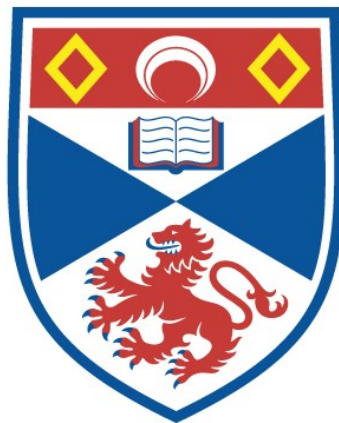


**An investigation into the relevance of
dual-specificity phosphatase 6
in colorectal cancer**

Hannah Louise Williams

A thesis submitted for the degree of PhD
at the
University of St Andrews



2019

Full metadata for this item is available in
St Andrews Research Repository
at:

<https://research-repository.st-andrews.ac.uk/>

Identifier to use to cite or link to this thesis:

DOI: <https://doi.org/10.17630/10023-18745>

This item is protected by original copyright

Candidate's declaration

I, Hannah Louise Williams, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 35,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in October 2013.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date 8th April 2019

Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date 8th April 2019

Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Hannah Louise Williams, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

I, Hannah Louise Williams, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

Embargo on all of print copy for a period of 3 years on the following ground(s):

- Publication would preclude future publication

Supporting statement for printed embargo request

Planning to submit two manuscripts for publication that use the data included in this thesis.

Electronic copy

Embargo on all of electronic copy for a period of 3 years on the following ground(s):

- Publication would preclude future publication

Supporting statement for electronic embargo request

Planning to submit two manuscripts for publication that use the data included in this thesis.

Title and Abstract

- I require an embargo on the abstract only.

Date 8th April 2019

Signature of candidate

Date 8th April 2019

Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Hannah Louise Williams, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date

8th April 2019

Signature of candidate

This thesis is dedicated to an inspirational woman.

Never forgotten.

Acknowledgments

I would firstly like to thank my supervisor; David Harrison for supporting me along this journey without whom I would not have the self-belief and confidence to go forward as a researcher today.

For their contribution to immunohistochemistry and digital image analysis of cohort 1. I would like to thank Nourjahan Khafaga and Sharma Ganesanathan; the doctors of tomorrow.

My sincere thanks goes to the members of Lab 248 and 249 for their advice, friendship and generosity throughout my PhD.

A massive thanks goes to my family and friends who have supported me during the ups and down. Most importantly to my parents, Rob and Dee for the provision of a second home, an ear to listen and an endless supply of coffee whilst completing my PhD.

Finally, thank you to my husband who proves to continue to support my dreams and aspirations.

This work was supported in part by NHS Lothian.

This work was supported in part by a grant from the Melville Trust for the Cure and Care of Cancer (M00109.0001/TZH/JGC).

Abstract

A negative regulator of ERK; dual-specificity phosphatase 6 (DUSP6), has been implicated in the pathogenesis of a variety of cancers including lung, pancreatic and oesophageal. Additionally DUSP6 expression has been associated with treatment response in breast, lung and ovarian cancer. Despite this range of investigations little is known about its involvement in colorectal cancer.

The main aims of this study were firstly to explore the landscape of DUSP6 protein expression across formalin fixed paraffin embedded (FFPE) samples representing the main stages of progression from normal mucosa to adenocarcinoma. Secondly, to assess the relationship between DUSP6 protein expression and *RAS* genotype in colorectal adenocarcinoma and finally to explore the impact of DUSP6 manipulation *in vitro* on response to the anti-EGFR monoclonal antibody Cetuximab.

Immunohistochemical analysis of DUSP6 in FFPE samples demonstrated a significant increased expression in adenoma, specifically those with high grade dysplasia in comparison to normal mucosa and adenocarcinoma. No significant association between *RAS* genotype and DUSP6 protein expression was observed based upon immunofluorescence assessment of an FFPE colorectal adenocarcinoma cohort. Expression of DUSP6 in both RG/C2/80 adenoma and C99 adenocarcinoma cell lines provides evidence for a functional relationship between DUSP6 and regulation of activated ERK. In preliminary studies using C99 cells, over-expression of DUSP6 increased sensitivity to Cetuximab treatment.

In conclusion, this series of investigations has highlighted a potential tumour suppressive role for DUSP6 in colorectal adenoma. It is hypothesised that up-regulation of DUSP6 in response to increased MAPK pathway activation attempts to reduce overall proliferation and survival of dysregulated cells which have the potential for malignant transformation. Further work is warranted to confirm the role of specific DUSP6 isoforms in colorectal pathogenesis.

List of figures

Figure 1. Colonic mucosa structure. A. Haematoxylin stained section of normal colonic mucosa.	3
Figure 2. Haematoxylin stained sections of adenomatous lesions.	4
Figure 3. The Hallmarks of Cancer.	5
Figure 4. Schematic of basic molecular aberrations in colorectal pathogenesis.	8
Figure 5. MAPK/ERK signalling pathway.	13
Figure 6. MAPK/ERK signalling pathway and the hallmarks of cancer.	14
Figure 7. Schematic of ERK regulation.	15
Figure 8. Dual specificity phosphatase 6 structure.	16
Figure 9. DUSP6 amino acid sequence.	17
Figure 11. Schematic of DUSP6 transcriptional and post-translational control.	19
Figure 12. Summary of therapies targeting MAPK/ERK pathway.	23
Figure 13. Epidermal growth factor receptor conformation upon Cetuximab binding.	24
Figure 14. Hoechst staining overview at 5x magnification.	41
Figure 15. AQUA experiment steps for DUSP6 semi-quantitation.	42
Figure 16. DUSP6 immunohistochemistry.	43
Figure 17. Cell classifier with positivity threshold applied.	44
Figure 18. DUSP6 plasmid map.	51
Figure 19. eGFP plasmid map.	52
Figure 20. Positive control for DUSP6 immunohistochemistry.	58
Figure 21. DUSP6 antibody optimisation and validation by western blot.	59
Figure 22. DUSP6 immunofluorescence optimisation: antibody dilution.	60
Figure 23. ERK immunohistochemistry positive control.	61
Figure 24. ERK antibody optimisation and validation by western blot.	62
Figure 25. p-ERK antibody optimisation and validation by western blot.	63
Figure 26. Output from X-Tile software for the detection of high and low thresholds for DUSP6 protein expression.	66
Figure 27. KRAS genotyping of adenoma cases.	67
Figure 28. Frequency of RAS mutations.	69
Figure 29. Puromycin optimisation in C99 and RG/C2/80 cells.	71

Figure 30. DNA sequencing of DUSP6 plasmid.....	72
Figure 31. Sequencing alignment results for DUSP6 plasmid DNA sequencing.....	73
Figure 32. Fluorescence signal in eGFP transduced cell lines. A: C99 cells. B: RG/C2/80 cells. ..	74
Figure 33. DUSP6 positive immunohistochemistry.....	83
Figure 34. Example of QuPath annotation for DUSP6 immunohistochemistry.	84
Figure 35. DUSP6 positivity across subgroups; normal, low grade dysplasia.	86
Figure 36. DUSP6 positivity.	87
Figure 37. DUSP6 positivity across adenoma histological class.	88
Figure 38. DUSP6 positivity across histological class of adenoma.	89
Figure 39. DUSP6 positivity across molecularly distinct adenocarcinoma groups.	90
Figure 40. DUSP6 positivity in colorectal adenocarcinoma.	91
Figure 41. p-ERK and ERK positive immunohistochemistry.	94
Figure 42. ERK and p-ERK expression across subgroups; normal, low grade dysplasia, high grade dysplasia and adenocarcinoma.	96
Figure 43. ERK and p-ERK immunohistochemistry.....	97
Figure 44. ERK and p-ERK expression across subgroups; normal, tubular, tubulovillous, villous, serrated, dMMR, KRAS mutant, TP53 mutant, dual mutant.	99
Figure 45. Ratio of p-ERK positivity and nuclear DUSP6 positivity in low and high grade dysplasia.	102
Figure 46. p-ERK and DUSP6 nuclear immunohistochemistry.	103
Figure 47. Ratio of p-ERK positive cells and DUSP6 positive cells in adenocarcinoma.	104
Figure 48. RG/C2/80 adenoma cells.....	105
Figure 49. Western blot of DUSP6 expression in RG/C2/80 cells.	106
Figure 50. Western blot of ERK and p-ERK expression in RG/C2/80 cell lines.....	107
Figure 51. Proposed mechanism of TP53 mediated upregulation of DUSP6.....	109
Figure 52. Range of nuclear and cytoplasmic DUSP6 protein expression values.	117
Figure 53. DUSP6 positivity in colorectal adenocarcinoma by immunofluorescence.	118
Figure 54. Cytoplasmic DUSP6 expression by T stage.....	119
Figure 55. DUSP6 AQUA expression by sub-cellular compartment.	121
Figure 56. Screening of colorectal adenocarcinoma cell lines for DUSP6 protein expression..	124
Figure 57. C99 adenocarcinoma cells.....	127
Figure 58. DUSP6 overexpression in C99 cells as determined by western blot.....	128

Figure 59. ERK and p-ERK protein expression in protein lysates of parental, eGFP transduced and DUSP6 transduced C99 cells.....	129
Figure 60. Dose-response for C99 parental lines.	130
Figure 61. Dose-response for Cetuximab treated C99 cells.....	131
Figure 62. Kaplan-Meier plot for cytoplasmic DUSP6 expression and overall survival.....	135
Figure 63. Kaplan-Meier plot for categorical cytoplasmic DUSP6 expression.	137
Figure 64. Kaplan-Meier for cytoplasmic DUSP6 expression in TNM stage 3 disease and overall survival.	139
Figure 65. Kaplan-Meier for cytoplasmic DUSP6 expression in TNM stage 3 disease and disease specific survival.	141
Figure 66. Schematic detailing activity of upregulated DUSP6 in colorectal adenoma.	147
Figure 67. Schematic demonstrating the ERK-DUSP6 axis for regulation of cell proliferation and survival.	149
Figure 68. Proposed model of ERK – DUSP6 axis for regulation of MAPK signalling.	150
Figure 69. Schematic detailing complementary action of DUSP6 with Cetuximab treatment.	152
Figure 69. Western blot repeats of ERK protein expression in RG/C2/80 cell lines.	174
Figure 70. Western blot repeats of p-ERK protein expression in RG/C2/80 cell lines.	175
Figure 71. Western blot repeats of ERK protein expression in C99 cell lines.	201
Figure 72. Western blot repeats of p-ERK protein expression in C99 cell lines.....	202

List of tables

Table 1. Summary of prognostic outcomes in serrated lesions.....	10
Table 2. Volumes for PCR set up.	33
Table 3. Thermal cycler program for target amplification.....	33
Table 4. Primer sequences.	34
Table 5. Pyrosequencing dispensation orders and sequence to analyse.....	35
Table 6. Limits of blank and limits of detection for pyrosequencing.	36
Table 7. Details of primary antibodies used. *Automated staining dilution used.....	38
Table 8. Details of secondary antibodies used.....	39
Table 12. Cell line details.....	46
Table 13. Summary of primary and secondary antibodies used for Western blotting.....	48
Table 14. Reagents for resolving and stacking gel preparation.	49
Table 15. Reagents for running and transfer buffers.....	49
Table 13. DAB intensity values for DUSP6 phenotypes.	64
Table 14. DAB intensity values for ERK and p-ERK positive phenotypes.	64
Table 15. Details of negative normalisation.....	65
Table 16. Details of fold change normalisation.....	66
Table 17. Cohort demographics.	82
Table 18. Summary of median ratio of DUSP6 positive cells for subgroups.....	85
Table 20. Summary of median ratio of ERK and p-ERK positive cells for subgroups.	95
Table 21. Mann Whitney test for difference in the ratio of ERK and p-ERK positive cells between subgroups.....	98
Table 22. Mann Whitney test for difference in the ratio of ERK and p-ERK positive cells between subgroups; normal, tubular, tubulovillous, villous, serrated, dMMR, KRAS mutant, TP53 mutant.	100
Table 23. Mann-Whitney and Kruskal-Wallis test results assessing relationship between <i>RAS</i> mutation status and DUSP6 expression.....	120
Table 24. Molecular and Cetuximab response data.....	126
Table 25. Summary of statistical analysis of change in growth for C99 (DUSP6 transduced) cells and other conditions.	132

Table 26. Cox-regression analysis with categorical cytoplasmic DUSP6 expression and overall survival.	134
Table 28. Cox-regression analysis of TNM stage and relationship of cytoplasmic DUSP6 expression with 5yr overall survival.	138
Table 29. Cox-regression analysis of TNM stage and relationship of cytoplasmic DUSP6 expression with 5yr disease specific survival.	140
Table 31. Genotyping of cohort by pyrosequencing.	167
Table 32. Ratio of target positive cells.	173
Table 33. Genotyping of adenocarcinoma cohort by pyrosequencing.	187
Table 34. Survival and Automated semi-quantitative (AQUA) analysis of DUSP6 protein expression in cohort 2.	199
Table 35. Summary of statistical analysis of clinicopathologic parameters and DUSP6 expression.	200
Table 36. Change in growth in C99 adenocarcinoma cells treated with Cetuximab.	203
Table 37. Change in growth of C99 cell lines treated with Cetuximab.	204

List of abbreviations

5-FU	Fluorouracil
ACF	Aberrant crypt foci
ALK	Anaplastic lymphoma kinase
APC	Adenomatous polyposis coli
AQUA	Automated-Semi Quantitative analysis
BCL2	B-cell lymphoma 2
BRAF	v-Raf murine sarcoma viral homolog B
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
COSMIC	Catalogue of somatic mutations in cancer
CpG	Cytosine-phosphate-guanine
CRC	Colorectal cancer
ctDNA	Circulating tumour deoxyribonucleic acid
DCC	Deleted in colorectal carcinoma
dMMR	deficient mis-match repair
DNA	Deoxyribonucleic acid
DSS	Disease specific survival
DUSP	Dual-specificity phosphatase
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EMT	Epithelial-mesenchymal transition
ER	Estrogen receptor

ERK	Extracellular regulated kinase
ESCC	Eosophageal squamous cell carcinoma
ETS	E-twenty-six
FAP	Familial adenomatous polyposis
FFPE	Formalin fixed paraffin embedded
FIT	Faecal immunochemical test
FOBT	Faecal occult blood test
GDP	Guanine-diphosphate
GFP	Green fluorescent protein
GTP	Guanine-triphosphate
HG	Human glioblastoma
HNPCC	Hereditary non-polyposis colorectal cancer
HNSCC	Head and neck squamous cell carcinoma
HR	Hazard ratio
IF	Immunofluorescence
IHC	Immunohistochemistry
IQR	Inter-quartile range
KIM	Kinase interaction motif
LC	Lung cancer
LOD	Limits of detection
LOH	Loss of heterozygosity
MAPK	Mitogen activated protein kinase
Md	Median
MEK	Mitogen activated protein kinase kinase

MLH1	MutL homolog 1
mm	Macroscopic measurements
MMR	Mis-match repair
MSI	Microsatellite instability
MSS	Microsatellite stable
mTOR	Mechanistic target of rapamycin
NC	Nasopharyngeal carcinoma
NES	Nuclear export signal
NGS	Next generation sequencing
NHS	National health service
NSCLC	Non-small cell lung cancer
OC	Ovarian cancer
OS	Overall survival
p21	Cyclin-dependent kinase inhibitor 1
PC	Pancreatic cancer
PCR	Polymerase chain reaction
p-ERK	phosphorylated extracellular regulated kinase
PGM	Personal genome machine
PI3K	Phosphatidylinositol-3-kinase
P-loop	Phosphate-binding loop
PTPase	Protein tyrosine phosphatase
RAS	Rat sarcoma virus
RIE	Royal infirmary of Edinburgh
RNA	Ribonucleic acid

SSA	Sessile serrated adenoma
TNM	Tumour, node, metastasis system
TP53	Tumour protein 53
TSA	Traditional serrated adenoma
VAF	Variant allele frequency
WNT	Wingless integration-1

Table of Contents

1. Introduction.....	1
1.1 Colorectal cancer.....	1
1.2 Colorectal cancer pathogenesis	2
1.2.1 Histology.....	2
1.2.2 Molecular pathology of CRC.....	5
1.3 The MAPK/ERK signalling pathway and cancer.....	13
1.3.1 MAPK pathway activation and regulation.....	14
1.4 Dual specificity phosphatases: negative regulators of ERK.....	16
1.4.1 Structure and function	16
1.4.2 Tempo-spatial regulation of ERK signalling.....	17
1.4.3 Regulation of DUSP6 expression	19
1.4.4 Evidence for a role of DUSP6 in cancer pathogenesis.....	21
1.5 MAPK/ERK pathway: a target for the treatment of CRC.....	23
1.5.1 The MAPK/ERK as a therapeutic target.....	23
1.5.2 Cetuximab (Erbix [®]) for the treatment of advanced colorectal cancer	24
1.5.3 DUSP6 and chemotherapeutic response	25
1.6 Investigation aims and hypothesis	27
2. Materials and Methods	29
2.1 Statistical analysis.....	29
2.2 Cohort identification	29
2.3 Tissue preparation.....	30
2.3.1 Tissue assessment	30
2.3.2 PCR Microtomy.....	30
2.4 FFPE DNA extraction	31
2.5 Sequencing.....	32
2.5.1 Pyrosequencing	32
2.5.2 Template preparation	32
2.5.3 Pyrosequencing and analysis setup.....	35
2.5.4 Pyrosequencing analysis	36
2.6 Protein expression in formalin fixed paraffin embedded tissue.....	38
2.6.1 Details of primary and secondary antibodies.....	38

2.6.2	Immunohistochemistry	39
2.6.3	Multiplex Immunofluorescence – three targets	39
2.7	Digital image analysis	41
2.7.1	Leica SCN 400	41
2.7.2	Automated semiQUantitative Analysis (AQUA) analysis	41
2.7.3	QuPath for digital image analysis.....	43
2.8	Cell culture	46
2.8.1	Details of cell lines.....	46
2.8.2	Cell passaging	46
2.8.3	Protein extraction from mammalian cells.....	46
2.8.4	Protein determination by Bicinchoninic Acid (BCA) assay	47
2.8.5	Western blotting for protein expression analysis	48
2.9	Genetic manipulation of mammalian cell lines.....	51
2.9.1	Puromycin optimisation	51
2.9.2	DUSP6 transduction of RG/C2/80 and C99 cells	51
2.9.3	Sequencing of DUSP6 plasmid.....	52
2.9.4	Generation of Zeocin™ agar plates	53
2.9.5	Lentiviral production in HEK293T cells.....	54
2.9.6	Dose-response experiments on manipulated cell lines	56
2.9.7	Sulforhodime B (SRB) growth assay	56
3.	Preliminary studies.....	57
3.1	Introduction	57
3.2	Results	58
3.2.1	Antibody optimisation and validation.....	58
3.2.2	Immunohistochemical data analysis	64
3.2.3	<i>Immunofluorescence data analysis</i>	65
3.2.4	X-Tile for the generation of DUSP6 protein expression high and low thresholds	66
3.2.5	Genotyping of cohorts by pyrosequencing	67
3.2.6	Generation of an DUSP6 overexpression model <i>in vitro</i>	70
3.3	Discussion.....	75
4.	The role of dual-specificity phosphatase 6 in colorectal pathogenesis	79
4.1	Introduction	79
4.2	Results	81
4.2.1	Clinicopathological characteristics of adenoma-carcinoma cohort.....	81

4.2.2	Dual-specificity phosphatase 6 (DUSP6) expression	83
4.2.3	Investigating DUSP6 in colorectal adenoma	86
4.2.4	Investigating DUSP6 in colorectal adenocarcinoma	90
4.2.5	Investigating <i>KRAS</i> mutation and associations with DUSP6 expression	92
4.2.6	Investigating DUSP6 and adenocarcinoma type	93
4.2.7	Extracellular regulated kinase (ERK) and phospho-ERK (p-ERK) expression	94
4.2.8	Exploring the relationship between DUSP6 and p-ERK protein expression.....	101
4.2.9	Characterisation of DUSP6 function <i>in vitro</i>	105
4.3	Discussion.....	108
5.	Investigating the role of DUSP6 in treatment response and prognosis in colorectal adenocarcinoma.....	113
5.1	Introduction	113
5.2	Results	115
5.2.1	Cohort clinicopathological dataset.....	115
5.2.2	DUSP6 expression in colorectal adenocarcinoma.....	117
5.2.3	DUSP6 expression and associations with clinicopathological dataset.....	119
5.2.4	DUSP6 expression and association with <i>RAS</i> mutation status.....	120
5.2.5	Investigating DUSP6 expression and function <i>in vitro</i>	123
5.2.6	DUSP6 as a biomarker of prognosis	133
5.2.7	Low DUSP6 expression is associated with poor 5yr. overall survival and disease specific survival	134
5.2.8	Low DUSP6 expression in Stage III colorectal cancer is associated with poor 5yr. overall survival and disease specific survival	138
5.3	Discussion.....	142
6.	Discussion.....	147
7.	References.....	153
8.	Appendix	165
8.1	Cohort 1.....	165
8.1.1	Genotyping of cohort by pyrosequencing.....	165
8.1.2	DUSP6 protein expression: QuPath analysis data	168
8.1.3	Assessment of ERK expression in RG/C2/80 cell lines by western blot (repeats)	174
8.1.4	Assessment of p-ERK expression in RG/C2/80 cell lines by western blot (repeats)	175
8.2	Cohort 2.....	176
8.2.1	Genotyping of cohort by pyrosequencing.....	176

8.2.2	DUSP6 protein expression: Survival and DUSP6 AQUA data	188
8.2.3	Statistical analysis of clinicopathologic features with DUSP6 protein expression	200
8.2.4	Assessment of ERK expression in C99 cell lines by western blot (repeats).....	201
8.2.5	Assessment of p-ERK expression in C99 cell lines by western blot (repeats) ...	202
8.2.6	Growth response to Cetuximab treatment – parental cell line	203
8.2.7	Growth response to Cetuximab treatment – all cell lines.....	204
8.3	Ethics approval	205
8.4	Tissue Governance approval	206
8.4.1	Cohort 1 tissue governance approval	206
8.4.2	Cohort 2 tissue governance approval	207
8.5	Standard operating procedures	208

1. Introduction

1.1 Colorectal cancer

Approximately 41,000 new cases of colorectal carcinoma (CRC) are diagnosed annually in the UK; 110 new cases each day. CRC accounts for 12% of all cancer incidence in the UK and is the 4th most common type of cancer. The incidence increases with increasing age from approx. 50yrs with diagnosis being most prevalent (44% of new cases) in people aged over 75¹.

The faecal immunochemical test (FIT) was first introduced in Scotland in 2017. This nationwide screening test is routinely sent to individuals aged 50-74 years. The FIT is a variation on the faecal occult blood test (FOBT) used in England, and specifically binds to haemoglobin to quantify the amount of blood in a patients stool sample². A large proportion of individuals will demonstrate abnormal results and be referred for colonoscopy. Of these patients adenomatous lesions (pre-neoplastic) are identified in 48% of men and 35% of women. This screening method has identified the highest proportion of individuals with CRC (40%). The introduction of routine screening may explain in large part the increased 10 year survival over the past decade. It is estimated that 57% of patients now diagnosed with CRC will survive 10 years. However, as half of all cases are diagnosed at a later stage CRC remains the 2nd most common cause of cancer deaths in the UK³.

A variety of environmental factors are thought to impact lifetime risk of developing CRC including diet, smoking, obesity and lack of physical exercise⁴. Approximately 90% of colorectal cancers are sporadic, but the remaining 10% have some form hereditary genetic basis i.e. Lynch syndrome (hereditary non-polyposis colorectal cancer, HNPCC) or familial adenomatous polyposis (FAP).

1.2 Colorectal cancer pathogenesis

1.2.1 Histology

1.2.1.1 Normal colonic mucosa

The ultrastructure of the colonic mucosa is composed of three main layers: epithelium, lamina propria and muscularis mucosa. Absorptive columnar cells are the outermost cell type of the colonic epithelium and interact with the luminal environment absorbing ions and water whilst goblet cells (located mid-crypt) secrete mucin which is secreted into the lumen of the large intestine. Undifferentiated stem cells are located at the base of the crypt and migrate up the crypt during differentiation⁵. The lamina propria is situated between crypts and contains a variety of cell types. Comprising of a thin smooth muscle layer the muscularis mucosa delineates the border of the colonic epithelium and lamina propria, Fig. 1.

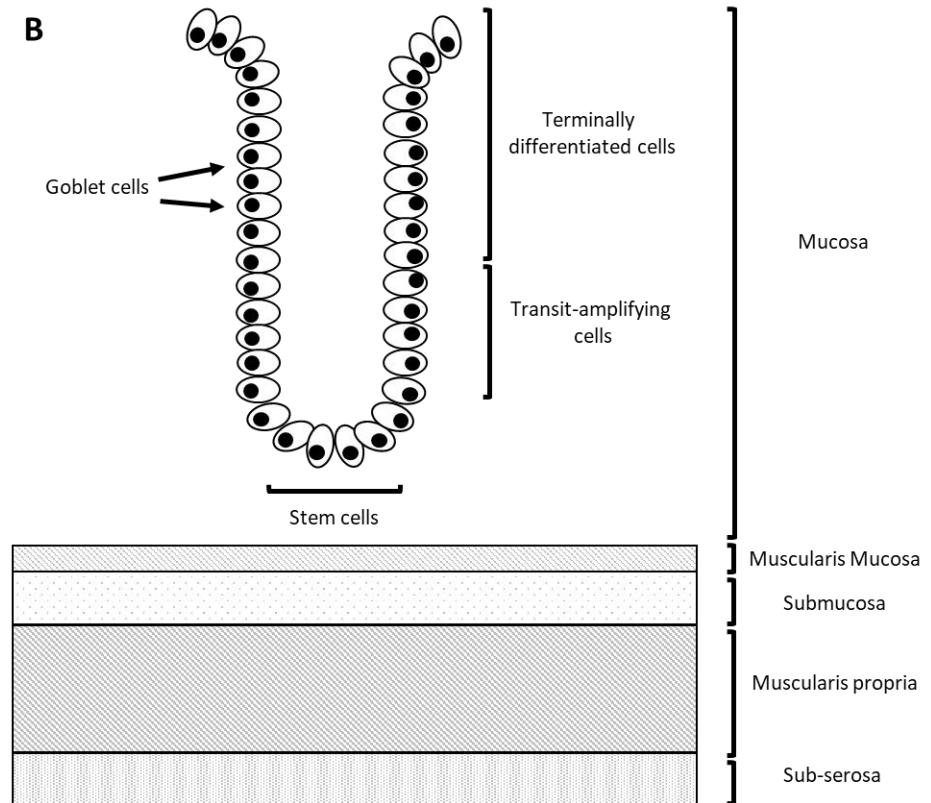
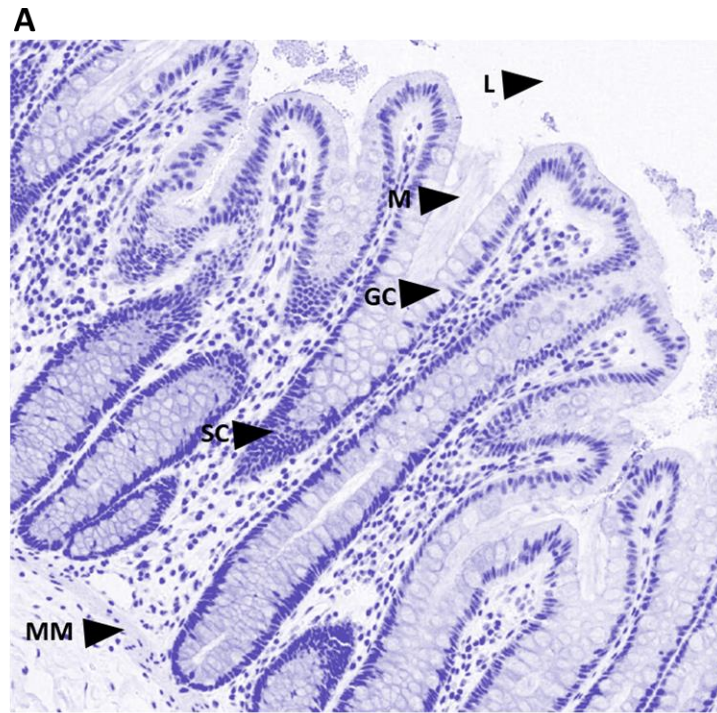


Figure 1. Colonic mucosa structure. A. Haematoxylin stained section of normal colonic mucosa. B. The three main sections of the colonic mucosa are epithelium, lamina propria and muscular mucosa. A range of cell types with different functions are present in the colonic crypt. Absorptive columnar cells on the luminal edge of the mucosa aid ion and water absorption whilst mucin producing and secreting goblet cells lubricate the intestinal lumen. Undifferentiated stem cells are located at the base of the crypt and migrate upward during differentiation. The lamina propria contains a variety of cells including fibroblasts whilst the muscular mucosa, a thin layer of smooth muscle delineates the colonic mucosa from the sub-serosa. L: intestinal lumen. M: mucin. GC: goblet cells. SC: stem cell region. MM: muscularis mucosa.

1.2.1.2 Adenoma and adenocarcinoma

Aberrant crypt foci (ACF) are the first observable features of colorectal carcinogenesis and are identified by the presence of enlarged crypt diameter, vesicular nuclei and variable mucin depletion⁶. Adenomatous lesions are thought to develop from ACF and present with a variety of histology including tubular, tubulovillous, villous and serrated features, Fig. 2.

Tubular adenomas are characterised histologically by a smooth epithelial boundary whilst villous adenomas are characterised by the presence of villous, finger-like projections on the outer surface of the colonic ultrastructure. Tubulovillous adenomas comprise a combination of the two structural forms. Serrated adenomas present with saw-tooth appearance⁷. Tubular, tubulovillous and villous adenomas are thought to follow similar molecular pathways of pathogenesis whereas serrated adenomas are considered both a phenotypically and molecularly distinct sub-group of adenomas.

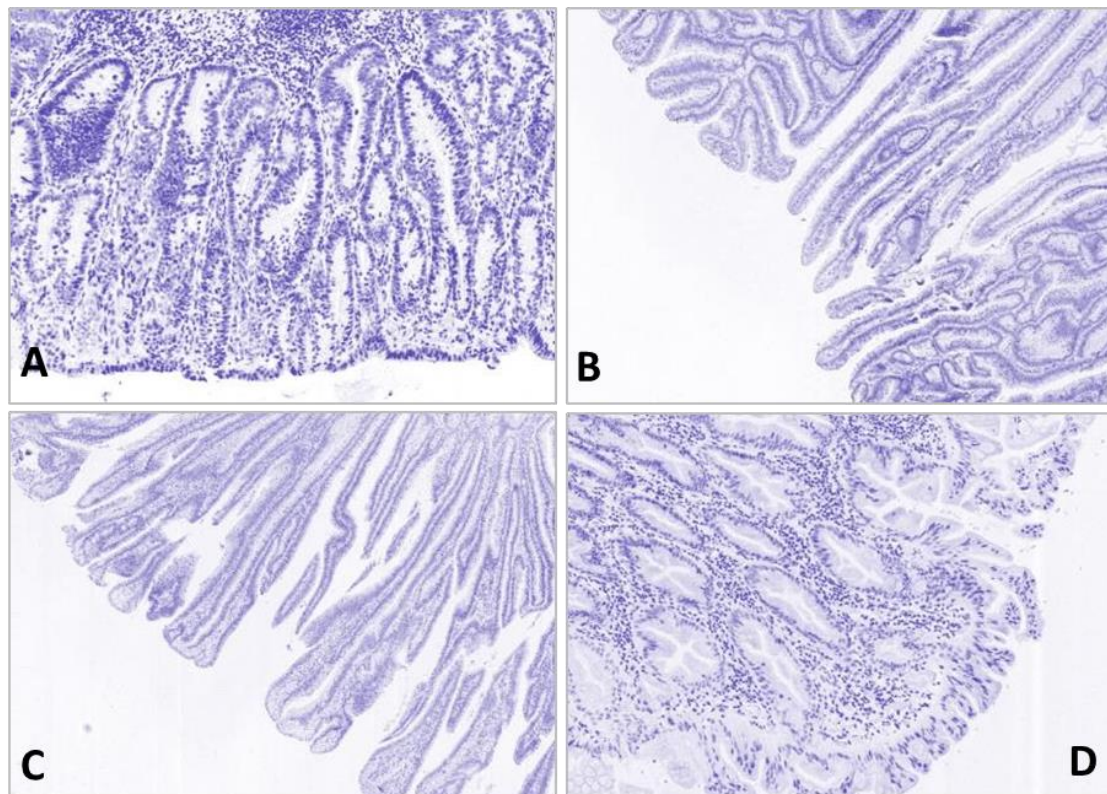


Figure 2. Haematoxylin stained sections of adenomatous lesions. Tubular adenomas present with a smoothed epithelial boundary whilst villous adenomas demonstrate finger like projections. Tubulovillous adenomas are comprised of a combination of both features. Serrated adenomas present with saw-tooth like appearance to the colonic crypts and are considered to follow a distinct molecular pathway from the previous adenomas. A: Tubular adenoma. B: Tubulovillous adenoma. C: Villous adenoma. D: Serrated adenoma.

Diagnosis of invasive carcinoma requires histological evidence of neoplastic cells exceeding the muscularis mucosa boundary⁷. Staging of colorectal adenocarcinoma is based upon a tumour, node, metastasis (TNM) system⁸. TNM stage I: evidence of neoplastic cells in the submucosa. TNM stage II: evidence of neoplastic cells in the muscularis propria. TNM stage III: neoplasm has grown into the sub-serosa. TNM stage IV: lesion has grown through all layers of the colon and may be present in other organs. TNM stage I and II fail to present with nodal involvement or metastasis. TNM stage III present with varying degrees of nodal involvement but no metastasis. TNM stage IV present with nodal involvement and metastasis.

1.2.2 Molecular pathology of CRC

As proposed by Hanahan and Weinberg^{9,10} the successful transformation of ‘normal’ cells to malignancy requires augmentation of the homeostatic cellular control by acquiring the ‘hallmarks of cancer’. At the core, the hallmarks of cancer comprise: “sustaining proliferative signalling, evading growth suppressors, activating invasion and metastasis, enabling replicative immortality, inducing angiogenesis and resisting cell death”¹⁰, Fig. 3.

