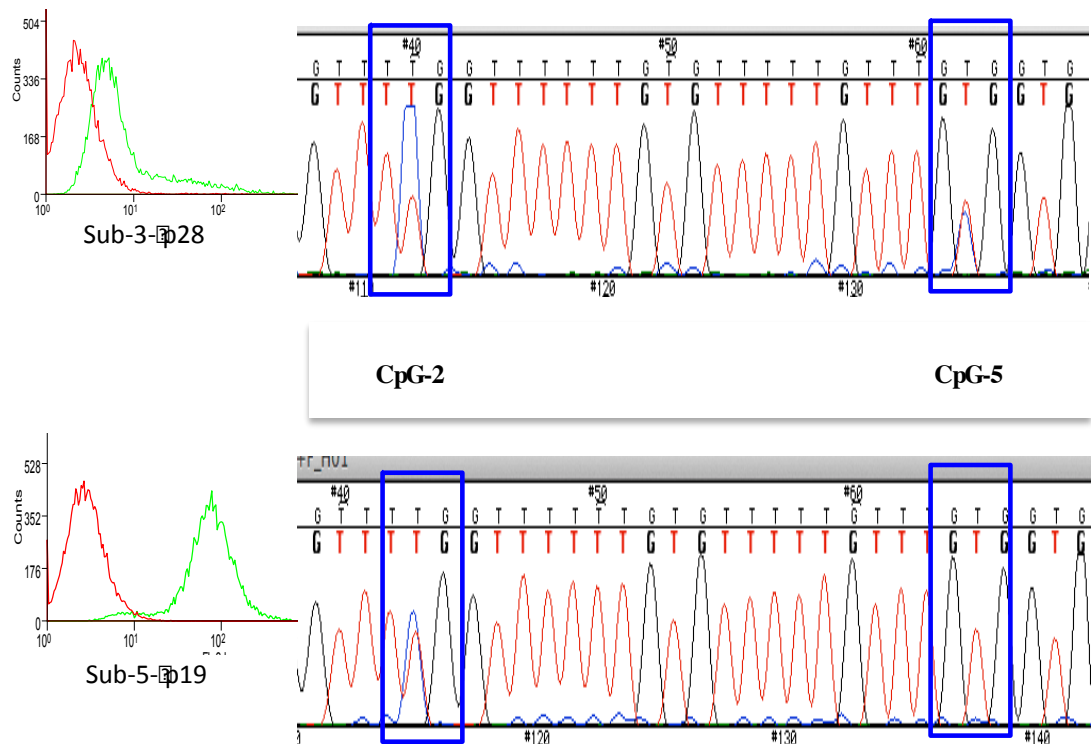


Fig 6.22: BS sequence results of Sub 3^{CD24-} and Sub 5^{CD24+}. Late passage (p28) of Sub 3 stabilised as CD24⁻ and Sub 5 stabilised as CD24⁺ (FACS profile of the left panel). BS-modified sequence shows CpG-2 and CpG-5 are the two positions that have methylation difference implying relative importance of these positions in methylation mediated suppression of CD24 in Sub 3. Blue peaks indicate methylated 'C's.

6.6.1 Methylation in CpG-2 and CpG-5 is mirrored by the expression of CD24

Further analysis provided more support for the relative importance of CpG-2 and CpG-5 as the driver/seed for methylation. It has been described in section 6.3.2 that Sub 3 of LS174T Clone 2 becomes CD24⁻ over time (Fig 6.6). Passage 11-13 of Sub 3 contains mostly CD24⁺ cells, whereas Passage 18 (p18) is bimodal with increasing CD24⁻ population. By passage 28 (p28), Sub 3 is mostly CD24⁻. Therefore, the idea was to compare Sub 3 at passage 13 (p13) and passage 28 (p28) to see if methylation status of the CpG positions mentioned above were correlated with the expression difference of early and late passage Sub 3.

When BS-modified and sequenced, the results provided further support for the importance of CpG-2 (and CpG-5 to a lesser extent) in the control of CD24 expression. In late passage Sub 3^{CD24⁻}, CpG-2 is around 80% methylated and CpG-5 is 45% methylated. However, in early passage Sub 3^{CD24⁺}, CpG-2 is around 25% methylated and CpG-5 is around 10% methylated (Fig 6.23) which mirrors the CD24 expression in them.

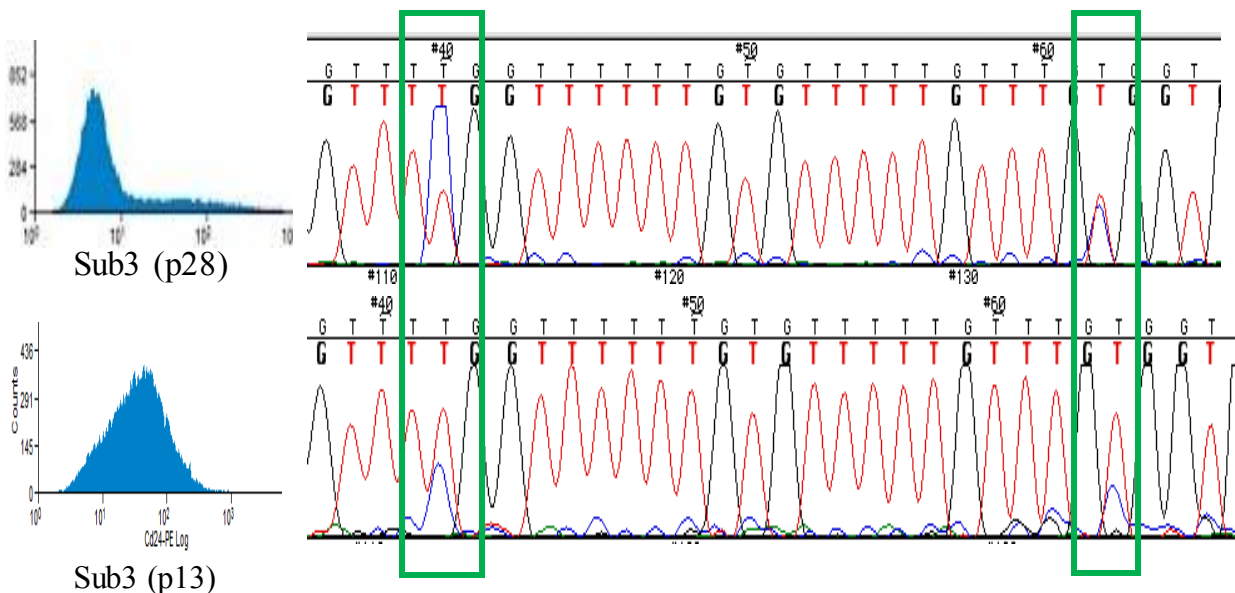


Fig 6.23: Higher degree methylation of CpG-2 (together with CpG-5) is associated with CD24 suppression in late passage Sub 3. Relative heights of blue and red peaks (indicating methylated 'C' and unmethylated 'C' respectively) is used to estimate the % methylation.

6.6.2 Relative importance of CpG-2 over CpG-5 to control CD24 expression

Analysis with the clones developed from SW480 cell line allowed the comparison of the relative importance of CpG-2 and CpG-5 positions to regulate CD24 expression. From the experiments with Sub 3 (Fig 6.23), methylation of CpG-2 showed better correlation with the change in CD24 expression. This was further supported from the study of CD24-positive Clone C2^{CD24+} and CD24-negative Clone D1^{CD24-} of SW480. When BS-modified and sequenced, it was found that only CpG-2 position was markedly different between them (Fig 6.24). This indicated that substantially higher levels of methylation at this position might be enough to shut down the expression of CD24.

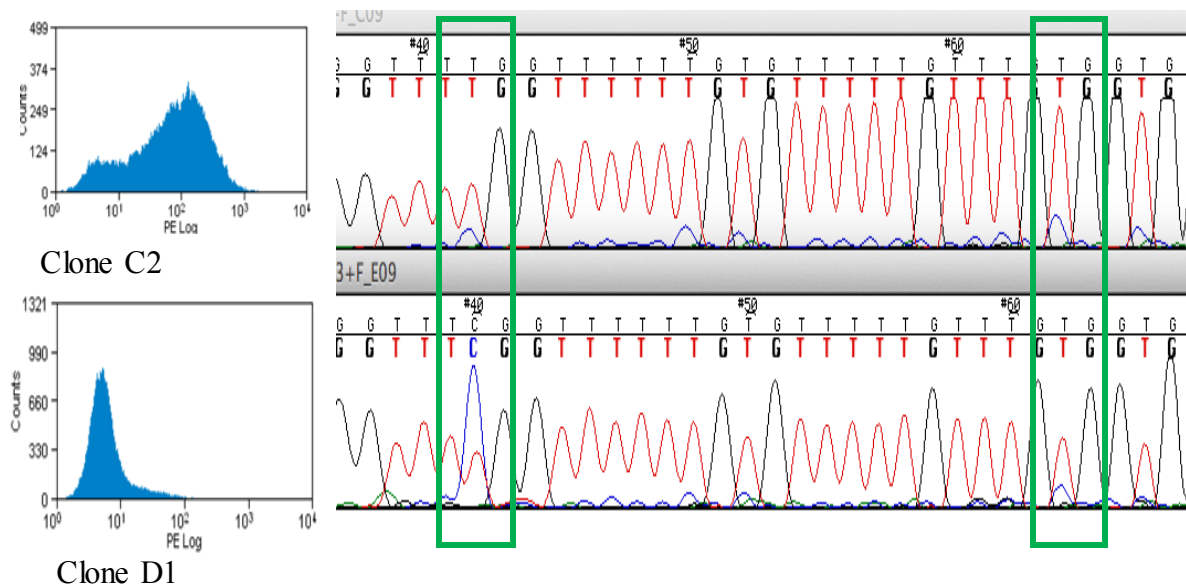


Fig 6.24: Clone C2 and D1 from SW480 shows that their CD24 level is directly linked with CpG-2 methylation. In Clone C2^{CD24+}, CpG-2 (left green box) is around 10% methylated whereas in Clone D1^{CD24-} this position is around 80% methylated. All other CpG positions are comparable between them. CpG-5 (right green box) does not show significance difference in methylation.

In summary, of 9 CD24⁻ cell lines, clones and subclones, 8 showed ~80% methylation in CpG-2 position, whereas, CpG-5 showed ~50% methylation in those same samples. Among 7 CD24⁺ cell lines, clones and subclones, 6 (except Sub 5) showed 0% methylation in CpG-2, and all of them showed 0% methylation in CpG-7 (Fig 6.25). Furthermore, clones and

subclones that shifted patterns of CD24 expression over time showed the predicted changes in methylation status that would be expected if only methylation of the key CpGs was directly involved in the regulation of *CD24*.

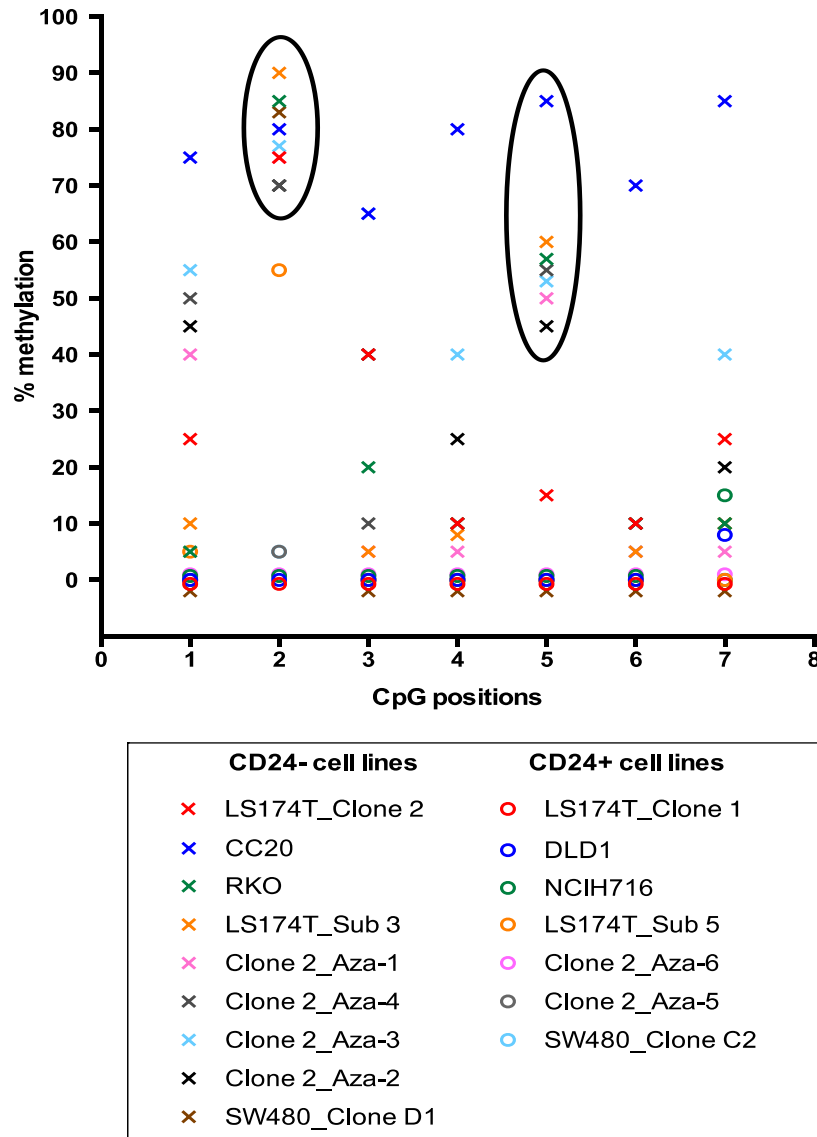


Fig 6.25: CpG-2 shows a consistently higher degree of methylation compared to other CpG positions indicating its seeding/driver role to control CD24 expression. The other position that strongly correlates with CD24 expression status is CpG-5. Cell lines and clones which are CD24⁻ and CD24⁺ are indicated by a cross and circle, respectively. Different colour represents different cell lines.

In conclusion, observations from this study strongly suggest that methylation status of CpG-2 and CpG-5 (together other 5 CpGs) play a regulatory role in the expression of *CD24*. However, the contribution of other CpGs positions for a more stable suppression of *CD24* should not be underestimated in this process.

6.7 Chapter Discussion

In this chapter, the mechanism of regulation of expression of CD24 in colorectal cancer (CRC) cell lines has been investigated. The experimental evidence presented here, established that the methylation of a series of 7 CpG positions in the promoter region may directly regulate CD24 expression in the CRC cell lines. A further extension of the investigation showed that the methylation status of 2 of the CpG positions is crucial for controlling the expression.

The scope of studying the regulation of CD24 came from the observations of previous chapters (chapter 4 and 5) where CRC cell lines (LS174T and SW480) had been demonstrated to be heterogeneous with the CD24 differentially expressing clones. The presence of these clones is stable in the parental cell lines and can be monitored by FACS analysis. Eventually both CD24⁺ and CD24⁻ clones were isolated by sorting single cells from the cell lines using FACS. Those clones became useful models to study the regulation of *CD24*.

In order to understand the expression and suppression of CD24 in the CD24⁺ and CD24⁻ clones respectively, both genetic and epigenetic causes were considered. There was no evidence for mutations and so methylation mediated regulation was investigated.

Changing gene expression by modifying the methylation status of CpGs of the promoter regions is a frequent event (Turker, 2002). This flexible mechanism gives cells the prospects of acquiring a certain phenotype without changing their genotype. That could explain the observed plasticity of CD24 expression in different clones mentioned above.

Dynamic changes in CD24 expression has been reported in pre-adipocytes (Smith, Fairbridge, Pallegar, & Christian, 2015), premature B cells (Ayre et al., 2015) and in breast cancer cell lines (Meyer et al., 2009). In the study of breast cancer cell lines, Meyer *et al.*, studied a 5' upstream region for evidence of methylation mediated regulation of *CD24*. However, they were unable to confirm any evidence of epigenetic regulation of *CD24*.

To our knowledge, dynamic change of CD24 expression in colorectal cancer cell lines or in colorectal cancer has not been reported before. Having demonstrated that these changes can occur in established clones of several colorectal cancer cell lines, the study continued to explore the methylation status of the CD24 differentially expressing clones, as well as a subset of colorectal cancer cell lines with different levels of CD24 expression. Unlike the study of Meyer *et al.*, where they examined only one set of primers for a particular region of CD24 promoter, a total of 5 kb region that covered the entire CpG island (as predicted on UCSC genome browser; Fig 6.13) and adjacent 2.5 kb upstream region was analysed during this study. This region covers the segment that Meyer *et al.* explored as well as additional surrounding lengths.

Methprimer (as well as Epidesinger) software was used to design primers that can amplify the bisulfite modified DNA regions of analysis. The resulting amplicons generated from bisulfite modified DNA were then sequenced and correlations between CD24 expression status and methylation status of individual and clusters of CpGs were sought. This extensive study, for the first time, shows the direct connection between methylation status of a region in the *CD24* promoter and its expression.

Using the clones of LS174T, this study has identified seven consecutive CpG positions (Fig 6.26) in the amplicon, which encompasses the genomic region between 106975560-106975834 on chromosome 6, that show clear differences between CD24-negative Clone 2^{CD24-} and CD24-positive Clone 1^{CD24+} of LS174T.

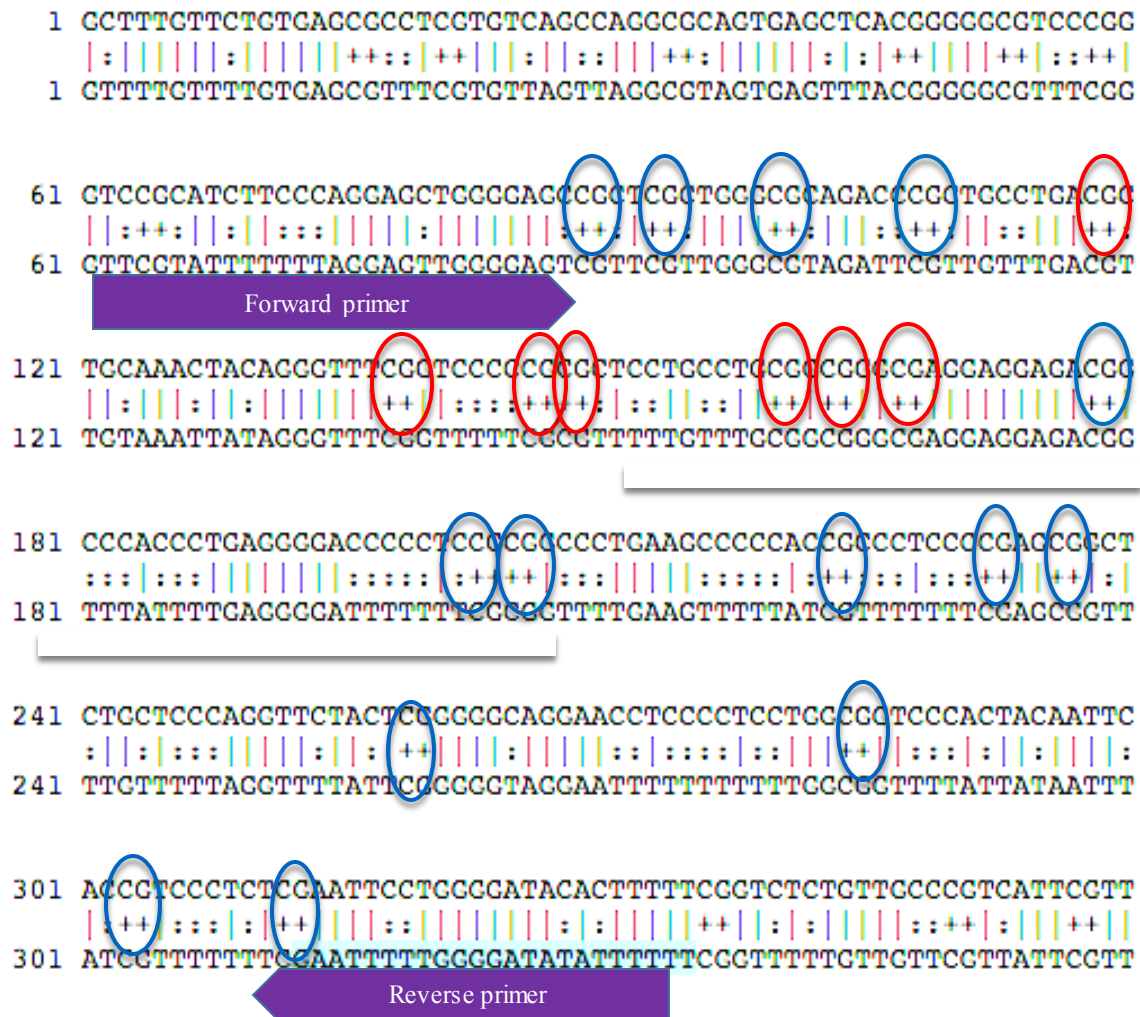


Fig 6.26: The amplicon of primer 4 covers 21 CpGs (shown in circles). Red circles indicate the seven CpG positions where CD24⁻ clones and cell lines are mostly different from CD24⁺ clones and cell lines. In the sequence track, the bottom line indicates bisulfite modified sequence, top line indicates the actual genomic sequence. Primer positions have been shown with violet arrows.

In these positions, LS174T Clone 2^{CD24-} has up to 75% methylation whereas Clone 1^{CD24+} lacks methylation completely. In order to identify whether the methylation differences in these positions correlate with CD24 expression in other colorectal cancer cell lines, the same genomic regions were sequenced in several CD24-positive and CD24-negative cell lines from our cell line panel. The results showed that cell lines such as DLD1, LS4111 and NCIH716 that express high levels of CD24 had no methylation in those positions, whereas

the CD24-negative cell lines RKO and CC20 have similar methylation in the CpGs as was seen in Clone 2^{CD24-}. These observations established that the seven CpG positions identified in the genomic region chr6:106975560-106975834 by primer 4 play a crucial role in CRC cell lines for the regulation of *CD24* by methylation.

This direct methylation mediated regulation of CD24 expression was further confirmed by sequencing CD24-negative clones which were obtained from 5-aza-2'-deoxycytidine treated Clone 2.

Upon treatment with 5-aza-2'-deoxycytidine, expression of CD24 was enhanced in LS174T Clone 2^{CD24-}. Six subclones were established by sorting cells from top 5% of the CD24⁺ expressing cells after 5-aza-2'-deoxycytidine treatment of Clone 2. Out of these six subclones, two showed a stable significant increase in CD24⁺ expression and the other four subclones were mostly negative for CD24 (Fig 6.18). Bisulfite sequencing of these six subclones showed that the two CD24-positive subclones lack methylation in all the seven CpG positions whereas, the four CD24-negative subclones were methylated similar to the parental untreated Clone 2^{CD24-}. These results corroborated that the seven CpG positions were clearly correlated with the expression of CD24.

Further studies enabled to look at the relative importance of the seven CpG positions in regulating CD24 expression. It has been shown that epigenetic silencing of genes can be achieved through methylation of specific key CpGs (Nile, Read, Akil, Duff, & Wilson, 2008). There is evidence to suggest that one or a few CpGs first become methylated and then act as a signal for methylation on other adjacent CpGs (Haerter, Lövkvist, Dodd, & Sneppen, 2014). These positions are considered as 'seed' CpGs for methylation. In addition to the 'seed' methylation event, other CpGs become methylated to make the suppression more stable (Bredberg & Bodmer, 2007). It has also been shown that methylation of just a few CpG positions are enough to prevent the expression of a gene. Claus et al., (2012) demonstrated that methylation of a single CpG position in *ZAP-70* gene can suppress its expression.

It was possible to explore the relative importance of individual CpGs by analysing the clones and subclones generated from the LS174T and SW480 cell lines and further correlating the methylation status of individual CpGs with CD24 expression profiles.

By subcloning the top 3% of CD24⁺ expressing cells of Clone 2^{CD24⁻} of LS174T, two subclones (Sub 3 and Sub 5; Figure 6.5) were established where just 2 CpG positions: CpG-2 and CpG-5 (which are 21 bp apart from each other) within the sequenced amplicon 4, could explain the CD24 expression difference between CD24-positive and CD24-negative subclones. By combining the sequence results of these two subclones with other CD24⁻ and CD24⁺ subclones and cell lines, it was assumed that CpG-2 is likely to be the seed/driver position for *CD24* methylation. This observation was further supported by results for the clones (Clone C2 and Clone D1, Figure 6.24) of SW480 where only CpG-2 methylation is different between the CD24⁺ Clone C2 and CD24⁻ Clone D1.

By combining the sequence data of a total of nine CD24-negative cell lines (Table 6.3), clones and subclones, on an average ~80% methylation has been observed in CpG-2 and ~50% methylation in CpG-5, together with varying degree of methylation in other 5 positions. On the contrary, a total seven CD24⁺ cell lines, clones and subclones did not have any methylation in those positions (Fig 6.27) (Sub 5 is an exception for just one CpG position which has been ignored in this statement).

Table 6.3: List of cell lines and clones used in this study and their CD24 expression status

Name of cell lines/Clones	CD24 status
LS174T_Clone 2	-
CC20	-
RKO	-
Sub3-p19	-
Clone 2_Aza-1	-
Clone 2_Aza-2	-
Clone 2_Aza-3	-
Clone 2_Aza-4	-
SW480_Clone D1	-
LS174T_CLone1	+
DLD1	+
NCIH716	+
Sub 5, p22	+
Clone 2_Aza-6	+
Clone 2_Aza-5	+
SW480_Clone C2	+

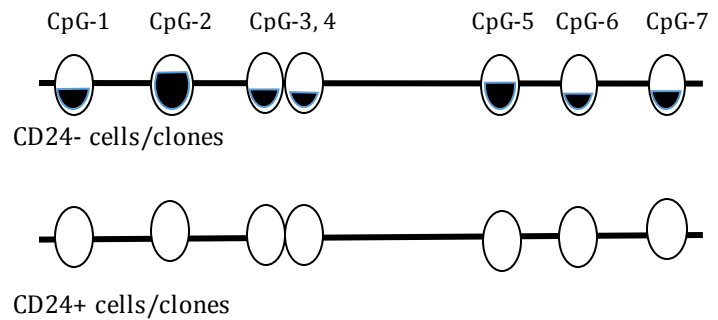


Fig 6.27: Average methylation percentage of the seven CpG positions. Black filling indicates average methylation in the seven CpG positions compared in 9 CD24-ve and 7 CD24+ve cell lines, clones and subclones.

The results presented here clearly shows the direct association of the methylation of seven CpG positions with the expression or lack of expression of CD24. There is convincing evidence that just one CpG position (CpG-2) may play a major role in the regulation of CD24 expression.

A closer look at the positions of the CpGs reveals that those seven CpG positions are actually at the flanking region of the 5' UTR of *CD24* (Fig 6.28) which indicates that this region is functionally important for as the promoter of *CD24* (Pass et al.,1998). Therefore, CpG methylation of these positions might directly affect the bindings of transcription of factors (TFs).

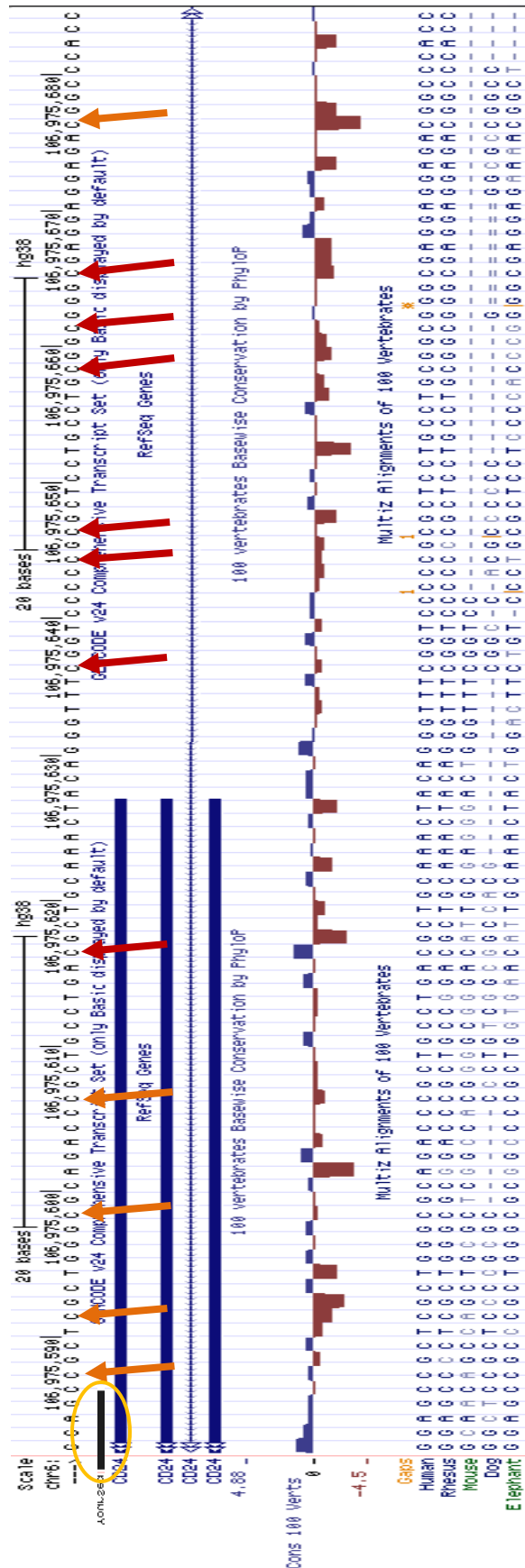


Fig 6.28: Positions of the seven CpGs in relation to the first exon of *CD24*. In the first panel the yellow circle (top left) indicates the position of the forward primer (shown as a black bar). In the sequence track, CpG positions have been indicated with arrows. Red arrows indicate the seven CpG positions that are mostly associated with *CD24* expression. Second red arrow is outside the start site of the first exons (showed by blue bars) which means that at least 6 of those seven CpG positions are within the flanking region of the 5'-UTR. At the bottom of the image, conservation in this region across species has been shown.

A predictive analysis with the sequence of this region of using PROMO 3.0 (a TF binding prediction tool available at: (http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3) gave hits for TFs that may bind in the region. Interestingly all the six CpGs out of seven in the flanking region of 5'-UTR of *CD24* gene, are part the predicted binding sites of transcription factors (Fig 6.29). An experimental scheme to confirm these predictions will involve mutagenesis of these CpG sites and follow how that affects the expression of CD24.

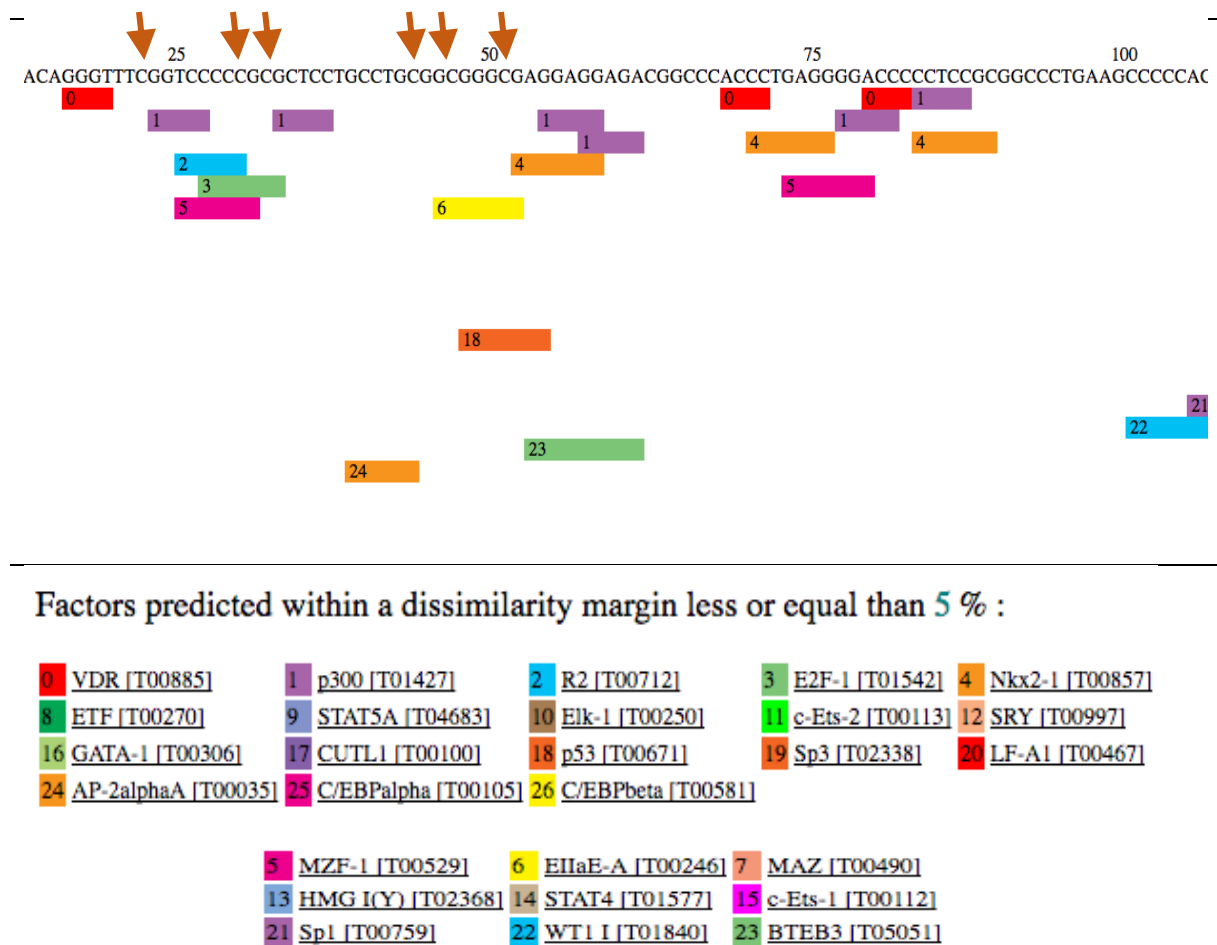


Fig 6.29: TFs binding prediction at the 5'-UTR flanking region. Six CpG positions are part of the region where TFs binding has been predicted. Different color indicates different TFs as shown in the bottom panel. The prediction tolerated 5% (or less) dissimilarity in sequence for binding by TFs.

However, it is important to remember that even if just one or two positions are more important for the regulation of *CD24*, methylation in the other CpG positions might be necessary for the stability of the suppression (Fig 6.27). Site directed mutagenesis can be carried out in order to confirm the relative importance of CpG-2 and other positions. This would allow to see if the mutagenesis results in continuous expression of CD24 in the CD24-negative clone(s)/cell lines because of the absence of methylation in that position and would help answer the question of whether CpG-2 plays a ‘seed’ role in the silencing CD24 expression or not.

Chapter-7

Evidence of Cooperation between Clones in Cell Line

Chapter 7

7.1 Chapter introduction

In chapter 4, it has been shown that CD24-differentially expressing clones are stably present in the LS174T cell line. A similar observation was made for the SW480 cell line (Chapter 5). A number of other CRC cell lines also indicated possible presence of clones with differential expression of CD24 (Section 5.3). The confirmation that CD24 expression is regulated by direct methylation (Chapter 6) provides an explanation for CD24 based dynamically heterogeneity in CRC cell lines.

In a heterogeneous environment, clones may have either negative-, no- or positive-interactions. In case of negative (or no) interaction between clones, it is mostly likely that the fastest growing clone will simply grow out, eventually eliminating the slow growing one from the culture environment. Such cell-autonomous interactions are generally due to mutations where cells with fitness-enhancing mutations are selected. However, apart from this competitive relationship, a positive interaction may also establish among cells in a heterogeneous environment in the form of complementation or synergism (Fig 7.1). A model proposed by Tomlinson and Bodmer (I. Tomlinson & Bodmer, 1997) has shown that only with positive interactions will two clones be stably maintained. In such cases, the two sets of cells with different properties interact in a positive way, so that each gains an advantage from the presence of the other.

According to this model, the stable presence of epigenetically maintained CD24⁺ and CD24⁻ clones in CRC cell lines would be supported only if they interact positively. There might be factors that the clones do not express separately but produce only when they come together. In such cases, the clones will either interact directly or communicate via factors secreted by one clone which is received by the other and vice versa.

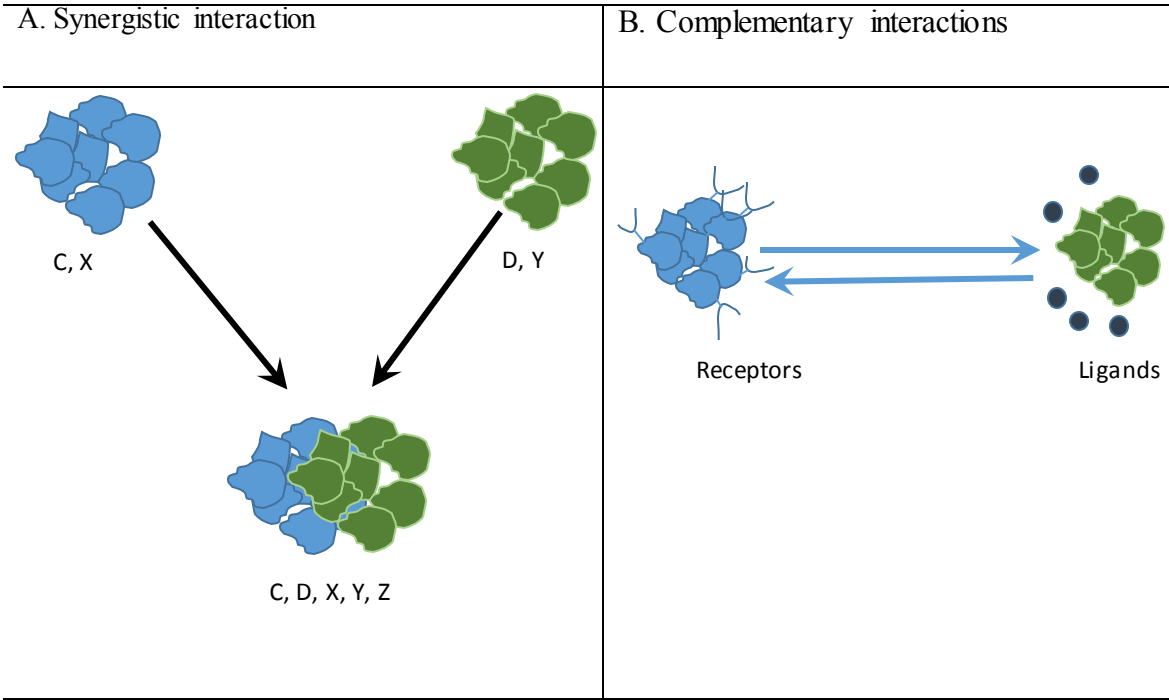


Fig 7.1: Positive interactions among cancer cells. A. In case of synergistic interaction, the mixed population expresses a new protein (z) compared to the pure clones. B. In case of complementary interaction, one clone may express a ligand that may be utilized by the other and vice and versa.

In order to demonstrate a positive interaction, during this thesis work, a simple mathematical model was developed which can predict the ratio of cells in a mixed culture if there is no (positive) interaction between the two populations.

$$R_T(C1, C2) = \frac{N_1 * 2^{\left(\frac{T}{Dt1}\right)}}{N_2 * 2^{\left(\frac{T}{Dt2}\right)}}$$

R_T = Ratio after T days

C1= Clone1

C2= Clone2

N1= Initial number of clone 1

N2= Initial number of clone 2

T= Total time of the co-culture

Dt1= doubling time of Clone 1

Dt2= doubling number Clone 2

Here the model predicts the ratio of Clone 1 and Clone 2 (R_T) after certain time (T) in culture given that their initial numbers (N1 and N2 respectively) and doubling time (Dt1 and Dt2, respectively) are known.

This model was applied to test the interaction between the clones of LS174T, namely LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-}. These clones are ideal to carry out a mixed culture experiment where they could easily be followed by their stable CD24 expression difference. It has been shown in Chapter 6 that LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-} are stable until late passage, which provides enough time to conduct the mixed culture experiment to test the proposed model and hypothesis.

The mixed culture experiment showed support for the positive interactions between the clones. Microarray data analysis of LS174T_Clone 1^{CD24+}, LS174T_Clone 2^{CD24-} and LS174T parental cell lines provided candidate genes that might be responsible for the interactions. To validate the microarray data and measure expressions in mixed culture and clones, qRT-PCR was conducted with convincing outcomes.

7.2 Mathematical model supports positive interactions between clones

LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-} had different doubling times, 23h and 30h respectively. From this doubling time, the number of divisions they will go through after a certain duration in the culture can be calculated (Table 7.1). This calculation shows that after 10 days in the culture, LS174T_Clone 1^{CD24+} will complete 2 more divisions than LS174T_Clone 2^{CD24-}. After ~6 weeks in culture, the difference will be 10 divisions.

Table 7.1: Doubling time and division number of LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-} after certain duration.

Clones	Doubling Time (hours)	Division after	
		10 days	41 days
LS174T_Clone 1 ^{CD24+}	22.84	10.52	43.15
LS174T_Clone 2 ^{CD24-}	29.32	8.18	33.56

Because of this difference in the growth rates of individual clones, the nature of the interaction between clones in the parental LS174T could be modeled in a mixed culture experiment where the clones are co-cultured for a specific duration. If there is no interaction between the clones, then the fast growing clone (LS174T_Clone 1^{CD24+}) would eliminate the slow growing one (LS174T_Clone 2^{CD24-}) from the mixed culture. However, if there is any cooperation between them, then LS174T_Clone 2^{CD24-} will be retained in the culture environment.

To follow the fate of individual clones in the mixed culture, the mathematical model mentioned in section 7.1 was applied. This model predicts the ratio of LS174T_Clone 1^{CD24+} to LS174T_Clone 2^{CD24-} in the culture based on the initial number of cells present in the mixed environment assuming that there is no interaction between them.

7.2.1 Modelling cooperation between Clone 1 and Clone 2 of LS174T: mixed culture experiment

To test the model, mixed cultures at three ratios (4:1, 1:1, and 1:4) were initiated by mixing CD24-positive LS174T_Clone 1^{CD24+} and CD24-negative LS174T_Clone 2^{CD24-}. Because of their distinct FACS profiles for CD24 expression the two clones were easily distinguishable (Fig 7.2).