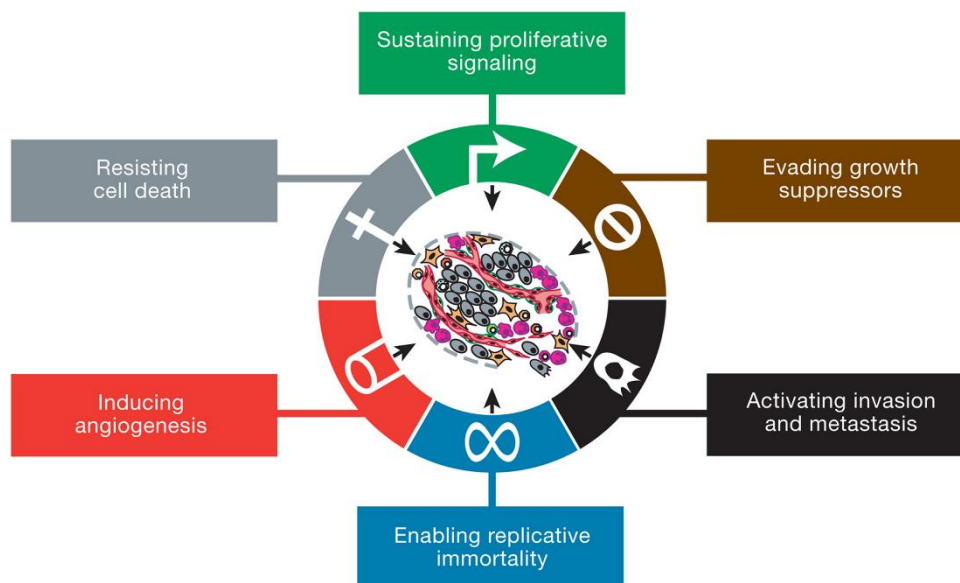


Figure 3. The Hallmarks of Cancer. The hallmarks of cancer proposed by Hanahan and Weinberg proposes that for the malignant transformation of normal tissue acquisition of a number of hallmarks is required. Image: Hanahan and Weinberg., 2011

Colorectal cancer (CRC) is a highly heterogeneous disease with a sequence of numerous molecular alterations implicated in its pathogenesis. Underpinning colorectal pathogenesis is the concept of a gradual accumulation of genetic and epigenetic aberrations enabling the transition of normal mucosa to adenocarcinoma. In this manner, dysregulated cells overcome the restraints of pathway regulation and acquire the hallmarks of cancer thus enabling malignant transformation.

The concept of a multi-step model of pathogenesis for colorectal cancer was first proposed by Fearon and Vogelstein¹¹ in 1990. Histologically, aberrant crypt foci (ACF) are considered to represent the initial observable transformation of normal cells to malignancy and are subdivided into two distinct forms; dysplastic and hyperplastic. Dysplastic ACF progress from adenomatous polyps to adenocarcinoma, whilst hyperplastic ACF progress through either sessile serrate (SSA) or traditional serrated adenomas (TSA) to serrated neoplasms (both benign and malignant). Three distinct molecular genetic and epigenetic features underpin these transitions.

1. Chromosomal instability: the most common form of genetic instability in colorectal carcinogenesis accounting for approx. 80% of sporadic CRC^{12,13}. CIN results in chromosomal aneuploidy, loss of heterozygosity (LOH) and genomic amplifications. Commonly in CRC, CIN is accompanied with a distinct signature of driver mutations however it is unclear whether CIN promotes this mutator phenotype or vice versa¹⁴.
2. Microsatellite instability: microsatellites comprise repeated nucleotide sequences dispersed throughout the genome and are a product of uncorrected polymerase errors during DNA synthesis¹⁵. In instances of loss of DNA mismatch repair (MMR) genes via hypermethylation of the MLH1 promoter or somatic gene mutations, surveillance and correction of replicative errors is compromised and microsatellite instability (MSI) occurs¹⁶.
3. CpG island methylation: Cytosine-phosphate-guanine (CpG) islands are regulatory sites comprised of CpG dinucleotide repeats or 'islands' located in gene promoter regions. Under normal conditions, CpG islands exist in an unmethylated state. Aberrant methylation of these islands results in transcriptional repression of the respective genes¹⁷. In approx. 15% of CRC, hypermethylation of CpG islands termed CpG island methylation phenotype-high (CIMP-H) occurs causing gene silencing.

The development of a colorectal adenocarcinoma can be characterised by a combination of any of the above molecular features. Figure 4 summarises the incorporation of these features to describe the pathogenic process of two main molecular pathways to malignancy.

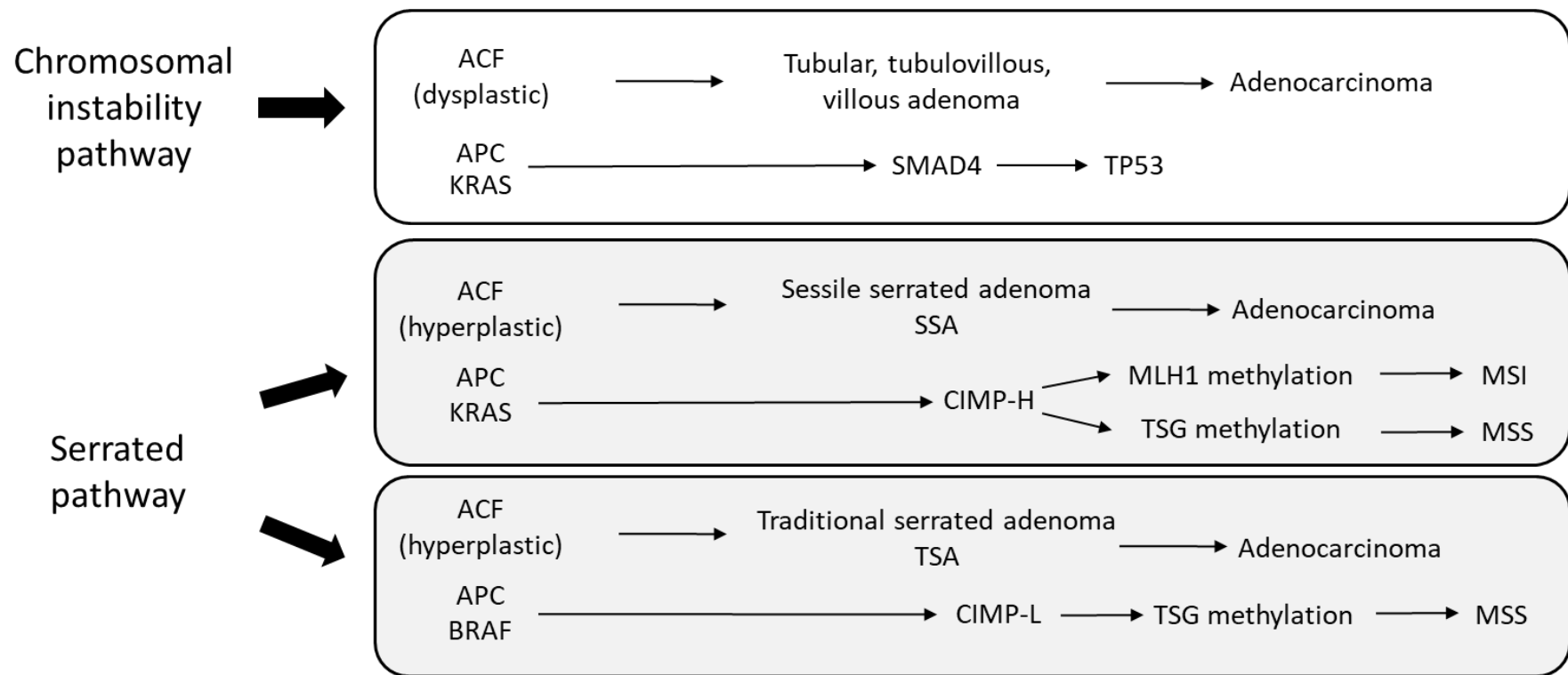


Figure 4. Schematic of basic molecular aberrations in colorectal pathogenesis. Two main pathways integrate histological and molecular features of CRC to describe the progression of 1. traditional adenomas to adenocarcinoma, termed the chromosomal instability pathway (CIN) and 2. serrated adenomas to serrated adenocarcinoma, termed the serrated pathway. It is important to note that whilst the majority of sporadic CRC develops from the CIN pathway, malignant transformation can be characterised by other molecular aberrations. The CIN model details early mutation events in *APC* and *RAS* initiating cellular transformation of normal mucosa to aberrant crypt foci (ACF) whilst later stages involve the accumulation of mutations and loss of heterozygosity in tumour suppressor genes (TSG) such as *TP53*. Hyperplastic ACF are considered to develop into serrated adenomas (sessile serrated adenoma (SSA) and traditional serrated adenoma (TSA)). Mutations in *APC*, *RAS* and *BRAF* are common early events. A high degree of CpG island methylation (CIMP) and microsatellite instability (MSI) being observed in SSA.

1.2.2.1 Chromosomal instability pathway

Chromosomal instability pathway describes the route to malignancy for the majority of sporadic CRC. In these cases, a high frequency of *adenomatous polyposis coli* (*APC*) mutations is observed in early adenomas (~40-70%) and is thought to be an important event in the progression of normal mucosa to adenoma^{18,19}. One study investigating the prevalence of allelic imbalances in a number of chromosomal locations frequently associated with CRC carcinogenesis identified allelic imbalance in 55% of early adenomas (1-3cm in size) at chromosome 5q, location of the *adenomatous polyposis coli* (*APC*)²⁰. Additional studies have identified mutations in the *APC* gene in early adenomas as small as 0.5cm.

In addition to the initial events in the adenoma carcinoma sequence, losses of heterozygosity in genes such as 18q (deleted in colorectal carcinoma, DCC locus) and 17p (*TP53* locus) in the mid-late stage adenomas confer inactivation of these genes thus enabling further progression toward malignant transformation. In reflection upon current evidence, CIN occurs in a large proportion of adenomas ultimately resulting in malignant transformation however the 'flavour' of allelic gains and losses on route can vary between lesions which may be reflected in subsequent prognostic outcomes and treatment responses^{21,22}.

1.2.2.2 Serrated pathway

The serrated pathway identifies a sub-group of histologically distinct colorectal lesions which differ in molecular pathogenesis from common adenomatous polyps previously described. Lesions developing from the serrated molecular pathway can be further sub-classified into sessile serrated (SSA) and traditional serrated adenomas (TSA) each with their own distinct phenotypic presentations^{23,24}. Both genetic and epigenetic changes are synonymous but not exclusive to the serrated pathway. At the genetic level, mutations are observed in *RAS* or *BRAF* and are indicative of distinct progression to SSA or TSA. At the epigenetic level, CpG island methylation (CIMP) and microsatellite instability (MSI) are associated with specific subgroups. In summary, *KRAS* mutations and CIMP-H are common to TSA whilst *BRAF* mutations and low CIMP are more common amongst SSA²⁵. The effect of the type of molecular aberration and how it impacts phenotype is evident in serrated lesions, Table 1.

Pathway	Prognosis	BRAF / KRAS mutant	CIMP status	Microsatellite status
SSA	Very poor	BRAF	High	Stable
SSA	Good	BRAF	High	High instability
TSA	Poor	KRAS	Low	Low instability
TSA	Poor	KRAS	Low	Stable

Table 1. Summary of prognostic outcomes in serrated lesions. It is evident from serrated lesions that the type and combination of molecular aberrations can significantly impact phenotype. In sessile serrated adenomas (SSA) the occurrence of microsatellite instability (MSI) by hypermethylation of *MLH1* contributes to the best prognostic outcome among all types of serrated lesions. Whereas, SSA with CpG island methylation (CIMP) in genes other than *MLH1* does not contribute as significantly to the MSI phenotype but appears to generate a more aggressive phenotype with patients demonstrating the poorest prognostic outcome in this molecular sub-group out of all serrated lesions.

SSA lesions with MSI demonstrate the best prognostic outcome among all serrated lesions. In these lesions MSI can be attributed to hypermethylation of DNA MMR genes such as *MLH1*. Conversely SSA lesions with microsatellite stability (MSS) have the poorest prognosis of all serrated lesions with aberrant methylation occurring in genes other than *MLH1*. It could be inferred that within SSA lesions, *MLH1* hypermethylation dependent MSI is a more potent contributor to an aggressive phenotype²⁵.

In an attempt to better capture the diverse molecular characteristics of colorectal adenocarcinoma a number of additional classification systems have been proposed; the Consensus Molecular Subtypes (CMS) and the CRC Intrinsic Subtypes (CRIS) systems.

The CMS system compiles a number of independent classification systems into a single system comprising four groups; CMS1, CMS2, CMS3 and CMS4^{26,27}.

- CMS1 comprises serrated polyps originating from the serrated pathway (approx. 14% of all CRC²⁸). Generally, lesions in this group are located in the caecum, ascending or transverse colon, a high incidence of *BRAF* mutation, CIMP high phenotype and an occurrence of a deficient mis-match repair system. These lesions have a tendency of high immune infiltrate in the tumour micro-environment.
- CMS2 lesions arise from the traditional CIN pathway as previously described (approx.. 37% of all CRC²⁸). These lesions demonstrate a high degree of somatic copy number variations with copy number gains in oncogenes and losses in tumour suppressor genes. CMS2 have a low mutation rate compared to CMS1.

- CMS3 which are also termed the metabolic subtype have similarities to CMS2 in terms of CIN however demonstrate a lesser degree of somatic copy number variations but a higher incidence of microsatellite instability (approx. 13% of all CRC²⁸). *KRAS* mutations are more frequently observed in CMS3 lesions compared to CMS2 and are thus considered hypermutated (approx. 30% of CMS3 lesions). When assessed, 90% of CMS3 lesions demonstrated molecular profiles favouring metabolic pathways such as glutamine and fatty acids metabolism.
- CMS4 group of lesions comprise a mesenchymal subtype (approx. 23% of all CRC²⁸). The molecular pathways observed in these lesions are akin to CMS1 but present with a CIMP low phenotype, microsatellite stable and the highest degree of somatic copy number variations of all the CMS subtypes. CMS4 lesions are termed mesenchymal due to the high expression of genes associated with among others, angiogenesis, cell cohesion and cancer-associated fibroblasts. The tumour micro-environment is highly inflammatory compared to CMS1 lesions. In addition, CMS4 lesions appear to associated with colitis associated CRC where loss of *TP53* occurs in the transition to dysplasia in comparison to CMS2 lesions where *TP53* inactivation is a later event for the transformation from adenoma to carcinoma.

The CRIS system of classification focuses on the transcriptional landscape of CRC patient derived xenografts (PDX). Five CRIS subtypes have been identified; CRIS-(A-E). These subtypes can be further grouped into two main classes; 1) CRIS-A and CRIS-B and 2) CRIS-C, CRIS-D and CRIS-E²⁹.

In group 1 microsatellite instability was a feature most commonly identified with CRIS-A and to a lesser extent CRIS-B. Additionally, group 1 lesions were predominantly located in the right colon and CIMP high phenotype and with a high prevalence of *BRAF* mutations (13 out of 15 lesions tested) as observed in CMS1 subtypes from the previous classification system. Those CRIS-A lesions which were microsatellite stable frequently presented with *KRAS* mutations a phenomenon which has been previously reported for traditional serrated pathway lesions²⁵.

Group 2 lesions generally demonstrated features of the CIN pathway with a high incidence of somatic copy number variation as observed in CMS2 subtype. CRIS-C and CRIS-E appeared to experience more chromosomal arm gains than CRIS-D. Specific gene amplifications were observed with group 2 lesions; CRIS-C demonstrated amplifications in the chromosomal location of *MYC* whilst CRIS-D demonstrated amplifications in *IGF2* which was accompanied with high

expression levels of this gene product. CRIS-C lesions was strongly enriched for *KRAS* wildtype genotype, whilst *TP53* mutations were strongly associated with CRIS-E lesions.

1.3 The MAPK/ERK signalling pathway and cancer

The MAPK/ERK pathway is a well characterised signalling cascade in human malignancies, Fig. 5. Responsible for the regulation of a variety of cellular process including proliferation and pro-survival activities, negating regulation of this pathway is an important step in the pathogenesis of many types of cancer. In addition to CRC, mutations in genes of the main effectors of this pathway, namely *RAS* and *RAF* are well documented for a variety of malignancies including pancreatic, prostate and NSCLC³⁰.

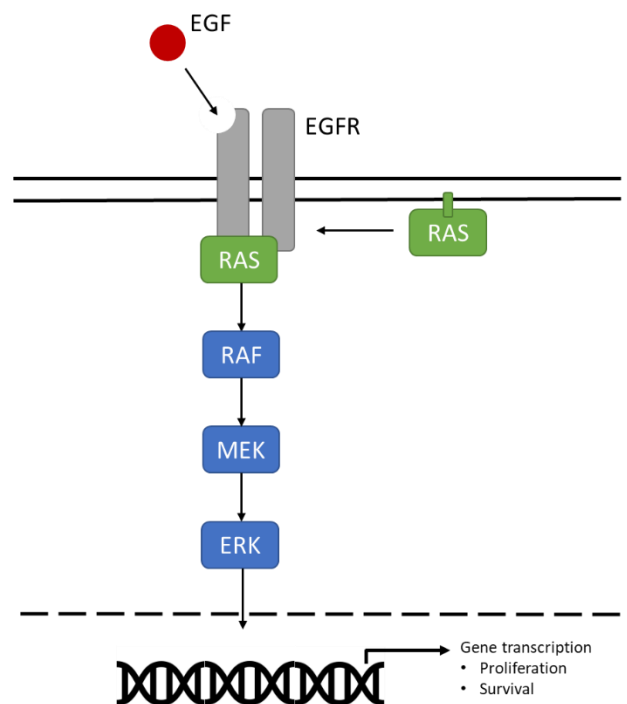


Figure 5. MAPK/ERK signalling pathway. One mechanism of activation over the MAPK/ERK signalling pathway involves the transmembrane epidermal growth factor receptor (EGFR) which, when activated by one of its ligands e.g. epidermal growth factor (EGF) associates with the Rat sarcoma virus (RAS) family of GTPases. This association sets off a cascade of phosphorylation events downstream culminating in transcription of genes involved in proliferation and survival.

Mutations in *Ras Sarcoma Virus (RAS)* family of GTPases are present in at least 90% of dysplastic ACF^{31,32} and are thus considered an early driver event in colorectal cancer. Mutation of the *RAS* gene in hotspot locations results in the constitutive activation of the protein³³. These genetic changes mean that RAS can interact with downstream effectors of the MAPK/ERK signalling cascade, negating the reliance upon upstream growth signals for activation. The high prevalence of mutations in *RAS* early on in disease development in CRC is suggestive that overcoming

regulation of this pathway enables proliferation and survival of dysregulated cells and thus favours the pathway to malignant transformation.

1.3.1 MAPK pathway activation and regulation

The epidermal growth factor receptor (EGFR) is a member of the tyrosine kinase family of receptors and an activator of the MAPK/ERK signalling pathway. Receptor activation occurs upon ligand binding of the EGFR inducing a conformational change in the receptor to enable homo/hetero-dimerization and autophosphorylation^{34,35}. This in turn, activates the effector protein RAS³⁶. A series of phosphorylation events ensue downstream of RAS ultimately terminating in the activation of ERK³⁷. For this axis of the pathway, ERK represents the cytoplasmic terminus of the signalling cascade. Once activated ERK is able to translocate to the nucleus and induce the upregulation of transcription factors such as the ETS family which augment the transcription of genes responsible for cell proliferation and survival^{38,39}. Additionally, ERK interacts with a variety of pathway activators and inhibitors which regulate a variety of cellular processes, Fig. 6⁴⁰.

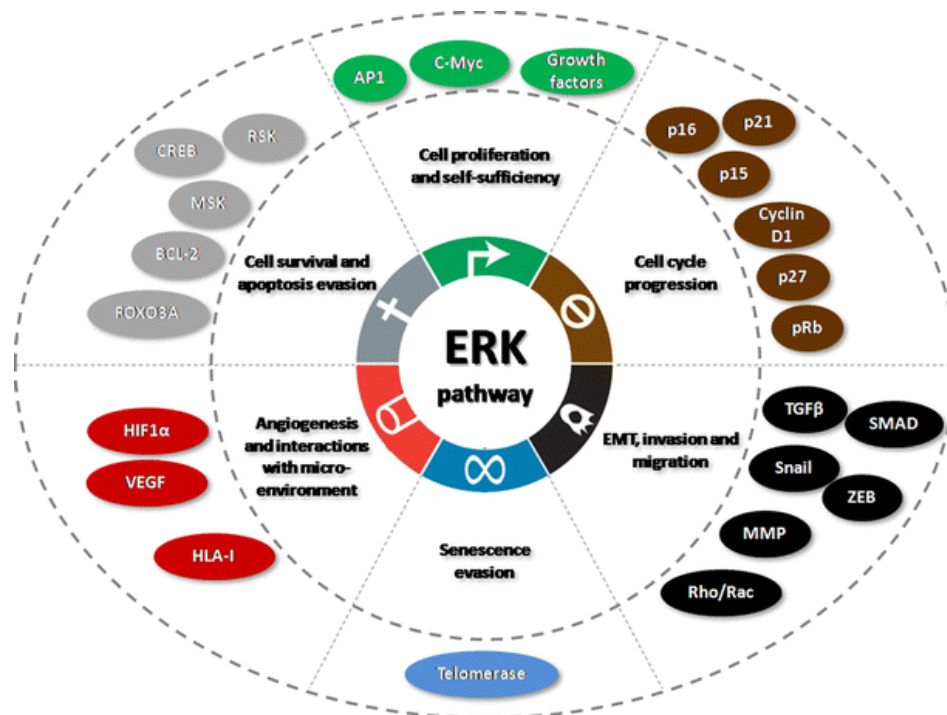


Figure 6. MAPK/ERK signalling pathway and the hallmarks of cancer. ERK interacts with a variety of mediators which regulate a variety of cellular processes including proliferation, cell cycle progression and survival. It is evident from the variety of interactions of ERK that aberrance in effectors of its activation and regulation can impact the tight regulation of signalling pathways maintaining the homeostatic control of the cell environment, thus favouring the survival of dysregulated cells. Image: Neuzillet et al., 2013.

With its pivotal role in the regulation of many cellular processes the MAPK/ERK pathway is tightly regulated in order to prevent the growth and advancement of dysregulated cells. One type of regulation is to attenuate the duration and intensity of ERK which is achieved by a variety of mechanisms both rapid and long term, Fig. 7³⁰.

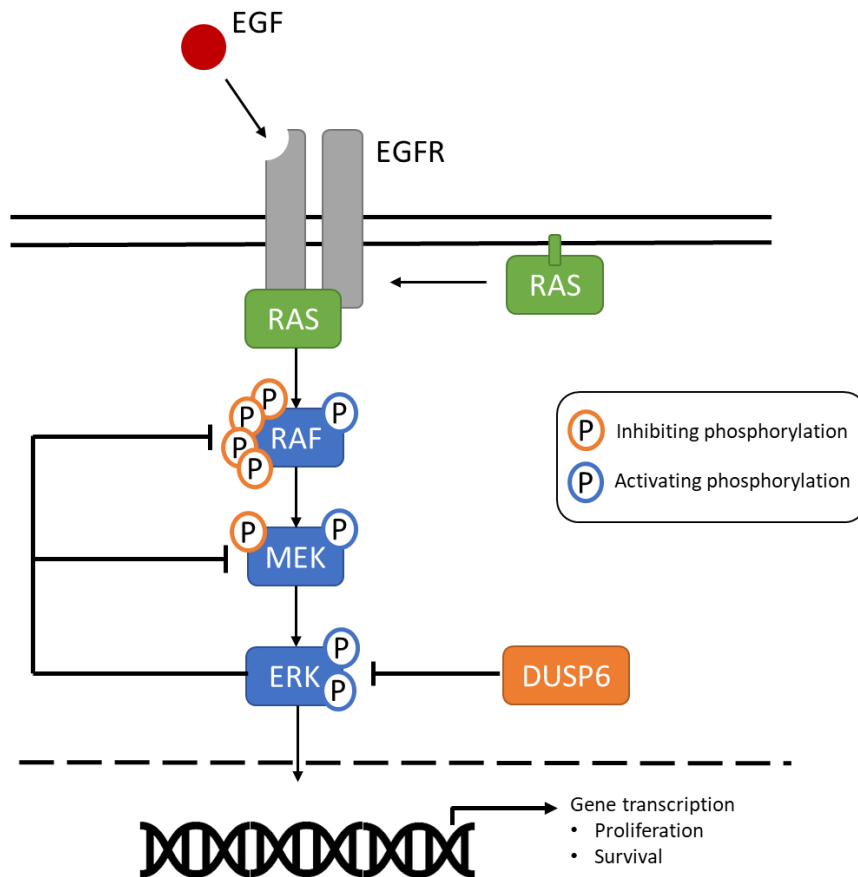


Figure 7. Schematic of ERK regulation. The MAPK/ERK pathway plays a vital role in the regulation of many cellular processes ensuring the growth and survival of normal cells. The MAPK/ERK pathway is regulated at a variety of points throughout the pathway. One type of regulation is the tempo-spatial regulation of ERK. Intrinsically, ERK self-regulates its activation by phosphorylating upstream effectors such as RAF and MEK, inhibiting their activity and subsequent phosphorylation of ERK. A second longer term method of regulation is carried out by the dual specificity phosphatases (DUSPs) which target both cytoplasmic and nuclear ERK. Figure adapted from Liu et al., 2018.

ERK acts to self-regulate its activation by phosphorylating its upstream effectors such as MEK and RAF resulting in their inhibition and subsequent absence of activity upon ERK. Direct, longer term regulation of ERK both cytoplasmic and nuclear is achieved by the dual specificity phosphatases (DUSPs).

1.4 Dual specificity phosphatases: negative regulators of ERK

1.4.1 Structure and function

The primary function of DUSPs as phosphatase enzymes is to de-phosphorylate and thus inhibit the activity of their target proteins. All DUSPs which interact with the ERK family of proteins (including DUSP6) share a large degree of structure homology comprising a C-terminal catalytic domain and an N-terminal non-catalytic domain. In the later reside a cluster of amino acids involved in MAPK recognition (kinase interaction motif (KIM))⁴¹⁻⁴³ and additional residues which determine sub-cellular location of the DUSP, Fig. 8. The presence of a nuclear export signal (NES) suggests the ability of DUSP6 to move between cytoplasmic and nuclear sub-cellular locations⁴⁴.

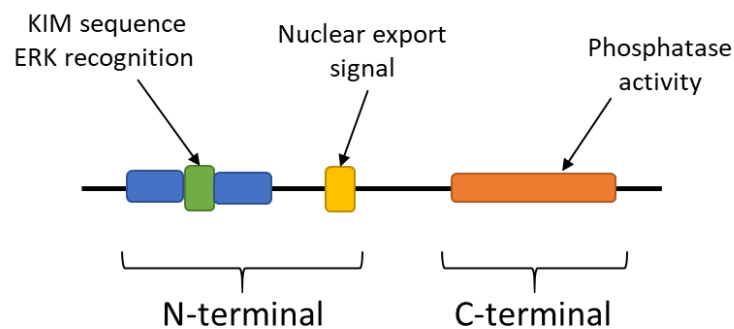


Figure 8. Dual specificity phosphatase 6 structure. The DUSP enzymes share a conserved structure of a C-terminal catalytic domain and an N-terminal non-catalytic domain which comprises a kinase interactive motif (KIM) enabling MAPK recognition and either a nuclear localisation signal (NLS) or nuclear export signal (NES) in order to determine sub-cellular location. Image adapted from Keyse et al., 2016.

The DUSP6 gene locus is on chromosome 12q21.33⁴⁵ and encompassed within 3 exons. Exon 1 encodes the KIM recognition sequence and exon 2 encodes the NES. The highly conserved protein tyrosine phosphatase (PTPase) domain spans part of exon 2 and 3. The catalytic site of the PTPase domain is located in exon 3⁴⁶, Fig. 9. Binding of DUSP6 to ERK1/2 can occur irrespective of the phosphorylation status of its substrate. The binding of negatively charged residues in the KIM domain of DUSP6 interact with positively charged residues of the common docking domain of ERK causing an activating conformational change in DUSP6 enabling its primary role in the dephosphorylation of ERK by the catalytic HCXXXXXR domain within the PTPase domain⁴⁷⁻⁴⁹.

10	20	30	40	50
MIDTLRPVPF	ASEMAISKTV	AWLNEQLELG	NERLLIMDCR	PQELYESSHI
60	70	80	90	100
ESAINVAIPG	IMLRRLOKGN	LPVRALFTRG	EDRDRFTRRC	GTDTVVLYDE
110	120	130	140	150
SSSDWNENTG	GESVLGLLK	KLKDEGCRAF	YLEGGFSKFQ	AEFSLHCETN
160	170	180	190	200
LDGSCSSSSP	PLPVLGLGGL	RISSDSSSDI	ESDLDRDPNS	ATDSDGSPLS
210	220	230	240	250
NSQPSFPVEI	LPFLYLGCCK	DSTNLVLEE	FGIKYILNVT	PNLPNLFENA
260	270	280	290	300
GEFKYQIPI	SDHWSQNSQ	FFPEAIFID	EARGKNCVGL	VHCLAGISRS
310	320	330	340	350
VTVTVAYLMQ	KLNLMSNDAY	DIVKMKKSN	SPNFNFMGQL	LDFERTLGLS
360	370	380		
SPCDNRVPAQ	QLYFTTPSNQ	NVYQVDSLQS	T	

Figure 9. DUSP6 amino acid sequence. The DUSP6 gene encompasses 3 exons. The KIM recognition sequence (grey highlighted, red text) is encoded by exon 1 (red). The NES (grey highlighted, blue text) is encoded by exon 2 (blue). The protein tyrosine phosphatase (PTPase) domain (grey highlighted, blue and green text) is encoded by exon 2 and 3 (green). The catalytic site of the PTPase (black italic text) is encoded by exon 3.

There are two documented isoforms of DUSP6⁵⁰. Isoform A is 381 amino acids in length and 42.3kDa in mass. This isoform is considered the canonical DUSP6 and most ubiquitously studied. Isoform B is a truncated version of A, which is 235 amino acids in length and 26.4kDa in mass⁵¹. Isoform B lacks the coding sequence for the NES and part of the phosphatase domain which may be indicative of an alternative function of isoform B however little is known about this isoform from literature.

1.4.2 Tempo-spatial regulation of ERK signalling

In addition to the dephosphorylating activities of DUSP6, interestingly, ERK is not required to be phosphorylated in order for its recognition and binding with DUSP6, identifying a potential role for DUSP6 in addition to ERK dephosphorylation. The spatial regulation of ERK by DUSP6 has been demonstrated previously. Upon its activation by phosphorylation from RAF, MEK acts to phosphorylate ERK resulting in its nuclear increase. DUSP6 acts to compete with MEK for ERK, binding ERK and retaining it within the cytoplasm, acting as a cytoplasmic anchor, thus reducing ERK mediated transcription in the nucleus. The presence of a NES within DUSP6 structure also implicates an ability for DUSP6 to sequester ERK from the nucleus to the cytoplasm. It has been demonstrated under basal conditions nuclear export dominates thus resulting in a cytoplasmic accumulation of DUSP6⁴⁴. This evidence of DUSP6 regulating ERK localisation as well as mediating its de-phosphorylation supports a potential role of DUSP6 as a tumour suppressor in

which it serves to reduce excessive ERK signalling. The mechanisms of DUSP6 control on ERK signalling is summarised in Fig. 10. There are 3 main ways in which DUSP6 regulates ERK signalling: 1. DUSP6 can bind ERK irrespective of phosphorylation state in the cytoplasm thus acting as a cytoplasmic anchor, preventing the nuclear translocation of p-ERK and subsequent transcription of pro-survival and proliferative genes, 2. DUSP6 can de-phosphorylate p-ERK thus preventing p-ERK nuclear translocation and 3. the presence of an NES enables DUSP6 to shuttle between the nucleus and cytoplasm whereby it can translocate p-ERK from the nucleus, thus reducing ERK directed gene transcription.

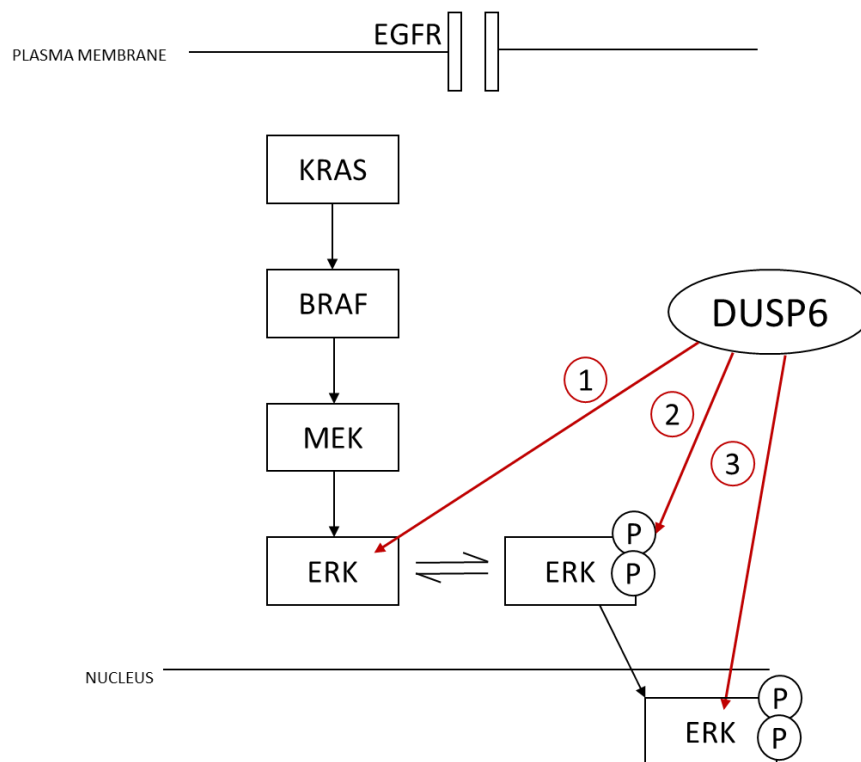


Figure 10. Summary of DUSP6 control on ERK signalling. DUSP6 has three main ways in which to regulate ERK signalling both temporally and spatially. 1. The KIM sequence enables recognition of ERK by DUSP6 which is able to bind ERK irrespective of phosphorylation state thus acting as a cytoplasmic anchor preventing movement of ERK to the nucleus where it can activate transcription of pro-survival and proliferative genes. The presence of a PTPase domain in the C terminal of DUSP6 enables dephosphorylation of ERK preventing its movement to the nucleus. Additionally, the presence of a nuclear export signal (NES) enables DUSP6 to shuttle between the cytoplasm and nucleus to remove p-ERK from the nucleus thus reducing its nuclear effects on gene transcription.

1.4.3 Regulation of DUSP6 expression

A number of MAPK pathway members have been implicated in the transcriptional and post-translational control of DUSP6, Fig. 11. Direct regulation of DUSP6 occurs upon the formation of the de-phosphorylation complex with ERK. This interaction subsequently results in the phosphorylation of DUSP6 on residues S159 and S197, targeting DUSP6 for proteasomal degradation indicating a negative feedback loop between ERK and DUSP6⁵².