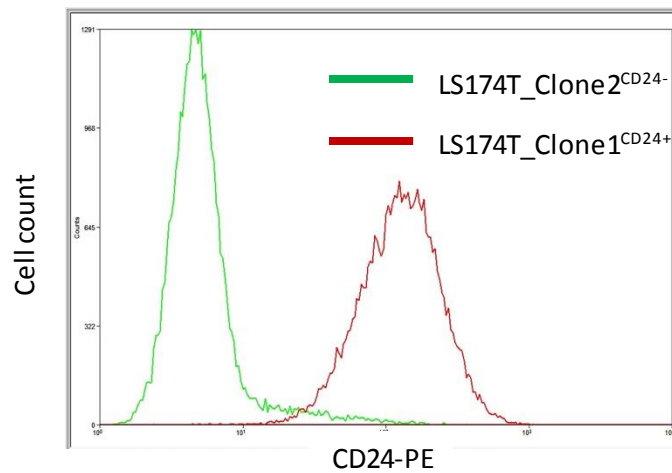


Fig 7.2: Clones of LS174T have distinguishable CD24 expression profile.

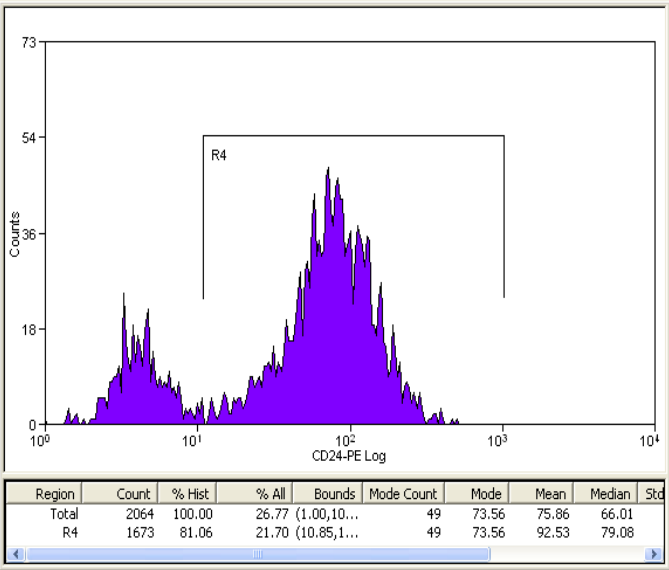
At the beginning of the mixed culture experiment (day 0), the profiles of the mixtures were checked to ensure that the culture started with the intended ratios (Fig 7.3). These profiles represent passage 0 (p0) for the mixed cultures.

A. 4 to 1 ratio:

LS174T_Clone 1^{CD24+} =
200000

LS174T_Clone 2^{CD24+} =
50000

CD24 expression profile of the 4:1 mixed culture at p0

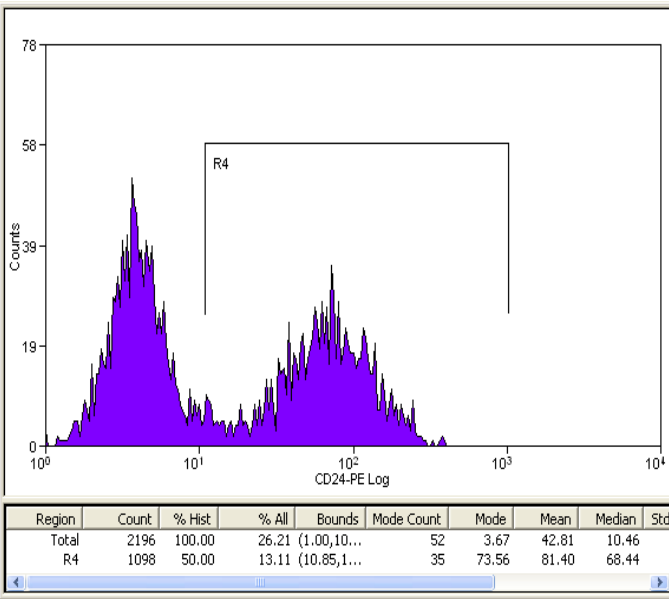


B. 1 to 1 ratio:

LS174T_Clone 1^{CD24+} =
125000

LS174T_Clone 2^{CD24+} =
125000

CD24 expression profile of the 4:1 mixed culture at p0



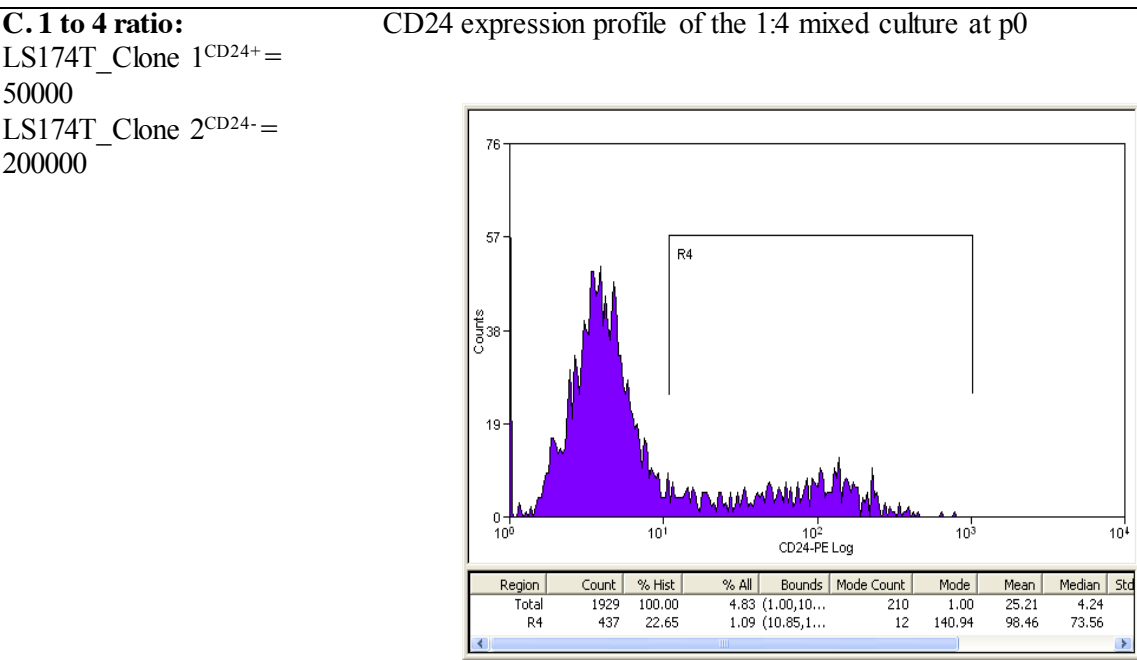


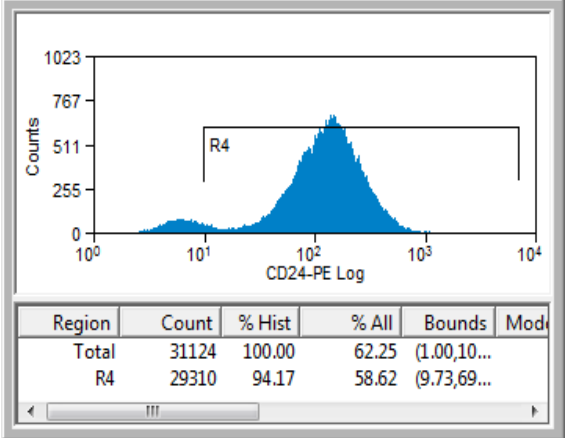
Fig 7.3: FACS profiles of clones LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-} in mixed cultures of (A) 4:1 ratio, (B) 1:1 ratio, and (C) 1:4 ratio on day 0. The ratios of CD24⁺ (indicated by the R4 bar) and CD24⁻ cells in the FACS diagrams represent their mixing ratios shown in numbers on the left. The % of CD24⁺ and CD24⁻ cells are shown at the bottom of each graph.

All three mixtures were maintained as continuous cultures and the ratio of the CD24⁺ cells (representing LS174T_Clone 1^{CD24+}) and CD24⁻ cells (representing LS174T_Clone 2^{CD24-}) were measured using FACS.

Due to the faster growth rate of LS174T_Clone 1^{CD24+}, the model of ‘no interactions’ presented earlier in section 7.1 would yield the following ratios for LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-} after 10 days:

Ratios (Clone 1: Clone 2)	Clones	Predicted %
4:1	Clone1	94
	Clone2	6
1:1	Clone1	84
	Clone2	16
1:4	Clone1	62
	Clone2	38

However, FACS analysis showed that more LS174T_Clone 2^{CD24⁻} persisted than the predicted ratios in the 1:1 and 1:4 mixed culture after 10 days (Fig 7.4).

Clone 1 Clone 2 ratio	Clones	Predicted %	Observed %	CD24 expression profile at Day 10
4:1	Clone1	94	94	
	Clone2	6	6	

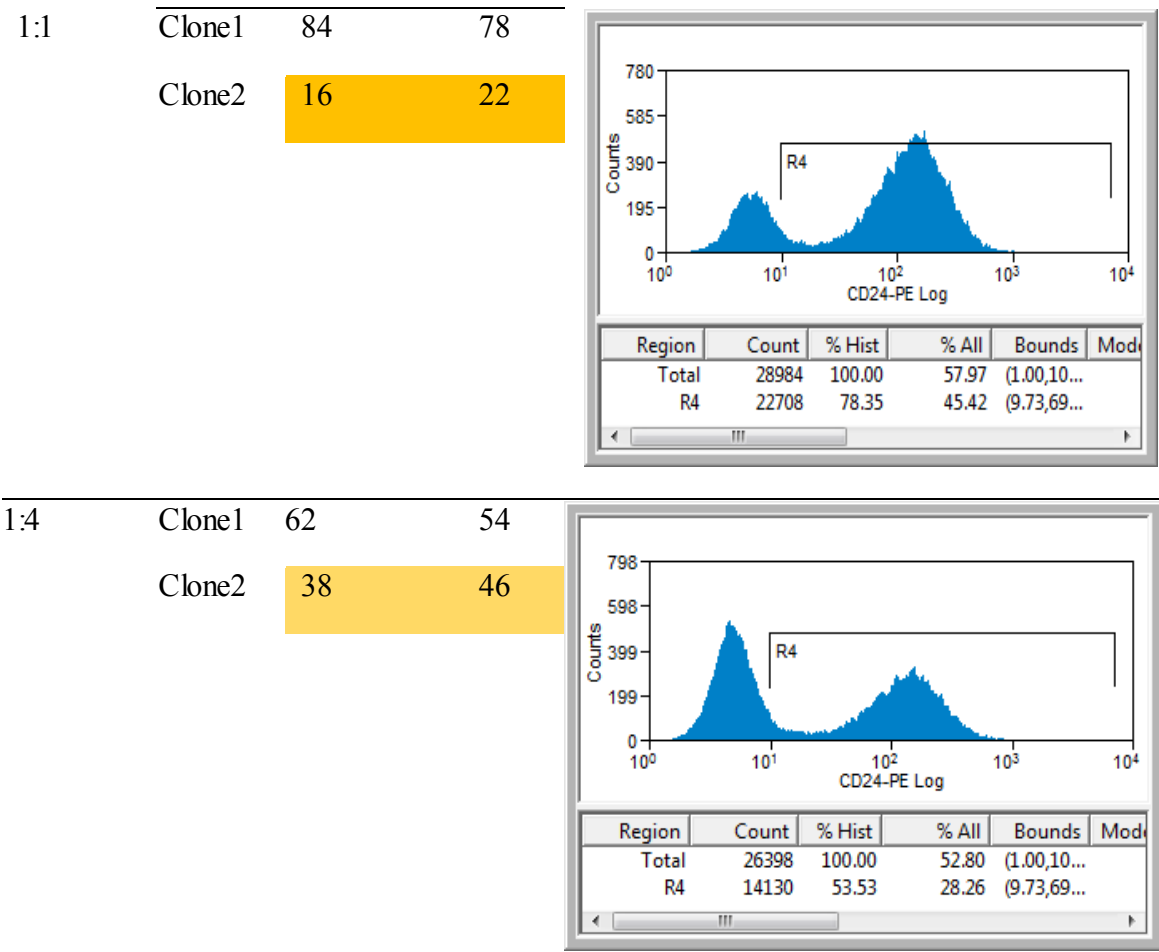


Fig 7.4: FACS profiles of the mixed cultures after 10 days. In the 4:1 mixed culture, the observed ratio matches the expected ratio of the model. However, in the 1:1 and 1:4 mixed cultures, more Clone 2 (LS174T_Clone 2^{CD24-}) is present (yellow highlights), indicating that its presence has been supported there.

The model predicts that after 41 days, LS174T_Clone 2^{CD24-} will be totally eliminated from all the mixed environment. However, the observed ratio showed significant number of LS174T_Clone 2^{CD24-} remained in the cultures of two (1:1 and 1:4) mixed environment (Fig 7.5).

Clone 1 Clone 2 ratio	Clones	Predicted %	Observed %	CD24 expression profile at Day 41
4:1	Clone1	99	99	
	Clone2	1	1	
1:1	Clone1	99	97	
	Clone2	1	3	

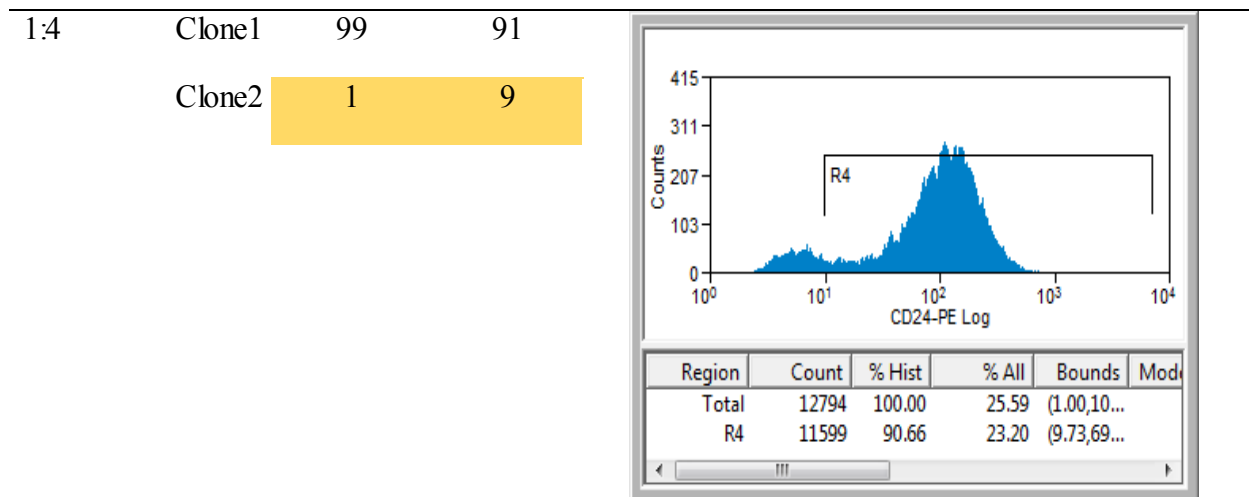


Fig 7.5: FACS profiles of the mixed cultures after 41 days. In the 1:1 and 4:1 mixed cultures, a more significant portion of Clone 2 (LS174T_Clone 2^{CD24-}) is present (yellow highlights) than predicted by the model. In the 4:1 mixed culture, the observed ratio matches the expected ratio of the model.

These results clearly show that the presence of LS174T_Clone 2^{CD24-} has been tolerated (and probably supported) by the LS174T_Clone 1^{CD24+} in the mixed cultures. It is noteworthy that, although LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-} co-existed in their 1:1 and 1:4 mixed cultures over the time, LS174T_Clone 2^{CD24-} got eliminated from the 4:1 culture, where LS174T_Clone 1^{CD24+} started with 4 times higher number. It indicates that sufficient amount of LS174T_Clone 2^{CD24-} needs to be in place to begin with in order for the cells to have enough time to establish an interaction through the time lag. However, the experiment will need to be repeated several more times to be sure of this assumption.

There might be a question whether the CD24⁻ populations found in the mixed cultures after 10 or 41 days actually represent LS174T_Clone 2^{CD24-} or they came from LS174T_Clone 1^{CD24+} (as this clone can give rise to a CD24⁻ subpopulation after a long time in culture (section 6.3.1)). To answer this questions, pure cultures of LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-} were maintained simultaneously with the mixed cultures. They were also profiled for CD24 expression at the same time the mixed

cultures were analysed. The analysis showed that the expression profiles of both of the clones remained unchanged throughout the experiment (Fig 7.6), confirming that in the mixed culture, the CD24⁺ or CD24⁻ populations sustained respectively from the LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-} that were mixed initially, and did not arise spontaneously from either of the clones.

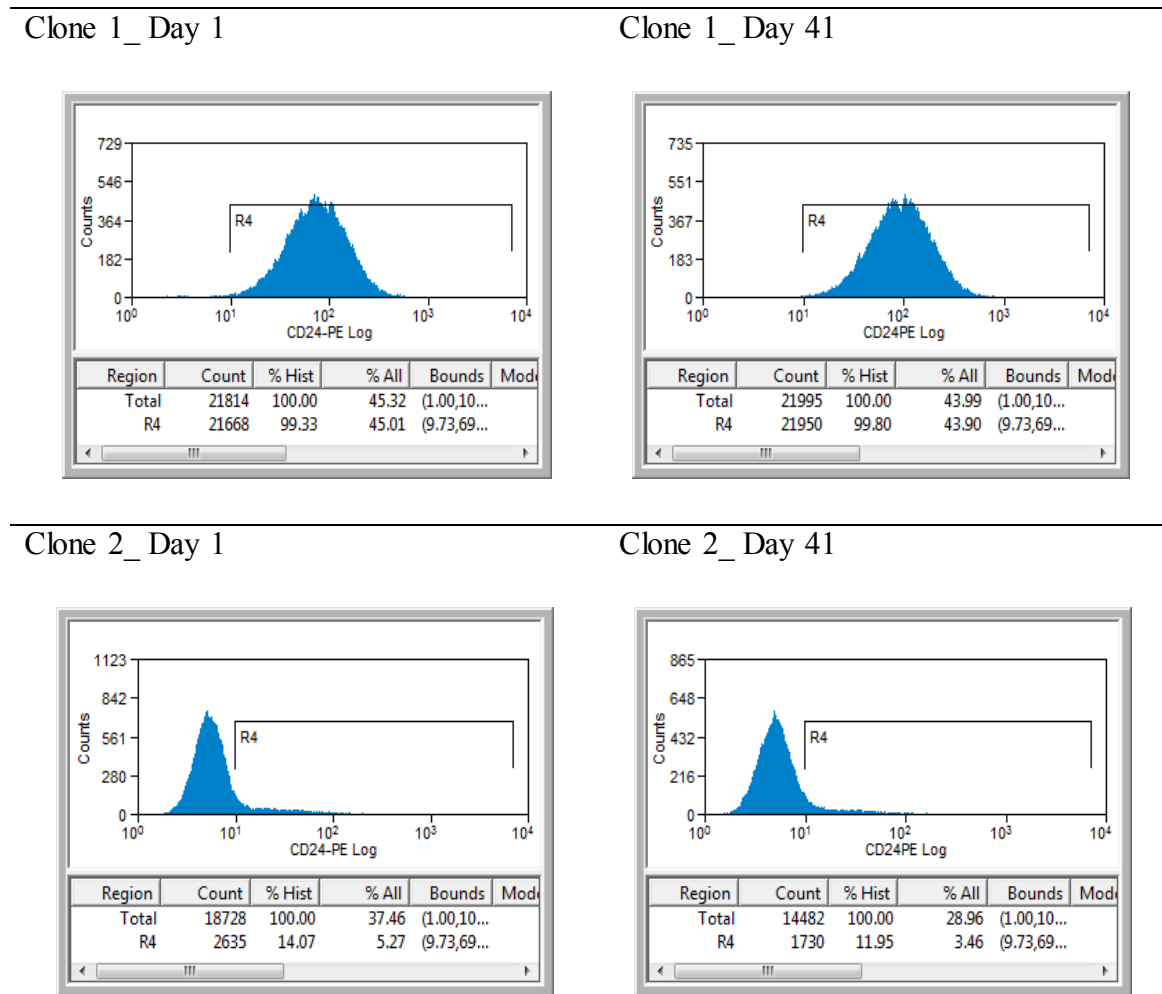


Fig 7.6: CD24 expression profiles of LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-} during the mixed culture experiment. The R4 bar areas indicate that the CD24 expression profile of the clones remained virtually unchanged during the experiment period.

The observation from the mixed culture experiment shows that at the beginning, the ratios of the clones are closer to the predicted ratios of the model (of no interactions), but with time, the observed ratios start to deviate more from the predicted ratios suggesting the interactions of the clones in the mixed culture. To investigate further which candidate genes may be participating in these interactions, microarray analysis and qRT-PCR of the clones and of the parental LS174T were conducted.

7.3 Evidence of positive interactions between the clones of LS174T

The results of the experiments described above provides theoretical support for the positive interactions between LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-} in the mixed cultures. As the LS174T cell line contains both clones, similar positive interactions are expected to exist in this cell line too.

To reveal such interactions, the microarray data of LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-} and their parent LS174T were examined to obtain lists of genes indicating either (i) synergistic or (ii) complementary interactions between the two clones.

7.3.1 Evidence of synergism: microarray analysis shows high expression of genes in the LS174T which are low in individual clones

In a synergistic relation, individual populations will not express certain gene(s), but when they come together, the gene(s) will be expressed in the mixed environment. To reveal whether this was the case, the microarray expression data of LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-} were compared with the parental LS174T to specifically look for the genes which are highly expressed in LS174T compared to both clones. In theory, such analysis should produce a list of genes that are expressed in the LS174T, supposedly due to the interactions of the two clones; separately, the clones either do not express them or express at a very low level.

From this analysis, a list of genes was obtained which are expressed several folds higher in the LS174T compared to individual clones (Table 7.2). All these genes are functionally relevant for cancer progression.

Table 7.2: Genes more highly expressed in LS174T compared to the clones and the functional role of these genes. Expression values are given on a linear scale.

GENE	Clone 1 (p8)	Clone 2 (p8)	LS174T (p125)	Functional significance of the gene in cancer
<i>VEGFA</i>	465	944	2703	<i>VEGFA</i> is known as a key mediator of angiogenesis and enhances cancer stem cell survival (Carmeliet, 2005; Grun, Adhikary, & Eckert, 2016).
<i>PYGB</i>	859	621	2002	<i>PYGB</i> overexpression helps cells to sustain glucose starvation (Philips et al., 2014).
<i>NEBL</i>	804	75	1942	<i>NEBL</i> is a member of the Nebulin family and fuses with MLL gene in leukemia (Cóser et al., 2010)
<i>BNIP3L</i>	531	625	1833	<i>BNIP3L</i> is induced by p53 under hypoxia and promotes tumour growth (Fei et al., 2004).
<i>SDC4</i>	955	670	1750	<i>SDC4</i> is involved in the migration of normal (Couchman, 2003) and tumour cells (O'Connell et al., 2009).
<i>REG1A</i>	30	30	1745	<i>REG1A</i> induces angiogenesis (Hara et al., 2015) and activates c-Jun through MAPK pathway (Wakita et al., 2015). It has been recognized as a prognostic marker of colorectal cancer (Astrosini et al., 2008).
<i>JAK1</i>	629	510	1700	<i>JAK1</i> is part of the JAK/STAT pathway which is overactive in cancer (Thomas, Snowden, Zeidler, & Danson, 2015).

This list of genes gives an indication of some candidates which might be expressed because of the co-existence of Clone 1^{CD24+} and Clone 2^{CD24-} in the LS174T cell line. *VEGFA*, *BNIP3L* and *REG1A* expression were selected for verification using qRT-PCR because of their expression difference and functional importance in cancer progression. However, qRT-PCR did not reproduce the observation of the microarray experiment for *VEGFA* and *BNIP3L* as there was no difference in cycle time (CT) between the clones and parental LS174T. These two genes were not considered for further analysis. Interestingly, qRT-PCR with *REG1A* confirmed that Clone 1 and Clone 2 does not express *REG1A* but LS174T does. (calculated from cycle times in qRT PCR, Fig 7.7). Here, the fold change was calculated from the deltaCT values where $2^{-[\text{deltaCT}]}$ is the fold difference of the *REG1A* compared to the reference gene *UBC* in individual samples.

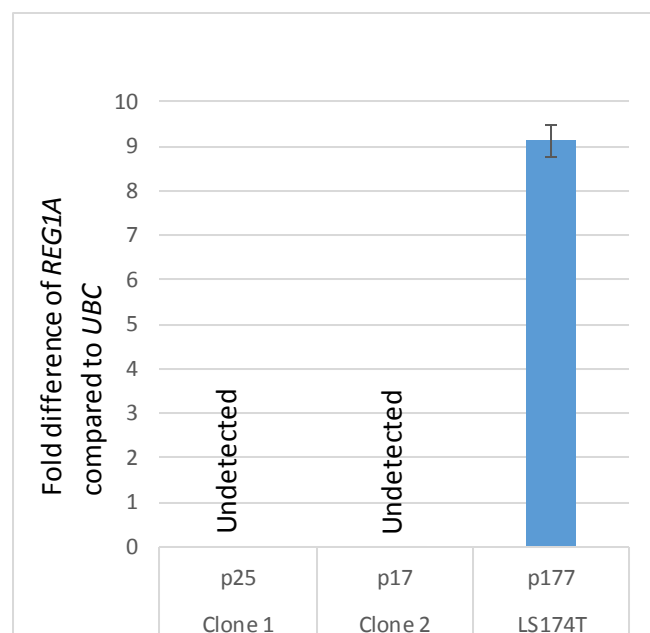


Fig 7.7: Fold difference in *REG1A* in qRT-PCR compared to *UBC* reference gene. Clone 1 and Clone 2 have been referred to as undetected as the cycle time too high (≥ 36). Parental LS174T has ~9 fold higher expression of *REG1A* compared to a regular gene *UBC* indicating its high abundance in the cell line. The error bars indicates the standard error of mean (S.E.M) the folds calculated for each experimental repeat.

This qRT-PCR confirmation for higher expression of REG1A in LS174T compared to the individual clones was similar to the observation from the microarray analysis, and so, REG1A was further examined.

7.3.2. Temporal influence on REG1A expression

The parental LS174T expresses very high level of REG1A whereas in individual clones it is essentially absent. If the expression in the parental LS174T is a result of the interaction between LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-}, then one may expect that when the clones are mixed and cultured together for a while, they would interact to re-express REG1A in the mixed culture. During the mixed culture experiments described above (section 7.2.1) *REG1A* level was measured by qRT-PCR in the individual clones and in the three different mixtures. However, the cycle time (CT) of *REG1A* was very high (≥ 36) in the mixtures up to 6 weeks, indicating no/very low expression of REG1A in the mixed culture environment (the higher the cycle time in the qRT-PCR, lower the expression of a gene).

Whether more time will increase the expression of REG1A in a mixed culture is an intriguing question. To investigate the time-dependent expression of REG1A, different passages of the clones were examined for REG1A expression. The results showed that in the late LS174T_Clone 1^{CD24+} (p62), REG1A is expressed as ~3 fold higher than reference gene *UBC*. However, the level of REG1A remained unexpressed in both early and late LS174T_Clone 2^{CD24-} (Fig 7.8).

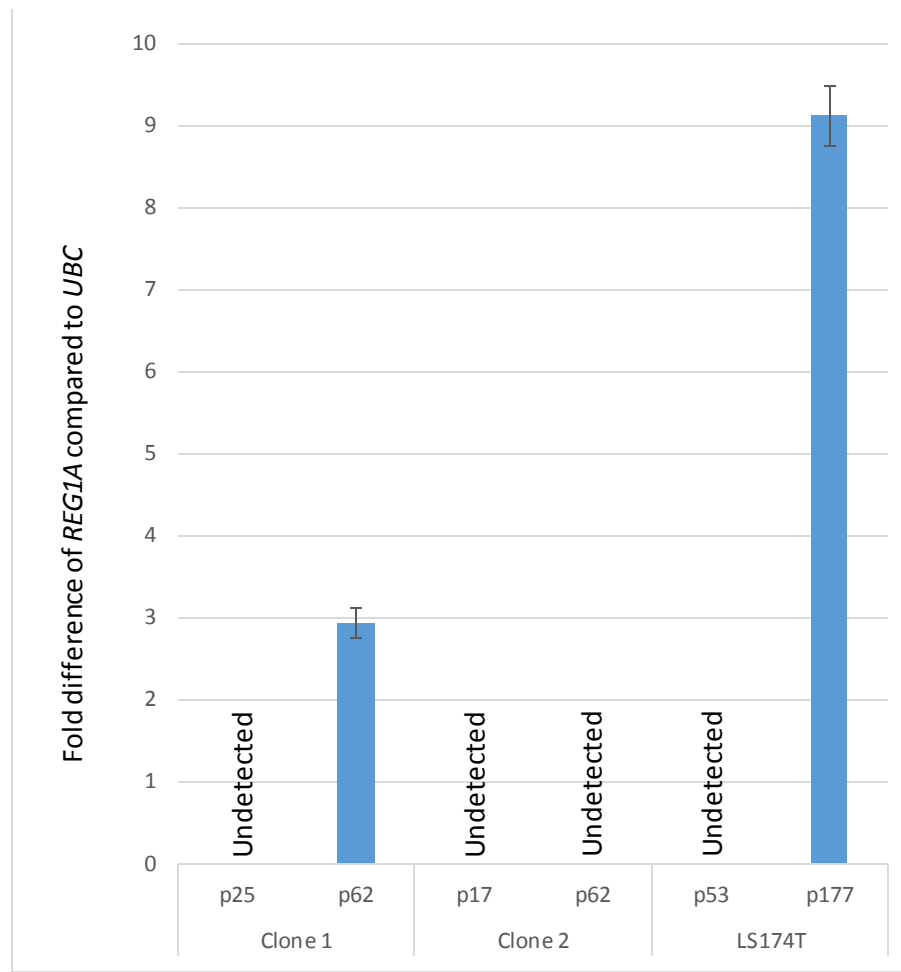


Fig 7.8: Expression of REG1A over time: LS174T_Clone 1^{CD24+} (indicated as Clone 1 and blue bars) changes the expression of REG1A from negative to very positive. Compared to UBC reference gene, late LS174T_Clone 1^{CD24+} has ~3 fold higher expression of REG1A. In early LS174T_Clone 1^{CD24+} REG1A is not expressed (shown as undetected). LS174T_Clone 2^{CD24-} does remain negative for REG1A expression across passages. LS174T late passage (P177) expresses REG1A but early passage (p53) does not. The error bars indicates the standard error of mean (S.E.M) the folds calculated for each experimental repeat.

An important observation here is the difference between the early and late passage LS174T. It shows that early passage LS174T does not express REG1A but late passage LS174T expresses very high amount although both of them contain CD24⁻ and CD24⁺ subpopulations. This suggests that time in the culture is important to influence to expression of REG1A. When for enough long the cells are in culture, the CD24⁺ subpopulation may contribute the expression of REG1A as seen from the difference between early and late Clone 1 and Clone 2 in Fig 7.8.

This temporal influence of REG1A expression is an interesting observation and need further investigation to understand the underlying regulation and consequence.

7.3.3 Complementation between clones: receptor-ligand relationship of CXCR4 and TFF2

In a complementary interaction, two or more subsets of cells may secrete factors that will be used by each other. In this case, one population of cells may secrete factors that will be received by other population and vice versa. In an extreme case of complementary relationship, cells that express the receptor may not express the ligand at all.

To find candidates for such complementary relations, microarray data were analysed to obtain a list of the receptor and ligand genes, whose expressions are most different between the clones of LS174T. This analysis revealed CXCR4 at the top of the receptor list, as it was expressed 40 times more in LS174T_Clone 1^{CD24⁺} than in LS174T_Clone 2^{CD24⁻}. Further, TFF2 was identified as the factor secreted in highest quantities, being twice as much in LS174T_Clone 2^{CD24⁻} than in LS174T_Clone 1^{CD24⁺}. The literature suggests that TFF2 binds as a ligand with CXCR4 to promote cancer growth (Dubeykovskaya, Dubeykovskiy, Solal-Cohen, & Wang, 2009).

qRT-PCR was performed to confirm the microarray data. It showed that *CXCR4* mRNA is undetermined in LS174T_Clone 2^{CD24⁻}, (indicating no/very low expression), whereas

in the LS174T_Clone 1^{CD24+}, the cycle time (CT) was 32.16, which indicates moderate expression. This observation is in line with the microarray data. However, the expression of TFF2 was not mutually exclusive in the clones, as the qRT PCR gave its cycle time as 35.77 in LS174T_Clone 1^{CD24+} and as 34.13 in LS174T_Clone 2^{CD24-} indicating a 3 times higher expression of TFF2 in LS174T_Clone 2^{CD24-} compared to LS174T_Clone 1^{CD24+}.

It was difficult to conclude the complementary interaction involving CXCR4 and TFF2 between clones. However, another experiment with CD24⁻ and CD24⁺ cell populations strengthened the assumption for such interaction. It has been mentioned in chapter 6 (section 6.3.2) that two subclones from LS174T_Clone 2^{CD24-} had different expression profiles regarding CD24: Sub 5 (p23) is CD24⁺ and Sub 3 (p20) is CD24⁻.

Sub 5 and Sub 3 represent, respectively, the CD24⁺ and CD24⁻ subpopulations in LS174T_Clone 2^{CD24-}. Therefore, it was argued that Sub 5 and Sub 3 may have a similar receptor-ligand relationship regarding CXCR4 and TFF2. Both subclones were examined for the expression of *CXCR4* and *TFF2* using qRT-PCR. The qRT-PCR data showed that the expressions of *CXCR4* and *TFF2* are mutually exclusive in those clones and are relatable to their CD24 expression (Fig 7.9). These expression profiles reinforce the observation made in the (CD24⁺) LS174T_Clone 1^{CD24+} and (CD24⁻) LS174T_Clone 2^{CD24-}.

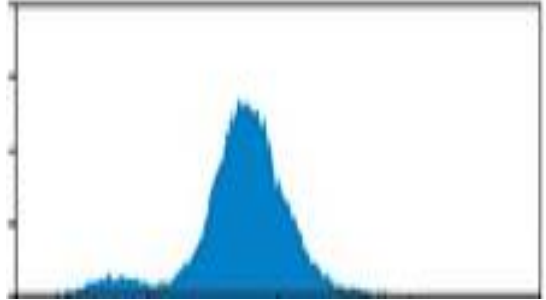
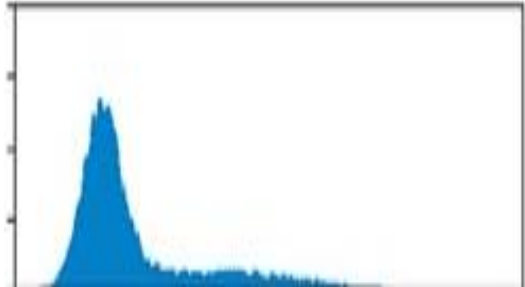
Subclones	Mean CT_ <i>CXCR4</i>	Mean CT_ <i>TFF2</i>	CD24 expression profile
Sub 5 (p23)	33.68	35.93 (extremely low expression)	 <p>CD24-PE</p>
Sub 3 (p23)	Undetermined	27.12	 <p>CD24-PE</p>

Fig 7.9: Mutually exclusive expressions of *CXCR4* and *TFF2*. qRT-PCR cycle time (CT) values indicate *CXCR4* mRNA is expressed in CD24⁺ Sub5 but not in CD24⁻ Sub 3. On the contrary, *TFF2* is almost absent in Sub 5 but highly expressed in Sub 3. FACS profiles of CD24 expressions are shown on the right.

It is difficult to come to a firm conclusion about the receptor-ligand complementary relationship between the LS174T_Clone 1CD24⁺ and LS174T_Clone 2CD24⁻ clones from qRT-PCR or microarray analyses. However, the results presented here provide some preliminary clues to indicate such interactions. Future research should focus on analysis at protein level to get more conclusive evidence.

It is important to mention here that the function of *CXCR4* is modulated by CD24. In pre-B cells and breast cancer cell lines, CD24 has been shown to alter *CXCR4* distributions in membrane lipid rafts (Schabath, 2006). Interestingly, all of the above observations indicate that *CXCR4* is expressed in clones and subclones which are CD24⁺.

7.4 Chapter Discussion

The observations in this chapter presented both theoretical experimental evidence in favour of the positive interactions among clones. The availability of CD24 differentially expressing clones LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-}, their stability in culture for a long time and their consistent presence in their parental cell line LS174T provided a convenient model to study the interactions *in vitro*.

It has been stated earlier (in chapter 4, section 4.2) that during culture, more cells of LS174T_Clone 2^{CD24-} are found in the suspension and that they attach to the culture flask relatively weakly compared to LS174T_Clone 1^{CD24+}. This raises the concern whether, during subculturing of the mixed culture, more LS174T_Clone 2^{CD24-} would be washed away. If so, that would affect the ratio in subsequent cultures. To test if this concern is justified, CD24 level was compared between cells in the suspension and cells that remained attached after a single wash with PBSA. The data showed that the ratio of CD24⁺ to CD24⁻ cells is the same in attached and suspension samples, indicating no bias from the attachment ability of the clones used in the mixed culture experiment (Fig 7.10).

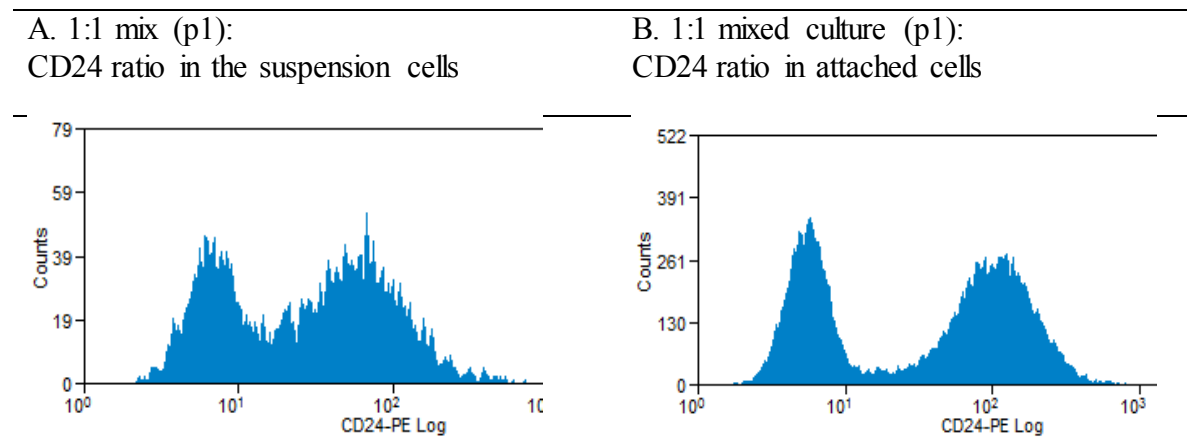


Fig 7.10: No bias from the attachment ability of clones used in the mixed culture: (A) Ratio of LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-} cells in suspension media the culture, (B) Ratio of LS174T_Clone 1^{CD24+} and LS174T_Clone 2^{CD24-} in attached cells.

The co-culture experiments presented in this chapter provide experimental support to the idea of non-autonomous interactions among cells, a theory of more than three decades old (Heppner, 1989; I. Tomlinson & Bodmer, 1997) microarray analysis of the gene expression in clones and their parent provided candidate genes that may mediate the interaction between the clones. qRT-PCR has confirmed the observation in clones and parental cell line.

An interesting observation during this study was the expression of REG1A in the late passage LS174T_Clone 1^{CD24+}. One speculation may be that the culture condition influenced such change over time. However, the fact that the late LS174T_Clone 2^{CD24-} did not express REG1A gives confidence that this has not been the case. Both of the clones should have expressed REG1A if culture condition was the inducer. However, the fact that the Late LS174T_Clone 1^{CD24+} achieved a CD24 expression profile very similar to that of parental LS174T possibly indicates the importance of the CD24 ratio among interacting populations for the expression of REG1A.

The ostensible proof for the clonal expression patterns of REG1A, CXCR4 and TFF2 indicates that they might be the potential candidates to investigate the non-autonomous interactions among cells proposed by Tomlinson & Bodmer (1997) model to explain angiogenic and autocrine/paracrine signaling between coexisting clones. More studies, such as blocking protein-protein interactions, gene silencing by knock down/out techniques and immunoprecipitations of interacting proteins can confirm some of the observations described here.

Chapter 8

General Discussion and Future direction

Chapter 8

8.1. Summary

The work presented in his thesis has - i) shown that several colorectal cancer (CRC) cell lines are heterogeneous with the co-existence of *CD24* differentially expressing clones; and - ii) revealed the methylation mediated regulation of *CD24* in colorectal cancer cells. Two cell lines, LS174T and SW480 were used extensively in the study and several clones were established from these parental cell lines by FACS sorting. Other CRC cell lines, such as RKO, CC20, HCT116, DLD1, NCIH716, LS411 were also studied when deemed relevant.

These studies have identified that the methylation of seven CpG positions within a small genomic region Chr6: 106975560-106975834 correlates with the expression of *CD24* in colorectal cancer cell lines. This work presents the first clear proof of methylation mediated regulation of *CD24*, which can explain the dynamic nature of the expression of this gene reported by other investigators (Meyer et al., 2009; Smith, Fairbridge, Pallegar, & Christian, 2015; Ayre et al., 2015).

8.2. Relevance of using cell line models for this work

The observation of *CD24*-based heterogeneity and the dynamic changes in *CD24* expression in the cell lines formed main body of this thesis. There are two routes that could have been followed after the initial observation of *CD24* expression change: i) functional study of the *CD24* and ii) understanding the regulation of the dynamic expression of *CD24* in colorectal cell lines. In this thesis, the second phenomenon has been elaborately studied.

The LS174T cell line and the clones gave a very convenient model to follow the dynamic regulation of CD24 and after studying colorectal cancer cell lines, clones and subclones, the conclusion about the methylation mediated regulation of CD24 was made.

Whether the regulation of CD24 follows the same dynamic change in primary cancer has not been investigated during this thesis work. However, based on the known facts that primary cancers go through global methylation change and CD24 expression is dynamically changed in breast and colorectal cancers, it can be confidently assumed that CD24 expression might be under methylation mediated regulation in primary cancers.

The consequence of this dynamic change might be subjected to further research.

8.3 Heterogeneity in CD24 expression in cancer cell lines and its implications

Cell lines are widely used model in cancer research to understand cancer biology, gene regulatory mechanisms, drug responses etc. Despite their limitations to mimic the actual biological system, cell lines provide convenient models in cancer study. In this thesis work, a relatively unexplored potential of cell lines, viz. the nature and evolution of heterogeneity have been revealed.