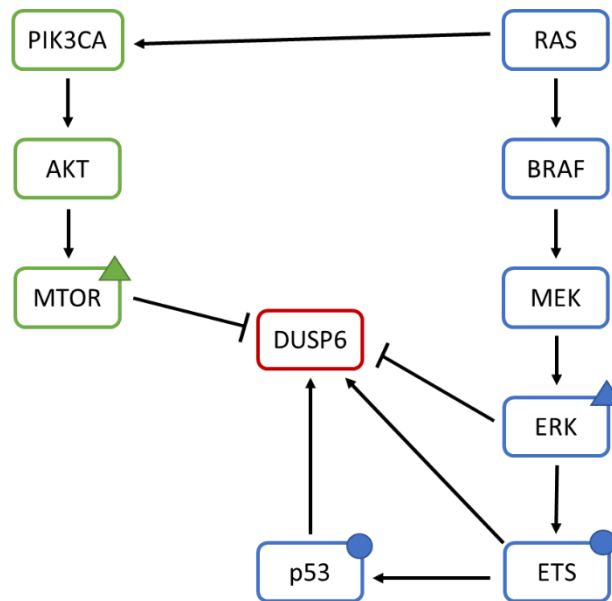


Figure 11. Schematic of DUSP6 transcriptional and post-translational control. A number of studies have demonstrated evidence for regulation of DUSP6 both transcriptionally and post-translationally. Post-translational: Following interaction with its substrate ERK, phosphorylation of residues S159 and S197 by ERK target DUSP6 for proteasomal degradation. In addition mTOR has been shown to bind DUSP6, phosphorylating S159 and targeting the protein for proteasomal degradation in a similar fashion to ERK. Transcriptional: ETS transcription factor binding motifs have been identified within the promoter region and intron 1 of the DUSP6 gene. Additionally p53 binding sites have been identified within the DUSP6 promoter. ▲ post-translational regulation ■ post-transcriptional regulation.

A well characterised target of ERK, the ETS family of transcription factors, Ets1 and Ets2 have been shown by a number of studies to elicit transcriptional control over DUSP6. In murine models an ETS binding motif; CGGAAATTCCT, was identified by Ekerot et al⁵³ within intron 1 of the DUSP6 gene. CpG islands within intron 1 of DUSP6 in pancreatic cell lines; MiaPaCa2 and PAN07JCK, have demonstrated a high degree of methylation. This has been associated with low DUSP6 RNA expression, implicit of a role of intron 1 in the regulation of DUSP6 expression⁵⁴. The promoter activity of intron 1 was confirmed by Furukawa et al⁵⁵ and that this activity depended specifically upon Ets2 binding. Further support for the role of Ets binding to intron 1 of DUSP6

was demonstrated in non-small cell lung cancer. In a study by Zhang et al⁵⁶, the application of the MAPK pathway inhibitor, Erlotinib, significantly suppressed promoter activity of DUSP6 as determined by luciferase reporter assay. Erlotinib treatment resulted in the decreased binding of Ets1 to the intron 1 promoter region of DUSP6. To further support these findings, siRNA knockdown of Ets1 led to a marked downregulation in DUSP6 expression.

Tumour protein 53 (p53), a tumour suppressor has also been implicated in the transcriptional regulation of DUSP6 expression. An interesting study by Piya et al⁵⁷ identified that overexpression of p53 in HCT116 colorectal cancer cells was associated with a significant increase in DUSP6 RNA expression in comparison to construct negative controls. In addition, treatment of HCT116 wildtype cells with Fluorouracil (5-FU), inducing genotoxic stress, resulted in an increase in endogenous levels of p53 and DUSP6. By contrast treatment of HCT116 p53^{-/-} cells demonstrated no increase in DUSP6 expression. Further investigation identified p53 binding sites within the promoter region of DUSP6 indicative of a functional relationship between p53 and DUSP6 in situations of genotoxic stress. Assessment of HCT116 parental lines following 5-FU treatment identified p53 occupation of p53 binding sites in the DUSP6 promoter. In response to 5-FU, over-expression of DUSP6 reduced cell viability as measured by an increase in Caspase-3 expression. Additionally, decreased activated ERK and BCL2 phosphorylation was observed. Conversely, shRNA silencing of DUSP6 increased cell viability accompanied by increased activated ERK and BCL2 expression. In support of a role of p53 mediating DUSP6 expression, a study by Zhang et al⁵⁸ observed in NRK-52E rat epithelial cells that decreased ERK was associated with increased DUSP6 expression. To investigate this further, GFP tagged DUSP6 was over-expressed in NRK-52E cells and demonstrated increase in cell size and β -galactosidase activity; both surrogate measures of senescence, in comparison to controls. In light of the fact that p53 plays an important role in the establishment of the senescent phenotype, cells were treated with a pharmacological senescent inducer, ETO and increased p53 and p21 levels were observed. Abolishment of ETO activity by pifithrin- α resulted in substantial depletion of p53 and p21 expression accompanied by downregulation of DUSP6 promoter activity and DUSP6 protein expression. Collectively these studies indicated a potential role of p53 in DUSP6 regulation.

Finally, mechanistic target of rapamycin (mTOR), a member of the phosphatidylinositol-3-kinase (PI3K) pathway has demonstrated post-translational control over DUSP6. In a study by Bermudez et al⁵⁹, induction of mTOR by pharmacological agonists of the mTOR pathway induced phosphorylation at S159⁵², a previously reported site for DUSP6 phosphorylation and

degradation. Treatment of R443 fibroblast derived cells with the mTOR inhibitor, rapamycin resulted in a significant increase in DUSP6 stability⁵⁹.

1.4.4 Evidence for a role of DUSP6 in cancer pathogenesis

DUSP6 function has been shown to be important in a variety of cancer types. Depending upon tissue context, DUSP6 demonstrates either a tumour suppressive or pro-oncogenic function. Evidence for its tumour suppressive role has been demonstrated in pancreatic, oesophageal squamous cell carcinoma (ESCC), nasopharyngeal carcinoma (NC), non-small cell lung cancer (NSCLC), human glioblastoma (HG) and ovarian carcinoma (OC).

Pancreatic cancer has been the most extensively studied disease group with respect to DUSP6 and its tumour suppressive role, but the protein has also shown importance in ESCC, NC, NSCLC and OC. Characteristic of many tumour suppressor genes including TP53, loss of heterozygosity of the DUSP6 gene locus has been frequently observed in PC^{50,60} to a lesser degree in NSCLC⁶¹ and is associated with reduced protein expression. Mutations in the DUSP6 gene have not been observed, however, a number of epigenetic studies have identified hypermethylation of CpG islands of intron 1 of the DUSP6 gene in PC, ESCC and NC and a potential role of histone deacetylation modification in the suppression of DUSP6 expression^{54,62}.

At transcriptional level, DUSP6 expression has been shown to be lower or lost in LC, ESCC and NC cell lines in comparison to normal epithelial cells^{61,62}. Converse to this, overexpression of DUSP6 in HG cells was observed and shown to result in reduced proliferation and an alteration in cell morphology whereby cells became more flattened in their appearance similar to features observed during cellular senescence⁶³. This may be indicative of a complex multi-faceted tumour suppressive function of DUSP6 in different disease sub-types.

At protein level, increased DUSP6 expression as measured by immunohistochemistry in primary pancreatic cell lines was significantly associated with mild-severe dysplasia and in-situ carcinoma whilst a decrease in DUSP6 expression was significantly associated with invasive carcinoma, particularly those with poorly differentiated histology⁶⁴. Similarities in the expression profile of DUSP6 has been observed in lung cancer (LC) whereby decreased DUSP6 expression with increasing growth activity and histological grade was identified⁶¹.

Further support for its tumour suppressive role is in the associations drawn between DUSP6 and the epithelial-mesenchymal transition (EMT) as identified by Wong et al⁶². In this study examining ESCC and NC, DUSP6 significantly impaired EMT-associated properties as measured

by a range of *in vitro* and in-vivo assays. DUSP6 expression was associated with polarization of spheroid formation resembling the epithelial polarity, a feature which is lost during EMT, suppression of cell migration and invasion and the upregulation of E-cadherin (epithelial marker⁶⁵) and downregulation of Vimentin (mesenchymal marker⁶⁶).

Despite the extensive investigation into the role of DUSP6 in the disease types mentioned above and the pertinence of MAPK/ERK pathway activation in colorectal pathogenesis, little has been explored in the context of CRC.

1.5 MAPK/ERK pathway: a target for the treatment of CRC

1.5.1 The MAPK/ERK as a therapeutic target

Evident from the type and prevalence of molecular aberrations involved in the malignant transformation of colorectal mucosa, the MAPK/ERK signalling cascade plays a pertinent role in the development of colorectal cancer. This makes members of the MAPK/ERK pathway an ideal target for treatment of the disease. A number of current therapies targeting the MAPK/ERK pathway with variable efficacy have been explored in a variety of malignancies, Fig. 12^{30,67}. Cetuximab (Erbix[®]) is an anti-epidermal growth factor receptor (EGFR) monoclonal antibody currently used for the first line treatment of metastatic colorectal cancer⁶⁸.

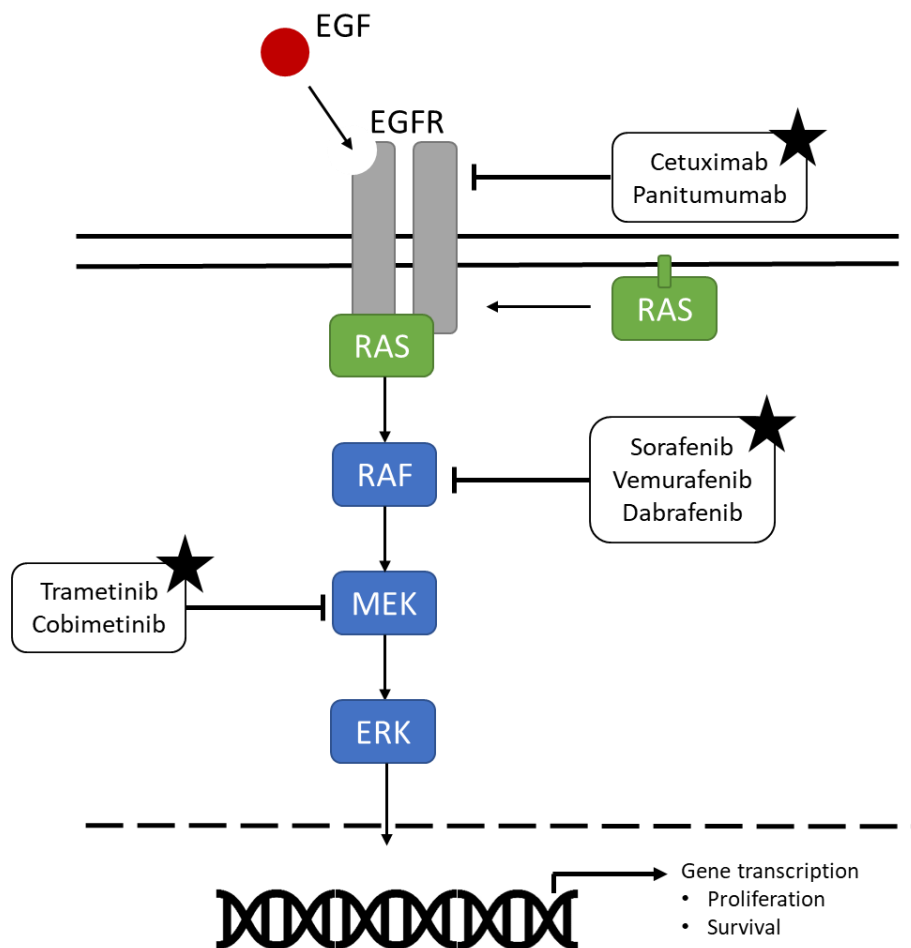


Figure 12. Summary of therapies targeting MAPK/ERK pathway. The MAPK/ERK signalling cascade is implicated in a variety of malignancies making it an ideal candidate for targeted therapies. A number of targeted treatments for different candidates in the MAPK/ERK pathway are currently used for the treatment of a variety of human cancers. Cetuximab (Erbix[®]) is a monoclonal antibody therapy which targets the epidermal growth factor receptor (EGFR). It is currently used for the first line treatment of advanced colorectal cancer.

1.5.2 Cetuximab (Erbix[®]) for the treatment of advanced colorectal cancer

1.5.2.1 Mechanism of action

A member of the tyrosine kinase inhibitor family of targeted therapies, Cetuximab competes with the endogenous EGFR ligand, epidermal growth factor (EGF) for the binding site of the EGFR. Cetuximab specifically binds to domain III of the EGFR and in doing so overlaps the binding site of EGF. In addition to competing for EGF binding site on the EGFR, when bound, Cetuximab prevents the receptor achieving the conformation required for dimerization with another EGFR or other HER family receptors thus preventing auto-phosphorylation and subsequent receptor activation⁶⁹, Fig. 13.

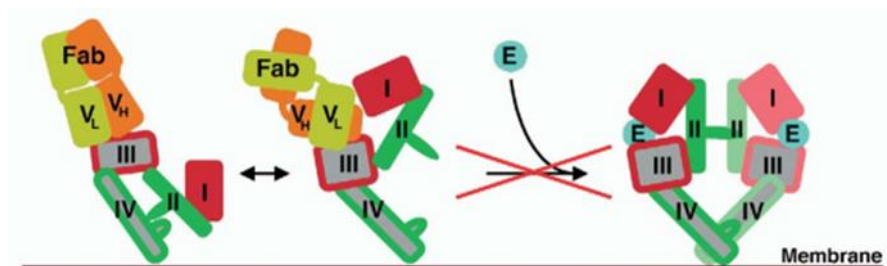


Figure 13. Epidermal growth factor receptor conformation upon Cetuximab binding. Li et al., 2005.

Cetuximab binding to the EGFR promotes receptor internalisation, thus reducing the number of receptors at the cell surface available for pathway activation⁷⁰. Cetuximab binding has been shown to reduce overall cellular proliferation. In A431 tumour xenografts in nude mice, Cetuximab was shown to inhibit tumour growth as determined by tumour measurements⁷¹. In a study by Huang et al, the effects of Cetuximab on proliferation were explored in squamous cell carcinoma cell lines from head and neck cancer patients. Treatment of cells with Cetuximab resulted in a reduction in growth in treated cells compared to controls. In these cells, treatment was shown to induce an accumulation of cells in G1 phase which was accompanied with a 2-3 fold reduction in the number of cells in S phase, this was also coupled with an increase in apoptosis as determined by flow cytometry which was previously observed in a study by Wu et al in DiFi colorectal cells⁷². Finally, the study showed that treatment of cells with Cetuximab results in increases in hyperphosphorylated retinoblastoma protein (Rb) and increases in p27KIP1 as determined by western blot also observed in a previous study by Peng et al in prostate cell lines⁷³. An increase in pro-apoptotic factors Bax and decreases in Bcl-2 expression support the flow cytometry findings for increased apoptosis⁷⁴.

1.5.2.2 Biomarkers of response

To identify individuals likely to benefit from Cetuximab treatment a range of preliminary studies have been carried out identifying hotspot mutations in exon 2 and 3 of *RAS*, a member of the MAPK/ERK signalling pathway as a negative predictor of Cetuximab response⁷⁵⁻⁷⁹. Further investigations into the relevance of extended *RAS* mutations demonstrated limited efficacy of Cetuximab in patients whose lesions comprised additional *RAS* mutations i.e. exon 4 (codon 117 and 146)⁸⁰⁻⁸⁵. With respect to the evidence presented, prescription of Cetuximab is dependent upon the clinical evaluation of *RAS* mutation status which should include assessment of hotspot locations for *KRAS* and *NRAS* in exon 2, 3 and 4⁶⁸.

1.5.2.3 The search for novel biomarker of response

Despite *RAS* qualifying as a clinical negative biomarker of response to Cetuximab treatment it remains that approx. 50% of individuals with *RAS* wildtype lesions are unresponsive to Cetuximab treatment⁸⁶. In light of the complexity of the MAPK/ERK signalling pathway and its fundamental role in regulating proliferation and cell-survival recent studies have attempted to identify additional surrogates of MAPK/ERK pathway dependence^{87,88}. Identification of surrogates of MAPK/ERK signalling pathway activation which represent neoplastic dependence upon ERK signalling cascade may prove more efficient in identifying responders to Cetuximab treatment. Despite investigations to identify predictors of Cetuximab response, *RAS* remains the only clinically useful biomarker of response.

Minimal investigation has been carried out into the utility of regulators of the MAPK/ERK pathway on response to Cetuximab treatment. Regulators such as DUSP6 could act as indicators of lesion dependency upon ERK and thus resistance to treatment. DUSP6 may prove more sensitive than *RAS* status alone in identifying these individuals.

1.5.3 DUSP6 and chemotherapeutic response

In addition to its role in the pathogenic processes of a variety of cancers, DUSP6 has an important role in determining sensitivity to chemotherapeutic and targeted treatments in a variety of cancer types. In NSCLC, decreased expression of DUSP6 *in vitro* promoted resistance to the ALK inhibitor, Crizotinib⁸⁹. Knockdown of DUSP6 in ovarian cancer cells increased resistance to Cisplatin treatment. In contrast to this, over-expression of DUSP6 in DUSP6 deficient ovarian cancer cells sensitised cells to Cisplatin treatment⁹⁰. Similarly, in human glioblastoma cells, DUSP6 over-expression also sensitised cells to Cisplatin treatment *in vitro* and *in vivo* whilst depletion increased resistance⁶³.

Two studies have examined the association of DUSP6 expression and Cetuximab response albeit to a limited degree. Oliveras-Ferraro et al⁹¹ investigated the cellular requirements for Cetuximab efficacy in *KRAS* wildtype A431 epidermoid cancer cells. Findings from bioinformatic gene expression analysis of KEGG pathway databases were suggestive that among a number of factors, Cetuximab function is dependent on an absence of negative feedback i.e. negative regulation by DUSP6, on ERK activation. A second study by Boeckx et al⁹² in head and neck squamous cell carcinoma (HNSCC) explored DUSP5 and 6 and Cetuximab resistance. Cetuximab resistant cell lines treated with Cetuximab demonstrated downregulation of DUSP6 gene expression and concluded this may be due to activated MAPK signalling. The question remains as to the impact of DUSP6 expression on Cetuximab treatment response in the context of colorectal cancer.

1.6 Investigation aims and hypothesis

It is evident from the literature presented above that DUSP6 plays a role in the pathogenesis of many cancer types. Despite this breadth of knowledge, the role of DUSP6 in CRC is yet to be elucidated.

The aims of this study are split into two arms of investigation and are explored as detailed:

1. DUSP6 and colorectal pathogenesis

Does DUSP6 protein expression change across stages of colorectal pathogenesis?

Evidenced in literature, it can be inferred that progression through the adenoma-carcinoma sequence relies, at least partly upon increased MAPK/ERK pathway activation and thus with progression through this sequence increased DUSP6 expression is required to negatively regulate ERK in response. Therefore, it is hypothesised that DUSP6 protein expression will increase with increasing progression through the adenoma-carcinoma sequence.

This hypothesis is explored through immunohistochemical analysis of DUSP6 protein expression across a cohort representing both histological and molecular features of colorectal pathogenesis. The functional relationship between DUSP6 and ERK in the context of colorectal tissue is explored *in vitro* by the transduction of DUSP6 in adenoma and adenocarcinoma cell lines.

2. DUSP6 and treatment response in colorectal adenocarcinoma

Is DUSP6 protein expression a surrogate of MAPK pathway dependence i.e. *RAS* mutation; and what impact does its expression have on the response of colorectal adenocarcinoma to treatment with Cetuximab?

It has been suggested that Cetuximab function is dependent upon a non-aberrant MAPK pathway. It is hypothesised that in colorectal adenocarcinoma, high DUSP6 expression will be a surrogate of MAPK pathway dependence and thus of Cetuximab resistance.

This hypothesis is explored firstly through the assessment of DUSP6 protein expression in a cohort of primary colorectal adenocarcinoma tissues by immunofluorescence. Secondly, by Cetuximab treatment of DUSP6 transduced Cetuximab sensitive colorectal adenocarcinoma cells and finally the utility of DUSP6 as a prognostic predictor is explored by association analysis of protein expression with 5 year survival outcomes.

2. Materials and Methods

2.1 Statistical analysis

All statistical analysis was carried out using SPSS v24 (IBM).

Normality tests were carried out prior to further statistical analysis. Non-parametric analyses was carried out for non-normally distributed data. Bonferroni corrections to alpha values were used for multiple comparisons. Monte-Carlo test of significance was used where applicable. In the instance of limited sample size Fisher's exact significance was applied.

Effect size for Chi² test for independence was carried out using Phi coefficient for the cases of 2x2 contingency tables and Cramer's coefficient for greater than 2x2. Effect size for Kruskal-Wallis, Mann-Whitney and Wilcoxon-signed rank testing was carried out using eta-squared (η^2) and interpreted using:

Lenhard, W. & Lenhard, A. (2016). *Calculation of Effect Sizes*. Retrieved from: https://www.psychometrica.de/effect_size.html. Dettelbach (Germany): Psychometrica. DOI: 10.13140/RG.2.1.3478.4245

The η^2 were interpreted as follows: 0.01 = small effect, 0.09 = medium effect and 0.25 = large effect. When transformed these value correspond with Cohens d effect thresholds for small, medium and large effect.

2.2 Cohort identification

Two cohorts both comprised of formalin fixed paraffin embedded (FFPE) material were identified from cases tested within molecular pathology at the Royal Infirmary of Edinburgh.

Cohort 1: This cohort comprises 147 FFPE colorectal specimens ranging from normal mucosa, adenoma to adenocarcinoma. The cohort is sub-divided into 8 discrete groups based upon histological and genotypic characterisation: normal mucosa, adenoma: tubular, tubulovillous, villous, serrated, adenocarcinoma: deficient mis-match repair (dMMR), *RAS* mutant and *TP53* mutant.

Cohort 2: A large cohort of 525 FFPE tissues from primary colorectal adenocarcinoma cases. These cases comprised sequential retrospective tissues from years 2012 and 2013 (to enable the

collation of 5 year survival data by the end of this study) whom had been previously tested within Molecular Pathology for molecular profiling for either *RAS* or *BRAF* somatic mutation assesment.

Eight hundred and five sequentially tested cases were returned. Duplicated samples, referred cases (due to an inability to obtain tissue), specimens not comprising resections (due to tissue availability) or non-primary adenocarcinoma specimens were removed resulting in a final study cohort of 525 cases.

Pathological dataset parameters were obtained for each case of both cohorts by manual assessment of parameter reported by pathologists in pathology reports.

The work required for molecular characterisation of both cohorts (tissue preparation to sequencing) was carried out by myself excluding a number of cases in cohort 2 (highlighted in red in appendix 8.2.1) which was carried out in molecular pathology at the Royal Infirmary of Edinburgh as part of routine clinical assessment.

2.3 Tissue preparation

2.3.1 Tissue assessment

Prior to cutting sections for DNA extraction, haematoxylin and eosin (H&E) sections of each sample were inspected microscopically. For each case, malignant cells suitable for genotyping were identified. Slides were marked with a fine pen to highlight tissue region suitable for macro-dissection in order to maximise the ratio of neoplastic to normal cells within the region to be tested. If a high tumour percentage throughout the section was observed, macro-dissection was not deemed necessary. The number of tissue sections cut for molecular analysis was determined by size and cellularity of the tissue. Highly cellular samples required only 1 x 10 μ m; with decreasing cellularity, the number of sections required increased.

2.3.2 PCR Microtomy

Ten micrometre sections were cut from formalin fixed paraffin embedded (FFPE) sections using a dedicated PCR microtome. The following precautions were observed to prevent contamination:

- Microtome cleaned with alcohol wipes thoroughly before and after use.
- Disposable nitrile gloves were worn when handling sections.
- A fresh portion of the microtome blade was used for each new block.

- Sections were handled using non-serrated micro-dissecting forceps which were cleaned thoroughly between blocks with alcohol wipes.
- Water bath was filled with fresh distilled water.

Clean nuclease-free microfuge tubes were used to hold curled sections, depending upon cellularity 1-6 sections were used for samples requiring downstream macro-dissection.

2.4 FFPE DNA extraction

DNA was extracted from formalin fixed paraffin embedded (FFPE) archival specimens using QIAmp DNA FFPE Tissue Kit (Qiagen #56404) or the RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE (ThermoFisher Scientific). Protocol was carried out as per the manufacturer's instructions. Paraffin removal by adding 1ml xylene to each sample, heating for 1min at 60°C followed by 2 washes with 1ml 100% EtOH was not used in this protocol as sufficient tissue digestion was achieved with protease digestion alone as evidenced by sufficient DNA concentrations.

2.5 Sequencing

Cohort 1 and 2 underwent molecular classification using pyrosequencing. In order to confirm the sensitivity and specificity of pyrosequencing as a method for genotype determination a subset of cases from cohort 2 were sequenced using next generation sequencing confirming concordance with pyrosequencing results and supported the use of this as a method for molecular classification of both cohorts in this study⁹³.

2.5.1 Pyrosequencing

Cohorts 1 and 2 were sequenced for the presence of variant alleles in hotspot locations using pyrosequencing.

2.5.2 Template preparation

Pyrosequencing of DNA from FFPE archival specimens was carried out on the Qiagen PyroMark Q96 and Q24 Instruments. Hotspots within *KRAS* exons 2, 3 and 4 were assessed using 3 separate PCR reactions followed by 4 separate sequencing reactions. Hotspots within *NRAS* exons 2 and 3 were assessed using 2 separate PCR reactions followed by 2 separate sequencing reactions. Hotspots within *BRAF* exon 15 are assessed using 1 PCR reaction followed by 1 sequencing reaction. Primers used for PCR reactions contain a biotin tag enabling immobilization of amplicons on Streptavidin Sepharose High Performance beads. Sequencing primers anneal with single-stranded DNA after-which sequencing ensues.

All PCRs were setup using the following protocol detailed in Table 2 with the inclusion of a negative control (H₂O) to identify false positives and all *RAS* wildtype control to identify interferences with the pyrosequencing run. Thermal cycler program detailed in Table 3.

Reagent	Volume/Reaction (μl)
2X HotStarTaq <i>Plus</i> Master Mix	12.5
CoralLoad Concentrate	2.5
Forward Primer	1
Reverse Primer	1
Water	3-7
DNA	1-5
Total	25

Table 2. Volumes for PCR set up.

Temperature (°C)	Time
95	5 minutes
Followed by 35 cycles of:	
94	30 seconds
58	30 seconds
72	30 seconds
Followed by:	
72	10 minutes

Table 3. Thermal cycler program for target amplification.

Primer sequences for gene specific regions sequenced are detailed in Table 4.

Primer name	Exon	Function	Sequence (5' to 3')
KRAS pyro 1213 F	2	PCR Forward	GGCCTGCTGAAAATGACTG
KRAS pyro 1213 R	2	PCR Reverse	Biotin-GCTGTATCGTCAAGGCACTCT
KRAS pyro 1213 Seq	2	Sequencing	TTGTGGTAGTTGGAGCT
KRAS pyro 61 F1	3	PCR Forward	Biotin-TGGAGAAACCTGTCTCTTGGATAT
KRAS pyro 61 R1	3	PCR Reverse	CTGGTCCCTCATTGCACTGTACTC
KRAS pyro 61 Seq	3	Sequencing	CCTCATTGCACTGTACTC
KRAS exon 4 F	4	PCR Forward	GGACTCTGAAGATGTACCTATGG
KRAS exon 4 R	4	PCR Reverse	TCAGTGTTACTTACCTGTCTTGT
KRAS 117 Seq	4	Sequencing	ACCTATGGTCCTAGTAGGAA
KRAS 146 Seq	4	Sequencing	AATTCCTTTTATTGAAACAT
NRAS pyro 1213 F	2	PCR Forward	CTTGCTGGTGTGAAATGACTG
NRAS pyro 1213 R	2	PCR Reverse	Biotin-TTCTGGATTAGCTGGATTGTCAAT
NRAS pyro 1213 Seq	2	Sequencing	GTGGTGGTTGGAGCA
NRAS pyro 61 F1	3	PCR Forward	Biotin-ACACCCCAGGATTCTTACAGA
NRAS pyro 61 R1	3	PCR Reverse	GCCTGTCCTCATGTATTGGTC
NRAS pyro 61 Seq	3	Sequencing	CATGGCACTGTACTCTTC
BRAF ex15 F	15	PCR Forward	AGGTGATTTTGGTCTAGCTACA
BRAF ex15 R	15	PCR Reverse	Biotin-AATCAGTGAAAAATAGCCTCAAT
BRAF ex15 Seq	15	Sequencing	GTGATTTTGGTCTAGCTAC

Table 4. Primer sequences. Primer sequences for both PCR and pyrosequencing reactions detailed. All primers were designed in-house within molecular pathology at the Royal Infirmary of Edinburgh.

Initial PCR setup was carried out with 1µl addition of DNA. If sequencing failed the PCR was repeated with 5µl addition of DNA. Should a second sequencing failure occur the specific amplicon region was considered a 'FAIL'.

2.5.3 Pyrosequencing and analysis setup

Two machines were used to carry out pyrosequencing analysis; PyroMark Q24 and Q96. Both machines require dispensation sequences to be determined prior to sequencing using IUPAC codes to ensure the detection of potential variants of interest. Dispensation orders and sequence to analyse are detailed in Table 5.

Gene	Codon	Position (base)	Genomic sequence	Dispensation order	Sequence to analyse
KRAS	12	34-36	GGTGGCGTA	TACGACTCAGATCGT AG	GNTGGCGTAGGC
	13	37-39			GGTGRCGTAGGC
	61	180-183	GGTCAAGAG	GCTCAGTCAGACT	CTCDTGACCTG
	117	351	ATAAATGT	CAGTACTGTG	ATAAHTGTG
		349			ATNAATGTG
	146	436	CAGCAAAGA C	GATCAGCTGAG	CAVCAAAGAC
		437			CAGBAAAGAC
		438			CAGCNAAGAC
	147	441			CAGCAAASAC
	NRAS	12	34-35	GGTGGTGT GGGAAAAGC	TACGACTCAGCATCG TAGAG
13		38	GGTGNTGTGGGAAAA GC		
59		175-176	TTGTCCAGCT GTAT	CAGTACGTCTATGTCT AGTA	TNGTCCAGCTGTAT
61		181-183			TTGTCCANCTGTAT
BRAF	600	1798-1800	AGTGAAATCT	GCACGTAACGTATCT	AGT/AGAAATCTG

Table 5. Pyrosequencing dispensation orders and sequence to analyse.

2.5.4 Pyrosequencing analysis

2.5.4.1 Assay limits: limits of blank, limits of detection

Due to the clinical application of the method, limit of blank and limit of detection studies had been carried out prior to use by a member of Molecular Pathology at the Royal Infirmary of Edinburgh and are detailed in Table 6.

Amplicon target	LOB	LOD
KRAS codon 12	2.05	>4
KRAS codon 13	1.74	
KRAS codon 61	3.18	>6
KRAS codon 117	7	>5
KRAS codon 146	9	>8
NRAS codon 12	4	>5
NRAS codon 13	4	
NRAS codon 59	6	>8
NRAS codon 61	4	

Table 6. Limits of blank and limits of detection for pyrosequencing.

Any variant alleles identified below the LODs for each assay were considered no mutation detected. Interferences from baseline blips on any of these pyrograms resulted in the sample be repeat tested to determine presence of variant allele.

2.5.4.2 Sequencing analysis of samples tested on PyroMark Q96

Samples tested using the PyroMark Q96 required manual analysis of pyrosequencing results as software for automated variant allele detection was unavailable. Following pyrosequencing of samples, peak heights were exported in addition to PDF files of pyrogram traces for visual interpretation. Expected peak heights for respective gene and codon to be analysed were assessed for the presence of variant allele based upon percentage changes in peak heights. The average single peak height for non-changing single peaks was calculated for each sample. If single peak height was higher than 20 (minimum peak height as per instrument guidelines) analysis proceeded. Peak heights per sample were transformed into ratios of the average peak

height. Variants were identified by percentage increases or decreases (within the limits of detection) from the expected peak height at the dispensation order in question.

Average peak heights below 20 were considered 'Failed' as per instrument guidelines. Due to the mutually exclusive nature of KRAS, NRAS and BRAF mutations; should an amplicon fail but test mutant for successful amplicons in other codons the failed amplicon was considered wildtype. If successful amplicons tested wildtype in addition to the failed amplicon, the amplicon was considered a 'Fail'.

2.6 Protein expression in formalin fixed paraffin embedded tissue

For immunohistochemical and immunofluorescence experiments, samples were prepared from FFPE blocks cut to 3µm onto Surgipath® Apex glass slides (Leica 3800080E) using microtomy protocol previously described.

2.6.1 Details of primary and secondary antibodies

Primary antibody details are summarised in Table 7. Secondary antibody details are summarised in Table 8.

Target	Cat number	Species	IHC Dilution	IF Dilution	IF visualisation	Incubation conditions
Dual-specificity phosphatase 6 (DUSP6)	NovusBio H00001848-M01	Mouse	1:500	1:1000	TSA FITC	30min room temp
Pan Cytokeratin	Dako Z0622	Rabbit	-	1:150	Streptavidin Alexa Fluor™ 555 conjugate	
Phospho-p44/42 MAPK (ERK1/2) (Thr202/Tyr204)	Cell Signal 9101	Rabbit	1:25 – 1:500*	-	-	4°C overnight OR 15mins at room temp
p44/42 MAPK (ERK1/2) (Thr202/Tyr204)	Cell Signal 9107	Mouse	1:400*	-	-	
P53 (DO-7 clone)	Dako M7001	Mouse	1:100	-	-	30min room temp

Table 7. Details of primary antibodies used. *Automated staining dilution used.

Fluorophore	Supplier	Cat number	Dilution	Incubation
TSA Fluorescein	Perkin Elmer	NEL701A001KT	1:50 – 1:100	10mins
Streptavidin, Alexa Fluor™ 750 conjugate	ThermoFisher	S21384	1:100	30mins
Streptavidin Alexa Fluor™ 555 conjugate		S32355		

Table 8. Details of secondary antibodies used.

2.6.2 Immunohistochemistry

*Note – PBS-T used at 0.1% for all immuno applications described below.

Sections were dewaxed in xylene (3 x 5min) and rehydrate for 2min each: 100% EtOH, 100% EtOH, 80% EtOH, 50% EtOH, distilled water. Sodium citrate solution (pH 6.0) was utilised for all antigen retrieval. Slides were cooled in running water for 20min and placed in PBS-T.

Three percent H₂O₂ peroxide block was applied for 5min and sections washed with PBS-T for 5min. Protein block (X090930-2, Agilent) was applied for 10min. Primary antibody was applied and incubated for designated time (Table. 7), following which sections were washed with PBS-T for 5min. Conjugated secondary anti-[primary antibody species] HRP (Rabbit K400311-2, Mouse K400111-2, Agilent) incubate for 30min, wash sections with PBS-T for 5min. Apply DAB chromogen for 10min and rinse slides with PBS-T. Apply Haematoxylin for 10secs, rinse in water for 10sec, wash in water for 2min and wash in Scotts Tap Water for 1min.

Dehydrate sections for 2min each: 50% EtOH, 80% EtOH, 100% EtOH, 100% EtOH and clear in xylene (3 x 5min). Mount sections using DPX (CellPath Cat:SEA-1304-00A).

Automated immunohistochemistry was carried out for pERK and ERK markers using the Leica BOND RX instrument (Leica Biosystems) using the above protocol – antigen retrieval is automated on the BOND RX.

2.6.3 Multiplex Immunofluorescence – three targets

Section dewaxing, rehydration, antigen retrieval and endogenous blocking was carried out as previously described (Immunohistochemistry).

Following endogenous blocking the 1st primary antibody was applied for designated time (antibody table), sections washed with PBS-T. A conjugated secondary anti-[primary antibody species] HRP was applied and incubated for 30min, sections washed with PBS-T. The 1st primary antibody visualisation was applied and incubated for 10min, sections washed with PBS-T.

Microwave stripping using sodium citrate heated in pressure cooker for 12min was carried out, sections placed in solution and heated with 'auto defrost function – meat 850g'. Sections washed in PBS-T following this.