The observation of morphological heterogeneity in LS174T (Fig 3.8) during the study of the E-Cadherin expression and mutation was considered a significant feature of this cell line. Prominent morphological heterogeneity in cell lines is not very common, though not completely unprecedented (Baguley & Leung, 2011; Hastings & Franks, 1983). Differential morphology provides a ready phenotypic feature in the study of tumour heterogeneity. The maintenance of such heterogeneity within cell lines can help to understand some features of tumour heterogeneity in primary cancer.

The experimental evidence of the presence of heterogeneity in the LS174T cell line motivates assumptions on the origin of such heterogeneity in a cell line. One suggestion is: if a primary cancer contains subclones which engage in positive interactions, then during the process of developing a cell line from the cancer, the relationship may be maintained among the subclones as described in the model in Fig 8.1. Once the clones are separated, the advantageous relationship is disrupted and the pure clone may be

subject to selective pressures that favour the re-establishment of a profile that approaches to their parental profile (late Clone 1 and Clone 2 in Fig 8.1). Methylation mediated regulation of genes can provide the rapidity for such dynamic changes as methylation is a faster process than mutation mediated changes.

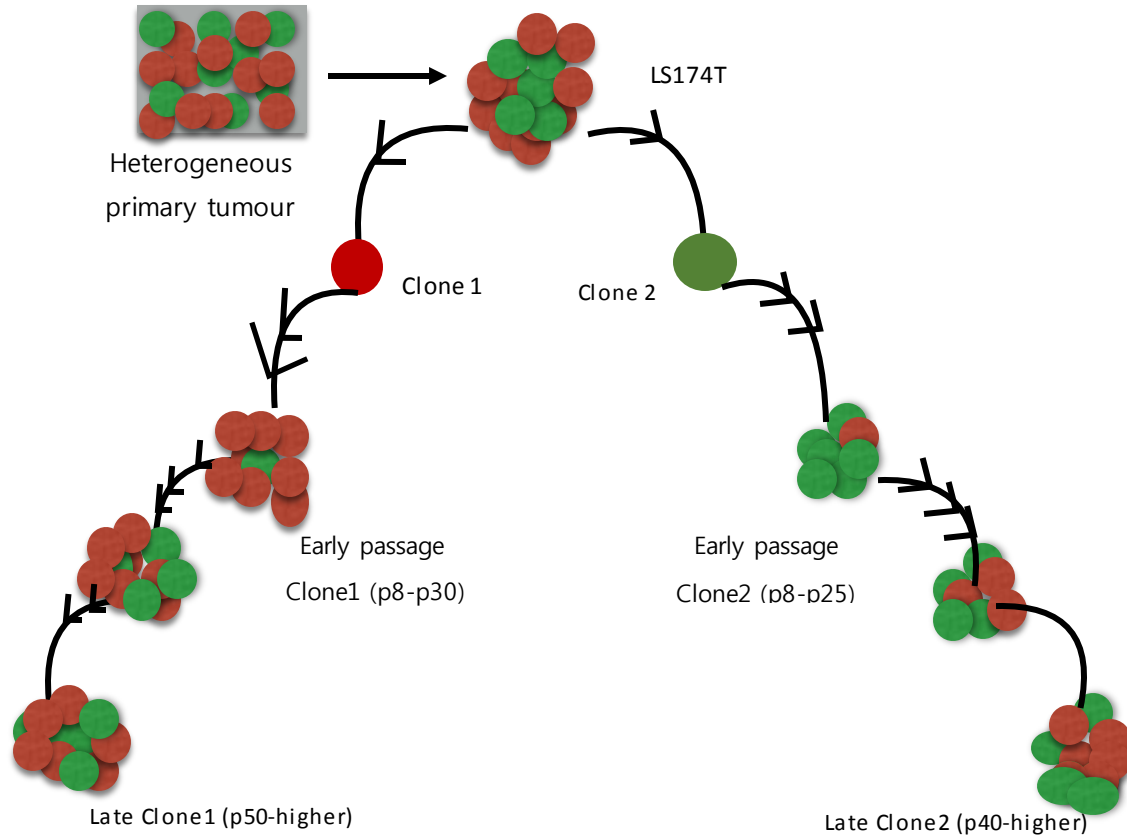


Fig 8.1: Possible route of heterogeneity in cell lines: if two (or more) clones were in a positive relationship in primary cancer, they may maintain the same relationship during the process of cell line development such that the resulting cell line might contain stably maintained subclones of the primary cancer. There might be a selective advantage of the co-existence of the subclones, as in the LS174T, mixture of CD24⁺ and CD24⁻ subpopulations are maintained. If separated, the two sub-populations evolve by supporting each other until they reach to a level for stable co-existence.

The model presented in Fig 8.1 may explain several examples where clones were approaching their parental profile regarding CD24 expression (Chapter 6, section 6.3.1). The appearance of a CD24⁻ subpopulation in the LS174T_Clone 1^{CD24+} makes its profile very similar to that of parental LS174T (Fig 6.1 A). Likewise, LS174T_Clone 2^{CD24-} (Fig

6.2 B) at passage 40 has begun to produce a minor population of CD24⁺ cells which may indicate movement towards the parental LS174T profile.

Another example of such movement towards parental profile of CD24 expression was found in SW480_Clone C2^{CD24+}. SW480 is mostly a CD24⁻ cell line. Early SW480_Clone C2^{CD24+} (up to p8) is CD24⁺. However, SW480_Clone C2^{CD24+}, when maintained in the culture, at passage 10, starts to lose its CD24 expression and by passage 19 became indistinguishable from parental SW480 regarding its CD24 profile. In contrast, SW480_Clone D1^{CD24-} which is CD24⁻ and similar to parental SW480 from the beginning, does not change its CD24 expression at all over a similar number of passages. It is noteworthy that the changes in SW480 clones happen much quicker than for LS174T clones.

These observations in the CRC cell lines have two-pronged significance: i) it confirms the dynamic nature of CD24 expression in CRC cell lines, a characteristic previously been observed in other cases (Ayre et al., 2015; Meyer et al., 2009; Smith et al., 2015); ii) it gives scope to capture the clones at different stages with regards to their CD24 expression to understand the regulatory aspect of *CD24*.

8.4 Methylation of a small part of a CpG island regulates CD24 expression

The dynamic nature of CD24 expression, though known for a long time, has never been mechanistically explained. However, it is unlikely that selection for sequential mutations could account for the dynamic changes in CD24 expression that were observed over relatively short periods of time in culture.

Further studies provided evidence that the methylation status of just two CpG positions can explain the expression or suppression of CD24 (Section 6.6) in CRC cell lines and clones. Much higher levels of methylation at these two positions correlates strongly with lower expression of CD24. However, it is striking that in the CD24-negative cell lines (such as of CC20 Figure 6.17) there is much more consistent methylation of all of the seven CpG positions in this region. It may be that when there is a **stable** lack of expression over very long periods of culture, a higher proportion of the relevant CpG

positions as well as those flanking them become methylated and the probability of reversion of expression is lower because it requires loss of methylation in more of the CpG sites. When there is partial methylation at these sites and some sites with almost no methylation, this may still be enough to block mRNA synthesis, but much less change in methylation status is then needed to switch on expression than in the more stable situation, as represented by the line CC20. Perhaps it is that lack of stability in the switching off of CD24 in the sub-clones, due to a lesser extent of overall methylation across the sequence upstream of the transcription start site, which leads to a higher frequency of reversion to CD24 expression.

This emphasises the need to have a much better understanding of the relationship between CpG patterns of methylation and mRNA synthesis.

8.5 CD24: Is it a marker to identify stem cells with differentiation ability?

Several studies have associated CD24 expression with stemness of cancer cells (Beça et al., 2013; King et al., 2012; Kristiansen et al., 2003; Yeung et al., 2010).

However, one interesting observation regarding CD24 was its association with several differentiation markers. In a microarray analysis across the panel of 95 cell lines in the Bodmer laboratory, *CD24* showed statistically significant positive correlation with mRNA expression of *CDX1*, *CEACAM5* and *KRT20* (yellow highlight in Table 8.1). *CDX1* is a well-known regulator of epithelial differentiation in colorectal tissue and cancer and the Bodmer lab has shown clearly that *KRT20* is transcriptionally regulated by *CDX1* (Chan et al., 2009; Jones et al., 2015; Soubeyran et al., 1999). *CEACAM5* and *KRT20* (and *KRT17*) are both well characterized and widely used differentiation markers in intestinal tissue (Merlos-Suárez et al., 2011). Besides these associations with differential markers, interestingly, in a similar analysis, *CD24* shows positive association with two stem cell markers (*PROM1*) and *EPHB2* (Table 8.1).

Table 8.1: Correlations between CD24 and differentiation regulators and markers (CDX1, KRT20, CEACAM5) and stemness markers (PROM1 and EPHB2). Expression in 95 CRC cell lines were compared in this analysis by calculating Pearson correlation coefficient.

	<i>CD24</i> <i>vs.</i> <i>CDX1</i>	<i>CD24</i> <i>vs.</i> <i>KRT20</i>	<i>CD24</i> <i>vs.</i> <i>CEACAM5</i>	<i>CD24</i> <i>Vs.</i> <i>PROM1</i>	<i>CD24</i> <i>Vs.</i> <i>EPHB2</i>
P (two-tailed)	< 0.0001	< 0.0001	< 0.0001	0.0236	0.0012
Number of Pairs	95	95	95	95	95

The explanation may be that, all the predicted “stem” cell markers are actually ‘marks’ for cells that retain the capacity to differentiate, and so “predict” for and will correlate with expression of differentiated markers. That would explain why whenever we have a cell line with higher expression of these “stem cell” markers, we also see high expression of differentiated markers such that there is a positive correlation between the stem cell marker and the differentiated cell marker. This implies that probably none of the stem cell markers we have detected are marking the type of stem cell that we see in the cell lines that do NOT have the capacity to differentiate. It can be noted here that Yeung et al. (Yeung et al., 2010) found that in HCT116, which is probably all stem cells, there was no correlation between tumour formation and CD44/CD24 status. That supports the above contention.

Microarray analysis just looks at global expression patterns, and in order to clarify and provide evidence for the hypothesis suggested above, one would have to separate the populations of cells that express the various candidate stem cell markers such as CD24, and then compare the expression of the known differentiated markers such as CDX1, KRT20 and CEACAM5 in that subpopulation versus the bulk population.

8.6 Morphology of cell is not related to CD24 expression

The morphological differences between the CD24-differentially expressing clones facilitated their initial isolation (Chapter 4, Section 4.2). LS174T_Clone 1 and LS174T_Clone 2, and SW480_Clone C2 and SW480_Clone D1, were primarily selected

for further study because of their morphological differences. So, it was of interest to see if there is any consistent relation between CD24 expression and morphological appearance of clones (and cell lines).

During this thesis work, it was possible to explore the possible link between CD24 expression and morphological differences, as some of the clones changed their CD24 expression during their course in the culture. However, observation of the morphology of different clones over time showed that there was no clear connection between cellular morphology and changes in CD24 expression. When CD24⁺ SW480_Clone C2 (p8) became CD24⁻ (p19), they did not change their overall morphology. Similarly, none of the changes in CD24 expression profiles associated with any of the other clones (and subclones) derived from LS174T were paralleled by morphological changes.

8.7 Limitations of using cell lines in this work

Although cancer cell lines are convenient *in vitro* models to study different aspects of cancer biology, they lack the actual environment of *in vivo* cancer. One of the key limitations is their time in the artificial culture system. It is a known phenomenon that because of their long exposure to culture environment, cell lines may change significantly over the time (Nelson-Rees, Owens, Arnstein, & Kniazeff, 1976). It might be the culture medium or the trypsinisation or mechanical detachment procedure together with the overall lack of microenvironment that influence the changes in the cell lines. Cell lines are prone to cross contamination (Gillet, Varma, & Gottesman, 2013; Nelson-Rees & Flandermeyer, 1977) too. Because of all these limitations, many studies, especially the drug response studies with cancer cell lines fail to reproduce in living system.

In this thesis work, the observation of the changes of CD24 expression in the clones over time in the culture should be taken with caution. Whether these observations will remain true in *in vivo* system requires further investigation.

8.8 Concluding remarks and future direction

The work presented in this thesis has convincingly revealed the methylation of CpGs in the flanking region of the 5'-UTR as a regulatory mechanism of CD24 expression. The dynamic regulation of *CD24* has unveiled a deeper challenge of coping with the

heterogeneity in cancer. It shows that cancer cells by modifying methylation status of a small number of CpG positions can change the expression of certain genes.

The role of individual CpG positions to control gene expression is still a new area to investigate. This study shows strong evidence of the association of two CpG positions with control the expression of CD24. A further extension of this work will be site directed mutagenesis of those CpG positions to reveal their direct involvement in CD24 expression.

From this study, there was no clear evidence that differential expression of CD24 is linked with response to monotherapy with common anticancer drugs. However, the very subtle differences that were observed may need further careful investigation (Chapter 4, section 4.10).

An important resolution of this thesis is the heterogeneity in CRC cell lines. This work has opened a new window to look at cancer cell lines for their use to study tumour heterogeneity *in vitro*. It will provide more flexibility and control to study this complex aspect of tumour biology. Understanding the evolution and maintenance of clones in cell lines can give important information about clonal evolution and heterogeneity in primary cancer.

Appendix and Reference

APPENDICES

Ax. 1: Sequence of the 5' flanking region of CD24 examined for CpG methylation

```
>hg38_dna range=chr6:106974606-106977143 5'pad=0 3'pad=0
strand=+ repeatMasking=none
AGCAGCCCCAGCCGAGCCTGGCCACCATTGCTCTGCCCATGTCCCCTCC
GTCGGTGCGCGGCGCGTCTAGCAGGATGCTGGGTGCTTGGAGAACCGCTG
GCTCCGGGCGGGGCGCAGGCAAGGTGGGGAGCGCGGCGAGCCGGCGAGACC
TTATATACCAGGAAAGCCACAATAGCCGTGACGTGGCGCCGCCCGCCAC
CCCGGCTCCCCCTCCGCCGCCGCCGCCGCTCCCCGCTTCCCCGCCCCG
CTCCGGGTTCCCCAGCGCCCCCGCCGCCGCCGCTGCCACCGCCGCTC
CGGCCCCTCGCGGGGCGGGAGCCCCCGGGCGGGCGCCGCTGCAGGTGGGC
GGGCGCCGCCGCTGCTGGTACCCGGCTGGTATCCCCGCGCCGGCTGCGG
GCTGCGGGCAGGCTGGCACAGGCGCCAGCGCCTCCGGCACAGGACCCGG
GGCACACACGCGAGCCACACACGCCGCGCTGGGGGCGGGGCCAAGTTTC
CTTTGTTTCCCGGGACCTGCCATCTTACCCCCAAAAGAAAAGTCCGCGC
CTCCGCTGCGGCGGGGTAGCCGCCTGTCTGGAGCAAGTGCATTGCCACT
CAGACTGCAAAATGCGCCCCGGGGACCCGAGAGGGGCTGCGAATTCCAGT
GCGACCGCGTGGGGGGCGGGCGCCGGGACCCCTCGGGCAGGGTTATCTCT
CGGCCCCCGCGCAGAGGAAAGGGGAACGCGGCGGGCGTGGCGCGGACGCG
GCCGACGCTACCCCTGGCCGACGGGACCCGGGCGCCCTAGCGCGAAC
CCTTCCCGGCTGCTCGGCGCTCCTCGCAGGCGGTGTCCGGGTCCGGAGCG
ATCTGCGCGCTCGGCCCCGCGGCCGCGCCCTCCCCGAAGCCCTTGCTTTG
TTCTGTGAGCGCCTCGTGTGAGCCAGGCGCAGTGAGCTCACGGGGGCGTC
CCGGGTCCGCATCTTCCCAGGAGCTGGGGAGCCGCTCGCTGGGCGCAGAC
CCGCTGCCTGACGCTGCAAACTACAGGGTTTCGGTCCCCCGCGCTCCTGC
CTGCGGCGGGGCGAGGAGGAGACGGCCCCACCCTGAGGGGACCCCTCCGCG
GCCCTGAAGCCCCACCGCCCTCCCCGAGCGGCTCTGCTCCCAGGTTCTA
CTCGGGGGCAGGAACCTCCCTCCTGGCGGTCCCACTACAATTACCGTC
CCTCTCGAATTCTTGGGGATACACTTTTTTCGGTCTCTGTTGCCCGTCATT
CGTTTAAAGAGCAGAACCGTTCGAGATAAAGGCATTTACTTATCCGTAGT
TGCCATTAGGCAATTTAATGAACCGGCACAAAGAGGAGATTCGTATCAC
TCTTTAGATTTTGGTGGAATACTTTTTTCCTATCCATCTTATTGTGAAAA
CTTTCAAGCATACAACAAAAGTTGAAGTAATTTTACTGGTGAGTATTTT
GAGTTTAAATTTTAATTGCAAGACTTAATCTTTTTATTGAAGGTTTAGCC
CAAGTGTTTTTTCAGTCTTTAAAGCAAACAGCAGAGACCAAACCTAGTATG
GTTGTAAATTTGTTTAAATTTTCAGTTTTGCAGAACACTCTCGTTAGTATT
TTTGGTTCAATTTTTTAATTGCAAGACTTAATACTTTTATTTAAGGTTTAG
CCCAAGCATTTTTTCGGTCTTTAAAGTAAACCACAGAGTAGTAAGATTCTA
AATTTGTCTAAATTTCACTTTTTCGTGGCAGTCTCATTAAAACAAAGTAA
GTTTCAGGGAAGCAAAGTGACAATGAAATCTGCAGTAGGTTTGTGACTG
AATAGTGACTGAATTGGTTTGTGCCTTCAAGGTCAGTCACTTTAGCTCAG
TAACTTTCCAAGCACGTTTAAGATGTGCTTGGGAACACAAGAATTTCTGT
CGAGCACCAGTGGACTTTCCTGTTTAACAGTGTACCACATTATATAGAAA
TACTTTGGCCGGGCGTGGTGGCTCACACCTATAATCCAGCACTTTAGGA
GGCTGAGGAGGGCTGATCACCTGAGGTCAGGAGTTCAGACCAACCTGAC
CAACATGGTGAAACCTAAAATCTAGTCTCTACTAAAATTTAGTCTCTACT
AAAAATCCAAAATTAGCCATGCATGGTTGCTCATACCTGTAATCCAGCT
ACTTGGGAGGCTGAGGCAGGAGAATAGCTTGAACCTGAGAGGTGGAGGTT
GCAGTGAGCAGAGATCAGGCCATTGCATTCCAGCCTGGGCAACAAGAGCA
AAACTCTGTCTCAAAAAAAAAAGAAAAAGAAATAGTCTGGCCCAGCTTC
TCCATTCTTTTTCCGGGTGAAAAAATGGACTGTTTGTGACTCAGGAGTG
CCGGGTTCCAAGGCAAGGAACATACTATGTCTCCACAAGCTGCTTGTG
AACTAACTGGTGGTATTCCTAGGATAAGACAGCCTCTAAAGACAGACTGC
CAGAGTTCAAACCGTGGCTCTGCTAATTTTGTGGAGACAAAAGCCTCA
GTTCTTCATCTATAAAATGGAGATAATAACATCATTA
```

Ax.2: SNPs analysis confirms the origin of Clone 1 and Clone 2 in the LS174T cell line

Cell line	Nucleotides in the 34 SNP positions compared for cell line identity																																	
LS174T	AG	G	G	CT	G	A	TC	A	GA	AG	A	G	C	T	T	GA	AG	G	G	AG	GA	CT	G	A	G	A	TC	G	G	GC	T	A	AG	DEL
CLONE 1	AG	G	G	CT	G	A	TC	A	GA	AG	A	G	C	T	T	GA	AG	G	G	AG	GA	CT	G	A	G	A	TC	G	G	GC	T	A	AG	DEL.G
CLONE 2	AG	G	G	CT	G	A	TC	A	GA	AG	A	G	C	T	T	GA	AG	G	G	AG	GA	CT	G	A	G	A	TC	G	G	GC	T	A	AG	DEL.G

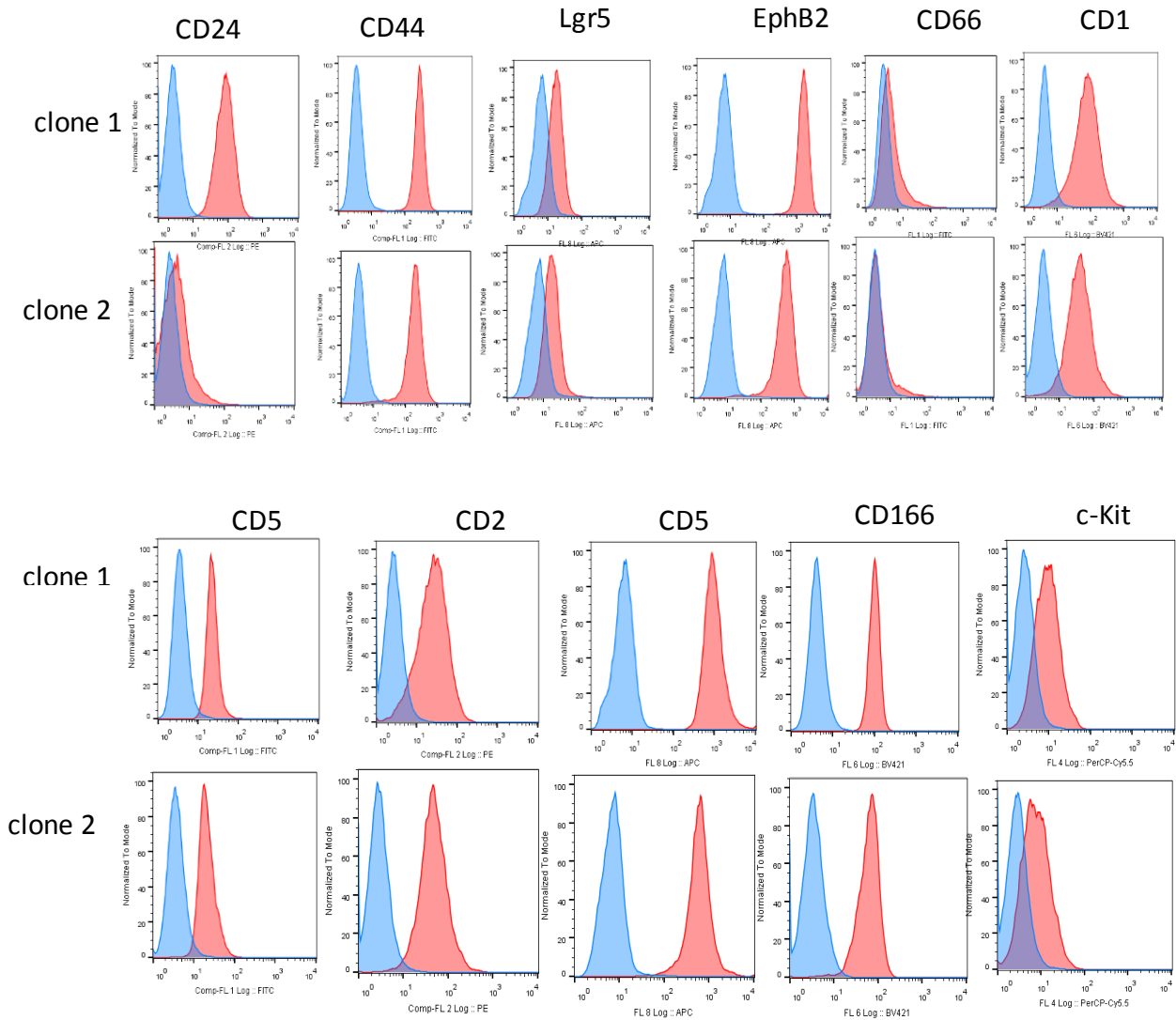
Ax.3 : A list of stemness markers compared between the clones of the LS174T

Fig Ax.3: Flow cytometry analysis of 11 makers in Clone 1 and Clone 2: Out of 11 makers, only CD24 showed difference between Clone 1 and Clone 2. Blue peaks: Isotype control, Red peaks: Antibody staining for respective markers. (courtesy: Dr Kamila Koprowska)

Ax.4: List of genes that are more highly expressed in Clone 1^{CD24+} than in Clone 2^{CD24-} of LS174T

Gene symbol	Fold high in Clone 1 compared to Clone 2
<i>LAPTM4B</i>	181.132
<i>CXCR4</i>	39.3365
<i>PPIC</i>	32.5887
<i>ELL2</i>	20.9648
<i>MT1M</i>	17.9295
<i>NEBL</i>	17.008
<i>FMR1NB</i>	16.3668
<i>CXCR4</i>	14.9167
<i>BLVRA</i>	14.4181
<i>FAM213A</i>	13.7084
<i>IL8</i>	13.1103
<i>FAM83B</i>	12.4361
<i>PMEPA1</i>	11.3773
<i>PDLIM1</i>	9.98435
<i>PKIB</i>	9.33209
<i>NT5E</i>	9.27141
<i>LAMC1</i>	8.93489
<i>H19 / MIR675</i>	8.41857
<i>ZNF462</i>	8.39154
<i>RCAN3</i>	8.26375
<i>TFPI</i>	7.95647
<i>IL17RD</i>	7.93632
<i>PLA2G16</i>	7.91652
<i>ASB4</i>	7.79792
<i>HSD3B1</i>	7.36248
<i>SERPINE2</i>	7.35459
<i>SORBS2</i>	7.0197
<i>TGFBI</i>	6.91737
<i>DKK1</i>	6.88089
<i>FOXQ1</i>	6.72313
<i>CD24</i>	6.63827
<i>BLVRA</i>	6.59517
<i>S100A4</i>	6.36198
<i>MEF2C</i>	6.2797
<i>EIF5A2</i>	6.27559
<i>LTBP1</i>	6.18733

Ax.5: List of genes compared between Clone 1 and Clone 2 of LS174T of sequence differences

<i>APC</i>	Adenomatous Polyposis Coli
<i>CTNNB1</i>	Catenin (Cadherin-Associated Protein), Beta 1, 88kDa
<i>TP53</i>	Tumor Protein P53
<i>KRAS</i>	Kirsten Rat Sarcoma Viral Oncogene Homolog
<i>BRAF</i>	V-Raf Murine Sarcoma Viral Oncogene Homolog B
<i>PIK3CA</i>	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase, Catalytic...
<i>PIK3C2B</i>	Phosphatidylinositol-4-Phosphate 3-Kinase, Catalytic Subunit...
<i>FBXW7</i>	F-Box And WD Repeat Domain Containing 7, E3 Ubiquitin Protein..
<i>TGFBR2</i>	Transforming Growth Factor, Beta Receptor II (70/80kDa)
<i>SMAD4</i>	SMAD Family Member 4
<i>ATM</i>	Ataxia Telangiectasia Mutated
<i>CDX1</i>	Caudal Type Homeobox 1
<i>RUNX3</i>	Runt-Related Transcription Factor 3
<i>CDH1</i>	Cadherin 1, Type 1, E-Cadherin (Epithelial)
<i>EPCAM</i>	Epithelial Cell Adhesion Molecule
<i>POLE</i>	Polymerase (DNA Directed), Epsilon, Catalytic Subunit
<i>MLH1</i>	MutL Homolog 1
<i>MLH3</i>	MutL Homolog 3
<i>MSH2</i>	MutS Homolog 2
<i>MSH3</i>	MutS Homolog 3
<i>MSH6</i>	MutS Homolog 6
<i>PMS2</i>	PMS2 Postmeiotic Segregation Increased 2 (S. Cerevisiae)
<i>PTEN</i>	Phosphatase And Tensin Homolog
<i>SOX9</i>	SRY (Sex Determining Region Y)-Box 9
<i>BCL9L</i>	B-Cell CLL/Lymphoma 9-Like
<i>ACVR2A</i>	Activin A Receptor, Type IIA
<i>AXIN2</i>	Axin 2
<i>FAT1</i>	FAT Atypical Cadherin 1
<i>CREBBP</i>	CREB Binding Protein
<i>TCF7L2</i>	Transcription Factor 7-Like 2 (T-Cell Specific, HMG-Box)
<i>RNF43</i>	Ring Finger Protein 43
<i>EP300</i>	E1A Binding Protein P300
<i>CD24</i>	CD24 Molecule
<i>CD44</i>	CD44 Molecule (Indian Blood Group)
<i>HLA-A</i>	6p22.1 Major Histocompatibility Complex, Class I, A

<i>HLA-B</i>	Major Histocompatibility Complex, Class I, B
<i>HLA-C</i>	Major Histocompatibility Complex, Class I, B
<i>B2M</i>	Beta-2-Microglobulin
<i>RUNX1</i>	Runt-Related Transcription Factor 1
<i>NRAS</i>	Neuroblastoma RAS Viral (V-Ras) Oncogene Homolog
<i>NOTCH1</i>	Notch 1
<i>BMI1</i>	BMI1 Polycomb Ring Finger Oncogene

Ax.6: Change in CD24 expression in the clones of DLD1 cell line

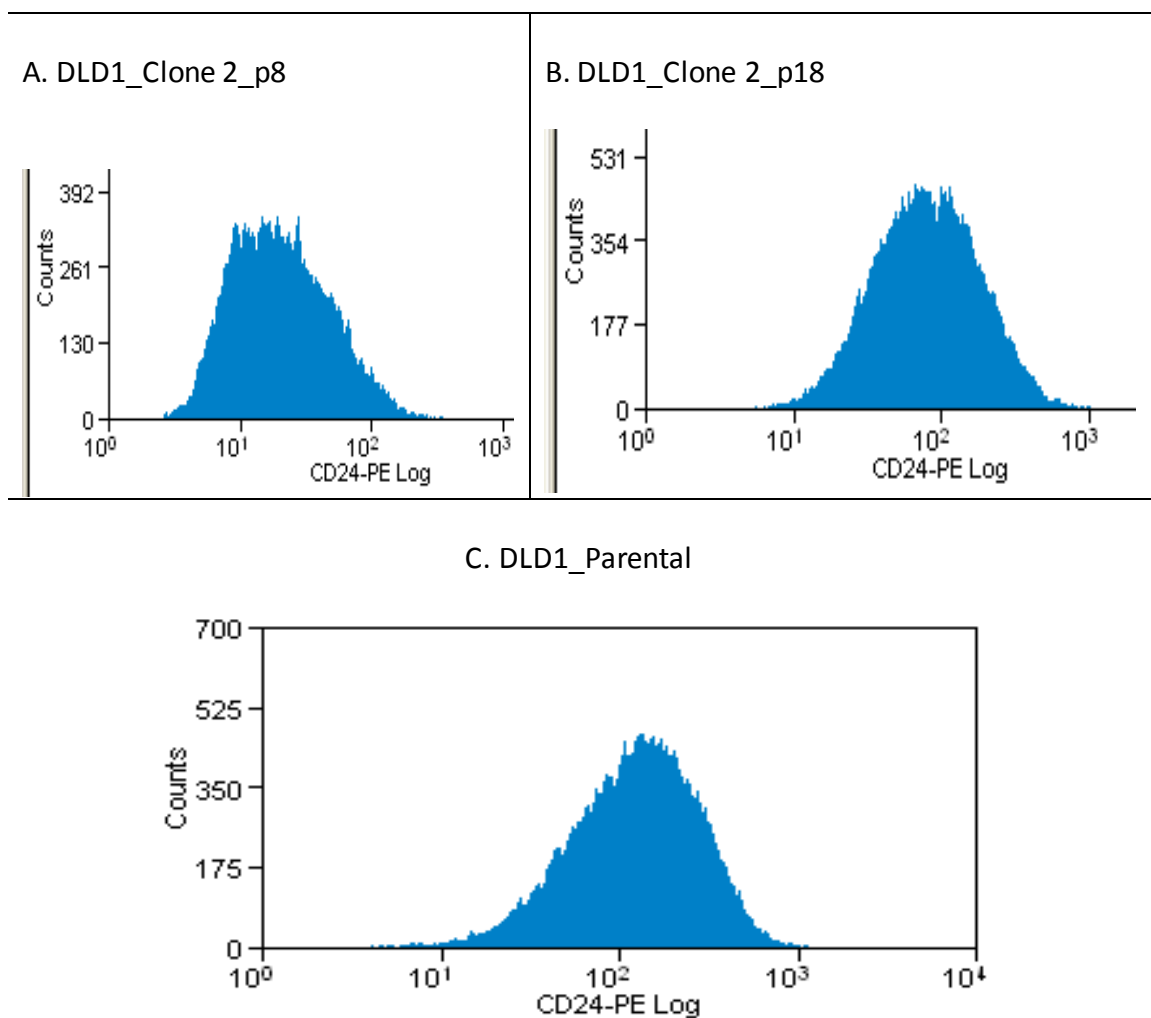
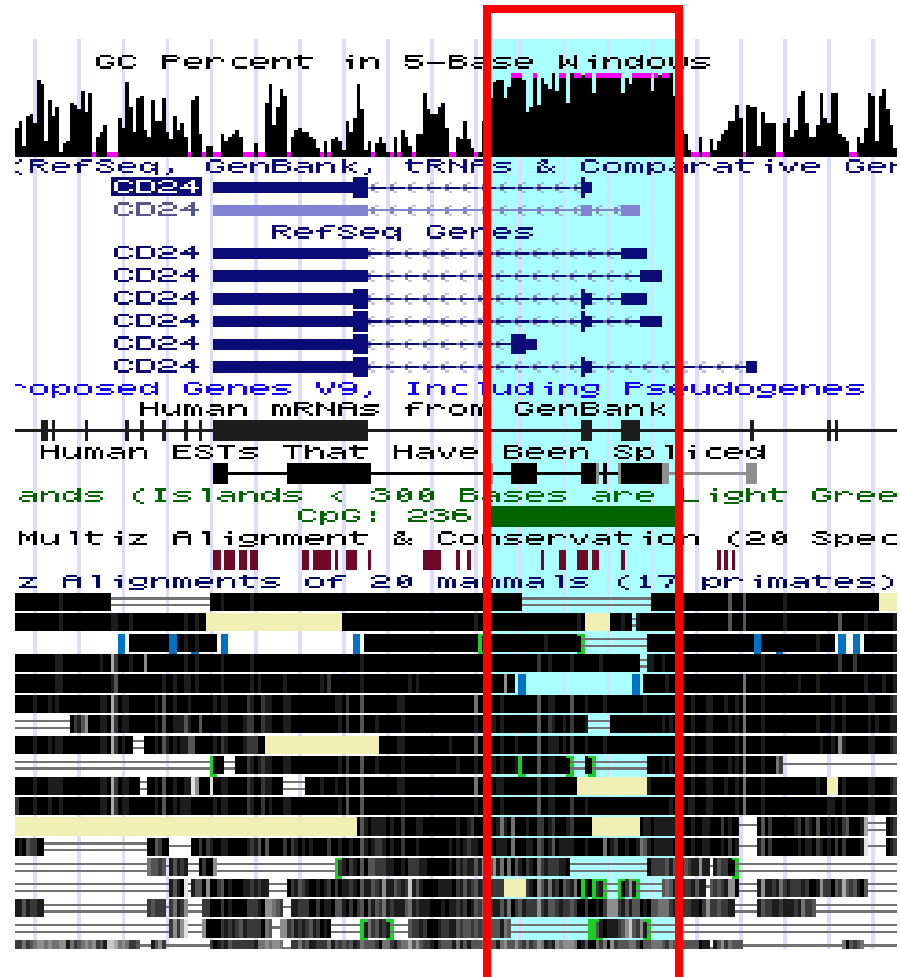


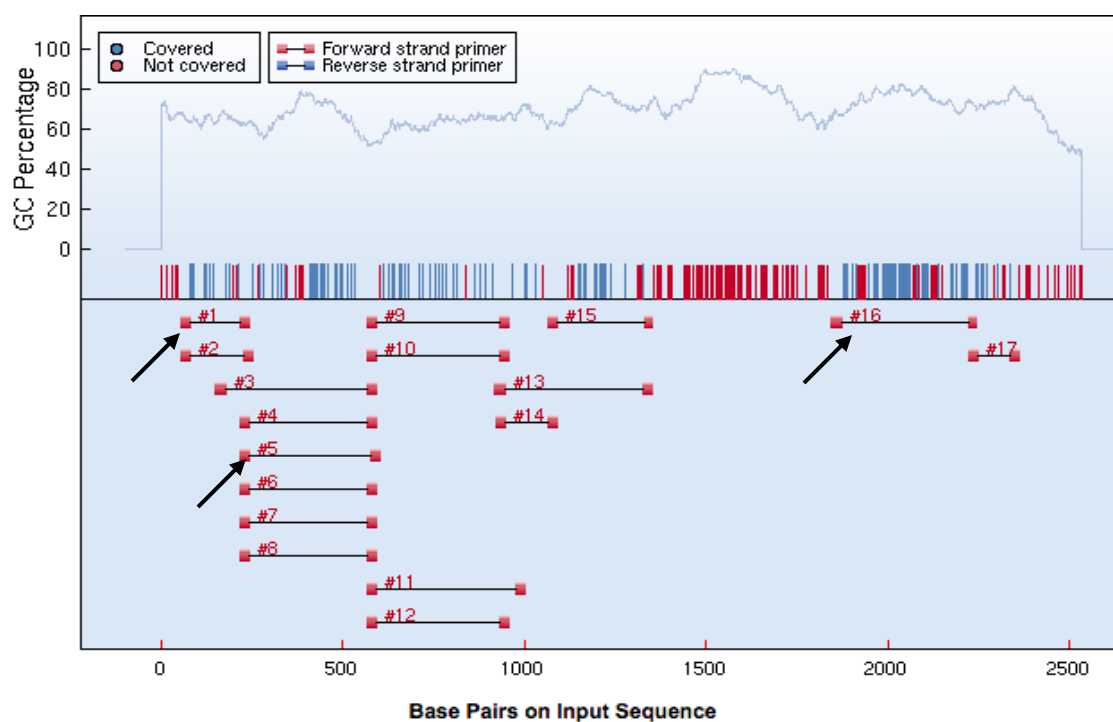
Fig Ax.6: The changes in CD24 expression in DLD1 clone approaching parental expression profile. Early DLD1_Clone 2 of passage 8 (A) has lower expression of CD24 than DLD1_Clone 2 of passage 18 (B) which looks now similar to parental DLD1 (C).

Fig Ax.7: Methylations in the CpGs of which are downstream to the Exon 1 region of CD24 does not differ between Clone1^{CD24+} and Clone 2^{CD24-} of LS174T.

A. The region which were examined for primer designed have been shown highlighted (in magenta). Most of the region is downstream to the Exon 1 (shown in red box)

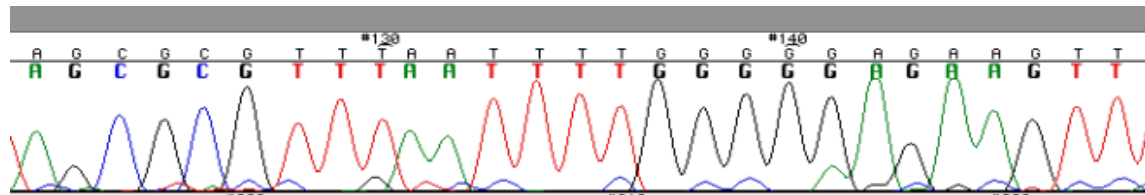


B. Primers predicted by Epidesigner were downstream in the Exon 1 region. Two primers 1, 5, 16 were sequenced after BS modification of Clone 1 and Clone 2 DNA.

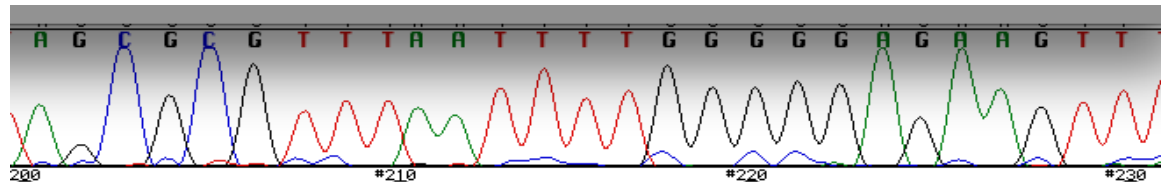


C1. There was no difference between Clone 1 and Clone 2 in those regions. Here representative part of amplicon 1 has been shown:

Clone 1

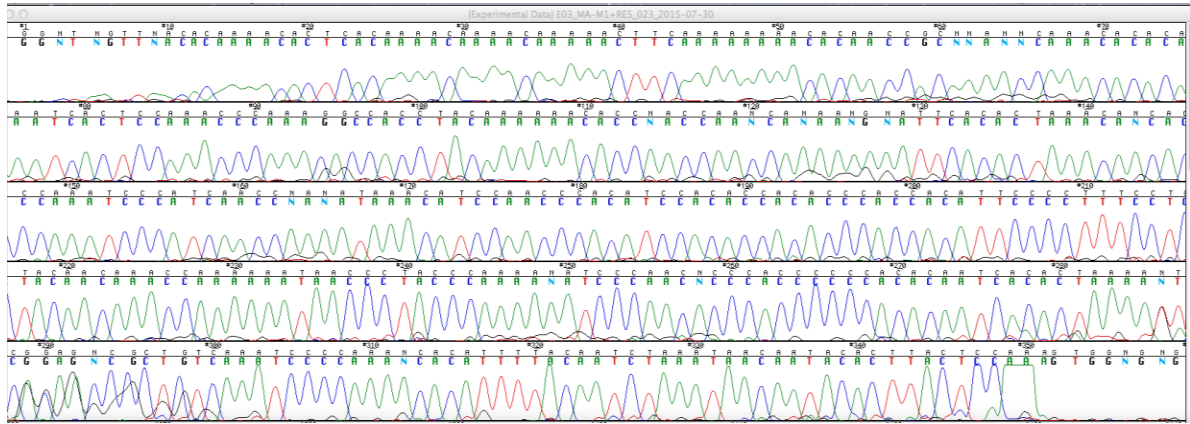


Clone 2

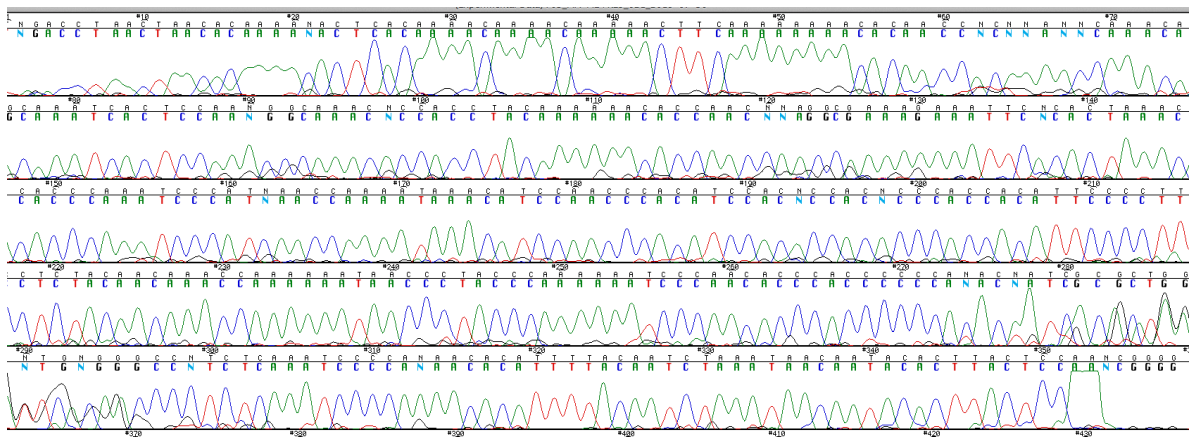


C2. The entire amplicon of primer 5 has been shown below. No difference can be found between the clones in this region.