2nd endogenous blocking was carried out as previously described, sections washed in PBS-T. The 2nd primary antibody was applied for designated time, sections washed with PBS-T. A conjugated secondary anti-[2nd primary antibody species] HRP was applied and incubated for 30min, sections washed with PBS-T. The 2nd primary antibody visualisation was applied and incubated for 10min, sections washed with PBS-T.

Microwave stripping using sodium citrate heated in pressure cooker for 12min was carried out, sections placed in solution and heated with 'auto defrost function – meat 850g'. Sections washed in PBS-T following this.

3rd endogenous blocking was carried out as previously described, sections washed in PBS-T. The 3rd primary antibody was applied plus appropriate species pan cytokeratin together for designated incubation time, sections washed in PBS-T for 5min. For 3rd primary antibody visualisation, antibody was diluted with antibody diluent and biotinylated secondary antibody for pan-cytokeratin visualisation and incubated for designated time. Sections washed in PBS-T for 5min. Streptavidin conjugated alexa fluor for pan-cytokeratin visualisation was applied and incubated for 30min, slides washed in PBS-T for 5min.

Sections incubated in Hoechst 33342 (1:20) for 10min. Slides washed in PBS-T 2 x 5min. Slides dehydrated in 80% EtOH for 1min and air dried. Mounted using Prolong anti-fade mounting medium (without DAPI).

Automated multiplex immunofluorescence was carried out for DUSP6 using the Dako Autostainer Link 48 (Agilent) using the above protocol – cases were removed between primary antibody runs in order to manually microwave strip.

2.7 Digital image analysis

2.7.1 Leica SCN 400

The Leica SCN400 was used to obtain whole section images for H&E and IHC slides. Overview images were obtained at x5 magnification followed by x20 magnification. Scanner operated as per SOP SaSoM-EQUIP-101 (Appendix 8.6).

2.7.2 Automated semiQUantitative Analysis (AQUA) analysis

AQUA⁹⁶ was used to semi-quantify (as this was not a direct quantification from protein, a tyramide signal amplification secondary antibody was used) protein expression by immunofluorescence for EREG and DUSP6. AQUA instrument was used as per SOP SaSoM-EQUIP-025 and SaSoM-EQUIP-026 (Appendix 8.6).

AQUA software enables the semi-quantitation of protein expression from digital images within specific sub-cellular compartments. The derived scores incorporate the sum of the target compartment pixel intensities as an expression of the area of the defined compartment of interest. For whole slide images between 10 and 30 fields of view (FOV) were selected depending upon tissue size, each FOV was treated as a pseudo-tumour microarray (TMA) core for each case, Fig. 14. AQUA experiments were setup to generate AQUA scores for cytoplasmic EREG (used as a surrogate for membrane expression), cytoplasmic and nuclear for DUSP6, Fig. 15.

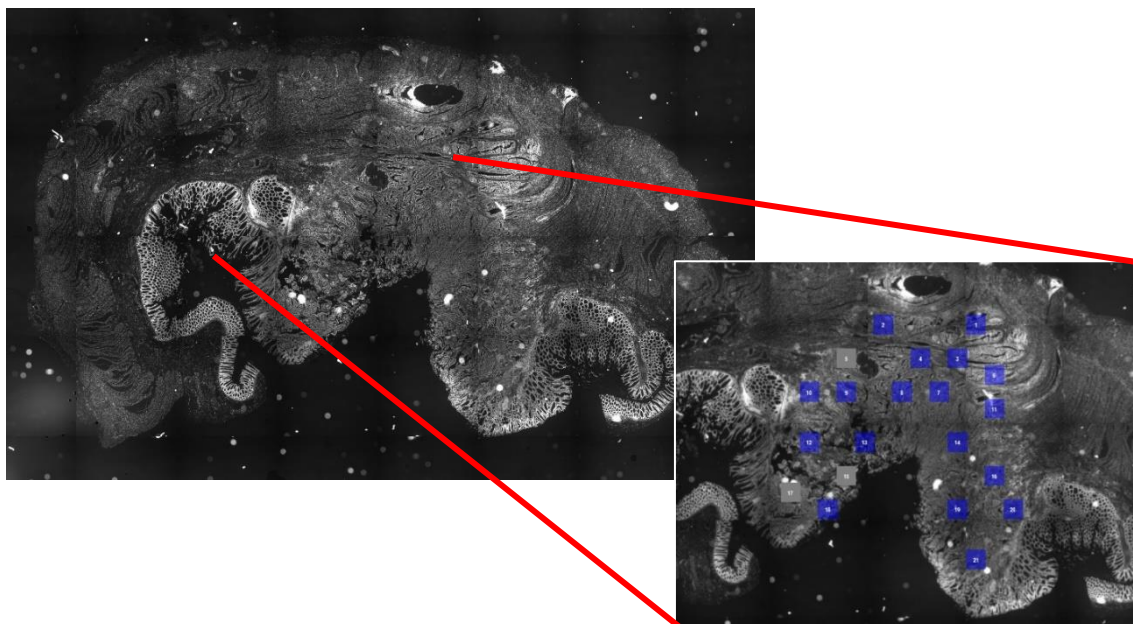


Figure 14. Hoechst staining overview at 5x magnification. 10 to 30 fields of view (FOV) were selected to represent adenocarcinoma for each sample. Case ID 3.

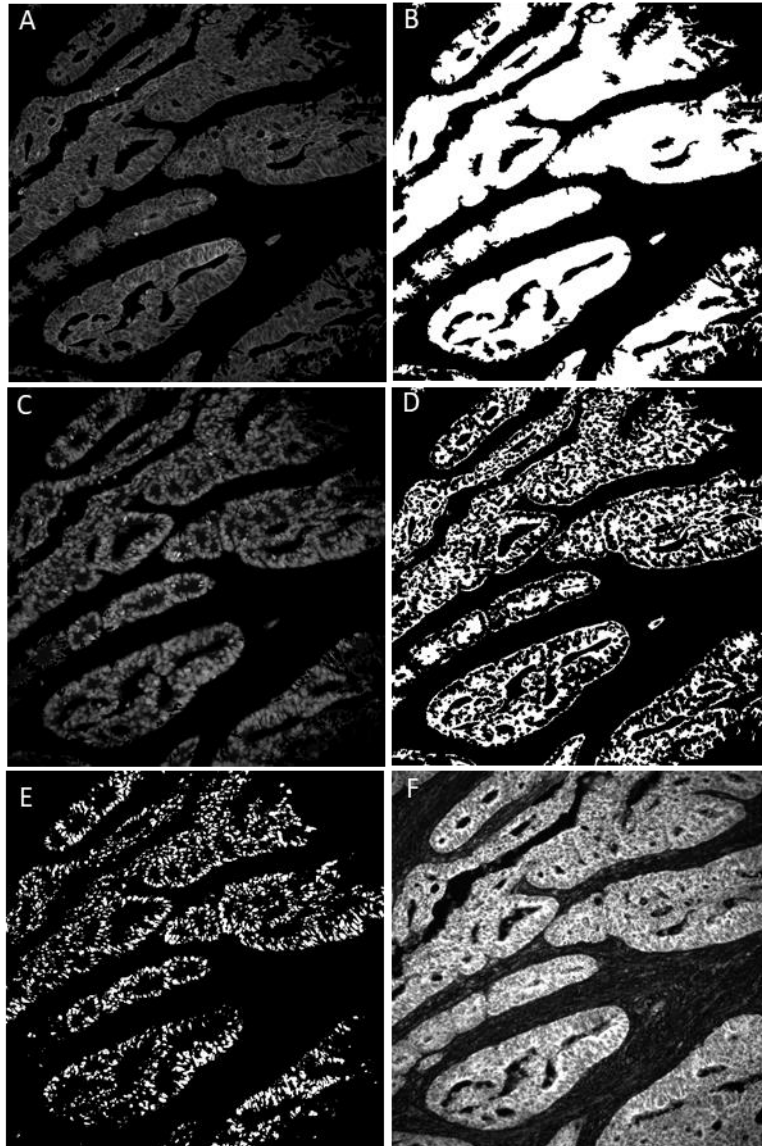


Figure 15. AQUA experiment steps for DUSP6 semi-quantitation. A: Pan-Cytokeratin expression (PanCK). PanCK is used to identify tumour epithelium. B: Tumour mask. Tumour mask is derived from PanCK expression. C: Hoechst expression. Hoechst identifies nuclei. D: PanCK and Hoechst expression are used to denote cytoplasmic (D) and nuclear (E) compartments within tumour mask (B). DUSP6 (F) pixel intensity is then quantified within denoted areas, nuclear and cytoplasmic). Sum of pixel intensity is then expressed as a proportion of compartment area to give final AQUA score.

2.7.3 QuPath for digital image analysis

QuPath⁹⁷ is an open source digital image analysis software which enables automated feature classification and marker quantification on whole slide immunohistochemically stained FFPE sections. Immunohistochemically stained FFPE sections for DUSP6, ERK and p-ERK were converted to digital images by the Leica SCN 400 scanner.

Intensity thresholds to determine phenotype classification for each marker were generated. A subset of cases were initially phenotypically assessed by eye as being low (1+), moderate (2+) or high (3+) for DAB staining, Fig. 16.

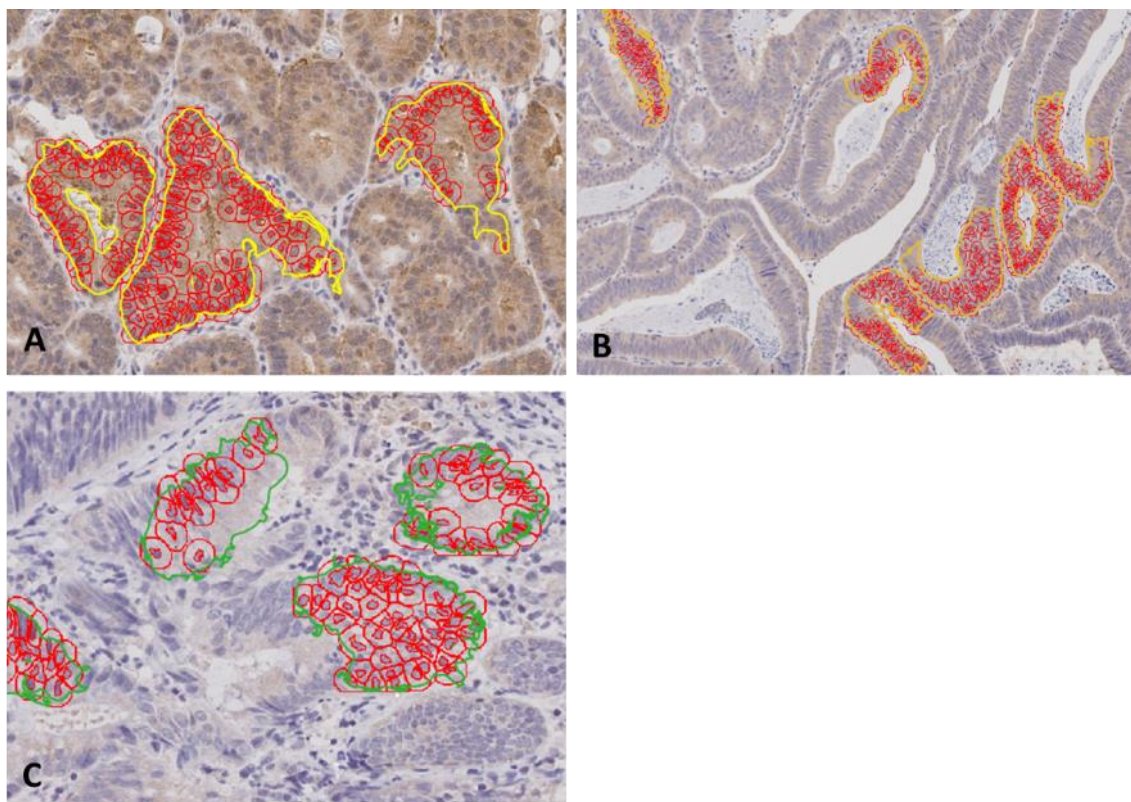


Figure 16. DUSP6 immunohistochemistry. In order to determine DAB intensity thresholds for DUSP6 classification a subset of cases were phenotypically assessed by eye as being low (1+), moderate (2+) and high (3+) for DUS6 expression. A: 3+ phenotype. B: 2+ phenotype. C: 1+ phenotype.

Image analysis was carried out on these cases to first segment cells into nuclear and cytoplasmic compartments and second to calculate the mean DAB intensity threshold for each compartment. A mean and max DAB value was collated for 400+ cells per phenotype i.e. 1+, 2+ and 3+.

DAB intensity thresholds for each phenotype (1+, 2+ and 3+) were determined by identifying a threshold between the maximum value of the intensity below and the mean DAB intensity of the phenotype for which a threshold is being derived. The determined threshold did no overlap with the maximum value for the phenotype below. For example, to determine 2+ phenotype DAB intensity threshold:

0.16 is the maximum value of 1+ phenotype, 0.165 is the mean DAB intensity value of 2+ phenotype therefore 0.165 was rounded to 1d.p. and 0.17 was derived as the threshold for 2+ phenotype.

Automated classification profiles were used to identify cells of interest i.e. dysplastic (adenomatous), adenocarcinoma or normal and previously detailed DAB thresholds were applied to identify the ratio of phenotypes within annotated region, Fig. 17. Annotation metrics were subsequently collated including total number of cells assessed and total number of cells classified under each phenotype.

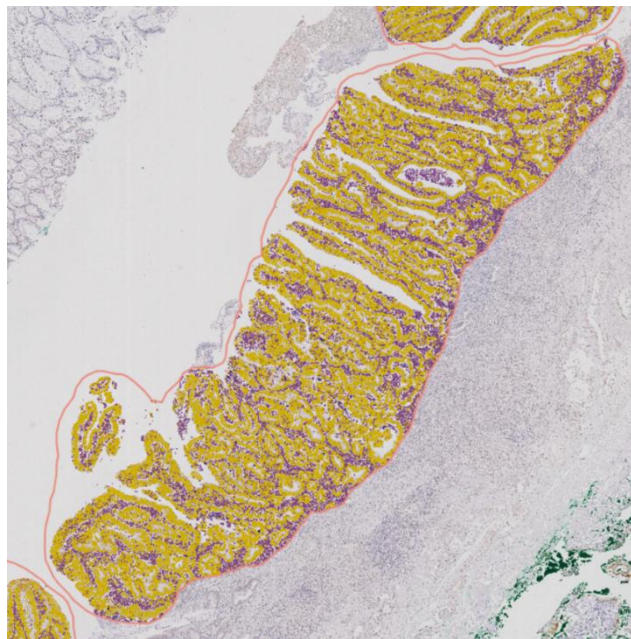


Figure 17. Cell classifier with positivity threshold applied. Regions of interest were annotated and classifiers applied to identify positive cells based upon pre-define thresholds. Yellow: positive cells.

The number of positive cells for each marker was expressed as a ratio of the total number of cells assessed. Subsequent statistical analysis was carried out using these values. In cases where direct comparison between the ratio of positive cells was required for example generating the

ratio of p-ERK positive cells : DUSP6 nuclear positive cells; ratio of positive cells for each marker was expressed as a proportion of 1000 and subsequent ratios derived.

2.8 Cell culture

2.8.1 Details of cell lines

Two cell lines were used during this study; 1. C99 derived from colorectal adenocarcinoma classified as TNM stage III 2. RG/C2/80 derived from tubular colorectal adenoma (Gifted from Prof. Williams at Bristol University). Culturing conditions are detailed in Table 12.

Cell line	Culture media
C99	IMDM, 10% FBS, PenStrep
RG/C2/80	DMEM (-) glutamine, 20% FBS, PenStrep, 2mM Glutamine, Hydrocortisone (1µg/mL), Insulin (0.2units/mL).

Table 9. Cell line details. C99 cells were derived from colorectal adenocarcinoma classified at TNM stage III. RG//C2/80 cells were derived from a tubular colorectal adenoma (Gfted from Prof. Williams at Bristol University) DMEM: Dulbecco's Modified Eagle Medium. IMDM: Iscove's Modified Dulbecco's Medium. FBS: Foetal bovine serum. PenStrep: Penicillin/Streptomycin.

2.8.2 Cell passaging

An aliquot of 0.05% Trypsin-EDTA was thawed and warmed to 37 °C. In the fume hood, media was removed from cells and washed in warmed sterile PBS. PBS was poured off and trypsin added to the flask/petri dish containing the attached cells (the volume required will depend on the size of flask/dish being used). The flask/dish was returned to the CO₂ incubator for 5-10 minutes. Flasks were checked under the microscope to ensure sufficient detachment of cells. Media was addedd to flask/dish to inactivate trypsin. Cell containing suspension was transferred to a labelled universal and centrifuged at 1200rpm for 5mins. Following centrifugation a pellet formed in the bottom of the centrifuge tube, the solution poured off and pellet resuspended using sterile syringe.

2.8.3 Protein extraction from mammalian cells

Prior to making working lysis buffer, one Complete Protease Inhibitor Tablet (Roche, 11836153001) was diluted in 1mL of milli-Q water, dispensed into 100µl aliquots and frozen. To make working cell lysis buffer the following were added to a 5ml tube: 150µl of 10X cell lysis buffer (Cell Signal, 9803), 5µl of Phosphatase Inhibitor Cocktail 2 (Sigma, P5726 1ml), 5µl of Phosphatase Inhibitor Cocktail 3 (Sigma, P0044 1ml), 10µl of Aprotinin (A6279), 1 aliquot of complete protease inhibitor

All samples and reagents were kept on ice throughout the procedure and centrifuge was chilled to 4°C prior to commencing procedure. Media was removed from dishes containing cells to be lysed and rinsed twice in cold 0.1% PBS. All residual liquid was removed with a pastette. 200µl of cell lysis buffer was added to each plate and using a cell lifter, all cells scraped to one side of the dish. Dish was placed on an angle in ice and left for 10mins, cell suspension collected at edge of dish. After 10mins, cell suspension was transferred to chilled centrifuge tube and spun at maximum speed for 20mins. Supernatant was removed and transferred to a clean labelled chilled centrifuge tube for storage of quantitation by BCA.

2.8.4 Protein determination by Bicinchoninic Acid (BCA) assay

Using protein lysates the protein content was determined by BCA assay. Prior to commencing protocol water bath was set to 60°C. Eight tubes were labelled A-H. Standards were made in 75mm borosilicate glass tubes (Fisher; Cat 14-961-26) from 1mg/ml protein standard in concentrations comprising; 0, 50, 75, 100, 250, 500, 750 and 1000µg/ml. 45µl of distilled water was added to a labelled sample tube and 5µl of sample added to this. A BCA solution was made by mixing 1 volume (e.g. 1ml) of Copper Sulphate solution to 50 volumes (e.g. 50ml) of Bicinchoninic Acid solution and mixed thoroughly. 1ml of this solution was added to all tubes and an additional 2ml to tube A, which was also used as a blank as well as the lowest standard. All tubes were vortexed and incubated in the pre-heated water bath for 15mins at 60°C following which samples were removed and allowed to cool. In a flat bottom 96-well plate, 200µl of standard 'A' was aliquoted into each well of the column 1 (wells A-H). 200µl of each of the following were also aliquoted: standard 'A' to wells A2-B2, standard 'B' to wells C2-D2 standard 'C' to wells E2-F2, standard 'D' to wells G2-H2, standard 'E' to wells A3-B3, standard 'F' to wells C3-D3, standard 'G' to wells E3-F3, standard 'H' to wells G3-H3. 200µl of sample 1 was added to wells A4-B4, followed by duplicates of remaining samples. The plate was read at 540nm on a microplate reader. Standard curve was plotted and data extrapolated.

2.8.5 Western blotting for protein expression analysis

2.8.5.1 Western blotting primary and secondary antibody dilutions

Details of dilutions of primary and secondary antibodies used, Table 13.

Primary antibody	Cat number	Species	Western blot dilution
Dual-specificity phosphatase 6 (DUSP6)	NovusBio H00001848- M01	Mouse	1:10,000
Phospho- p44/42 MAPK (ERK1/2) (Thr202/Tyr204)	Cell Signal 9101	Rabbit	1:25 – 1:500*
p44/42 MAPK (ERK1/2) (Thr202/Tyr204)	Cell Signal 9107	Mouse	1:400*
β-Actin (8H10D10)	Cell Signal 3700	Mouse	1:10,000
Pan-Actin	Cell Signal 4968	Rabbit	1:10,000
Secondary antibody	Cat number	Species	Western blot dilution
IRDye® 800CW Goat anti-Mouse	926-32210	Goat	1:10,000
IRDye® 680RD Donkey anti-Mouse	926-68072	Donkey	
IRDye® 800CW Donkey anti-Rabbit	926-32213	Donkey	
IRDye® 680RD Goat anti-Rabbit	926-68071	Goat	

Table 10. Summary of primary and secondary antibodies used for Western blotting.

2.8.5.2 SDS-PAGE and protein transfer to membrane

For a 10% resolving gel and 3.6% stacking gel use the following reagents, Table 14 (do not add TEMED or AMPS until ready to pour the gel):

Reagent	Resolving gel (mL)	Stacking gel (mL)
Acrylamide	13.5	3.6
Tris	15*	10**
10% SDS	0.4	0.3
dH ₂ O	11.1	16
TEMED	0.1	0.1
AMPS	0.1	0.1

Table 11. Reagents for resolving and stacking gel preparation. *Use IM Tris pH8.85. **Use 0.375M Tris pH6.8. N.B. These volumes will make 4 mini-gels or 2 large gels.

Prepare running and transfer buffer as follows, Table 15:

Reagent	Running buffer	Transfer buffer
Tris Base	3.03g	3.03g
Glycine	14.42g	14.42g
10% SDS	10mL	-
dH ₂ O	Up to 1L	Up to 1L

Table 12. Reagents for running and transfer buffers.

The gel apparatus was assembled as per SOP SaSoM/EQUIP/034 (Appendix 8.6). Prior to pouring resolving gel, TEMED and AMPS was added and tub inverted gently to mix. Gel was poured between glass plates and Isopropanol (propan-2-ol) added to level top of gel. Once set, isopropanol was poured off and cleaned with dH₂O. Excess dH₂O was drained off with blotting paper. TEMED and AMPS was added to stacking gel tube and pouring procedure repeated as detailed previously. Comb was inserted and bubbles removed. Once gel was set, gel plates were removed from casting rig and running tank set up. Running buffer was added in between the two plates. Well comb was removed and prepared samples loaded. For 1 gel the running time was 35mins at 35mA then 60mA for 3hrs (or until proteins have reached required distance from

origin). For 2 gels the running time was 30mins at 70mA then 120mA for 3hrs (or until proteins have reached required distance from origin).

One membrane per gel was prepared for transfer, PVDF membrane was cut to appropriate size and activated in 100% methanol for 15secs, washed in dH₂O for 2mins and equilibrated in transfer buffer until required. In addition, 4 pieces of blotting paper and 2 foam pads per gel were prepared. Once samples were at an appropriate distance through gel, the gel was removed from rig. The transfer cassette was placed with black side facing the bench. One foam pad was soaked in transfer buffer and place on black side of cassette. This was repeated for two pieces of blotting paper which were placed on top of the foam pad. Gel was carefully removed from plate and placed on top of blotting paper. One transfer membrane was placed on top of gel and bubbles removed. To this two pieces of blotting paper and one foam pad were added. The cassette was secured shut and placed in transfer rig with black side of cassette facing the black cathode (-). Rig was topped up with transfer buffer to ensure cassette was fully immersed and a magnetic stirring bar was placed inside the tank which was connected to power supply and run for 30V overnight.

2.8.5.3 Semi-quantitation of protein expression in cell lysates using Li-Cor odyssey

Prepare sufficient Li-Cor Odyssey Blocking Buffer (TBS) diluted 50:50 in TBS for blocking of membrane and dilution of primary and secondary antibodies. Following transfer of proteins to membrane block membrane in the prepared Li-Cor Odyssey Blocking Buffer (TBS) for 1hr at room temp. Dilute primary antibody in prepared Li-Cor Odyssey Blocking Buffer (TBS) and incubate membrane overnight at 4°C. Following overnight incubation; wash membranes 3x5mins in TBS-T at room temperature on rocker. Dilute secondary antibodies in Li-Cor Odyssey Blocking Buffer (TBS) and incubate membranes in secondary antibodies for 45mins at room temp on rocker. Wash the membrane 3x5mins in TBS-T then repeat washes using TBS. Place membrane on clean blotting paper and air dry in the dark. Once dry scan the membrane on the Li-Cor Odyssey scanner, SOP SaSoM/EQUIP/037, (Appendix 8.6).

2.9 Genetic manipulation of mammalian cell lines

2.9.1 Puromycin optimisation

Downstream procedures for genetic manipulation of mammalian cell lines C99 and RG/C2/80 cells required mammalian selection of insert positive cells via Puromycin. For each cell line, cells were plated into 6 well plates. Increasing concentrations of puromycin (0, 0.5, 1, 1.5, 2 and 3 µg/ml) were added to cells in suspension and assessed following 4 days exposure.

2.9.2 DUSP6 transduction of RG/C2/80 and C99 cells

DUSP6 was transduced in 2 cell lines (RG/C2/80 and C99) using a pLEX-MCS plasmid containing full length DUSP6 (Plasmid #27975, Addgene⁹⁸), Fig. 18. In addition to this, a pLJM1-eGFP (enhanced green fluorescent protein) plasmid (Plasmid #19319, Addgene⁹⁹) was used alongside to act as a transduction control, Fig. 19.

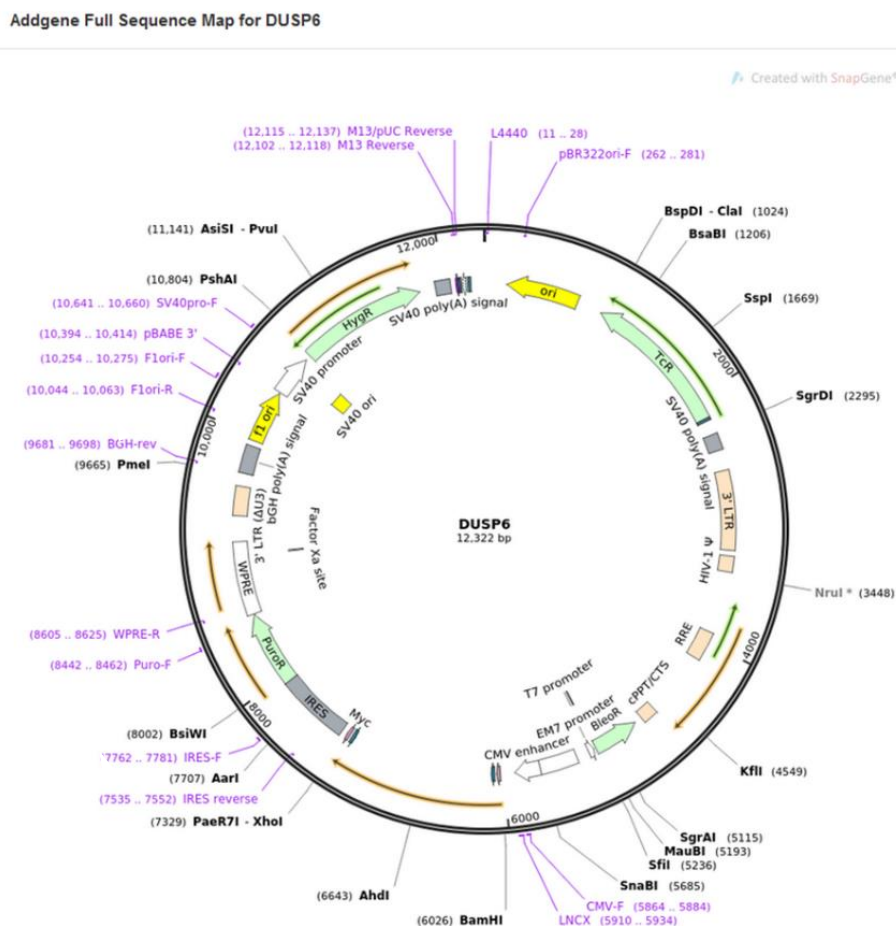


Figure 18. DUSP6 plasmid map. DUSP6 plasmid was obtained from Addgene (plasmid #27975, Bagnyukova et al., 2013). Plasmid comprises pLEX-MCS backbone grown in STBL3 bacteria. Plasmid contains Zeocin™ resistance elements for bacterial selection and Puromycin resistance elements for mammalian selection.

Depositor Full Sequence Map for pLJM1-EGFP

Created with SnapGene®

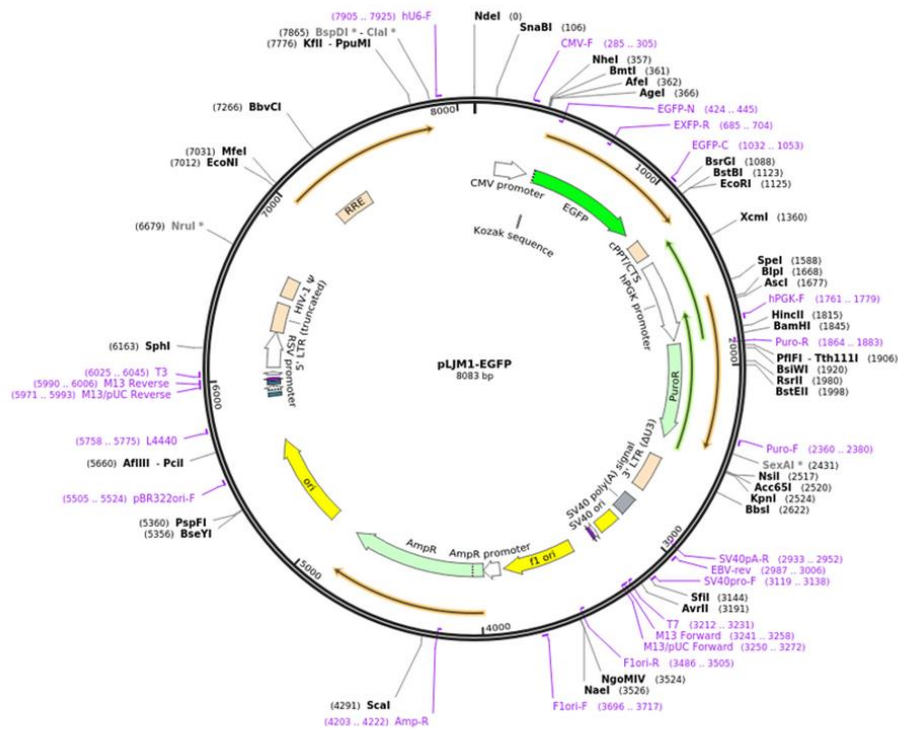


Figure 19. eGFP plasmid map. eGFP (enhanced green fluorescent protein) plasmid was obtained from Addgene (plasmid #27975, Sancak et al., 2008). Plasmid comprises pLJM1 backbone with eGFP grown in STBL3 bacteria. Plasmid contains ampicillin resistance elements for bacterial selection and Puromycin resistance elements for mammalian selection.

2.9.3 Sequencing of DUSP6 plasmid

To confirm the inclusion of DUSP6 in the selected plasmid, plasmid DNA was extracted from bacterial cultures and sent for DNA sequencing using the following primers (Life Technologies Ltd):

CMV-Forward: CGCAATGGGCGGTAGGCGTG

IRES-Reverse: CCTCACATTGCCAAAAGACG

The presence of DUSP6 in both forward and reverse directions was confirmed by the alignment of each sequence to the DUSP6 reference (NG_033915.1) using the nucleotide basic sequence alignment tool¹⁰⁰.

2.9.4 Generation of Zeocin™ agar plates

In order to enable bacterial selection of the DUSP6 plasmid low-salt agar plates containing Zeocin™ was required. Following autoclave of low-salt L-Broth with agar (Sigma L3272), Zeocin™ selection reagent (R25001, Thermo Fisher Scientific) was added to result in a final concentration of 70µg/mL. This solution was poured into 10cm petri dishes (Sigma P5606-400EA) and left to cool. Plates were stored in the dark at 4°C.

2.9.4.1 *Culturing of DUSP6 plasmid positive STBL3 bacteria*

DUSP6 plasmid was received in bacterial agar stab. A pipette tip was used to obtain bacteria and plated out onto Zeocin™ proficient low-salt agar plates. Bacteria were cultured in incubator overnight at 37°C.

2.9.4.2 *Selection of plasmid positive STBL3 bacteria*

Following overnight incubation plates were transferred to 4°C until required. For culturing of DUSP6 plasmid positive STBL3 bacteria a culturing solution comprising low-salt L-Broth (Sigma L3397) and Zeocin™ at a final concentration of 50µg/mL was prepared. Using a pipette tip single colonies of bacteria were selected and placed into 100mL of prepared culture media. Incubation was carried out with agitation overnight at 37°C.

2.9.4.3 *Glycerol stocks and maxi prep*

Following overnight incubation glycerol stocks of bacterial culture were made (750µl bacterial culture + 250µl 80% glycerol).

Maxiprep to extract DNA was carried out using Qiagen EndoFree® Plasmid Maxi Kit (12362, Qiagen). The remaining culture media was transferred to a 500ml screw capped bottle and centrifuged at 6000g for 15mins at 4°C. The supernatant was carefully discarded and pellet resuspended in 10ml Buffer P1. 10ml Buffer P2 was added and mixed thoroughly by inverting tube 4-6 times and incubated at room temperature for 5min. 10ml of chilled Buffer P3 was added to tube and mixed immediately and thoroughly by inverting. This was not incubated on ice. The lysate was poured into the prepared QIAfilter Cartridge (cap engaged) and incubated at room temperature for 10min. Following incubation, cap was removed and plunger inserted into QIAfilter Cartridge, the filtered lysate emptied into a 50ml tube. This was filtered until all of lysate has passed through cartridge. 2.5ml of Buffer ER was added to tube containing filtered lysate and mixed by inverting approx. 10 times then incubated on ice for 30mins. Meanwhile, QIAGEN-tip 500 was equilibrated by applying 10ml Buffer QBT, allowing column to empty by

gravity flow. Once all solution had passively passed through cartridge, the incubated filtered lysate was applied to the QIAGEN-tip and allowed to pass through by gravity flow. The QIAGEN-tip was washed twice with 30ml Buffer QC. DNA was eluted in a 30ml tube with 15ml Buffer QN.

10.5ml isopropanol was added, precipitating eluted DNA. This was mixed and centrifuged at 15000g for 30mins at 4°C. The supernatant was decanted. Pellet was washed with 5ml of 70% ethanol, centrifuged at 15,000g for 10mins. The supernatant was decanted without disturbing pellet and the pellet was air-dried for 10mins. Finally DNA was redissolved by adding 200µl of endotoxin-free dH₂O and placed in fridge overnight.