Clone 1



Clone 2



Ax.8: Colony morphology of E-Cadherin^{low} cell lines in matrigel.

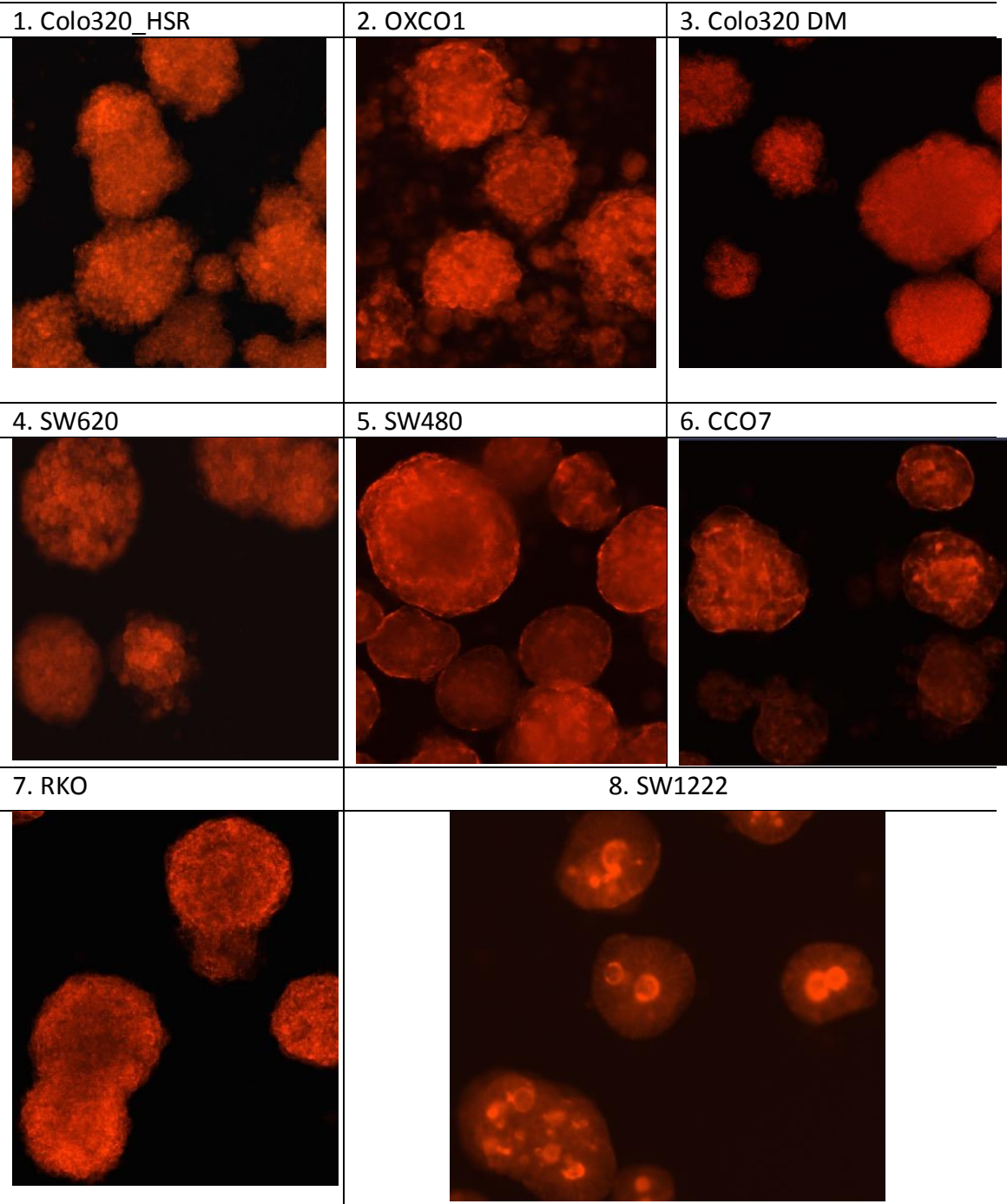


Fig Ax.8: All E-Cadherin low cell lines in the panel of 95 cell lines in the Bodmer laboratory are unable to form well defined lumens in matrigel. Image 1-7 are of mesenchymal like cell lines, Image 8 is of SW1222 which is a representative of a group of E-Cadherin high cell lines that retain the capacity to differentiate in matrigel and shows contrast between non-lumen vs lumen containing colonies in matrigel. Red: TRITC-Phalloidin for F-actin staining.

Ax.9: mRNA levels of *TTL* in the CRC cell lines

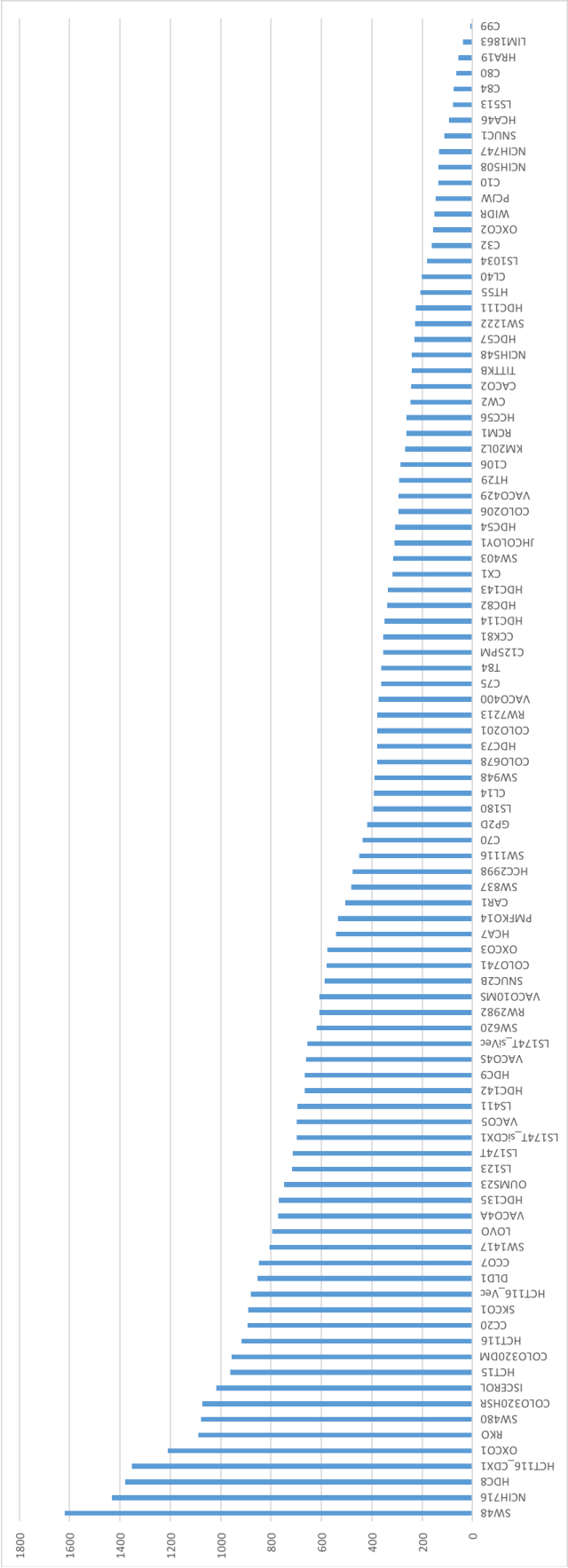


Fig Ax.9: Expression of *TTL* in CRC cell line panel. X axis: Cell lines, Y-axis: linear value of *TTL* mRNA expression.

LIST OF REFERENCES

- Abdel-Rahman, W. M., Katsura, K., Rens, W., Gorman, P. A., Sheer, D., Bicknell, D., ... Edwards, P. A. (2001). Spectral karyotyping suggests additional subsets of colorectal cancers characterized by pattern of chromosome rearrangement. *Proceedings of the National Academy of Sciences of the United States of America*, 98(5), 2538–43. <https://doi.org/10.1073/pnas.041603298>
- Agrawal, S., Kuvshinov, B. W., Khoury, T., Yu, J., Javle, M. M., LeVea, C., ... Gibbs, J. F. (2007). CD24 Expression is an Independent Prognostic Marker in Cholangiocarcinoma. *Journal of Gastrointestinal Surgery*, 11(4), 445–451. <https://doi.org/10.1007/s11605-007-0091-5>
- Ahmed, M. A. H., Al-Attar, A., Kim, J., Watson, N. F. S., Scholefield, J. H., Durrant, L. G., & Ilyas, M. (2009). CD24 shows early upregulation and nuclear expression but is not a prognostic marker in colorectal cancer. *Journal of Clinical Pathology*, 62(12), 1117–1122. <https://doi.org/10.1136/jcp.2009.069310>
- Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., & Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, 100(7), 3983–8. <https://doi.org/10.1073/pnas.0530291100>
- Ashley, N., Yeung, T. M., & Bodmer, W. F. (2013). Stem cell differentiation and lumen formation in colorectal cancer cell lines and primary tumors. *Cancer Research*, 73(18), 5798–809. <https://doi.org/10.1158/0008-5472.CAN-13-0454>
- Auman, J. T., & McLeod, H. L. (2010). Colorectal cancer cell lines lack the molecular heterogeneity of clinical colorectal tumors. *Clinical Colorectal Cancer*, 9(1), 40–7. <https://doi.org/10.3816/CCC.2010.n.005>
- Ayre, D. C., Elstner, M., Smith, N. C., Moores, E. S., Hogan, A. M., & Christian, S. L. (2015). Dynamic regulation of CD24 expression and release of CD24-containing microvesicles in immature B cells in response to CD24 engagement. *Immunology*, 146(2), 217–33. <https://doi.org/10.1111/imm.12493>
- Baguley, B. C., & Leung, E. (2011). Heterogeneity of Phenotype in Breast Cancer Cell Lines.

- Barker, N., van Es, J. H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., ... Clevers, H. (2007). Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature*, *449*(7165), 1003–1007.
<https://doi.org/10.1038/nature06196>
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., ... Garraway, L. A. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, *483*(7391), 603–307.
<https://doi.org/10.1038/nature11003>
- Battula, V. L., Evans, K. W., Hollier, B. G., Shi, Y., Marini, F. C., Ayyanan, A., ... Mani, S. A. (2010). Epithelial-mesenchymal transition-derived cells exhibit multilineage differentiation potential similar to mesenchymal stem cells. *Stem Cells (Dayton, Ohio)*, *28*(8), 1435–45. <https://doi.org/10.1002/stem.467>
- Beça, F. F. de, Caetano, P., Gerhard, R., Alvarenga, C. A., Gomes, M., Paredes, J., & Schmitt, F. (2013). Cancer stem cells markers CD44, CD24 and ALDH1 in breast cancer special histological types. *Journal of Clinical Pathology*, *66*(3), 187–191.
<https://doi.org/10.1136/jclinpath-2012-201169>
- Bendardaf, R., Elzagheid, A., Lamlum, H., Ristamäki, R., Collan, Y., & Pyrhönen, S. (2005). E-cadherin, CD44s and CD44v6 correlate with tumour differentiation in colorectal cancer. *Oncology Reports*, *13*(5), 831–5. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15809746>
- Bodmer, W. F. (2006). Cancer genetics: colorectal cancer as a model. *Journal of Human Genetics*, *51*(5), 391–6. <https://doi.org/10.1007/s10038-006-0373-x>
- Bredberg, A., & Bodmer, W. (2007). Cytostatic drug treatment causes seeding of gene promoter methylation. *European Journal of Cancer*, *43*(5), 947–954.
<https://doi.org/10.1016/j.ejca.2006.12.003>
- Buono, R. Del, Pignatelli, M., & Hall, P. A. (1991). Control of differentiation in a rectal adenocarcinoma cell line: The role of diffusible and cell-associated factors. *The Journal of Pathology*, *164*(1), 59–66. <https://doi.org/10.1002/path.1711640111>
- Burrell, R. A., McGranahan, N., Bartek, J., & Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, *501*(7467), 338–345. <https://doi.org/10.1038/nature12625>
- Cairns, J. (1975). Mutation selection and the natural history of cancer. *Nature*, *255*(5505), 197–200. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1143315>

- Campbell, P. J., Pleasance, E. D., Stephens, P. J., Dicks, E., Rance, R., Goodhead, I., ... Stratton, M. R. (2008). Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 105(35), 13081–6.
<https://doi.org/10.1073/pnas.0801523105>
- Chen, G.-Y., Tang, J., Zheng, P., & Liu, Y. (2009). CD24 and Siglec-10 selectively repress tissue damage-induced immune responses. *Science (New York, N.Y.)*, 323(5922), 1722–5. <https://doi.org/10.1126/science.1168988>
- Christman, J. K. (2002). 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene*, 21(35), 5483–5495. <https://doi.org/10.1038/sj.onc.1205699>
- Claus, R., Lucas, D. M., Stilgenbauer, S., Ruppert, A. S., Yu, L., Zucknick, M., ... Byrd, J. C. (2012). Quantitative DNA Methylation Analysis Identifies a Single CpG Dinucleotide Important for ZAP-70 Expression and Predictive of Prognosis in Chronic Lymphocytic Leukemia. *Journal of Clinical Oncology*, 30(20), 2483–2491.
<https://doi.org/10.1200/JCO.2011.39.3090>
- David, A. R., & Zimmerman, M. R. (2010). Cancer: an old disease, a new disease or something in between? *Nature Reviews Cancer*, 10(10), 728–733.
<https://doi.org/10.1038/nrc2914>
- Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development*, 25(10), 1010–22. <https://doi.org/10.1101/gad.2037511>
- DeVita, V. T., Lawrence, T. S., & Rosenberg, S. A. (2011). *DeVita, Hellman, and Rosenberg's cancer: principles & practice of oncology*. Wolters Kluwer Health/Lippincott Williams & Wilkins.
- Duppel, U., Woenckhaus, M., Schulz, C., Merk, J., & Dietmaier, W. (2016). Quantitative detection of TUSC3 promoter methylation?— a potential biomarker for prognosis in lung cancer. *Oncology Letters*, 12(4), 3004–3012.
<https://doi.org/10.3892/ol.2016.4927>
- Efstathiou, J. A., Liu, D., Wheeler, J. M., Kim, H. C., Beck, N. E., Ilyas, M., ... Bodmer, W. F. (1999). Mutated epithelial cadherin is associated with increased tumorigenicity and loss of adhesion and of responsiveness to the motogenic trefoil factor 2 in colon carcinoma cells. *Proceedings of the National Academy of Sciences of the United States of America*, 96(5), 2316–21.
<https://doi.org/10.1073/PNAS.96.5.2316>

- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., ... Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer*, 136(5), E359-86. <https://doi.org/10.1002/ijc.29210>
- Fishel, R., Kay Lescoe, M., Ft Rao, M. S., Copeland, N. G., Jenkins, N. A., Garber, J., ... Kolodner, R. (1993). The Human Mutator Gene Homolog MSHP and Its Association with Hereditary Nonpolyposis Colon Cancer. *Cell*, 75, 1027–1038.
- Fujii, H., Marsh, C., Cairns, P., Sidransky, D., & Gabrielson, E. (1996). Genetic divergence in the clonal evolution of breast cancer. *Cancer Research*, 56(7), 1493–7. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8603391>
- Fujii, H., Yoshida, M., Gong, Z. X., Matsumoto, T., Hamano, Y., Fukunaga, M., ... Shirai, T. (2000). Frequent genetic heterogeneity in the clonal evolution of gynecological carcinosarcoma and its influence on phenotypic diversity. *Cancer Research*, 60(1), 114–20. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10646862>
- Garza-Treviño, E. N., Said-Fernández, S. L., Martínez-Rodríguez, H. G., Thomassen, I., Gestel, Y., Lemmens, V., ... Choudhuri, T. (2015). Understanding the colon cancer stem cells and perspectives on treatment. *Cancer Cell International*, 15(1), 2. <https://doi.org/10.1186/s12935-015-0163-7>
- González-García, I., Solé, R. V., & Costa, J. (2002). Metapopulation dynamics and spatial heterogeneity in cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20), 13085–9. <https://doi.org/10.1073/pnas.202139299>
- Greaves, M., & Maley, C. C. (2012). Clonal evolution in cancer. *Nature*, 481(7381), 306–313. <https://doi.org/10.1038/nature10762>
- Gryfe, R., Kim, H., Hsieh, E. T. K., Aronson, M. D., Holowaty, E. J., Bull, S. B., ... Gallinger, S. (2000). Tumor Microsatellite Instability and Clinical Outcome in Young Patients with Colorectal Cancer. *New England Journal of Medicine*, 342(2), 69–77. <https://doi.org/10.1056/NEJM200001133420201>
- Guarino, M., Rubino, B., & Ballabio, G. (2007). The role of epithelial-mesenchymal transition in cancer pathology. *Pathology*, 39(3), 305–18. <https://doi.org/10.1080/00313020701329914>
- Haerter, J. O., Lövkvist, C., Dodd, I. B., & Sneppen, K. (2014). Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states.

- Nucleic Acids Research*, 42(4), 2235–44. <https://doi.org/10.1093/nar/gkt1235>
- Hanahan, D., & Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell*, 100(1), 57–70. [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9)
- Hanahan, D., Weinberg, R. A., Adams, J. M., Cory, S., Aguirre-Ghiso, J. A., Ahmed, Z., ... Christofori, G. (2011). Hallmarks of cancer: the next generation. *Cell*, 144(5), 646–74. <https://doi.org/10.1016/j.cell.2011.02.013>
- Hastings, R. J., & Franks, L. M. (1983). Cellular heterogeneity in a tissue culture cell line derived from a human bladder carcinoma. *British Journal of Cancer*, 47(2), 233–44. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/6572066>
- Hay, E. D. (1995). An overview of epithelio-mesenchymal transformation. *Acta Anatomica*, 154(1), 8–20. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8714286>
- Hu, T., Yao, Y., Yu, S., Han, L., Wang, W., Guo, H., ... Nan, K. (2014). SDF-1/CXCR4 promotes epithelial-mesenchymal transition and progression of colorectal cancer by activation of the Wnt/ β -catenin signaling pathway. *Cancer Letters*, 354(2), 417–426. <https://doi.org/10.1016/j.canlet.2014.08.012>
- Hubbe, M., & Altevogt, P. (1994). Heat-stable antigen/CD24 on mouse T lymphocytes: evidence for a costimulatory function. *European Journal of Immunology*, 24(3), 731–737. <https://doi.org/10.1002/eji.1830240336>
- Humphries, A., Cereser, B., Gay, L. J., Miller, D. S. J., Das, B., Gutteridge, A., ... Graham, T. A. (2013). Lineage tracing reveals multipotent stem cells maintain human adenomas and the pattern of clonal expansion in tumor evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 110(27), E2490–9. <https://doi.org/10.1073/pnas.1220353110>
- Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D., & Perucho, M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature*, 363(6429), 558–561. <https://doi.org/10.1038/363558a0>
- Jaggupilli, A., Elkord, E., Jaggupilli, A., & Elkord, E. (2012). Significance of CD44 and CD24 as Cancer Stem Cell Markers: An Enduring Ambiguity. *Clinical and Developmental Immunology*, 2012, 1–11. <https://doi.org/10.1155/2012/708036>
- Jiao, W., Miyazaki, K., & Kitajima, Y. (2001). Exogenous expression of E-cadherin in gallbladder carcinoma cell line G-415 restores its cellular polarity and differentiation. *International Journal of Oncology*.

- <https://doi.org/10.3892/ijo.19.6.1099>
- Johannsdottir, J. T., Jonasson, J. G., Bergthorsson, J. T., Amundadottir, L. T., Magnusson, J., Egilsson, V., & Ingvarsson, S. (2000). The effect of mismatch repair deficiency on tumorigenesis; microsatellite instability affecting genes containing short repeated sequences. *International Journal of Oncology*, 16(1), 133–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10601558>
- Kai Yoshimi Arima AE Toshio Kamiya AE Hideyuki Saya, K. A. (n.d.). Breast cancer stem cells. <https://doi.org/10.1007/s12282-009-0176-y>
- Kalluri, R., & Neilson, E. G. (2003). Epithelial-mesenchymal transition and its implications for fibrosis. *The Journal of Clinical Investigation*, 112(12), 1776–84. <https://doi.org/10.1172/JCI20530>
- Karatzas, G., Karayiannakis, A. J., Syrigos, K. N., Chatzigianni, E., Papanikolaou, S., Riza, F., & Papanikolaou, D. (1999). E-cadherin expression correlates with tumor differentiation in colorectal cancer. *Hepato-Gastroenterology*, 46(25), 232–5. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10228798>
- Kay, R., Takei, F., & Humphries, R. K. (1990). Expression cloning of a cDNA encoding M1/69-J11d heat-stable antigens. *Journal of Immunology (Baltimore, Md. : 1950)*, 145(6), 1952–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2118158>
- Khoursheed, M. A., Mathew, T. C., Makar, R. R., Louis, S., Asfar, S. K., Al-Sayer, H. M., ... Al-Bader, A. (2003). Expression of E-cadherin in human colorectal cancer. *The Surgeon : Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland*, 1(2), 86–91. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15573626>
- Kim, H., Jen, J., Vogelstein, B., & Hamilton, S. R. (1994). Clinical and pathological characteristics of sporadic colorectal carcinomas with DNA replication errors in microsatellite sequences. *The American Journal of Pathology*, 145(1), 148–56. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8030745>
- Kleene, R., Yang, H., Kutsche, M., & Schachner, M. (2001). The Neural Recognition Molecule L1 Is a Sialic Acid-binding Lectin for CD24, Which Induces Promotion and Inhibition of Neurite Outgrowth. *Journal of Biological Chemistry*, 276(24), 21656–21663. <https://doi.org/10.1074/jbc.M101790200>
- Kleinman, H. K., & Martin, G. R. (2005). Matrigel: basement membrane matrix with biological activity. *Seminars in Cancer Biology*, 15(5), 378–86. <https://doi.org/10.1016/j.semcancer.2005.05.004>
- Kleinman, H. K., McGarvey, M. L., Liotta, L. A., Robey, P. G., Tryggvason, K., &

- Martin, G. R. (1982). Isolation and characterization of type IV procollagen, laminin, and heparan sulfate proteoglycan from the EHS sarcoma. *Biochemistry*, 21(24), 6188–93. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/6217835>
- Kneip, C., Schmidt, B., Seegebarth, A., Weickmann, S., Fleischhacker, M., Liebenberg, V., ... Dietrich, D. (2011). SHOX2 DNA Methylation Is a Biomarker for the Diagnosis of Lung Cancer in Plasma. *Journal of Thoracic Oncology*, 6(10), 1632–1638. <https://doi.org/10.1097/JTO.0b013e318220ef9a>
- Knudson, A. G., & Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America*, 68(4), 820–3. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/5279523>
- Konishi, N., Hiasa, Y., Matsuda, H., Tao, M., Tsuzuki, T., Hayashi, I., ... Lin, J.-C. (1995). Intratumor cellular heterogeneity and alterations in ras oncogene and p53 tumor suppressor gene in human prostate carcinoma. *The American Journal of Pathology*, 147(4), 1112–22. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7573356>
- Kristiansen, G., Winzer, K.-J., Mayordomo, E., Bellach, J., Schlüns, K., Denkert, C., ... Dietel, M. (2003). CD24 expression is a new prognostic marker in breast cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 9(13), 4906–13. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14581365>
- Krueger, F., Kreck, B., Franke, A., & Andrews, S. R. (2012). DNA methylome analysis using short bisulfite sequencing data. *Nature Methods*, 9(2), 145–151. <https://doi.org/10.1038/nmeth.1828>
- Lavelle, D., Sauntharajah, Y., & DeSimone, J. (2008). DNA methylation and mechanism of action of 5-azacytidine. *Blood*, 111(4).
- Lee, H. J., Kim, D.-I., Kwak, C., Ku, J. H., Moon, K. C., Shuch, B. M., ... al., et. (2008). Expression of CD24 in Clear Cell Renal Cell Carcinoma and Its Prognostic Significance. *Urology*, 72(3), 603–607. <https://doi.org/10.1016/j.urology.2008.01.061>
- Leedham, S. J., Rodenas-Cuadrado, P., Howarth, K., Lewis, A., Mallappa, S., Segditsas, S., ... Tomlinson, I. P. M. (2013). A basal gradient of Wnt and stem-cell number influences regional tumour distribution in human and mouse intestinal tracts. *Gut*, 62(1), 83–93. <https://doi.org/10.1136/gutjnl-2011-301601>
- Li, C., Heidt, D. G., Dalerba, P., Burant, C. F., Zhang, L., Adsay, V., ... Simeone, D. M.

- (2007). Identification of Pancreatic Cancer Stem Cells. *Cancer Research*, 67(3).
- Li, L.-C., & Dahiya, R. (2002). MethPrimer: designing primers for methylation PCRs. *Bioinformatics (Oxford, England)*, 18(11), 1427–31. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12424112>
- Li, O., Zheng, P., & Liu, Y. (2004). CD24 expression on T cells is required for optimal T cell proliferation in lymphopenic host. *The Journal of Experimental Medicine*, 200(8), 1083–9. <https://doi.org/10.1084/jem.20040779>
- Li, Y., Iglehart, J. D., Richardson, A. L., & Wang, Z. C. (2012). The amplified cancer gene LPTM4B promotes tumor growth and tolerance to stress through the induction of autophagy. *Autophagy*, 8(2), 273–274. <https://doi.org/10.4161/auto.8.2.18941>
- Lo, J.-F., Yu, C.-C., Chiou, S.-H., Huang, C.-Y., Jan, C.-I., Lin, S.-C., ... Yu, Y.-H. (2011). The Epithelial-Mesenchymal Transition Mediator S100A4 Maintains Cancer-Initiating Cells in Head and Neck Cancers. *Cancer Research*, 71(5).
- Loeb, L. A. (1991). Mutator phenotype may be required for multistage carcinogenesis. *Cancer Research*, 51(12), 3075–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2039987>
- Maddams, J., Utley, M., & Møller, H. (2012). Projections of cancer prevalence in the United Kingdom, 2010-2040. *British Journal of Cancer*, 107(7), 1195–202. <https://doi.org/10.1038/bjc.2012.366>
- Makino, S. (1956). Further evidence favoring the concept of the stem cell in ascites tumors of rats. *Annals of the New York Academy of Sciences*, 63(5), 818–830. <https://doi.org/10.1111/j.1749-6632.1956.tb50894.x>
- Maley, C. C., Galipeau, P. C., Finley, J. C., Wongsurawat, V. J., Li, X., Sanchez, C. A., ... Reid, B. J. (2006). Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nature Genetics*, 38(4), 468–473. <https://doi.org/10.1038/ng1768>
- Marjanovic, N. D., Weinberg, R. A., & Chaffer, C. L. (2013). Cell Plasticity and Heterogeneity in Cancer. *Clinical Chemistry*, 59(1).
- Marusyk, A., Tabassum, D. P., Altrock, P. M., Almendro, V., Michor, F., & Polyak, K. (2014). Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature*, 514(7520), 54–58. <https://doi.org/10.1038/nature13556>
- Meacham, C. E., & Morrison, S. J. (2013). Tumour heterogeneity and cancer cell plasticity. *Nature*, 501(7467), 328–37. <https://doi.org/10.1038/nature12624>
- Medema, J. P. (2013). Cancer stem cells: The challenges ahead. *Nature Cell Biology*,

- 15(4), 338–344. <https://doi.org/10.1038/ncb2717>
- Meng, F., Chen, X., Song, H., & Lou, G. (2015). LPTM4B Down Regulation Inhibits the Proliferation, Invasion and Angiogenesis of HeLa Cells & In Vitro. *Cellular Physiology and Biochemistry*, 37(3), 890–900. <https://doi.org/10.1159/000430216>
- Meyer, M. J., Fleming, J. M., Ali, M. A., Pesesky, M. W., Ginsburg, E., & Vonderhaar, B. K. (2009). Dynamic regulation of CD24 and the invasive, CD44posCD24neg phenotype in breast cancer cell lines. *Breast Cancer Research : BCR*, 11(6), R82. <https://doi.org/10.1186/bcr2449>
- Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K. W., ... Nakamura, Y. (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Research*, 52(3), 643–5. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1310068>
- Mohamet, L., Hawkins, K., & Ward, C. M. (2011). Loss of Function of E-Cadherin in Embryonic Stem Cells and the Relevance to Models of Tumorigenesis. *Journal of Oncology*, 2011, 1–19. <https://doi.org/10.1155/2011/352616>
- Mora, J., Cheung, N. K., & Gerald, W. L. (2001). Genetic heterogeneity and clonal evolution in neuroblastoma. *British Journal of Cancer*, 85(2), 182–9. <https://doi.org/10.1054/bjoc.2001.1849>
- Mouradov, D., Sloggett, C., Jorissen, R. N., Love, C. G., Li, S., Burgess, A. W., ... Sieber, O. M. (2014). Colorectal Cancer Cell Lines Are Representative Models of the Main Molecular Subtypes of Primary Cancer. *Cancer Research*, 74(12).
- Mueller-Klieser, W. (n.d.). Three-dimensional cell cultures: from molecular mechanisms to clinical applications.
- Nile, C. J., Read, R. C., Akil, M., Duff, G. W., & Wilson, A. G. (2008). Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis. *Arthritis and Rheumatism*, 58(9), 2686–93. <https://doi.org/10.1002/art.23758>
- Nishisho, I., Nakamura, Y., Miyoshi, Y., Miki, Y., Ando, H., Horii, A., ... Hedge, P. (1991). Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. *Science*, 253(5020).
- Nowell, P. (1976). The clonal evolution of tumor cell populations. *Science*, 194(4260).
- Nowell, P. C. (1986). Mechanisms of Tumor Progression. *Cancer Research*, 46, 2203–2207.

- Orkin, R. W., Gehron, P., McGoodwin, E. B., Martin, G. R., Valentine, T., & Swarm, R. (1977). A murine tumor producing a matrix of basement membrane. *Journal of Experimental Medicine*, 145(1).
- Pantou, D., Rizou, H., Tsarouha, H., Pouli, A., Papanastasiou, K., Stamatellou, M., ... Bardi, G. (2005). Cytogenetic manifestations of multiple myeloma heterogeneity. *Genes, Chromosomes and Cancer*, 42(1), 44–57. <https://doi.org/10.1002/gcc.20114>
- Pils, D., Horak, P., Vanhara, P., Anees, M., Petz, M., Alfanz, A., ... Krainer, M. (2013). Methylation status of *TUSC3* is a prognostic factor in ovarian cancer. *Cancer*, 119(5), 946–954. <https://doi.org/10.1002/cncr.27850>
- Qiao, Y., Jiang, X., Lee, S. T., Karuturi, R. K. M., Hooi, S. C., & Yu, Q. (2011). FOXQ1 Regulates Epithelial-Mesenchymal Transition in Human Cancers. *Cancer Research*, 71(8), 3076–3086. <https://doi.org/10.1158/0008-5472.CAN-10-2787>
- Ramchandani, S., Bhattacharya, S. K., Cervoni, N., & Szyf, M. (1999). DNA methylation is a reversible biological signal. *Proceedings of the National Academy of Sciences of the United States of America*, 96(11), 6107–12. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10339549>
- Richman, P. I., & Bodmer, W. F. (1988). Control of differentiation in human colorectal carcinoma cell lines: Epithelial—mesenchymal interactions. *The Journal of Pathology*, 156(3), 197–211. <https://doi.org/10.1002/path.1711560305>
- Ryland, G. L., Doyle, M. A., Goode, D., Boyle, S. E., Choong, D. Y. H., Rowley, S. M., ... Gorringer, K. L. (2015). Loss of heterozygosity: what is it good for? *BMC Medical Genomics*, 8(1), 45. <https://doi.org/10.1186/s12920-015-0123-z>
- Sammar, M., Aigner, S., & Altevogt, P. (1997). Heat-stable antigen (mouse CD24) in the brain: dual but distinct interaction with P-selectin and L1. *Biochimica et Biophysica Acta*, 1337(2), 287–94. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9048906>
- Sauter, G., Moch, H., Gasser, T. C., Mihatsch, M. J., & Waldman, F. M. (1995). Heterogeneity of chromosome 17 and erbB-2 gene copy number in primary and metastatic bladder cancer. *Cytometry*, 21(1), 40–46. <https://doi.org/10.1002/cyto.990210109>
- Schabath, H. (2006). CD24 affects CXCR4 function in pre-B lymphocytes and breast carcinoma cells. *Journal of Cell Science*, 119(2), 314–325. <https://doi.org/10.1242/jcs.02741>
- Schneider, M., Hansen, J. L., & Sheikh, S. P. (2008). S100A4: a common mediator of

- epithelial–mesenchymal transition, fibrosis and regeneration in diseases? *Journal of Molecular Medicine*, 86(5), 507–522. <https://doi.org/10.1007/s00109-007-0301-3>
- Schneikert, J., & Behrens, J. (2007). The canonical Wnt signalling pathway and its APC partner in colon cancer development. *Gut*, 56(3), 417–25. <https://doi.org/10.1136/gut.2006.093310>
- Shen, Z. (2011). Genomic instability and cancer: an introduction. *Journal of Molecular Cell Biology*, 3(1), 1–3. <https://doi.org/10.1093/jmcb/mjq057>
- Shipitsin, M., Campbell, L. L., Argani, P., Weremowicz, S., Blushtain-Qimron, N., Yao, J., ... Polyak, K. (2007). Molecular Definition of Breast Tumor Heterogeneity. *Cancer Cell*, 11(3), 259–273. <https://doi.org/10.1016/j.ccr.2007.01.013>
- Shoemaker, R. H. (2006). The NCI60 human tumour cell line anticancer drug screen. *Nature Reviews. Cancer*, 6(10), 813–23. <https://doi.org/10.1038/nrc1951>
- Smith, B. D., Martin, G. R., Miller, E. J., Dorfman, A., & Swarm, R. (1975). Nature of the collagen synthesized by a transplanted chondrosarcoma. *Archives of Biochemistry and Biophysics*, 166(1), 181–186. [https://doi.org/10.1016/0003-9861\(75\)90378-1](https://doi.org/10.1016/0003-9861(75)90378-1)
- Smith, N. C., Fairbridge, N. A., Pallegar, N. K., & Christian, S. L. (2015). Dynamic upregulation of CD24 in pre-adipocytes promotes adipogenesis. *Adipocyte*, 4(2), 89–100. <https://doi.org/10.4161/21623945.2014.985015>
- Springer, T., Galfre, G., Secher, D. S., & Milstein, C. (1978). Monoclonal xenogenic antibodies to murine cell surface antigens: identification of novel leukocyte differentiation antigens. *European Journal of Immunology*, 8(8), 539–551. <https://doi.org/10.1002/eji.1830080802>
- Sun, X., & Yu, Q. (2015). Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment. *Acta Pharmacologica Sinica*, 36(10), 1219–27. <https://doi.org/10.1038/aps.2015.92>
- Tan, Y., Zhao, M., Xiang, B., Chang, C., & Lu, Q. (2016). CD24: from a Hematopoietic Differentiation Antigen to a Genetic Risk Factor for Multiple Autoimmune Diseases. *Clinical Reviews in Allergy & Immunology*, 50(1), 70–83. <https://doi.org/10.1007/s12016-015-8470-2>
- Thibodeau, S. N., Bren, G., & Schaid, D. (1993). Microsatellite instability in cancer of the proximal colon. *Science (New York, N.Y.)*, 260(5109), 816–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8484122>
- Tom, B. H., Rutzky, L. P., Jakstys, M. M., Oyasu, R., Kaye, C. I., & Kahan, B. D.

- (1976). Human colonic adenocarcinoma cells. I. Establishment and description of a new line. *In Vitro*, 12(3), 180–91. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1262041>
- Tomlinson, I. P., & Bodmer, W. F. (1997). Modelling the consequences of interactions between tumour cells. *British Journal of Cancer*, 75(2), 157–60. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9010019>
- Tomlinson, I. P., Novelli, M. R., & Bodmer, W. F. (1996). The mutation rate and cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 93(25), 14800–3. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8962135>
- Turker, M. S. (2002). Gene silencing in mammalian cells and the spread of DNA methylation. *Oncogene*, 21(35), 5388–5393. <https://doi.org/10.1038/sj.onc.1205599>
- Uriel, S., Labay, E., Francis-Sedlak, M., Moya, M. L., Weichselbaum, R. R., Ervin, N., ... Brey, E. M. (2009). Extraction and assembly of tissue-derived gels for cell culture and tissue engineering. *Tissue Engineering. Part C, Methods*, 15(3), 309–21. <https://doi.org/10.1089/ten.tec.2008.0309>
- van Roy, F., & Berx, G. (2008). The cell-cell adhesion molecule E-cadherin. *Cellular and Molecular Life Sciences*, 65(23), 3756–3788. <https://doi.org/10.1007/s00018-008-8281-1>
- Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., ... Bos, J. L. (1988). Genetic Alterations during Colorectal-Tumor Development. *New England Journal of Medicine*, 319(9), 525–532. <https://doi.org/10.1056/NEJM198809013190901>
- Walter, E., & Kissel, T. (1995). Heterogeneity in the human intestinal cell line Caco-2 leads to differences in transepithelial transport. *European Journal of Pharmaceutical Sciences*, 3(4), 215–230. [https://doi.org/10.1016/0928-0987\(95\)00010-B](https://doi.org/10.1016/0928-0987(95)00010-B)
- Weigelt, B., & Reis-Filho, J. S. (2009). Histological and molecular types of breast cancer: is there a unifying taxonomy? *Nature Reviews Clinical Oncology*, 6(12), 718–730. <https://doi.org/10.1038/nrclinonc.2009.166>
- Wendt, M. K., Taylor, M. A., Schiemann, B. J., & Schiemann, W. P. (2011). Down-regulation of epithelial cadherin is required to initiate metastatic outgrowth of breast cancer. *Molecular Biology of the Cell*, 22(14), 2423–2435. <https://doi.org/10.1091/mbc.E11-04-0306>
- Wheeler, J. M. D., Beck, N. E., Kim, H. C., Tomlinson, I. P. M., Mortensen, N. J. M., &

- Bodmer, W. F. (1999). Mechanisms of inactivation of mismatch repair genes in human colorectal cancer cell lines: The predominant role of hMLH1. *Genetics*, 96, 10296–10301.
- Wilding, J. L., & Bodmer, W. F. (2014). Cancer Cell Lines for Drug Discovery and Development. *Cancer Research*, 74(9).
- Wilding, J. L., McGowan, S., Liu, Y., & Bodmer, W. F. (2010). Replication error deficient and proficient colorectal cancer gene expression differences caused by 3'UTR polyT sequence deletions. *Proceedings of the National Academy of Sciences*, 107(49), 21058–21063. <https://doi.org/10.1073/pnas.1015604107>
- Williams, L. A., Hock, B. D., & Hart, D. N. (1996). Human T lymphocytes and hematopoietic cell lines express CD24-associated carbohydrate epitopes in the absence of CD24 mRNA or protein. *Blood*, 88(8), 3048–55. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8874203>
- Williams, M. T., & Hord, N. G. (2005). The role of dietary factors in cancer prevention: beyond fruits and vegetables. *Nutrition in Clinical Practice: Official Publication of the American Society for Parenteral and Enteral Nutrition*, 20(4), 451–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/16207684>
- Wishart, D. S. (2015). Is Cancer a Genetic Disease or a Metabolic Disease? *EBioMedicine*. <https://doi.org/10.1016/j.ebiom.2015.05.022>
- Woodford-Richens, K. L., Rowan, A. J., Gorman, P., Halford, S., Bicknell, D. C., Wasan, H. S., ... Tomlinson, I. P. (2001). SMAD4 mutations in colorectal cancer probably occur before chromosomal instability, but after divergence of the microsatellite instability pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 98(17), 9719–23. <https://doi.org/10.1073/pnas.171321498>
- World Cancer Report. (2014). Retrieved from <http://publications.iarc.fr/Non-Series-Publications/World-Cancer-Reports/World-Cancer-Report-2014>
- Yamada, K. M., Cukierman, E., Affolter, M., Bellusci, S., Itoh, N., Shilo, B., ... Matsudaira, P. (2007). Modeling tissue morphogenesis and cancer in 3D. *Cell*, 130(4), 601–10. <https://doi.org/10.1016/j.cell.2007.08.006>
- Yao, D., Dai, C., & Peng, S. (2011). Mechanism of the Mesenchymal-Epithelial Transition and Its Relationship with Metastatic Tumor Formation. *Molecular Cancer Research*, 9(12), 1608–1620. <https://doi.org/10.1158/1541-7786.MCR-10-0568>

- Yeung, T. M., Gandhi, S. C., & Bodmer, W. F. (2011). Hypoxia and lineage specification of cell line-derived colorectal cancer stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 108(11), 4382–7. <https://doi.org/10.1073/pnas.1014519107>
- Yeung, T. M., Gandhi, S. C., Wilding, J. L., Muschel, R., & Bodmer, W. F. (2010). Cancer stem cells from colorectal cancer-derived cell lines. *Proceedings of the National Academy of Sciences of the United States of America*, 107(8), 3722–7. <https://doi.org/10.1073/pnas.0915135107>
- Zellmer, V. R., Zhang, S., Kleppe, M., Levine, R., Hanahan, D., Weinberg, R., ... Swanton, C. (2014). Evolving concepts of tumor heterogeneity. *Cell & Bioscience*, 4(1), 69. <https://doi.org/10.1186/2045-3701-4-69>
- Zeuner, A., Todaro, M., Stassi, G., & De Maria, R. (2014). Colorectal Cancer Stem Cells: From the Crypt to the Clinic. *Cell Stem Cell*, 15(6), 692–705. <https://doi.org/10.1016/j.stem.2014.11.012>

The end