2.9.5 Lentiviral production in HEK293T cells

HEK293T cells were removed from liquid nitrogen and put in a T175 flask to allow recovery and growth. When sufficient cells were present, the cell passaging protocol was followed as previously described (2.9.2). Following syringing of cell suspension cells were counted with a haemocytometer and diluted in media ensuring 4×10^6 cells per 10cm dish (per condition i.e. eGFP vector and DUSP6 vector). Diluted cell suspension was dropped into a 10cm dish containing 8mL media and placed in the incubator overnight. For transfection of HEK293T cells with construct; OptiMEM (Thermo Fisher Scientific, Cat: 31985062) was pre-warmed in water bath and 1.5mL placed in a 15mL tube per condition. To the 15mL tube 27µl Mirus LT1 (Mirus, Cat: MIR2304) was added and left for 20mins at room temperature. In a 1.5ml tube, 2µl pVSVg and 3µl psPAX2 packaging plasmids were mixed plus either 6.8µl of eGFP plasmid or 2.8µl DUSP6 plasmid. Following the 20mins at room temperature the plasmid mixture was added to the OptiMEM and Mirus mix and left at room temperature for 30mins. HEK293T dishes were removed from incubator (media changed if necessary) and the transfection mixture (1.5ml per dish) was dropped over the plate. The plate was placed in the virus incubator overnight.

Working in the viral hood, the media was removed from each dish and 4ml of fresh media added (added to the side of plate to avoid disturbing cells). Dishes were placed in viral incubator overnight. To collect viral particles from the HEK293T cells, dishes were firstly checked for colour change (media will turn orange when viral particles have been released). The media/virus solution was collected in a 15ml tube. 4ml of media was added to the plate and the plate placed back into the virus incubator. The media/virus solution was stored in a sealed box at 4°C. New 10cm dishes were prepared with recipient cells (C99 = 4×10^6 , RG/C2/80 = 1.7×10^6) and placed in incubator.

Viral dishes and media/virus solution tubes were retrieved. 8µl Polybrene transfection reagent (Sigma, Cat:TR-1003-G) was placed in a new 50ml collection tube (this will result in a 1:1000 dilution in 8ml viral solution). 4ml of viral solution was collected from the viral dish along with 4ml of the viral solution from the previous collection and filter syringed to remove cellular debris. Recipient cell dishes were removed from the incubator. 4ml of viral solution were added to each plate, rocked gently to distribute solution across the dish and dishes placed back into the viral incubator for 4 hours (rocking occasionally). After 4 hours, the dishes were removed, media aspirated to waste and 5ml media added to each plate and aspirated to wash remaining viral particles. 10ml media was added and incubated for 48hrs in the viral incubator.

Following 48 hours incubation cells were passaged as previously described (2.9.2) and resuspended cell suspension (5ml) placed into a T75 flask containing 20ml media (total 25ml volume). To this flask was added 5µl puromycin (10mg/ml) to equal a final volume of 2µg/ml concentration in 25ml. The flask was placed in the virus incubator. Four days following puromycin treatment only resistant cells remained. Cells were passaged twice before using the non-viral incubator.

2.9.6 Dose-response experiments on manipulated cell lines

To assess the effects of DUSP6 transduction on C99 cells dose-response experiments were carried out. A plate for each of the 3 cell lines; C99 (parental), C99 (eGFP transduced) and C99 (DUSP6 transduced) was set up. Cells were treated 24hrs post seeding for 4 days. Six concentrations of Cetuximab (Selleckchem A2000) (0.001, 0.01, 0.1, 1, 10 and 100 μ g/mL) were prepared in media. In addition, a media only control was included for each cell line. Each condition was set up in sixuplicate. At the end of the treatment period an SRB assay was used to assess the percentage change in growth in the Cetuximab treated wells compared to the no media control.

2.9.7 Sulforhodime B (SRB) growth assay

Following treatment of cells with drug add 50 μ l of 25% Trichloroacetic acid (TCA) to each well and place at 4 $^{\circ}$ C for 1hr. Remove plate from refrigeration and pour TCA solution down the sink. Wash each row of the plate with running tap water (repeat for 10 washes). Blot the plate dry and place in oven at 50 $^{\circ}$ C until dry. Add 50 μ l of SRB solution to each well and tap the plate to ensure the surface of each well is covered. Leave at room temp for 30mins. Drain of the SRB solution down the sink and wash each row with 1% glacial acetic acid (repeat for 4 washes) ensuring all surplus SRB solution is removed. Blot the pate dry and leave to dry in oven. Once dry, add 150 μ l of 10mM (pH 10.5) Tris buffer solution and tap plate to ensure each well is covered. Leave on the rocker at room temperature for 1hr. Read the plate at 540nm on a plate reader.

3. Preliminary studies

3.1 Introduction

In order to explore DUSP6 in colorectal cancer in both tissue and *in vitro* a number of preliminary studies were carried out to prepare materials and generate models in order to explore its expression at protein level and its function in-vitro. The results from this series of preliminary studies are detailed in this chapter.

3.2 Results

3.2.1 Antibody optimisation and validation

Prior to use in routine experiments antibodies for dual-specificity phosphatase 6 (DUSP6), extracellular regulated kinase (ERK) and phospho-ERK (p-ERK) were optimised. The results are presented herein.

3.2.1.1 *Dual-specificity phosphatase 6 antibody optimisation and validation*

Two dilutions were assessed for the DUSP6 antibody (NovusBio, H00001848-M01), 1 in 200 and 1 in 500. The specificity of this antibody for DUSP6 was confirmed by positive immunohistochemistry in normal human pancreas (Fig. 20) and western blot using protein lysates from colorectal adenoma (RG/C2/80) and adenocarcinoma (C99) cell lines overexpressed with DUSP6, Fig. 21c. The dilution of 1 in 500 was determined for application on colorectal tumour tissue, Fig. 21b.

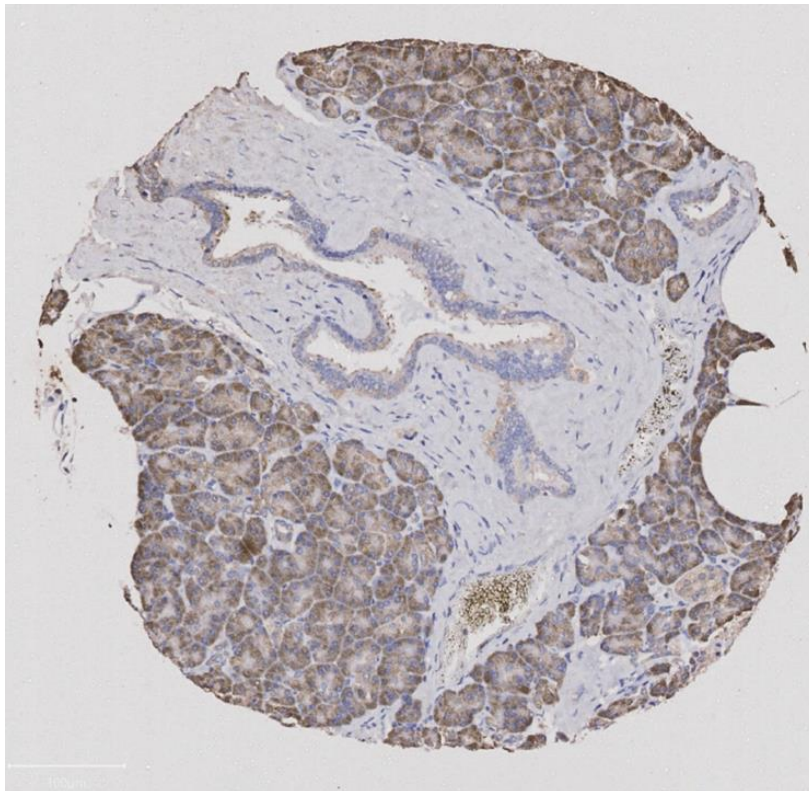


Figure 20. Positive control for DUSP6 immunohistochemistry. Image shows normal pancreas positive for DUSP6 (1:500 dilution). DUSP6 is highly expressed in normal human pancreas. The normal pancreas core on the tissue microarray was used to confirm positivity for DUSP6 immunohistochemistry. x20 mag, scale 100 μ m.

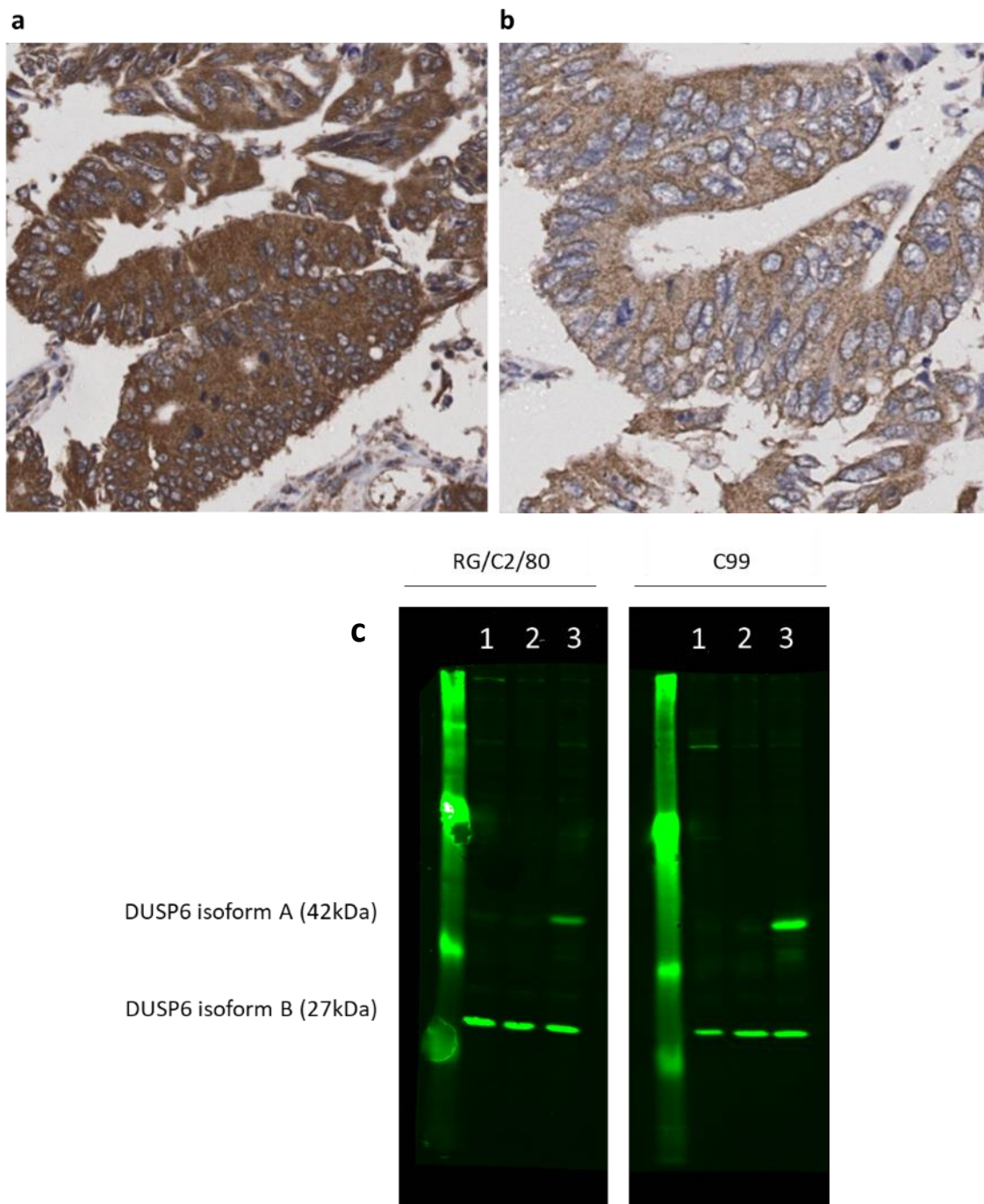


Figure 21. DUSP6 antibody optimisation and validation by western blot. A. 1 in 200 and B: 1 in 500. Representative core from antibody optimisation TMA, colorectal adenocarcinoma. 20x magnification. C. Western blot was carried out on protein lysates of colorectal adenoma (RG/C2/80) and colorectal adenocarcinoma (C99) cells transduced with DUSP6. 1. Parental line. 2. eGFP transduction control. 3. DUSP6 transduction. The DUSP6 antibody was highly specific. Primarily in both parental lines only isoform B was identified. Transduction of both lines with DUSP6 confirmed ability of antibody to bind to isoform A epitopes as evidenced by the presence of a strong fluorescence in both lines on western blot.

This antibody was further optimised for immunofluorescence. Dilutions for use with TSA direct detection methods were optimised using dilutions: 1 in 500 and 1 in 1000, Fig. 22. The dilution; 1 in 1000 proved to limit the level of background and was further optimised with FITC TSA detection (1 in 100).

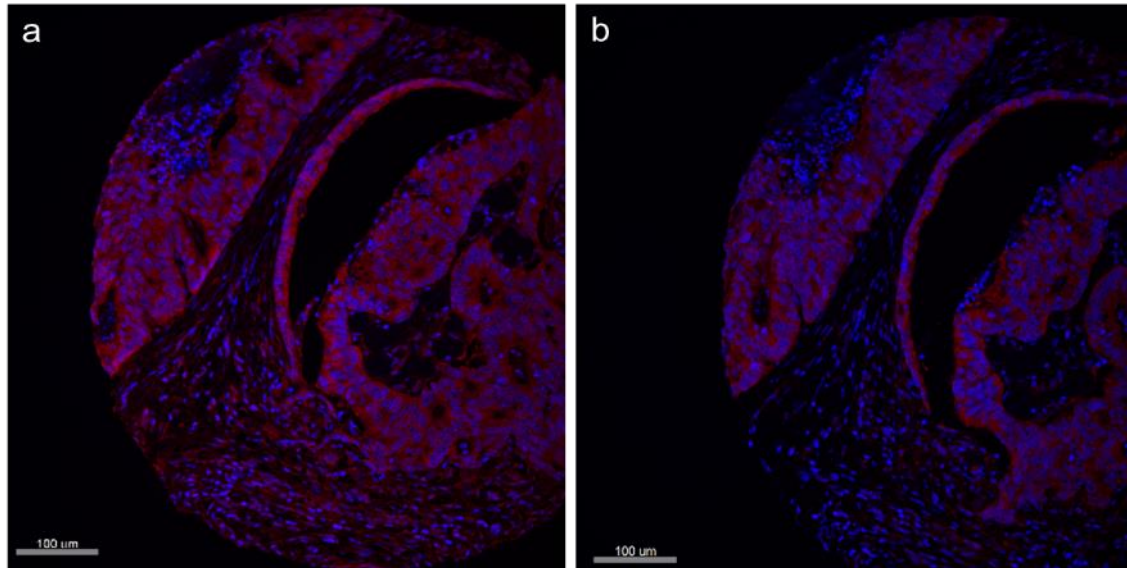


Figure 22. DUSP6 immunofluorescence optimisation: antibody dilution. DUSP6 antibody dilution for immunofluorescence was optimised on a tissue microarray containing a variety of cores from colorectal normal and cancer tissue. Dilutions a: 1 in 500 and b: 1 in 1000 were applied. 1 in 1000 was used for immunofluorescence application.

3.2.1.2 ERK immunohistochemistry optimisation and validation

Human prostate tumour tissue was used as a positive control¹⁰¹ for ERK immunohistochemistry and included on each run of immunohistochemical staining, Fig. 23. Two dilutions were assessed using the ERK antibody (Cell Signal, 9107), 1 in 250 and 1 in 500, Fig. 24a and 24b. The dilution 1 in 500 was used. The antibody was further validated by western blot, Fig. 24c.

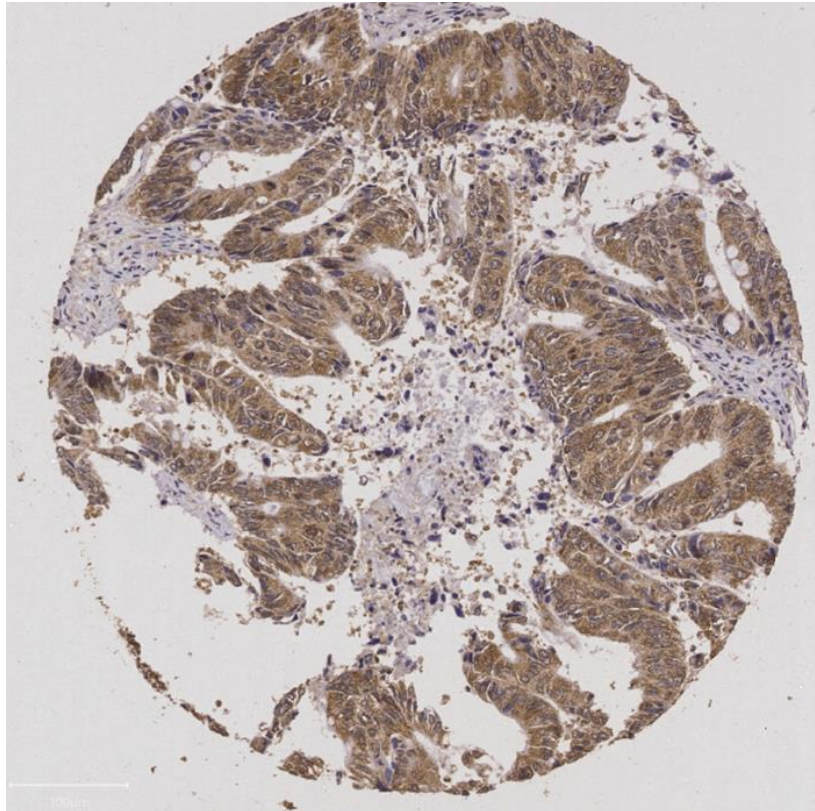


Figure 23. ERK immunohistochemistry positive control. A positive control (human prostate tumour tissue) was used to optimise the dilution of ERK antibody and also included in each subsequent immunohistochemistry run as a positive control. ERK antibody dilution 1 in 500. X20 mag, scale 100 μ m.

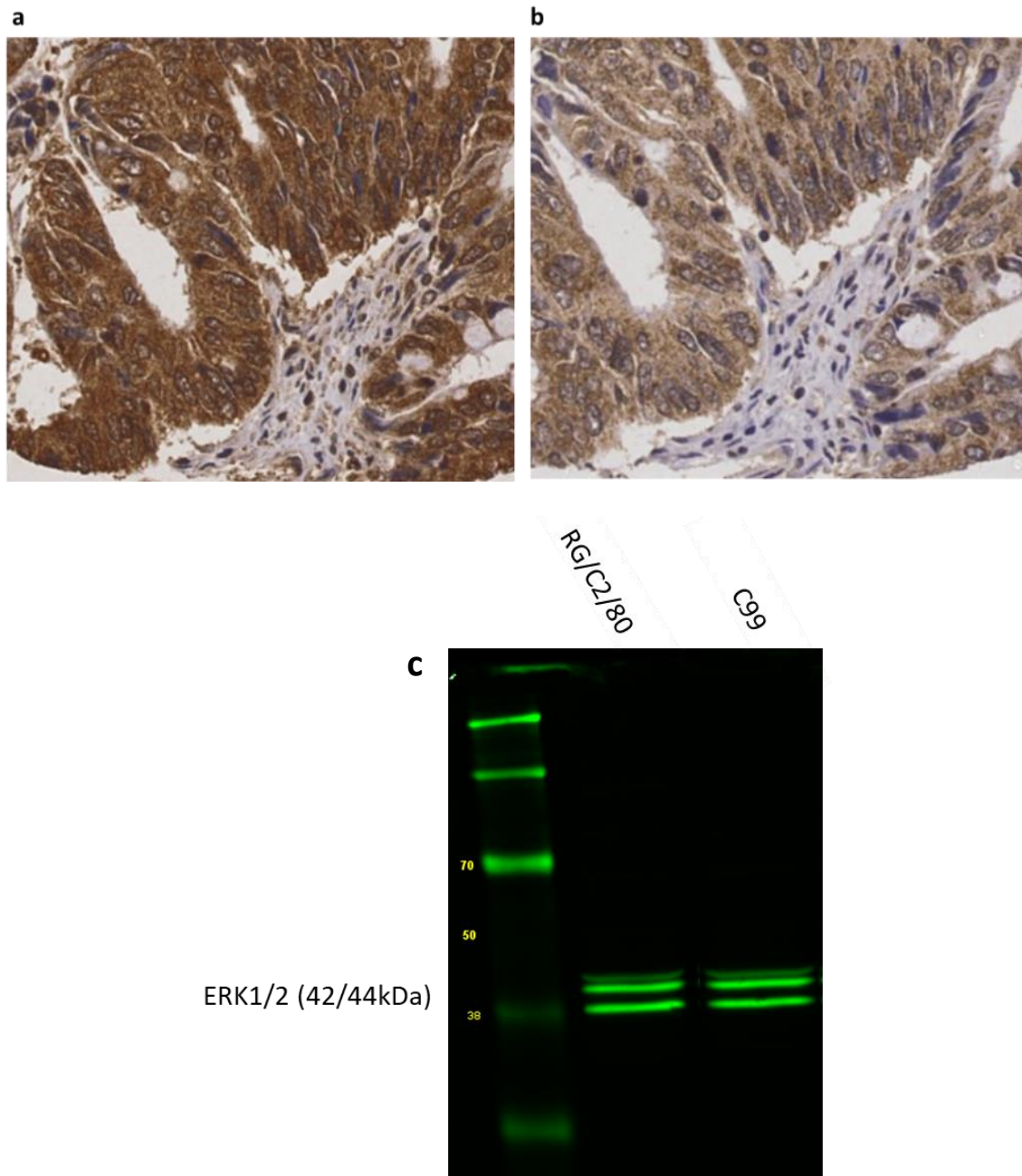


Figure 24. ERK antibody optimisation and validation by western blot. A. 1 in 250 and B: 1 in 500. Representative core from antibody optimisation TMA, colorectal adenocarcinoma. 20x magnification. C. Western blot was carried out on protein lysates of colorectal adenoma (RG/C2/80) and colorectal adenocarcinoma (C99) cells with ERK. The ERK antibody was highly specific giving bands only at the predicted molecular weight.

3.2.1.3 p-ERK immunohistochemistry optimisation and validation

Two dilutions were assessed for the p-ERK antibody (Cell Signal, 9101), 1 in 25 and 1 in 50, Fig. 25a and 25b. The dilution 1 in 50 was used. The antibody was further validated by western blot, Fig. 25c.

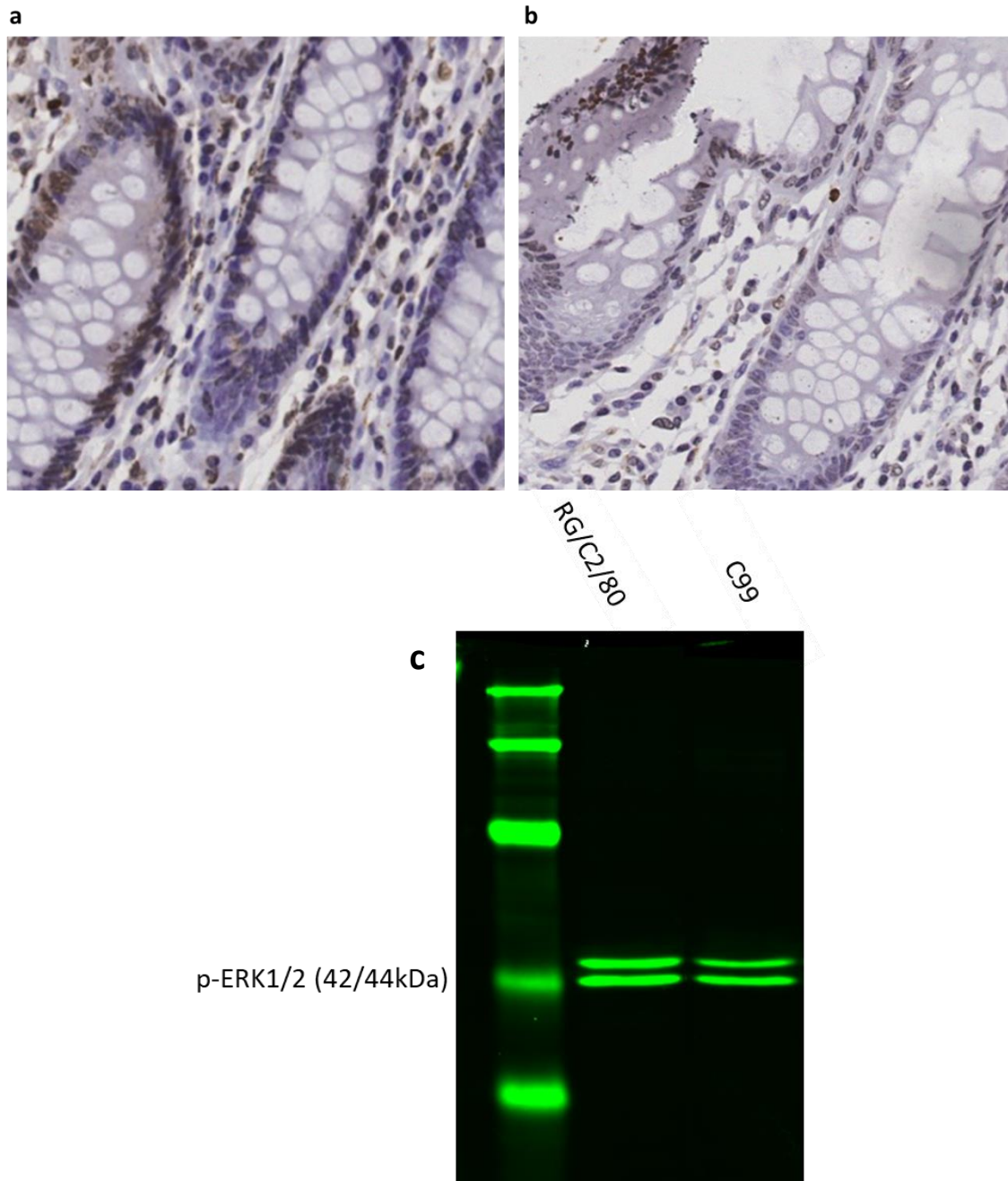


Figure 25. p-ERK antibody optimisation and validation by western blot. A. 1 in 25 and B: 1 in 50. Representative core from antibody optimisation TMA, colorectal adenocarcinoma. 20x magnification. C. Western blot was carried out on protein lysates of colorectal adenoma (RG/C2/80) and colorectal adenocarcinoma (C99) cells with p-ERK. The p-ERK antibody was highly specific giving bands only at the predicted molecular weight.

3.2.2 Immunohistochemical data analysis

3.2.2.1 Digital image analysis target threshold optimisation

QuPath⁹⁷ was used to analyse immunohistochemical outputs for DUSP6, ERK and p-ERK. To identify target positive cells, optimised target DAB intensity thresholds for DUSP6, ERK and p-ERK were generated as detailed in methods 2.8.3. Mean and maximum DAB intensity values were collated for 400+ cell per 1+, 2+ and 3+ phenotype for DUSP6, table 13.

DAB intensity value for all cells assessed	DUSP6 intensity phenotype		
	3+	2+	1+
Max	0.334	0.217	0.16
Mean	0.245	0.165	0.121

Table 13. DAB intensity values for DUSP6 phenotypes. Image analysis was carried out on the subset of cases which had been phenotypically determined as low (1+), moderate (2+) and high (3+) for DUSP6 expression. The mean and maximum intensity values for each category was collated for all cells assessed (400+ per classification). A threshold for each phenotype was determined ensuring that each threshold did not overlap with the maximum value of the phenotype below.

The DAB intensity value of >0.17 was determined sufficiently high to identify true positive DUSP6 expression excluding identification of 1+ cells.

Intensity thresholds for ERK and p-ERK were identified as detailed in materials and methods and results displayed in table 14. Thresholds for ERK and p-ERK were determined upon a positive or negative classification.

Target	DAB intensity value for target		
	Mean	Low	High
ERK	0.35	0.16	0.54
pERK	0.31	0.17	0.43

Table 14. DAB intensity values for ERK and p-ERK positive phenotypes. Image analysis was carried out on a subset of cases for ERK and p-ERK expression. The mean, high and low DAB intensity values were calculated from the cells assessed. Intensity thresholds for positive cell classification were determined by the average DAB intensity value.

The DAB intensity value of >0.35 for ERK and >0.3 for p-ERK were determined sufficient to positively classify cells.

3.2.3 Immunofluorescence data analysis

Following whole section immunofluorescent labelling of DUSP6 for cohort 2 (comprising 525 colorectal adenocarcinoma tissues) each slide was converted into a digital image using the AQUA system detailed in methods 2.8.2.

In order to prepare the data for analysis a number of normalisation steps were taken. A negative control accompanied each multiplex immunofluorescent run whereby primary antibody was replaced with antibody diluent. The average AQUA score was calculated from a tissue microarray block comprising of a variety of tissue types including colon adenocarcinoma. All data points were normalised by subtracting the appropriate AQUA negative control value from the raw AQUA data, Table 15.

Target	Compartment	Average negative AQUA score
DUSP6	Nuclear	34.1
	Cytoplasmic	32.7

Table 15. Details of negative normalisation. Each AQUA data point derived from cohort was normalised to the average AQUA score from colorectal cores in negative control (no primary antibody). These values were subtracted from each data point prior to further antibody lot normalisation.

As multiple lot numbers of antibodies were used across 14 process runs normalisation to account for between lot variability was carried out. A positive control using the same TMA as previously described was incorporated in each run. The AQUA score for colorectal cores within this TMA were obtained. The average AQUA score from positive control using the first lot number of each antibody was calculated. The AQUA score from each run was subsequently expressed as a ratio of the calculated average. This was termed the fold change. All AQUA data points from cohort were subsequently divided by the run specific fold change prior to statistical analysis, Table 16.

Target	Target compartment	Run number					
		1	2	3	4	5	6
DUSP6	Nuclear	0.88	1.14	1.10	0.86	1.02	0.36
	Cytoplasmic	0.90	1.18	1.05	0.87	1.00	0.39
		7	8	9	10	11	12
DUSP6	Nuclear	0.51	0.59	0.43	0.46	0.64	0.45
	Cytoplasmic	0.56	0.63	0.39	0.45	0.58	0.43

Table 16. Details of fold change normalisation. Following negative control normalisation each data point was further normalised to the fold change of each antibody to account for changes in antibody lot numbers.

3.2.4 X-Tile for the generation of DUSP6 protein expression high and low thresholds

In order to categorise patients based upon high and low DUSP6 expression X-Tile software v3.6.1¹⁰² was used. Low and high thresholds for cytoplasmic DUSP6 protein expression were determined by identification of a significant threshold of protein expression associated with stratification of survival outcome.

The determined threshold with a Miller-Sigmund p-value, $p < 0.001$ for significantly classifying groups into good and poor survival was 217.92 (AQUA value). Cases with DUSP6 AQUA values below this value were classified as 'low' and above this value classified as 'high', Fig. 26.

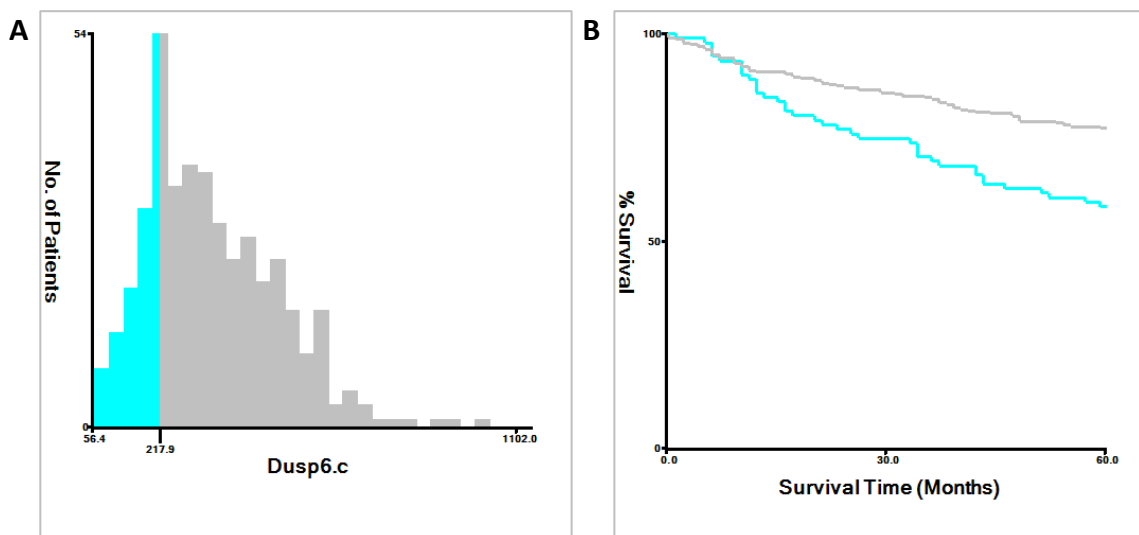


Figure 26. Output from X-Tile software for the detection of high and low thresholds for DUSP6 protein expression. X-Tile software v3.6.1 (Rimm et al., 2004) was used to identify a threshold for DUSP6 AQUA values which significantly stratified survival outcome, A threshold of 217.92 was determined. DUSP6 protein expression below this value were classified 'low' and above this threshold 'high'.

3.2.5 Genotyping of cohorts by pyrosequencing

In order to assess associations between DUSP6 protein expression and *RAS* mutation status, an important oncogene in CRC, both cohorts were assessed for the presence of *RAS* mutations using pyrosequencing, methods 2.5.

3.2.5.1 Cohort 1

Eighty of the eighty-seven adenomas were successfully sequenced for hotspot mutations in *KRAS* exon 2 and 3, Appendix 8.1.1. A wildtype frequency of 67.5% was observed amongst successfully sequenced samples. Of the *KRAS* mutant lesions (32.5%), a large majority of mutations were observed in codon 12 of the gene, Fig. 27.

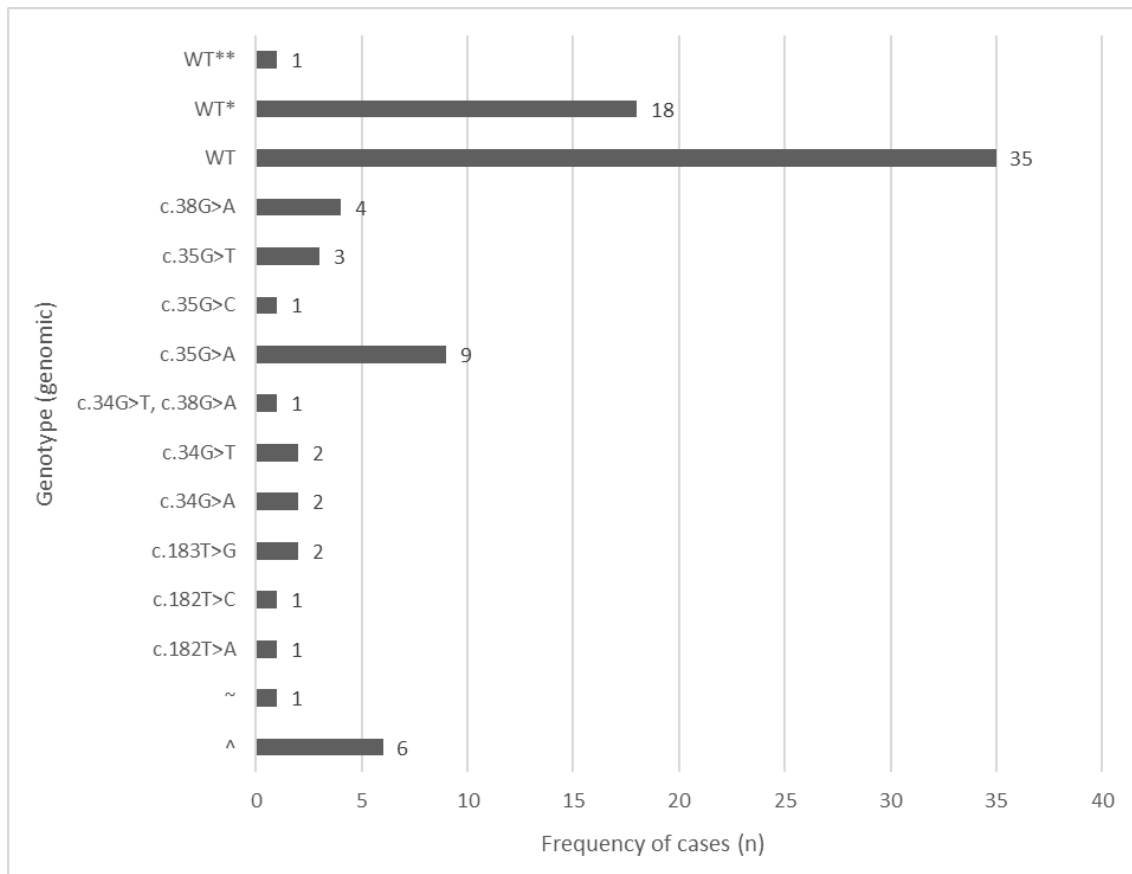


Figure 27. *KRAS* genotyping of adenoma cases. **KRAS* exon 1 only successfully sequenced. ***KRAS* exon 2 only successfully sequenced. ~Not tested due to insufficient tissue. ^Failed sequencing. A wildtype frequency of 67.5% was observed amongst successfully sequenced samples. Of the *KRAS* mutant lesions (32.5%), a large majority of mutations were observed in codon 12 of the gene. This is consistent with the observed frequency of *KRAS* exon 2 and 3 mutants in the adenocarcinoma cohort assessed in Chapter 4.

3.2.5.2 Cohort 2

Genotyping for hotspot positions of *KRAS*, *NRAS* and *BRAF* genes (as per routine clinical practice in Molecular Pathology, Royal Infirmary of Edinburgh) was carried out for each case (*n*525), case specific results in Appendix 8.2.1.

A total *RAS* mutant population of 36.1% was identified. *KRAS* mutant individuals comprised 33.1% of the study cohort and 91.6% of the *RAS* mutant population. *NRAS* mutant individuals comprised 3% of the study cohort and 5.3% of the *RAS* mutant population, Table 17.

Genotypic assessment		Study cohort frequency
RAS status	Wildtype	334
	Mutant	190
BRAF status	Wildtype	457
	Mutant	67

Table 17. Summary of RAS and RAF mutant cases. RAS mutant cases comprised 36.3% of study cohort. BRAF mutant cases comprised 12.8% of study cohort.

Thirteen different *KRAS* variants and five different *NRAS* variants were identified in the study cohort.

The most frequently mutated variants were in *KRAS* codon 12/13. Most notably, *KRAS* c.35G>A p.(Gly12Asp) and *KRAS* c.35G>T p.(Gly12Val) were the most commonly mutated variants as is comparative with frequencies observed on the Catalogue of Somatic Mutations (COSMIC) database¹⁰³, Fig. 28. Mutations in *NRAS* were less frequent, the most commonly mutated variant being c.35G>A p.(Gly12Asp).

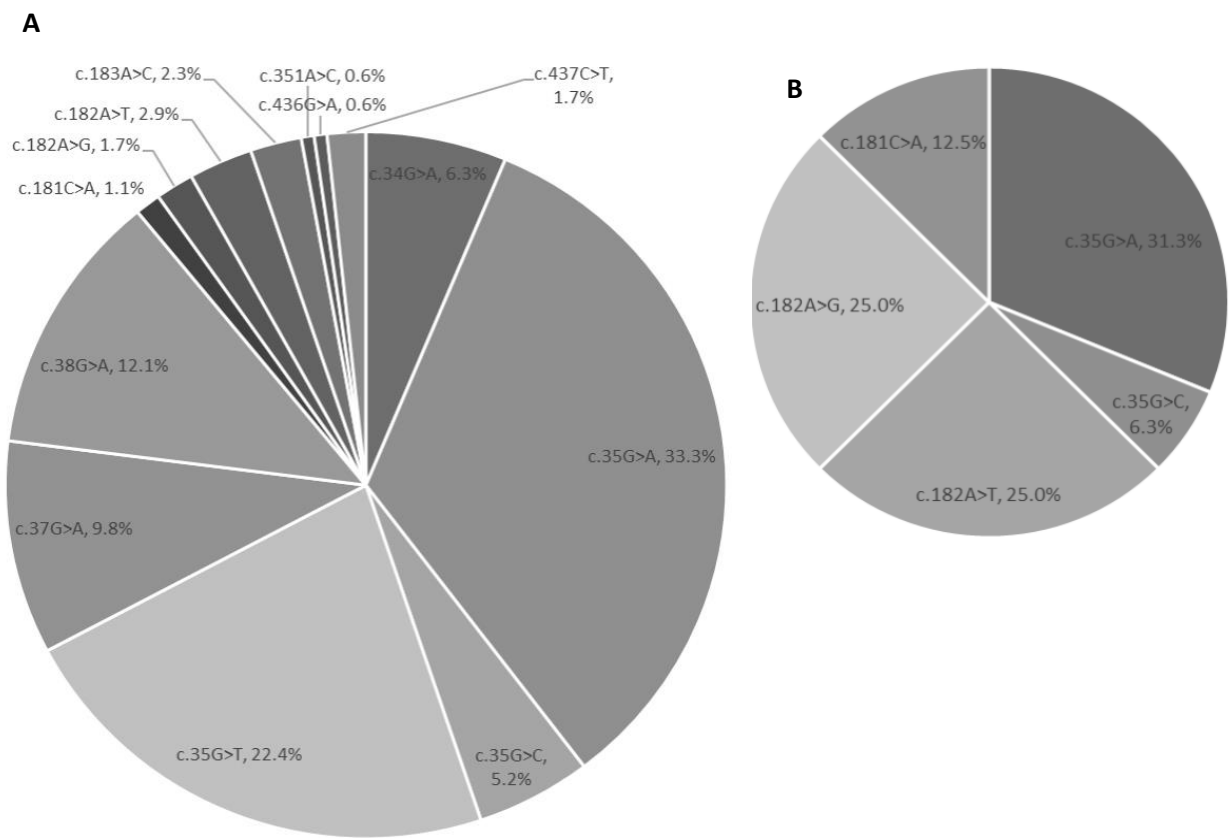


Figure 28. Frequency of RAS mutations. A: Frequency of KRAS variants within study cohort. *KRAS* c.35G>A p.G13D was most frequently mutated in study cohort (33.3% of all *KRAS* mutants). B: Frequency of NRAS variants within study cohort. *NRAS* mutations are less common than *KRAS* mutations in colorectal cancer. The most frequently mutated *NRAS* variant was analogous to *KRAS*, i.e. c.35G>A p.G13D (31.3% of all *NRAS* mutants).

3.2.6 Generation of an DUSP6 overexpression model *in vitro*

In order to explore the effects of DUSP6 on ERK and p-ERK expression in colorectal models, DUSP6 was overexpressed in both RG/C2/80 colorectal adenoma cells and C99 colorectal adenocarcinoma cells. The results from this process are detailed herein.

3.2.6.1 Puromycin optimisation

Downstream procedures for genetic manipulation of mammalian cell lines C99 and RG/C2/80 cells required mammalian selection of insert positive cells via Puromycin. Protocol was carried out as detailed in methods 2.10.1, (Fig. 29), it was determined that 2µg/ml was sufficient to kill all cells.

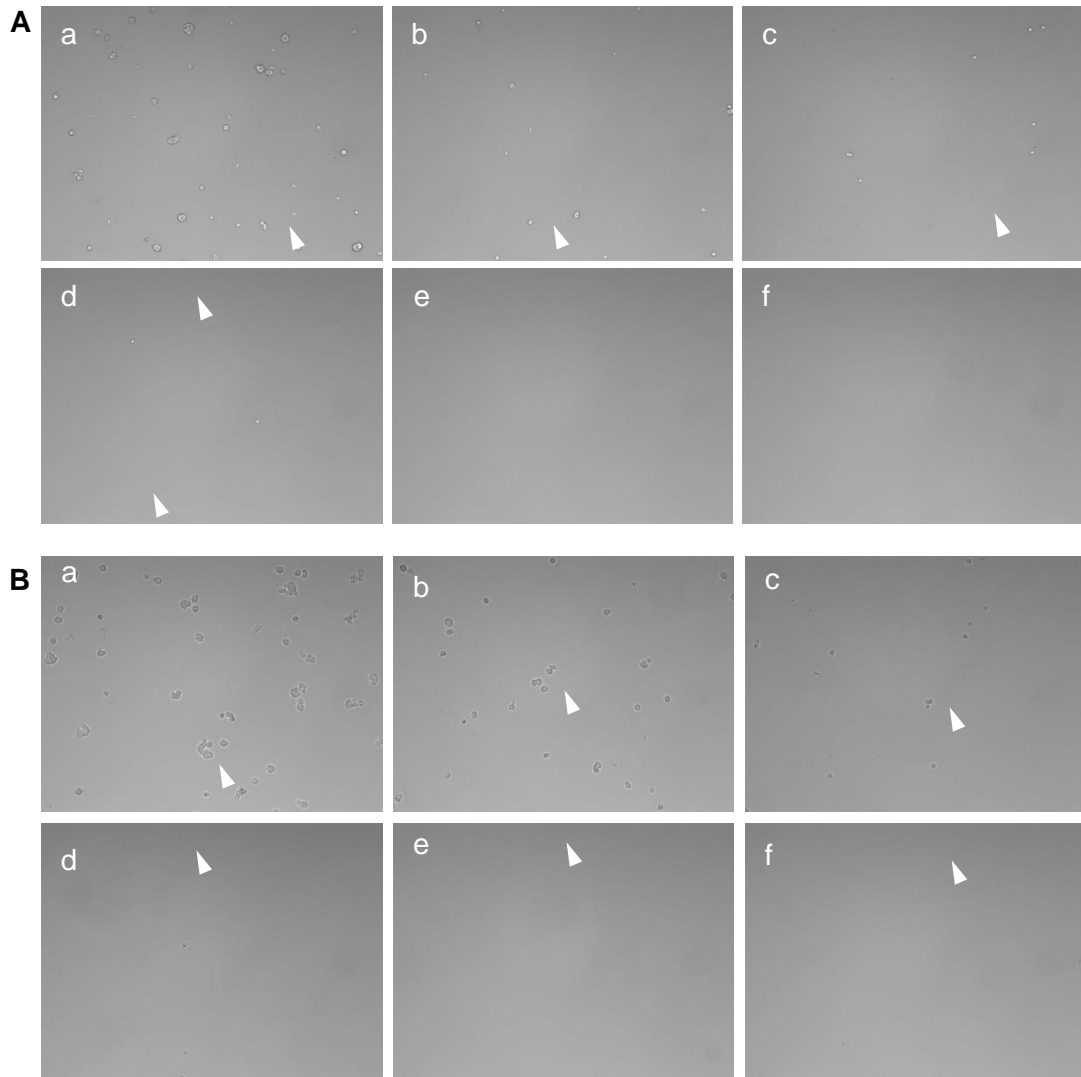


Figure 29. Puromycin optimisation in C99 and RG/C2/80 cells. A: C99 adenocarcinoma cells. B: RG/C2/80 adenoma cells. Puromycin was applied to cells in suspension at the following concentrations: a: 0µg/ml, b: 0.5µg/ml, c: 1µg/ml, d: 1.5µg/ml, e: 2µg/ml and f: 3µg/ml. Cells were assessed following 4 days exposure. No cells were visible at 2µg/ml Puromycin concentration. Cells denoted by white arrow.

3.2.6.2 Confirmation of DUSP6 plasmid

A DUSP6 plasmid, pLEX-MCS containing full length DUSP6 (Plasmid #27975, Addgene⁹⁸) was obtained in a bacterial stab relieving the requirement of bacterial transformation for this specific plasmid. A control plasmid, pLJM1-eGFP (enhanced green fluorescent protein, Plasmid #19319, Addgene⁹⁹) did require bacterial transformation, methods 2.10.2.

Firstly in order to confirm that the bacterial stab received contained the DUSP6 gene, DNA was extracted from transformed bacteria and sequenced using sites flanking the DUSP6 insert (CMV and IRES, methods 2.10.3).

Sequencing results were inspected using BioEdit¹⁰⁴ v7.0.5 to confirm the quality of sequencing, Fig. 30.

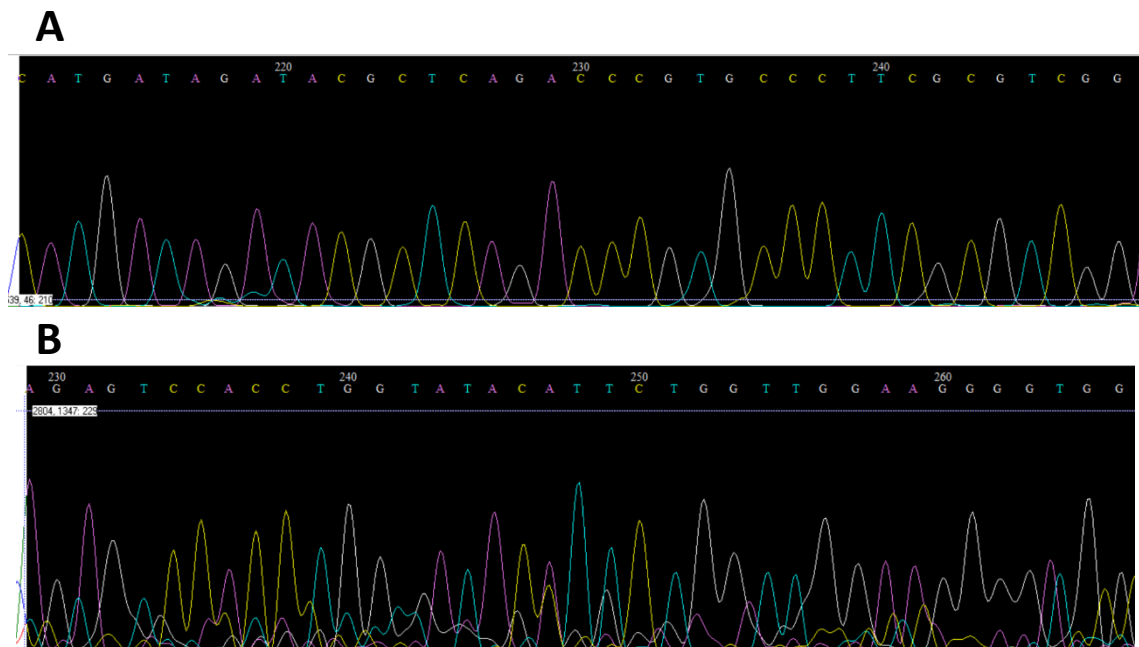


Figure 30. DNA sequencing of DUSP6 plasmid. A: CMV-Forward sequencing. B: IRES-Reverse sequencing. The quality of the forward sequencing was better than reverse sequencing as evidenced by the low background in A.

The presence of DUSP6 in both forward and reverse directions was confirmed by the alignment of each sequence to the DUSP6 reference (NG_033915.1) using the nucleotide basic sequence alignment tool¹⁰⁰, Fig. 31 A and B.

A

	Score	Expect	Identities	Gaps	Strand
	734 bits(397)	0.0	401/403(99%)	0/403(0%)	Plus/Plus
Query 1	CATGATAGATACGCTCAGACCCGTGCCCTTCGCGTCGGAAATGGCGATCAGCAAGACGGT				60
Sbjct 1					60
Query 61	GGCGTGGCTCAACGAGCAGCTGGAGCTGGGCAACGAGCGGCTGCTGCTGATGGACTGCCG				120
Sbjct 61					120
Query 121	GCCGACGAGCTATACGAGTCGTGCGACATCGAGTCGGCCATCAACGTGGCCATCCCGGG				180
Sbjct 121					180
Query 181	CATCATGCTGCGGCGCTGCAGAAGGGTAACCTGCCGGTGCAGCGCTCTTCACGCGCGG				240
Sbjct 181					240
Query 241	CGAGGACCGGACCGCTTCACCCGGCGCTGTGGCACCACACAGTGGTGTCTACGACGA				300
Sbjct 241					300
Query 301	GAGCAGCAGCGACTGGAACGAGAATACGGGCGGCGAGTCGGTGTCTGGGCTGCTGCTCAA				360
Sbjct 301					360
Query 361	GAAGCTCAAGGACAGGGCTGCCGGGTGTTCTACCTGGAAGGT			403	
Sbjct 361				403	

B

	Score	Expect	Identities	Gaps	Strand
	499 bits(270)	4e-145	290/300(97%)	0/300(0%)	Plus/Plus
Query 1	AGAGTCCACCTGGTATACTTCTGGTTGGAAGGGGTGGTAAAATACAGCTGCTGTGCTGG				60
Sbjct 1					60
Query 61	AACCCTGTTGTCACATGGGCTGCTGATCCCAGCGTCTCTCCAAGTCCATCAGCTGACC				120
Sbjct 61					120
Query 121	CATGAAGTTGAAATTAGGGGATATGTTGGATTTTTTCATTTTGACAATGTCCTAGGCATC				180
Sbjct 121					180
Query 181	GTTTCATCGACAGATTGAGCTTCTGCATAAGGTAAGCCACAGTCACAGTGACTGATCGGCT				240
Sbjct 181					240
Query 241	TATGCCAGCCTAGCAATGTACCAAGACACCACAGTCTTGCCCCGGGCTTCGTCTATGAA				300
Sbjct 241					300

Figure 31. Sequencing alignment results for DUSP6 plasmid DNA sequencing. To confirm that the DUSP6 plasmid contains the correct DUSP6 insert DNA was extracted from DUSP6 containing bacterial cultures and sequenced in both forward and reverse directions. Using the nucleotide basic sequence alignment tool (BLAST) it was confirmed that both forward and reverse sequences matched the DUSP6 reference sequence (NG_033915.1). A: Forward sequencing results. B: Reverse sequencing results.

3.2.6.3 Confirmation of eGFP plasmid overexpression

The observation of a fluorescence signal in eGFP transduced cell lines provided evidence of a successful transduction process of the eGFP plasmid for both C99 and RG/C2/80 cell lines, Fig. 32.

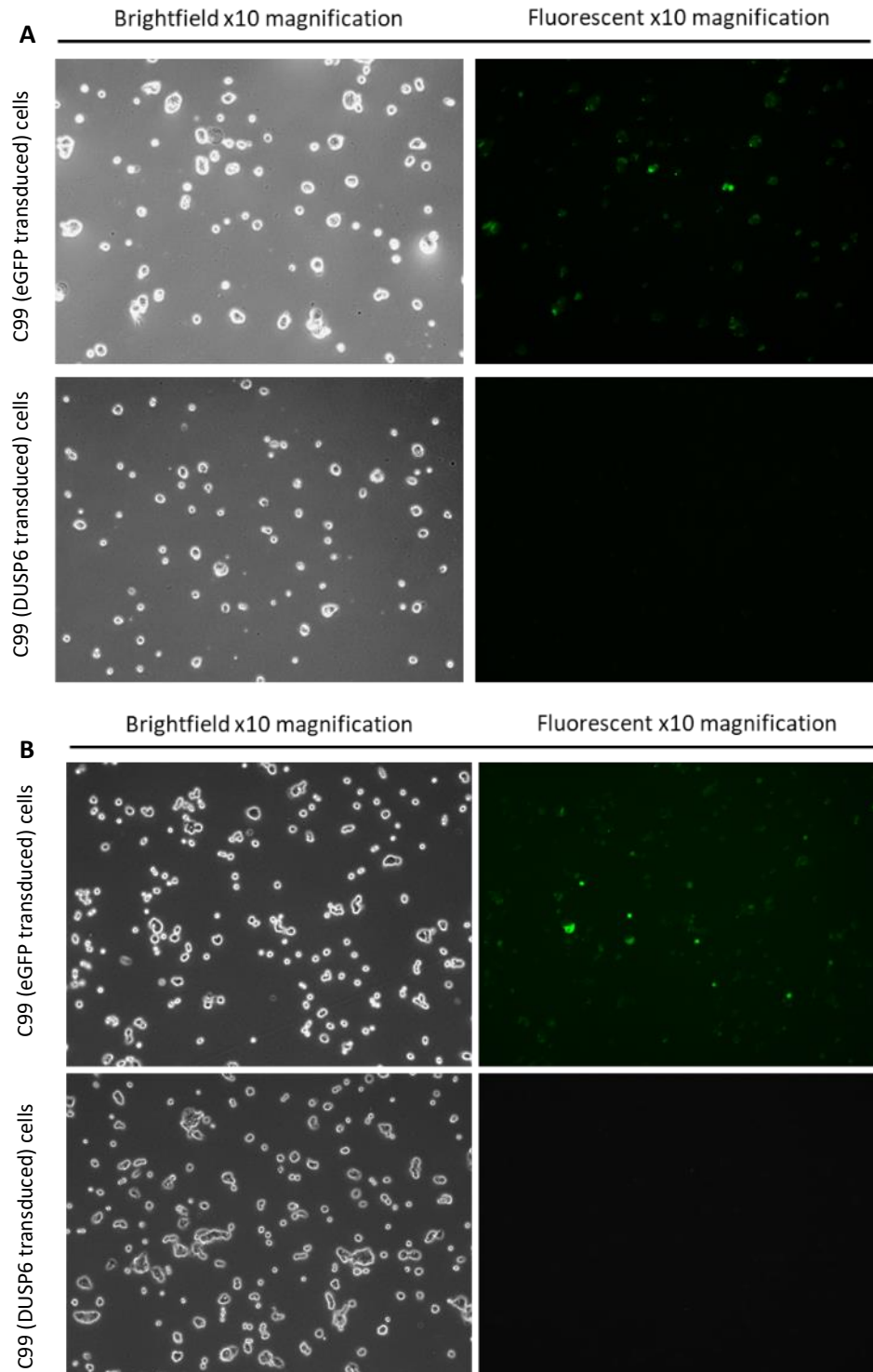


Figure 32. Fluorescence signal in eGFP transduced cell lines. A: C99 cells. B: RG/C2/80 cells. The observation of a fluorescence signal in eGFP transduced cells for both lines suggest successful transduction process.

3.3 Discussion

The preliminary work presented in this chapter enabled the investigation into the expression of DUSP6 at protein level and its function with respect to ERK and p-ERK expression *in vitro*.

The optimisation of antibodies prior to large scale assessment of cohorts is of primary importance to ensure the specificity and sensitivity of the antibody for the target protein. The validation of antibodies is achieved through a combination of positive control tissue whereby the target is known to be expressed at high levels and western blotting of cell lines either endogenously expressing high levels of the target protein or targets having been exogenously introduced as was the case in this study for the validation of the DUSP6 antibody.

In this study to ensure specificity of antibodies toward target proteins, where possible positive control tissue was used. In addition to tissue controls, the use of manipulated cell lines overexpressing targets of interest is a common method of validation for antibody specificity and was thus utilised in antibody optimisation for this study. The presence of clean western blotting membrane with single bands suggests the antibodies used are highly specific for their targets. The location of each target protein in western blot was confirmed by size i.e. DUSP6 isoform A at 42kDa, DUSP6 isoform B at 27kDa, ERK1/2 at 42/44kDa and p-ERK1/2 at 42/44kDa which further enabled confirmation of antibody specificity. In order to address sensitivity of each antibody a number of dilutions were used to achieve the appropriate balance between staining intensity and background staining. As evidenced in the figures presented, the optimised antibodies produced minimal background staining. An additional factor in antibody validation is to negate for cross-reactivity whereby an antibody is able to recognise two specific antigens¹⁰⁵. To accommodate for this the DUSP6 antibody sequence was assessed for homology to sequences other than the target using NCBI BLAST¹⁰⁰ and returned sequence homology solely for DUSP6.

In addition to antibody optimisation and validation a number of steps were carried out to ensure consistency in staining intensity at the specified dilutions between antibody lots. Due to the nature of antibodies being generated in live organisms a degree of variability can occur in the binding affinities of antibodies for the target protein¹⁰⁵. To ensure all data between immunohistochemical staining runs was comparable and to overcome lot to lot variability a positive control slide was included in each run and staining intensity compared between runs and lots of antibody. A ratio in terms of fold change in staining intensity was calculated with

reference to the primary lot used and the staining intensity values normalised to this accordingly.

Two methods of digital image analysis were employed for the assessment of DUSP6 protein expression in the two cohorts, QuPath⁹⁷ and AQuA⁹⁶. QuPath was used to quantify the number of target positive cells from immunohistochemical staining whilst the AQuA system was applied to quantify the signal intensity of target proteins in immunofluorescent labelled whole sections. QuPath can be applied for the interrogation of fluorescently labelled tissue however as DUSP6 expression was explored in cohort one using a DAB chromogenic reporter which identifies one protein target e.g. DUSP6, the QuPath software was used for its ability to classify cells based upon cellular shape without the need for an additional antibody to label epithelial cells as a 'tumour mask'. Additionally, the Leica SCN scanner available during this study was only able to scan brightfield images making QuPath the ideal application for analysis of cohort 1.

In contrast to this, the advantages of using AQuA to analyse the data from immunofluorescent studies in cohort 2 were that a more specific identification of cells of interest by pan-cytokeratin labelling could be achieved compared to QuPath as epithelial cells were labelled with a pan-cytokeratin antibody and fluorescently labelled. The use of multiple spots of interest per section acting as a 'pseudo tissue microarray' alleviated analysis biases as regions from the whole section were able to be analysed and an average signal intensity taken for each case. Additionally, the biomarker cut off software X-Tile which accompanies AQuA data could be utilised to generate biomarker signal intensity cut offs to explore patient survival dynamics in relation to DUSP6 expression.

The identification of optimal biomarker cut-offs with respect to signal intensity values is a challenging task in protein biomarker analysis. A number of values can be explored with respect to optimal cut-offs to stratify patients, median, mean and percentiles are all valid. However, by using these statistical, all be it simplistic means of generating thresholds, due to the nature of patient data being 'not normally' distributed there is the potential for these cut offs to miss the true trends of data with respect to protein expression and survival¹⁰⁶. For these reasons, the X-tile¹⁰² software that stratifies patients based upon survival outcome was used to generate an optimal threshold for DUSP6 expression. However, the thresholds derived from this method are based upon the patient population in the cohort studied and may not be representative of other populations. Due to the lack of a physical validation cohort, statistical methods of random

sampling i.e. bootstrapping was employed to ensure the thresholds represent a true population and not just the sample population, see Chapter 5.

Overall, a number of quality control steps have been employed to ensure comprehensive validation and optimisation of antibodies used and sufficient normalisation and analysis techniques for data derived from whole cohort immunohistochemical staining and immunofluorescent labelling.

Pyrosequencing was carried out for both cohorts in order to assess associations between DUSP6 protein expression and *RAS* mutation status as *RAS* is frequently mutated in CRC and utilised as a negative biomarker for the targeted treatment, Cetuximab which is investigated in this series of studies. Pyrosequencing is an inexpensive and quick method for the detection of hotspot variants within short templates making it highly applicable to FFPE specimens which often contain highly fragmented DNA^{107,108}. Despite its applicability to FFPE specimens sequencing failures were observed during the characterisation of both cohorts. For cohort 2 the highest proportion of failures for pyrosequencing were in *KRAS* exon 4 which may be attributed to the length of the PCR product prior to sequencing. The *KRAS* exon 4 PCR product is 153 base-pair (bp) in length in comparison to *KRAS* exon 2 and 3 which are significantly shorter, 78 and 69 bp respectively. In order to confirm the specificity and sensitivity of pyrosequencing for the detection of *RAS* variants a number of cases were additionally assessed for variants by next generation sequencing (NGS) as detailed in methods 2.5. Upon this additional validation, the performance of pyrosequencing was deemed comparable to NGS and most suitably applied to this study.

At the variant frequency level, comparisons were drawn between both study cohorts and the COSMIC database which details a compilation of large-scale cohorts. The reported frequency of *RAS* mutant cases is comparative for both cohorts with large scale cohorts detailed on COSMIC database, 32.7% *KRAS* mutant, 3.7% *NRAS* mutant¹⁰⁹. The two most common variants within the study cohort were c.35G>A p.Gly12Asp and c.35G>T p.Gly12Val which was consistent with both COSMIC and Lothian 2014 cohort data. Alterations in variants located in amino acids 10-16 alter the interaction between the phosphate-binding loop (P loop) of GAP and *RAS* which is required for the transformation of guanine-triphosphate (GTP) bound *RAS* to guanine diphosphate (GDP) bound. This subsequently results in an accumulation of active GTP-bound *RAS* within the cell. GTP-bound *RAS* undergoes a conformational change, resulting in a higher affinity for its effectors such as RAF¹¹⁰. Variants occurring at position 61 alter the ability of the *RAS* protein to hydrolyse

GTP, conferring constitutive activation of the protein¹¹¹. Studies have shown *KRAS* exon 4 mutant cells demonstrate higher RAS-GTP activity than non-mutated cells however lower RAS-GTP activity than exon 2 variants¹¹². This suggests *KRAS* exon 2 variants may have a greater pathogenic impact than exon 4 variants and may explain why these are observed at a higher frequency across the study populations.

With respect to the generation of two *in vitro* models for investigation into the functional impact of DUSP6 on ERK and p-ERK expression in adenoma and adenocarcinoma a number of quality control steps were employed to ensure a robust model was generated. As the DUSP6 plasmid was commercially generated, sequencing of bacterial DNA containing the plasmid was carried out confirming the plasmid contained a full length DUSP6 insert.

For each cell line an eGFP containing plasmid was generated alongside the DUSP6 plasmid in order to confirm that the transduction process was successful by the presence of a fluorescent signal in the eGFP transduced lines and lack of signal in the DUSP6 transduced lines. Additionally, this control cell line could act as a control for biological outputs providing evidence that changes observed in protein expression could be attributed to the over-expression of DUSP6 and not merely the process of introducing a non-endogenous protein. Criticism could be made as to the use of an eGFP insert and not using an empty plasmid however both types of control involve the introduction of a non-endogenous construct into an *in vitro* model, thus both could induce an element of stress response in the model system. To further control for these effects additional experimentation using pharmacological manipulation of the target protein to recapitulate the results observed with *in vitro* models could be employed however these additional experiments were not within the scope of this series of studies.

In conclusion, sufficient measures have been made to ensure the correct construct has been introduced into both cell lines and a control model has been generated to enable the appropriate conclusions to be derived from downstream experiments as to the impact of DUSP6 on ERK and p-ERK expression.

4. The role of dual-specificity phosphatase 6 in colorectal pathogenesis

4.1 Introduction

The development of colorectal adenocarcinoma is viewed as a multi-step process beginning with the development of adenoma and progressing to adenocarcinoma with the aid of gradual accumulation of aberrations such as loss of tumour suppressor gene function i.e. *TP53* and the gain of function of oncogenes i.e. *KRAS*^{11,113}. Development of this model over the past 3 decades has highlighted the molecular complexity and diversity of not only the later stages of disease but also in the initiating stages. A simplified description of the malignant transformation of normal mucosa begins with mutant inactivation of the adenomatous polyposis coli (APC) gene resulting in the activation of Wnt signalling pathway enabling cells to overcome restraints on cellular growth. This results in the formation of adenoma, which can take a variety of forms and histological subtypes. Growth of adenomas can be accompanied by activation of oncogenes such as *KRAS* or *BRAF*, further loss of tumour suppressor genes such as *TP53* result in malignant transformation. During this malignant transition a number of additional genetic and epigenetic aberrations can result in the lesions growth sequence following a number of different pathways including CpG island methylation (CIMP) or serrated pathways¹¹⁴.

The role of dual specificity phosphatase 6 (DUSP6) in the pathogenesis of disease has been investigated across a variety of cancer types. In pancreatic cancer, increased DUSP6 protein expression was significantly associated with mild-severe dysplasia whilst a decrease was observed in invasive carcinoma⁶⁴. In lung cancer, similar DUSP6 expression profiles have been observed whereby increasing growth activity and histological grade is coupled with decreasing protein expression of DUSP6⁶¹. In addition to this line of evidence, DUSP6 has been shown to impair epithelial-mesenchymal transition (EMT) associated properties in oesophageal squamous cell carcinoma and nasopharyngeal carcinoma^{62,65,66}. Despite its extensive study across these cancer subtypes no studies have investigated the role of DUSP6 in colorectal cancer (CRC).

The aims of this chapter were to investigate whether DUSP6 plays a tumour suppressor role in colorectal pathogenesis, with particular focus at the stage of adenomatous lesions. In order to explore this a cohort of 147 colorectal formalin fixed paraffin embedded (FFPE) samples was compiled with cases representing the stages of normal mucosa, adenoma and adenocarcinoma. Protein expression across the cohort was investigated with immunohistochemistry and expression quantified by digital image analysis.

4.2 Results

4.2.1 Clinicopathological characteristics of adenoma-carcinoma cohort

4.2.1.1 *Cohort identification strategy*

In order to identify a cohort that represents a variety of structural and molecular aspects of the adenoma-carcinoma sequence a search of previously tested patients within Pathology at the Royal Infirmary of Edinburgh was carried out to identify 147 patient samples comprising: normal colorectal mucosa, tubular, tubulovillous, villous and serrated adenoma, adenocarcinoma with deficient mis-match repair (dMMR) as determined by immunohistochemistry, adenocarcinoma with *KRAS* mutation, adenocarcinoma with *TP53* mutation and adenocarcinoma with both *KRAS* and *TP53* mutation. Clinicopathological data was collated for each case and is detailed in Table 18.

A

Age range		32-87	
Gender[§]	Female	68	
	Male	77	
Classification			
Normal		Adenoma	Adenocarcinoma[^]
Number of cases		15	87
Site	Left	4	12
	Transverse	0	8
	Right	2	40
	Rectosigmoid	4	0
	Rectum	5	27

B

Sub-groups	Adenoma		
	Histological type	Serrated	10
		Tubular	25
		Tubulovillous	33
		Villous	19
	Dysplasia	LGD	64
		HGD	23
	KRAS mutation status⁺	Wildtype	54
		Mutant	26
	Adenocarcinoma		
Molecular classification	dMMR	15	
	KRAS mutant only ^{**}	11	
	TP53 mutant only ^{^^}	15	
	Dual KRAS and TP53 mutant	4	

C

Adenocarcinoma	TNM stage*	1	0
		2	17
		3	21
		4	4
	Differentiation*	Well	0
		Moderate	28
		Poor	14
	Adenocarcinoma type*	Adenocarcinoma	25
		Medullary	8
		Mucinous	6
Signet Ring		3	

Table 18. Cohort demographics. Table details clinicopathological data for 147 formalin fixed paraffin embedded (FFPE) cases. Normal cases comprised normal colorectal mucosa. Adenocarcinoma group was classified based upon having *KRAS* and or *TP53* mutation. [§]2 not specified. ^{**}3 not tested. ^{^^}8 not tested, 2 failed sequencing. ^{*}3 not specified.

4.2.2 Dual-specificity phosphatase 6 (DUSP6) expression

4.2.2.1 *DUSP6 immunohistochemistry across the adenoma-carcinoma spectrum*

Immunohistochemistry against DUSP6 (methods 2.7.3) was carried out to explore its expression profile across the adenoma-carcinoma spectrum using the 147 cases from the cohort described previously, Fig. 33. Individual expression data presented in Appendix 8.1.2.

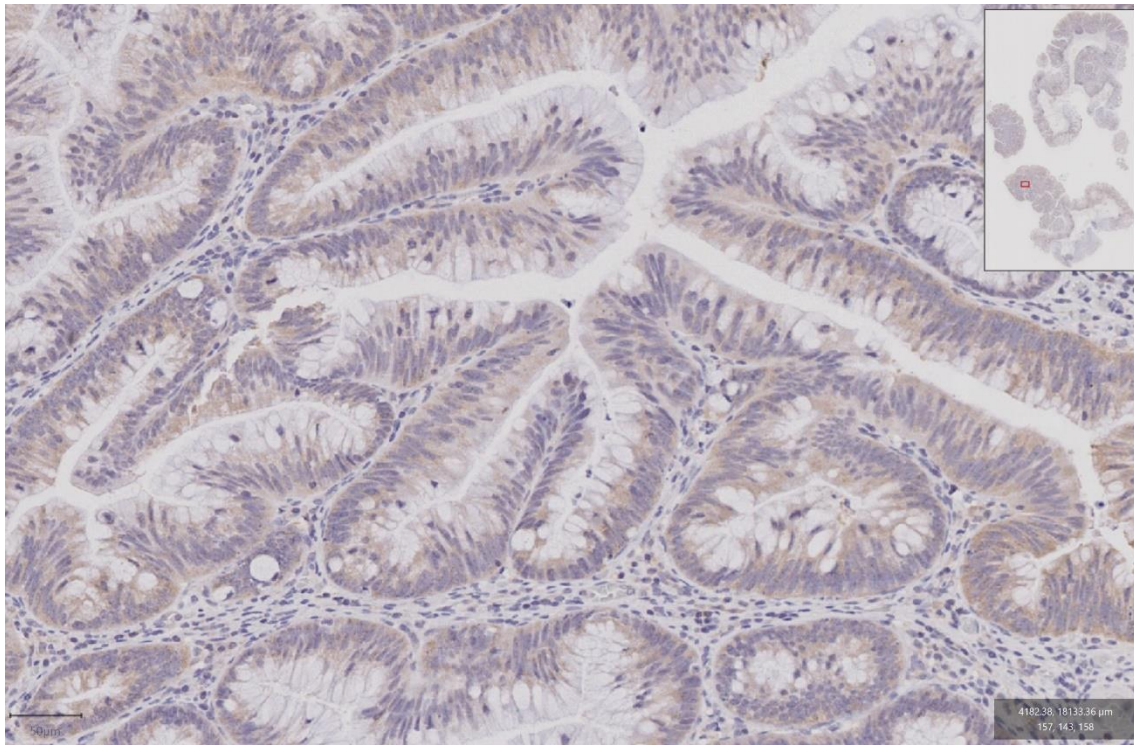


Figure 33. DUSP6 positive immunohistochemistry. Representative image of DUSP6 positive immunohistochemistry in tubular adenoma. Scale 50µm. Case 56.

For each case, DUSP6 expression (both nuclear and cytoplasmic) was classified by the ratio of cells (out of total assessed) with DUSP6 DAB intensity above an optimised threshold as described in materials and methods 2.8.4. Nuclear and cytoplasmic compartments were explored with respect to DUSP6 protein expression due to the ability of DUSP6 to shuttle between both compartments and the potential functional implications of expression demonstrating a dynamic range between the two, Fig. 34.

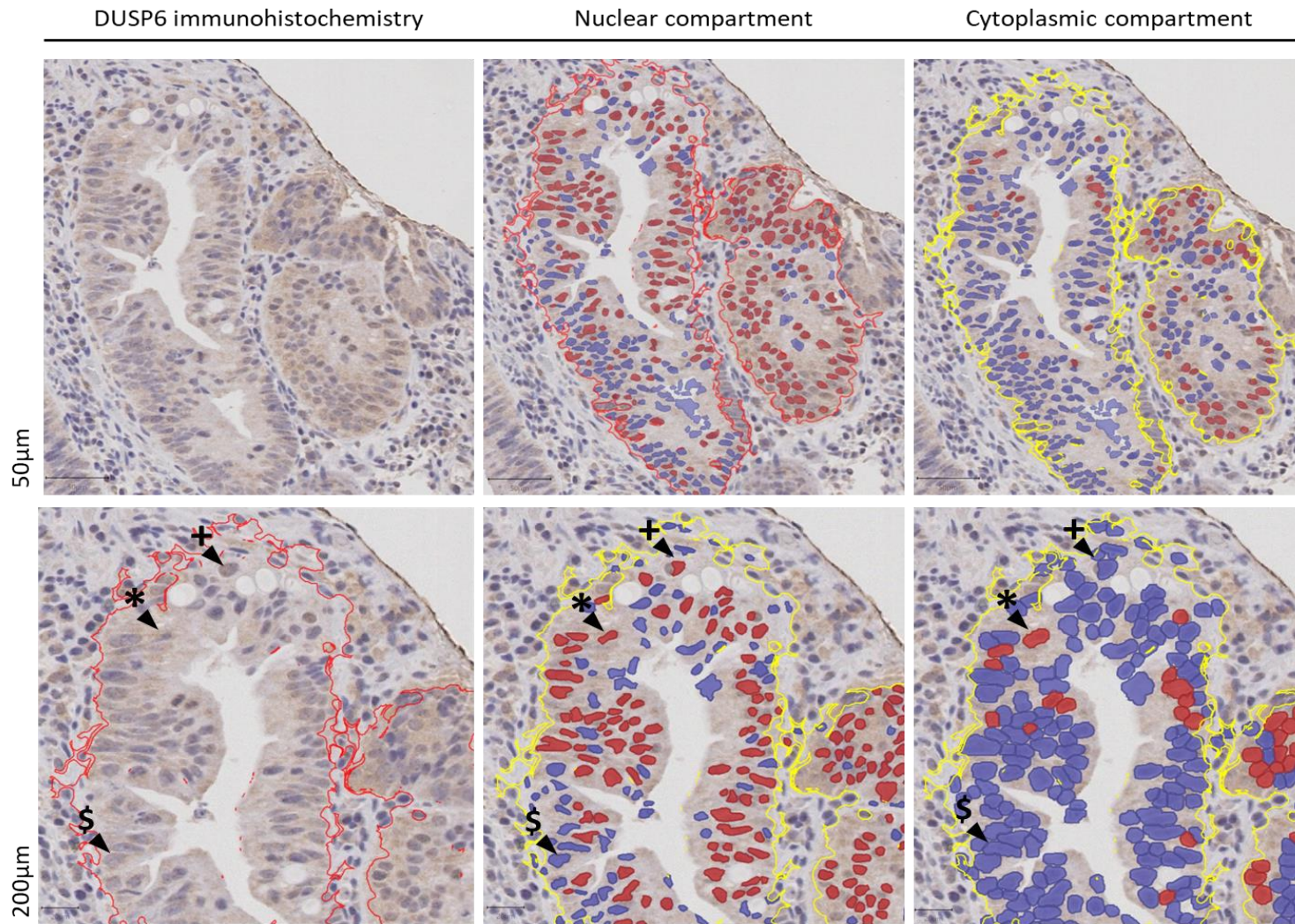


Figure 34. Example of QuPath annotation for DUSP6 immunohistochemistry. Positivity for DUSP6 immunohistochemistry was assessed in both nuclear and cytoplasmic compartments. *Nuclear and cytoplasmic compartment positive. +Nuclear compartment positive only. \$Cytoplasmic compartment positive only.

4.2.2.2 DUSP6 expression between histological classes

In the following analysis, DUSP6 positivity represents cells which express DUSP6 above this threshold. This value was representative of the degree of strong expression across the lesion, Table 19. Normality tests were carried out on the data set. The data set was not normally distributed as determined by Kolmogorov-Smirnov test, $p < 0.001$. Non-parametric statistical analysis was employed at the following levels of classification:

- Normal, adenoma and adenocarcinoma
- Normal, low grade dysplasia, high grade dysplasia and adenocarcinoma
- Normal, tubular, tubulovillous, villous and serrated adenoma, deficient MMR (dMMR), KRAS mutant, TP53 mutant and dual KRAS and TP53 mutant adenocarcinoma

Classification	Nuclear DUSP6 positivity		Cytoplasmic DUSP6 positivity	
	Median ratio	Inter-quartile range (IQR)	Median ratio	Inter-quartile range (IQR)
Normal	0.37	0.38	0.05	0.35
Adenoma	0.30	0.41	0.12	0.31
Adenocarcinoma	0.15	0.66	0.07	0.26
Subgroup 1				
Normal	0.37	0.38	0.05	0.12
Low grade dysplasia	0.26	0.39	0.09	0.19
High grade dysplasia	0.61	0.44	0.38	0.45
Adenocarcinoma	0.15	0.66	0.07	0.26
Subgroup 2				
Normal	0.37	0.38	0.05	0.12
Villous	0.51	0.49	0.15	0.31
Tubulovillous	0.30	0.40	0.14	0.27
Tubular	0.30	0.50	0.09	0.33
Serrated	0.18	0.28	0.01	0.06
KRAS mutant only	0.10	0.22	0.05	0.27
dMMR	0.07	0.30	0.01	0.08
TP53 mutant only	0.79	0.48	0.44	0.65
Dual KRAS and TP53 mutant	0.02	0.06	0.01	0.14

Table 19. Summary of median ratio of DUSP6 positive cells for subgroups. Median ratio of DUSP6 positive cells for both nuclear and cytoplasmic compartments plus inter-quartile range (IQR) is summarised. Median ratio per histological class is derived from the number of marker positive cells / total number of cells assessed per case.

4.2.3 Investigating DUSP6 in colorectal adenoma

4.2.3.1 DUSP6 positivity changes with degree of dysplasia

Colorectal adenomas demonstrate a diverse pathology with varying degrees of dysplasia. Adenomas in this study cohort were characterised by levels of dysplasia; low and high grade and changes in DUSP6 positivity were explored, Fig. 35. Statistical analysis for between group comparisons of grades of dysplasia in adenomas and degree of DUSP6 positivity for both nuclear and cytoplasmic compartments was carried out using a Mann-Whitney test. A significantly higher ratio of DUSP6 positive cells was observed in cases with high grade dysplasia (n18) in comparison to those with low grade dysplasia (n69) for both nuclear and cytoplasmic compartments, p0.007 (eta squared (η^2)=0.083: medium effect) and p0.001 (η^2 =0.133: medium effect) respectively, Fig. 36.

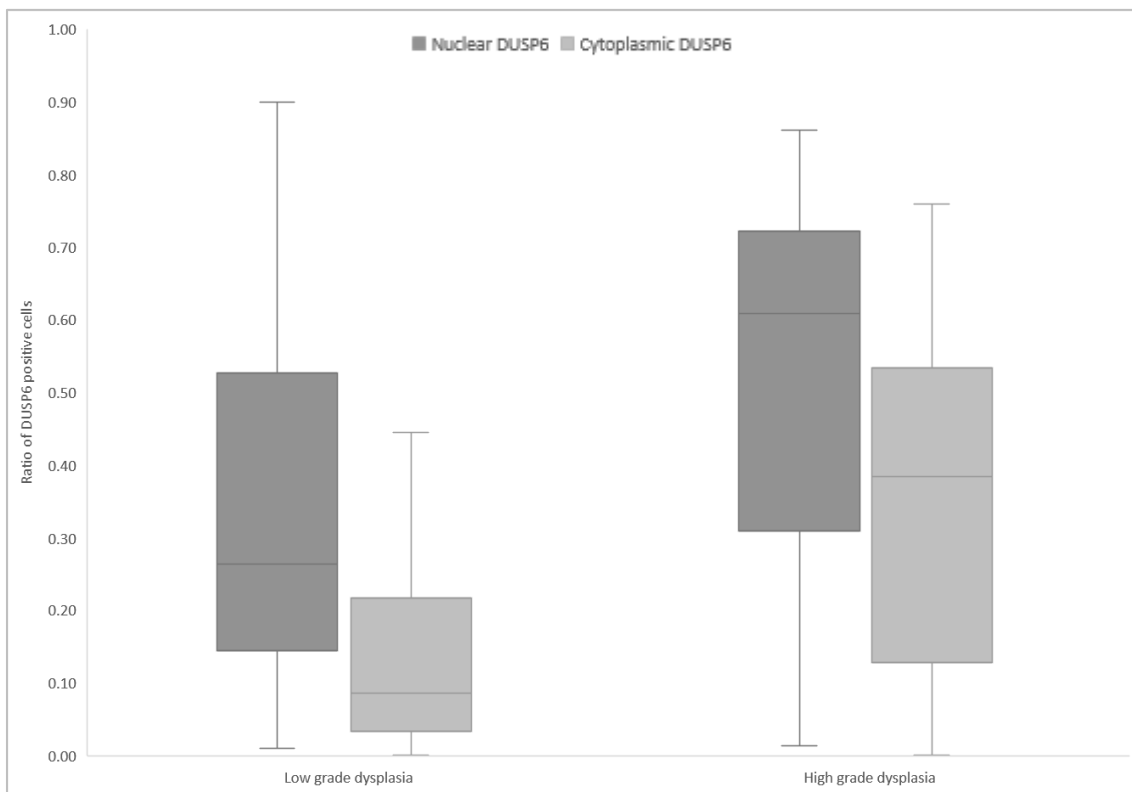


Figure 35. DUSP6 positivity across subgroups; normal, low grade dysplasia. The median ratio of DUSP6 positive cells (nuclear and cytoplasmic) by grade of dysplasia is plotted. Inter-quartile range (IQR) is represented. Statistical analysis for between group comparisons of dysplasia in adenomas and degree of DUSP6 positivity for both nuclear and cytoplasmic compartments was carried out using Mann-Whitney test. Significantly higher DUSP6 expression was observed in cases with high grade dysplasia (n18) than those with low grade dysplasia (n69) for nuclear; p0.007, IQR 0.44 and 0.39 respectively (eta squared (η^2)=0.083, medium effect) and cytoplasmic DUSP6; p0.001, IQR 0.45 and 0.19 respectively (η^2 =0.133, medium effect).

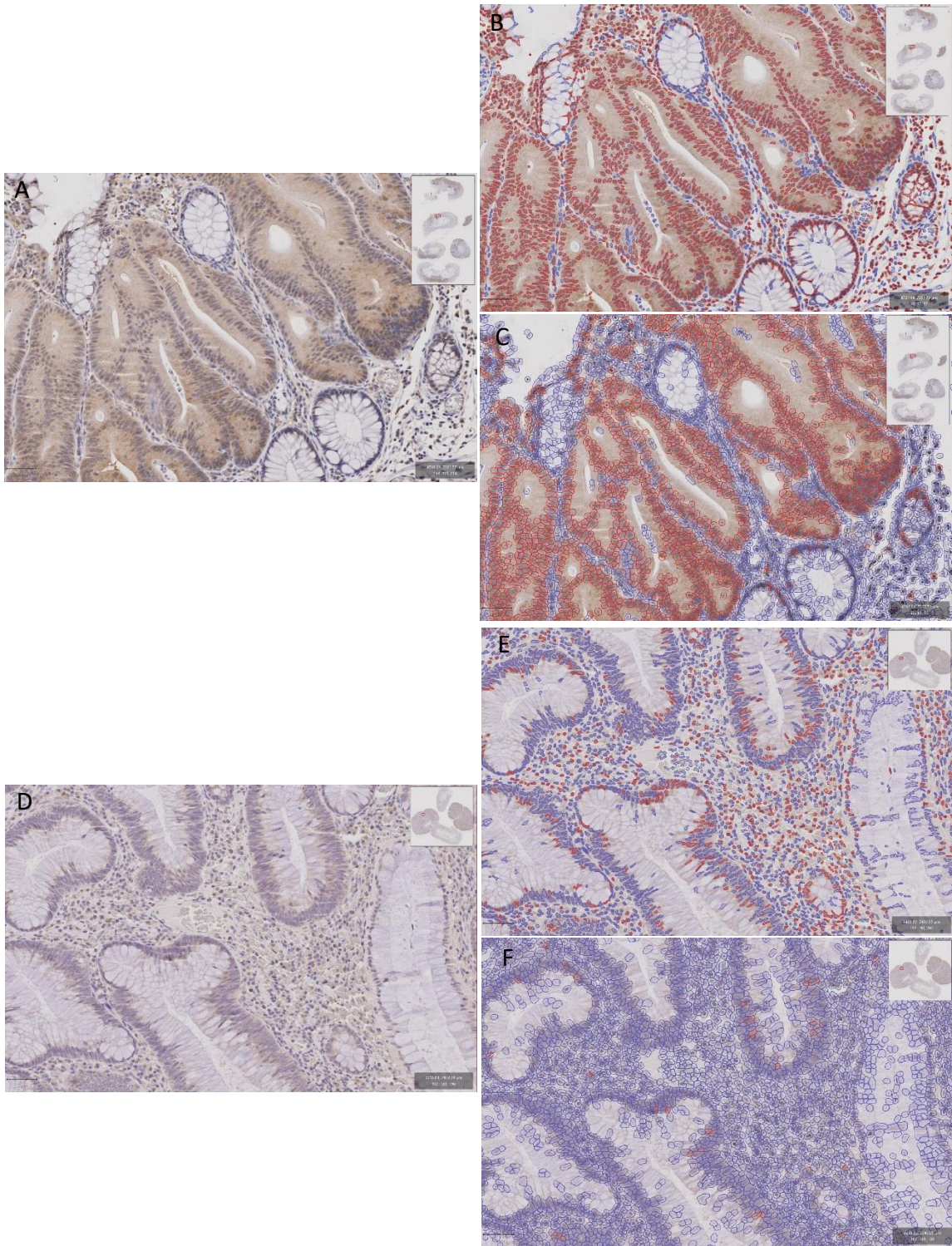


Figure 36. DUSP6 positivity. A-C: Case 7; tubular adenoma with high grade dysplasia. D-F: Case 20; tubular adenoma with low grade dysplasia. A & D: Immunohistochemical staining with primary antibody for DUSP6. B & E: QuPath nuclear compartment identification. C & F: Cytoplasmic compartment identification. High degree of nuclear and cytoplasmic DUSP6 positivity present in high grade dysplasia (red). Low proportion of nuclear and cytoplasmic positivity in low grade dysplasia.

4.2.3.2 Cytoplasmic DUSP6 positivity is significantly lower in serrated adenomas

Colorectal adenomas were characterised based upon histology i.e. tubular, tubulovillous, villous and serrated. Statistical analysis was carried out to determine differences in DUSP6 positivity (nuclear and cytoplasmic compartments) between histological classes using Kruskal-Wallis test. A significant difference between groups for both nuclear and cytoplasmic DUSP6 positivity was identified, $p=0.026$ and $p=0.003$ respectively, Fig. 37 and 38.

Analysis with Mann-Whitney test identified a significantly lower cytoplasmic DUSP6 positivity in serrated adenomas (n10) compared to tubular ($p=0.002$, n25, $\eta^2=0.09$, medium effect), tubulovillous ($p=0.001$, n33, $\eta^2=0.254$: large effect) and villous ($p=0.001$, n19, $\eta^2=0.408$: large effect) adenomas.

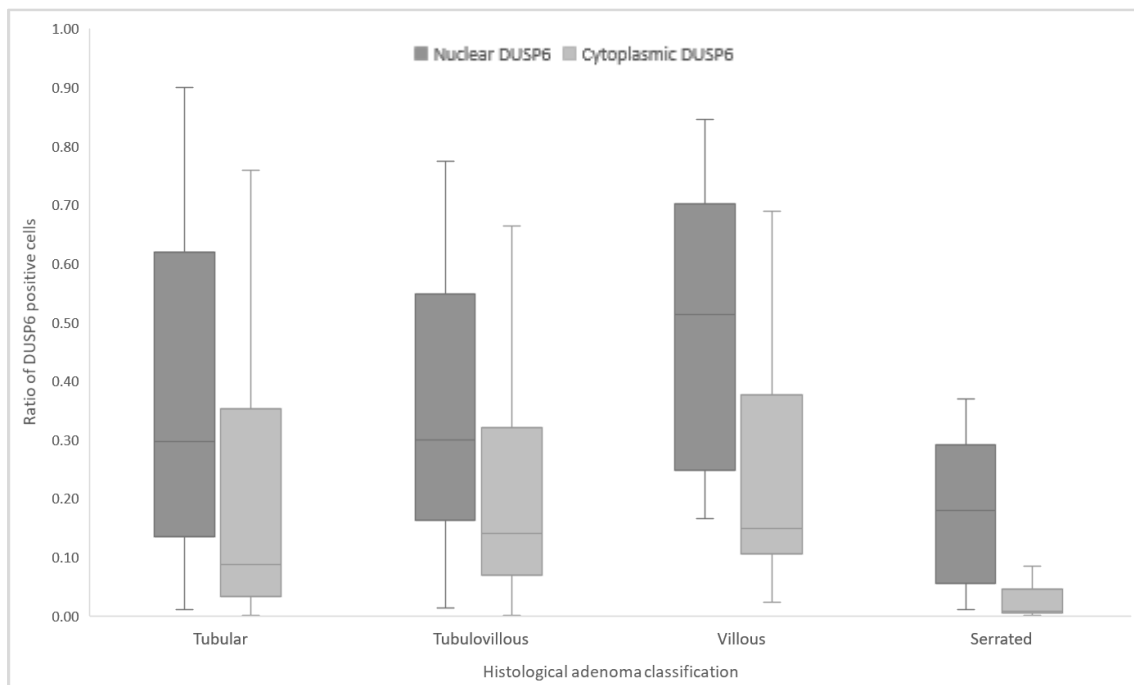


Figure 37. DUSP6 positivity across adenoma histological class. The median ratio of DUSP6 positive cells (nuclear and cytoplasmic) by histological class is plotted. Inter-quartile range (IQR) is represented. Statistical analysis for between group comparisons of adenoma histological class and degree of DUSP6 positivity for both nuclear and cytoplasmic compartments was carried out using Mann-Whitney test. A significantly lower ratio of cytoplasmic DUSP6 positivity was observed in serrated adenoma (n10) than tubular ($p=0.002$, n25, IQR 0.33, eta squared (η^2)=0.09: medium effect), tubulovillous ($p=0.001$, n33, IQR 0.27, $\eta^2=0.254$: large effect) and villous ($p=0.001$, n19, IQR 0.31, $\eta^2=0.408$: large effect).

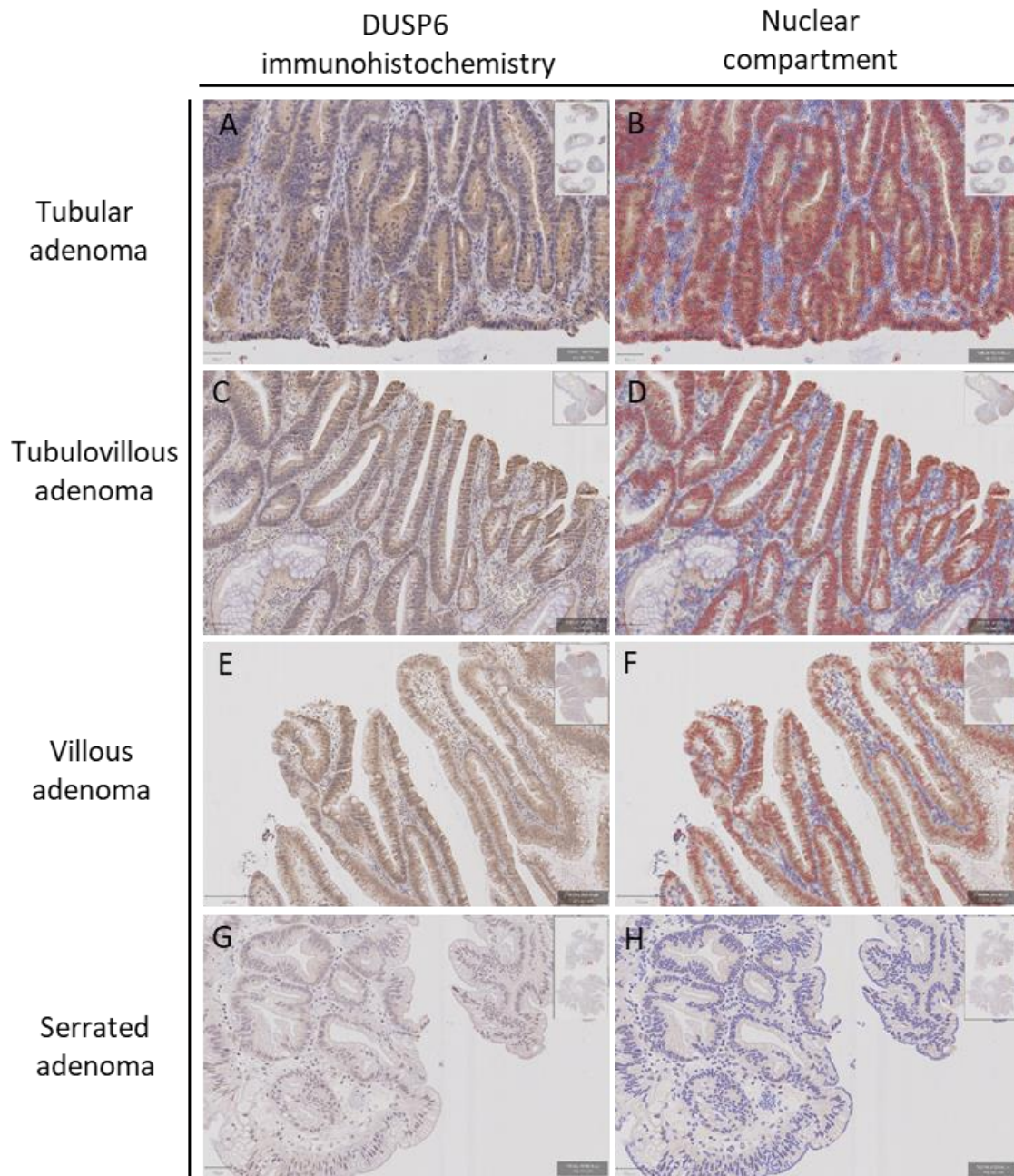


Figure 38. DUSP6 positivity across histological class of adenoma. A & B: Tubular adenoma, case 7. C & D: Tubulovillous adenoma, case 76. D & E: Villous adenoma, case 90. G & H: Serrated adenoma, case 50. Left: DUSP6 immunohistochemistry. Right: cytoplasmic compartmentation of cells; red denotes DUSP6 positivity. A significantly lower ratio of DUSP6 positive cells was identified in serrated adenomas compared with tubular, tubulovillous and villous adenomas.

4.2.4 Investigating DUSP6 in colorectal adenocarcinoma

4.2.4.1 *DUSP6* positivity is significantly higher in mutant *TP53* adenocarcinoma

No significant difference in DUSP6 positivity (nuclear or cytoplasmic) was observed between adenocarcinoma and adenomatous lesions. However, statistical analysis of DUSP6 positivity between molecular subtypes of adenocarcinoma as determined by Mann-Whitney test, identified a significantly higher DUSP6 positivity (nuclear and cytoplasmic) in *TP53* mutant adenocarcinomas compared to both dMMR ($p < 0.001$, $\eta^2 = 0.512$: large effect and $p < 0.001$, $\eta^2 = 0.534$: large effect respectively), *KRAS* mutant ($p = 0.008$, $\eta^2 = 0.264$: large effect and $p = 0.015$, $\eta^2 = 0.224$: large effect respectively) and dual *KRAS* and *TP53* mutant ($p = 0.001$, $\eta^2 = 0.443$: large effect and $p = 0.004$, $\eta^2 = 0.384$: large effect), Fig. 39 and 40.

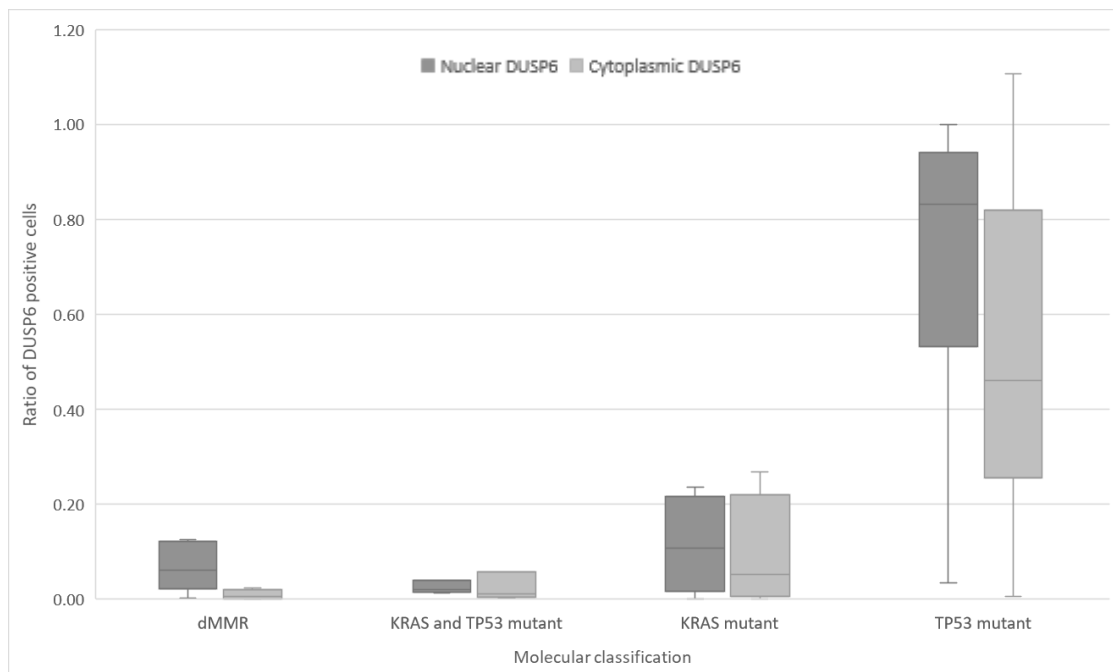


Figure 39. DUSP6 positivity across molecularly distinct adenocarcinoma groups. The median ratio of DUSP6 positive cells was assessed between molecular sub-groups of adenocarcinoma; deficient mismatch repair (dMMR), *KRAS* mutant and *TP53* mutant. A significantly higher ratio of DUSP6 positivity for both nuclear and cytoplasmic compartments of *TP53* lesions (median ratio of DUSP6 positive cells: nuclear 0.79, inter-quartile range 0.48 and cytoplasmic 0.44, inter-quartile range 0.65) compared to dMMR (nuclear, $p < 0.001$, median ratio of DUSP6 positive cells 0.07, inter-quartile range 0.30, $\eta^2 = 0.512$: large effect and cytoplasmic, $p < 0.001$, median ratio of DUSP6 positive cells 0.01, inter-quartile range 0.08, $\eta^2 = 0.534$: large effect), *KRAS* mutant (nuclear, $p = 0.008$, median ratio of DUSP6 positive cells 0.07, inter-quartile range 0.19, $\eta^2 = 0.264$ and cytoplasmic, $p = 0.015$, median ratio of DUSP6 positive cells 0.02, inter-quartile range 0.18, $\eta^2 = 0.224$: large effect) and dual *KRAS* and *TP53* mutant (nuclear, $p = 0.001$, $\eta^2 = 0.443$ and cytoplasmic $p = 0.004$, $\eta^2 = 0.384$: large effect) adenocarcinoma was identified by Mann-Whitney test.

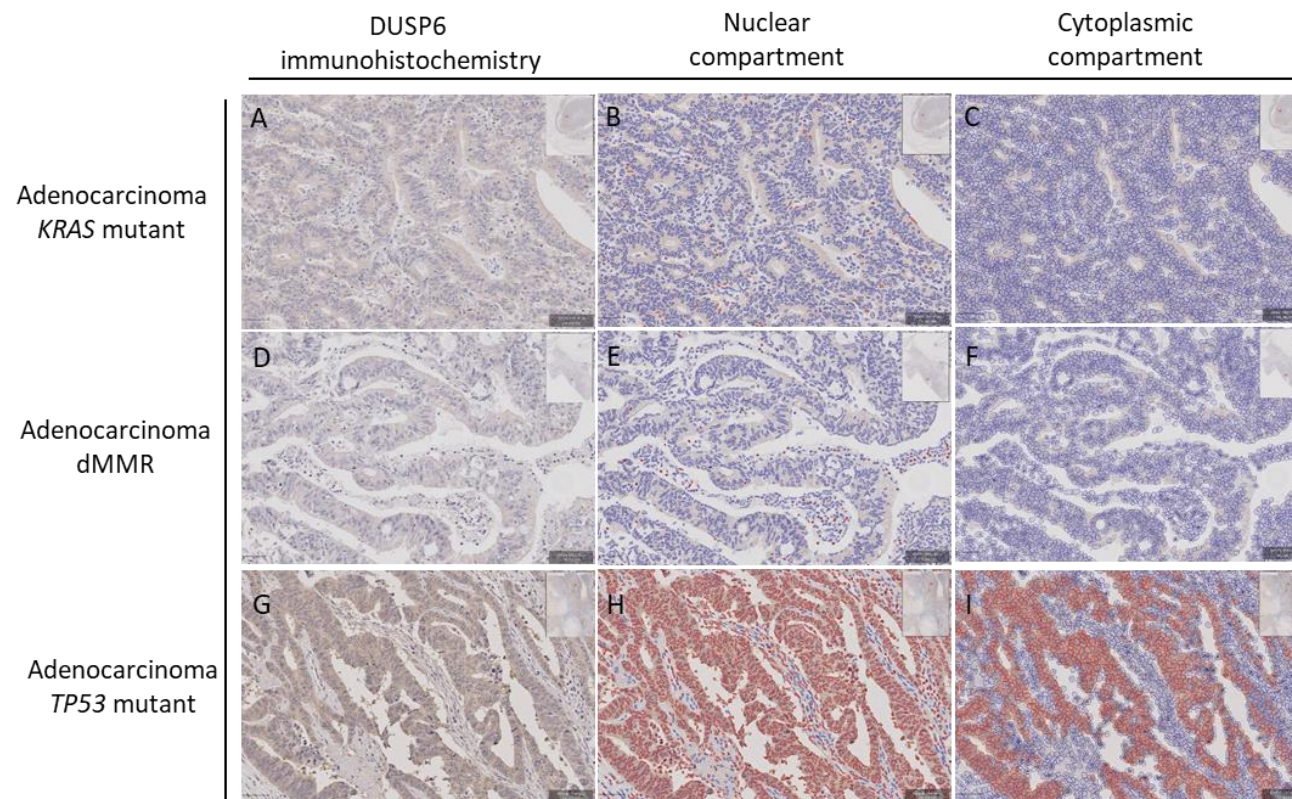


Figure 40. DUSP6 positivity in colorectal adenocarcinoma. A-C: Case 103, adenocarcinoma with dMMR. D-F: Case 118, adenocarcinoma with *KRAS* mutation. G-I: Case 134, adenocarcinoma with *TP53* mutation. Left column: DUSP6 immunohistochemistry. Middle column: nuclear DUSP6 positivity with annotations. Right column: cytoplasmic DUSP6 positivity with annotations. Adenocarcinomas with *TP53* mutation demonstrated the highest ratio of nuclear and cytoplasmic DUSP6 positivity compared to those with deficient mis-match repair (dMMR) or *KRAS* mutation.

4.2.5 Investigating *KRAS* mutation and associations with DUSP6 expression

4.2.5.1 *DUSP6* positivity is significantly lower in *KRAS* mutant adenocarcinoma

Adenocarcinoma cases were classified based upon *KRAS* mutation status (*mutant* *n*15, *wildtype* *n*27) and statistical analysis into the association of *KRAS* mutation and the ratio of DUSP6 positive cells was carried out using Mann-Whitney test.

A significantly lower ratio of nuclear DUSP6 positive cells was identified in *KRAS* mutant cases (median ratio of nuclear DUSP6 positive cells 0.07, inter-quartile range 0.68) compared to *KRAS* wildtype (median ratio of nuclear DUSP6 positive cells 0.48, inter-quartile range 0.82), *p*0.046, $\eta^2=0.096$: medium effect.

With evidence in literature for a relationship between RAS mutation and ERK it was not surprising to identify that the ratio of ERK positive cells was significantly higher in *KRAS* mutant individuals (median ratio of ERK positive cells 0.35, inter-quartile range 0.43) compared to wildtype (median ratio of ERK positive cells 0.07, inter-quartile range 0.27), *p*0.004, $\eta^2=0.197$: large effect. To explore whether this phenomenon is present in adenoma, pyrosequencing of *KRAS* exon 2 and 3 was carried out on the adenoma cases in this study cohort.

4.2.5.2 *DUSP6* positivity shows no association with *KRAS* mutation status in adenoma

The adenoma cohort had been previously assessed for the presence of *KRAS* mutations by pyrosequencing, Chapter. 3. As performed in adenocarcinoma, a Mann-Whitney test was carried out to assess an association between the ratio of DUSP6 positive cells and *KRAS* mutation status for which no significant difference was identified indicating no association between DUSP6 expression and *KRAS* mutation status.

4.2.6 Investigating DUSP6 and adenocarcinoma type

In addition to histological subtypes of colorectal adenoma, the adenocarcinoma cases in this cohort were classified by distinct histological subtypes: adenocarcinoma, mucinous adenocarcinoma, medullary adenocarcinoma and signet ring cell adenocarcinoma. Investigations were carried out to assess association between DUSP6 expression and these subtypes.

Initial analysis was carried out to assess associations between DUSP6 expression in adenocarcinoma and adenocarcinoma special type (mucinous, medullary and signet ring cell) using Mann-Whitney test. A significantly higher DUSP6 expression was identified in both nuclear and cytoplasmic compartments for adenocarcinoma compared to 'special type', nuclear: $p=0.025$, $\eta^2=0.17$: medium effect, median 0.5 vs 0.07 respectively and cytoplasmic: $p=0.002$, $\eta^2=0.22$: medium effect, median 0.3 vs 0.08 respectively. To explore this association further Mann-Whitney test was carried out for DUSP6 expression between adenocarcinoma and each 'special type', Table. 20. A significantly lower cytoplasmic DUSP6 expression was identified in mucinous adenocarcinoma (n6, median 0.003) compared to adenocarcinoma (n25, median 0.25) with a medium effect ($\eta^2=0.22$) however additional cases would be required to confirm this finding due to the limited numbers of 'special type' adenocarcinoma in this study.

	Mucinous	Medullary	Signet ring cell
Nuclear	0.53	0.33	0.08
Cytoplasmic	0.007* (0.22)	0.098	0.05

Table 20. Statistical analysis of DUSP6 expression between adenocarcinoma and adenocarcinoma 'special type'. A significantly lower DUSP6 expression was identified in mucinous adenocarcinoma compared to adenocarcinoma $p=0.007$. Due to the limited sample size of mucinous adenocarcinoma additional cases are required to confirm these findings. Bonferroni corrected alpha value 0.016. *Significant association at corrected alpha value. () effect size.

4.2.7 Extracellular regulated kinase (ERK) and phospho-ERK (p-ERK) expression

4.2.7.1 ERK and p-ERK immunohistochemistry across the adenoma-carcinoma sequence

In order to explore the relationship of DUSP6 with ERK and p-ERK across the adenoma-carcinoma sequence, ERK and p-ERK immunohistochemistry was carried out on whole sections, Fig. 41. The expression of both markers was assessed by QuPath image analysis software, Appendix 8.1.2.

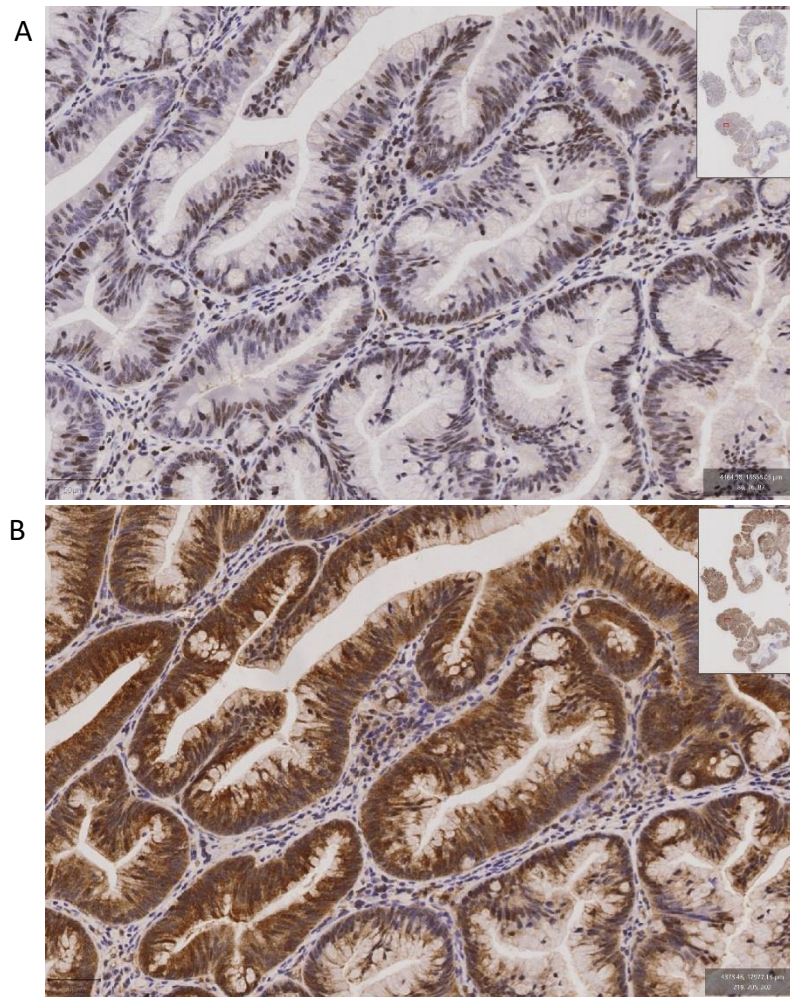


Figure 41. p-ERK and ERK positive immunohistochemistry. A: Representative image of p-ERK positivity in nuclei of tubular adenoma. B: ERK immunohistochemistry in cytoplasmic of tubular adenoma. Scale 50µm.

4.2.7.2 ERK and p-ERK expression between histological classes

The median ratio of marker positive cells for ERK within cytoplasmic and p-ERK within nuclear compartments was assessed for each subgroup, Table 21. Statistical analysis was carried out at the following levels of classification:

- Normal, adenoma and adenocarcinomas
- Normal, low grade dysplasia, high grade dysplasia and adenocarcinoma
- Normal, tubular, tubulovillous, villous and serrated adenoma, dMMR, *KRAS* mutant, *TP53* mutant adenocarcinoma and dual *KRAS* and *TP53* mutant adenocarcinoma.

Classification	ERK positivity		p-ERK positivity	
	Median ratio	Inter-quartile range (IQR)	Median ratio	Inter-quartile range (IQR)
Normal	0.53	0.45	0.02	0.27
Adenoma	0.86	0.23	0.29	0.38
Adenocarcinoma	0.16	0.47	0.01	0.03
Normal	0.53	0.45	0.02	0.27
Low grade dysplasia	0.86	0.22	0.29	0.39
High grade dysplasia	0.84	0.28	0.27	0.42
Adenocarcinoma	0.16	0.47	0.01	0.03
Normal	0.53	0.45	0.02	0.27
Villous	0.89	0.15	0.25	0.56
Tubulovillous	0.83	0.24	0.30	0.36
Tubular	0.89	0.18	0.30	0.28
Serrated	0.72	0.46	0.07	0.19
<i>KRAS</i> mutant	0.35	0.52	0.01	0.04
dMMR	0.03	0.11	0.00	0.01
<i>TP53</i> mutant	0.19	0.55	0.02	0.06
Dual <i>KRAS</i> and <i>TP53</i> mutant	0.34	0.50	0.04	0.63

Table 21. Summary of median ratio of ERK and p-ERK positive cells for subgroups. Median ratio of ERK and p-ERK positive cells ((ERK) cytoplasmic and (p-ERK) nuclear) plus inter-quartile range is summarised. Ratio expressed as number of marker positive cells / total number of cells assessed.

4.2.7.3 ERK and p-ERK positivity is significantly higher in adenoma but is unchanged between degree of dysplasia

The ratio of ERK (cytoplasmic) and p-ERK (nuclear) positive cells was assessed between normal, adenoma and adenocarcinoma tissue, Fig. 42. Adenomas demonstrated the highest ratio of ERK and p-ERK positivity compared to normal mucosa and adenocarcinoma, Fig. 43. This ratio was not statistically different between low and high grade dysplasia, Table 22.

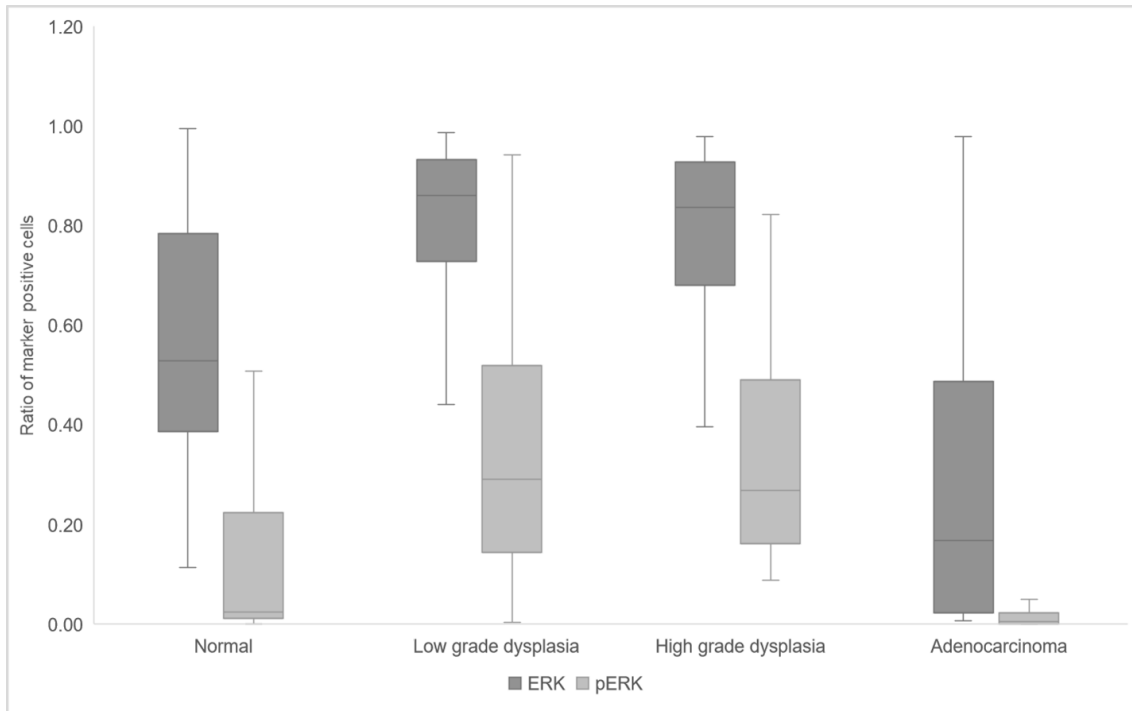


Figure 42. ERK and p-ERK expression across subgroups; normal, low grade dysplasia, high grade dysplasia and adenocarcinoma. The median ratio of cytoplasmic ERK and nuclear p-ERK positive cells per group. IQR is represented. Kruskal-Wallis test identified a significant difference in the ratio of cytoplasmic ERK and nuclear p-ERK positive cells across subgroups ($p < 0.001$ and $p < 0.001$ respectively). Adenomas demonstrated the highest ERK and p-ERK positivity.

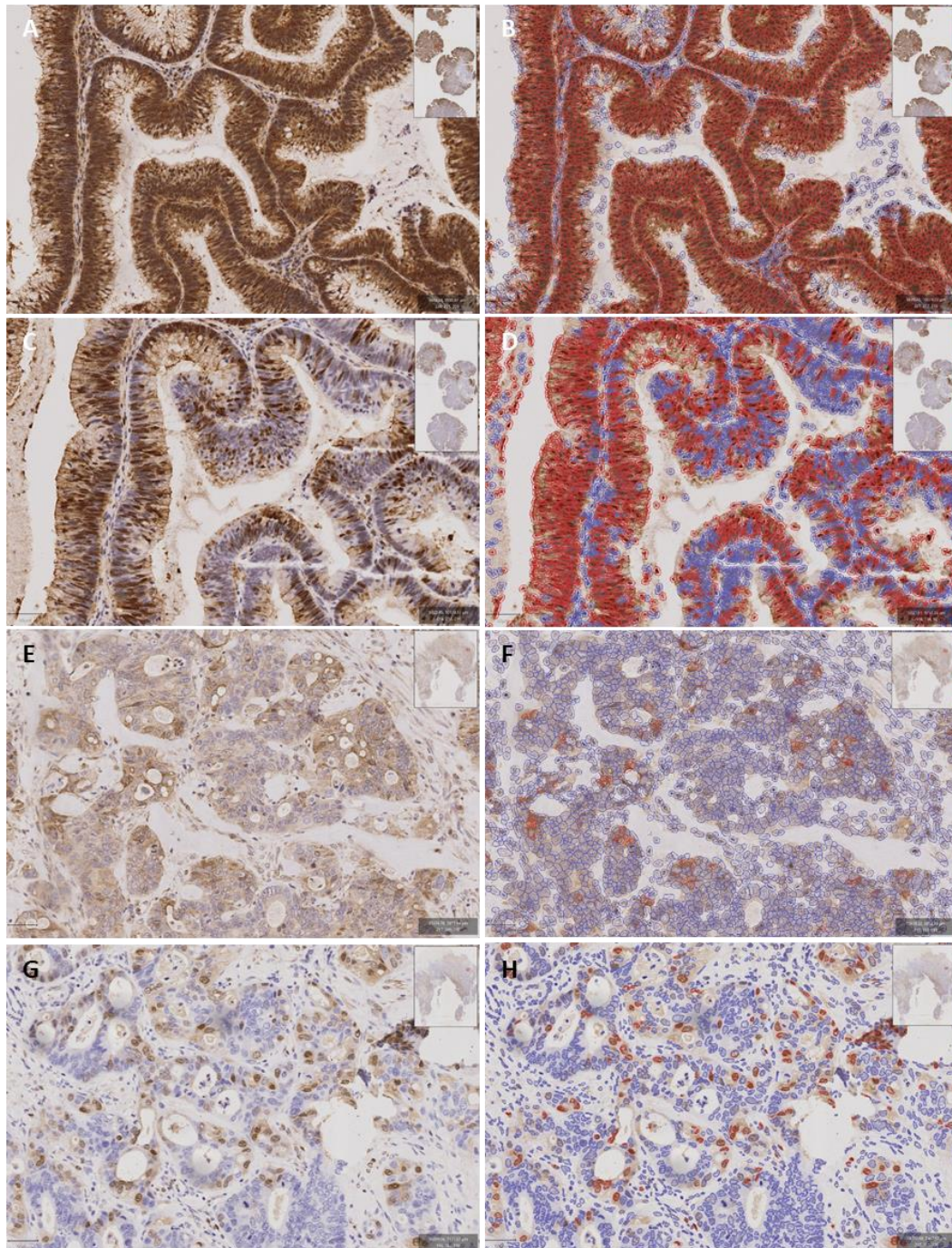


Figure 43. ERK and p-ERK immunohistochemistry. A-D: Case 14, tubulovillous adenoma. E-H: Case 126, KRAS mutant adenocarcinoma. Left column: immunohistochemistry. Right column: marker positive cells with annotations. B & F: ERK positive cells (cytoplasm). D & H: p-ERK positive cells (nuclear). Positive cells = red, negative cells = blue. Adenoma cases demonstrated a significantly higher ERK and p-ERK positivity than adenocarcinoma cases. The ratio of ERK and p-ERK positive cells in adenocarcinoma was lower than both adenoma and normal mucosa.

ERK				
Subgroup	Normal	Low grade dysplasia	High grade dysplasia	Adenocarcinoma
Normal		0.004* (0.097)	0.033	0.001* (0.171)
Low grade dysplasia	0.004* (0.097)		0.69	<.001* (0.443)
High grade dysplasia	0.033	0.69		<.001* (0.366)
Adenocarcinoma	0.001* (0.171)	<0.001* (0.443)	<0.001* (0.366)	

p-ERK				
Subgroup	Normal	Low grade dysplasia	High grade dysplasia	Adenocarcinoma
Normal		0.003* (0.097)	0.005* (0.235)	0.007* (.098)
Low grade dysplasia	0.003* (0.097)		0.863	<0.001* (0.481)
High grade dysplasia	0.005* (0.235)	0.863		<0.001* (0.504)
Adenocarcinoma	0.007* (0.098)	<0.001* (0.481)	<0.001* (0.504)	

Table 22. Mann Whitney test for difference in the ratio of ERK and p-ERK positive cells between subgroups. A significant difference between subgroups; normal, low grade dysplasia, high grade dysplasia and adenocarcinoma was identified by Kruskal-Wallis test. To further investigate these statistical differences Mann-Whitney test was applied to each pair of subgroups. Significant differences in the ratio of ERK and p-ERK positive cells was identified between all groups excluding low grade dysplasia-high grade dysplasia comparison. * Bonferroni adjusted alpha level 0.0125. () eta squared (η^2) effect size: 0.01 = small effect, 0.09 = medium effect, 0.25 = large effect.

4.2.7.4 Serrated adenomas have lower p-ERK positivity than other adenoma groups

The ratio of ERK positive cells was comparable across all adenoma groups. Serrated adenomas demonstrated a significantly lower ratio of p-ERK positive cells than tubular: $p < 0.001$ ($\eta^2 = 0.3$, large effect), and tubulovillous: $p < 0.001$ ($\eta^2 = 0.3$, large effect) adenomas, Table 23. The ratio of p-ERK positive cells in serrated adenoma was comparable to adenocarcinoma groups, Fig. 44.

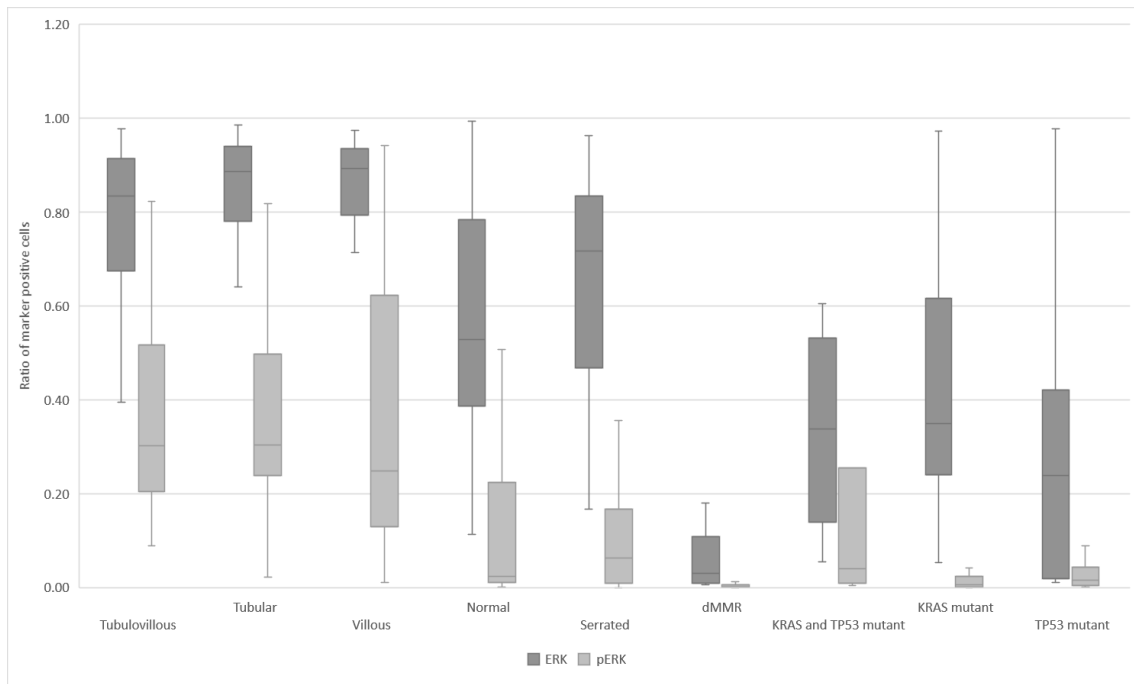


Figure 44. ERK and p-ERK expression across subgroups; normal, tubular, tubulovillous, villous, serrated, dMMR, KRAS mutant, TP53 mutant, dual mutant. The median ratio of ERK(c) and p-ERK(n) positive cells per group. IQR is represented. Kruskal-Wallis test identified a significant difference in the ratio of ERK and p-ERK positive cells between groups ($p < 0.001$, $\eta^2 = 0.3$, large effect and $p < 0.001$, $\eta^2 = 0.3$, large effect).

ERK								
Histological class	Normal	Villous	Tubulovillous	Tubular	Serrated	KRAS mutant	dMMR	TP53 mutant
Normal		0.008	0.011	0.006* (0.2)	0.338	0.187	0.001* (0.5)	0.011
Villous	0.008		0.216	0.803	0.045	<0.001* (0.4)	<0.001* (0.7)	<0.001* (0.4)
Tubulovillous	0.011	0.216		0.275	0.157	<0.001* (0.3)	<0.001* (0.6)	<0.001* (0.3)
Tubular	0.006* (0.2)	0.803	0.275		0.083	<0.001* (0.4)	<0.001* (0.6)	<0.001* (0.4)
Serrated	0.338	0.045	0.157	0.083		0.041	<0.001* (0.5)	0.008
KRAS mutant	0.187	<0.001* (0.4)	<0.001* (0.3)	<0.001* (0.4)	0.041		<0.001* (0.4)	0.148
dMMR	0.001* (0.5)	<0.001* (0.7)	<0.001* (0.6)	<0.001* (0.6)	<0.001* (0.5)	<0.001* (0.4)		0.126
TP53 mutant	0.011	<0.001* (0.4)	<0.001* (0.4)	<0.001* (0.4)	0.008	0.148	0.126	

p-ERK								
Histological class	Normal	Villous	Tubulovillous	Tubular	Serrated	KRAS mutant	dMMR	TP53 mutant
Normal		0.006* (0.2)	0.001* (0.2)	0.004* (0.2)	0.892	0.106	0.001* (0.4)	0.186
Villous	0.006* (0.2)		0.790	0.896	0.007	<0.001* (0.5)	<0.001* (0.5)	<0.001* (0.5)
Tubulovillous	0.001* (0.2)	0.790		0.863	<0.001* (0.3)	<0.001* (0.5)	<0.001* (0.6)	<0.001* (0.5)
Tubular	0.004* (0.2)	0.896	0.863		0.001* (0.3)	<0.001* (0.5)	<0.001* (0.7)	<0.001* (0.5)
Serrated	0.892	0.007	<0.001* (0.3)	0.001* (0.3)		0.196	0.016	0.259
KRAS mutant	0.106	<0.001* (0.5)	<0.001* (0.5)	<0.001* (0.5)	0.196		0.089	0.715
dMMR	0.001* (0.4)	<0.001* (0.5)	<0.001* (0.6)	<0.001* (0.7)	0.016	0.089		0.041
TP53 mutant	0.186	<0.001* (0.5)	<0.001* (0.5)	<0.001* (0.5)	0.259	0.715	0.041	

Table 23. Mann Whitney test for difference in the ratio of ERK and p-ERK positive cells between subgroups; normal, tubular, tubulovillous, villous, serrated, dMMR, KRAS mutant, TP53 mutant. Mann-Whitney test identified a significant difference in the ratio of ERK and p-ERK positive cells between a number of groups. A large proportion of comparisons between adenocarcinoma groups and adenoma groups were identified as significant with a large effect. * Bonferroni adjusted alpha level .00625. () eta squared (η^2) effect size. 0.01: small effect, 0.09: medium effect, 0.25: large effect.

4.2.8 Exploring the relationship between DUSP6 and p-ERK protein expression

As evidenced in literature; DUSP6 is a direct negative regulator of ERK. One aspect of this relationship is the shuttling of activated ERK (p-ERK) from the nucleus to the cytoplasm. The ratios of nuclear DUSP6 positive cells and p-ERK positive cells were plotted in order to investigate whether an association between the ratio is evident in this study cohort, Fig 44.

4.2.8.1 Nuclear DUSP6 positivity is significantly higher than p-ERK in high grade dysplasia

The ratio (derived from previous analysis in 3.2.2 and 3.2.6.1) of nuclear DUSP6 positive cells was plotted against the ratio of p-ERK positive cells for both low and high grade dysplasia. No significant difference was identified in low grade dysplasia between the two targets, Fig. 45a. In high grade dysplasia, nuclear DUSP6 positivity was significantly higher than p-ERK positivity in 77% of cases as determined by Wilcoxon-Signed Ranks test, $p=0.026$, $\eta^2=0.6$ (large effect) (1-tailed Monte-Carlo significance), Fig. 45b and 46.

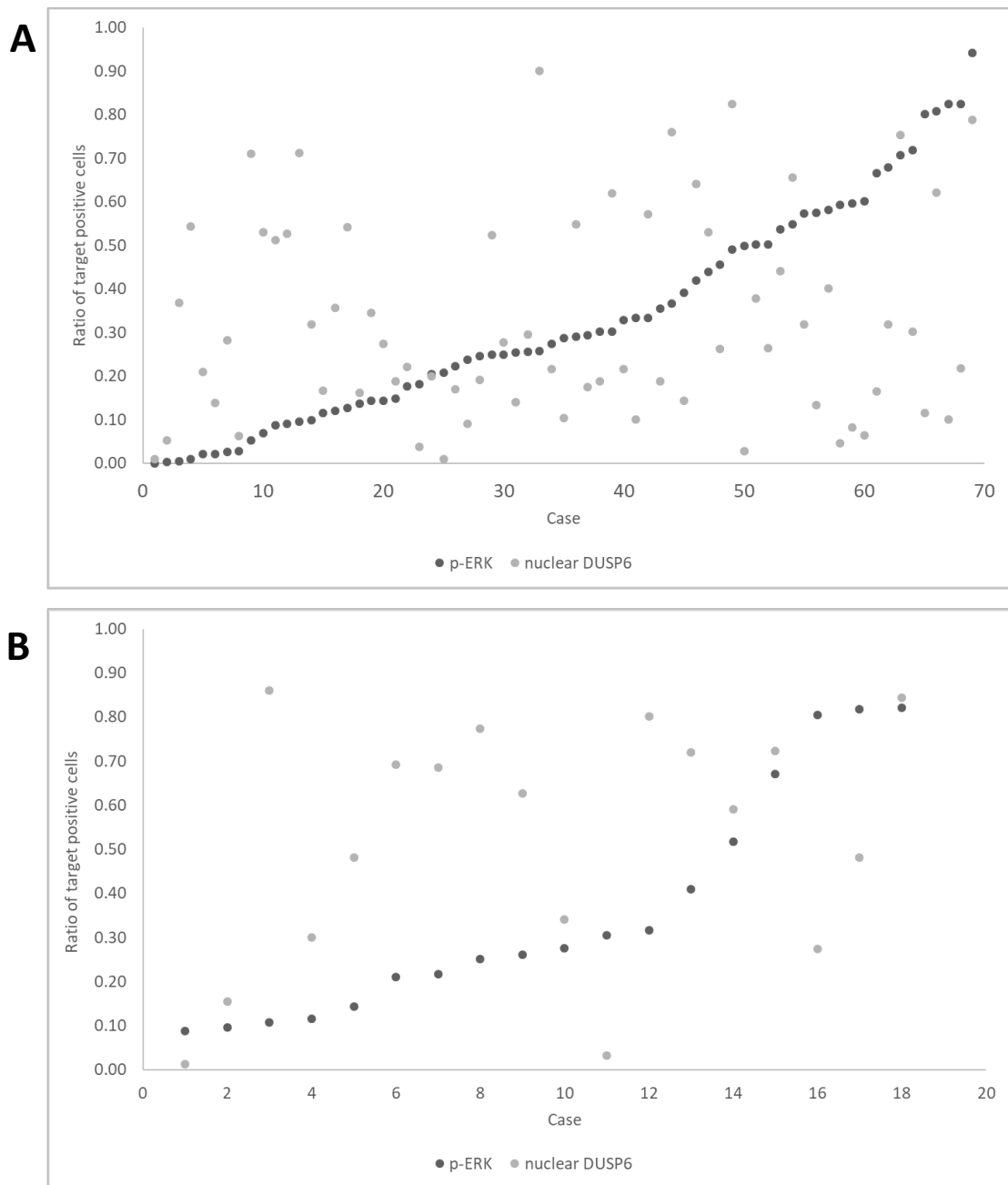


Figure 45. Ratio of p-ERK positivity and nuclear DUSP6 positivity in low and high grade dysplasia. The ratio of p-ERK positive cells was plotted against the ratio of nuclear DUSP6 positive cells per case for both (A) low and (B) high grade dysplasia. Ratios from previous analysis of individual markers was used. No significant difference between the ratio of p-ERK positive cells and nuclear DUSP6 cells was identified in low grade dysplasia. In high grade dysplasia, the ratio of p-ERK positive cells exceeded nuclear DUSP6 in 77% of cases, this was statistically significant as determined by Wilcoxon Signed Rank test, p.026 (1-tailed Monte-Carlo significance).

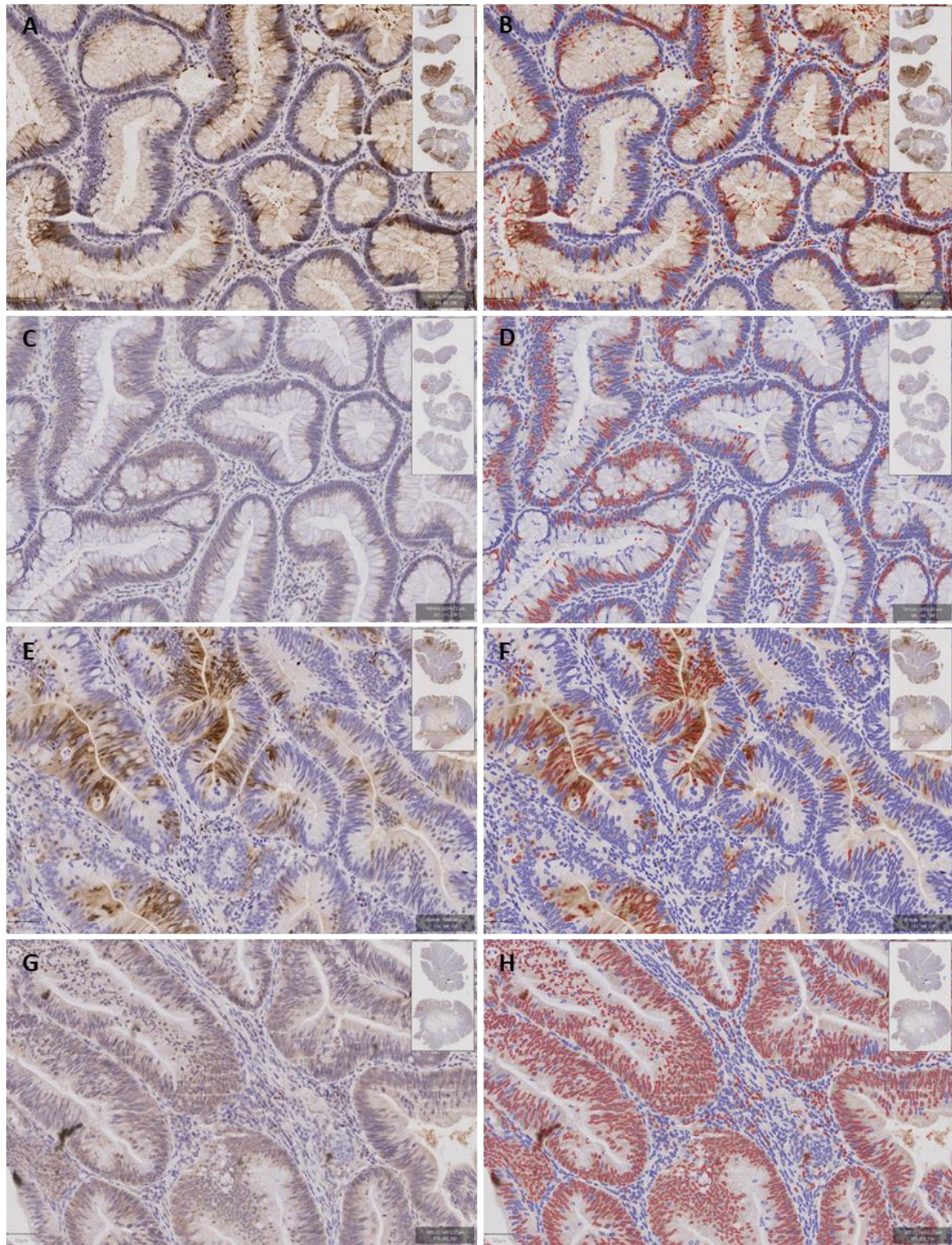


Figure 46. p-ERK and DUSP6 nuclear immunohistochemistry. A-D: Case 26; villous adenoma with low grade dysplasia. E-H: Case 8; tubulovillous adenoma with high grade dysplasia. Left column: immunohistochemistry. Right column: Marker positivity with annotations. Red = marker positive cells. Low and high grade dysplasia demonstrated comparative ratios of p-ERK positivity however, high grade dysplasia had a higher nuclear DUSP6 positivity in comparison to low grade dysplasia. This trend was not statistically significant.

4.2.8.2 Nuclear DUSP6 positivity is higher than p-ERK in adenocarcinoma

To explore whether the findings in high grade dysplasia were evident in adenocarcinoma, the ratio of p-ERK positive cells was plotted against the ratio of nuclear DUSP6 positive cells for all adenocarcinoma cases, Fig. 47.

The ratio of nuclear DUSP6 positive cells was significantly higher than the ratio of p-ERK positive cells for 84% of cases, $p < 0.001$, $\eta^2 = 0.5$ (large effect) (Wilcoxon-Signed Rank test, 1-tailed Monte-Carlo significance). In contrast to adenoma cases (median ratio of p-ERK positive cells 0.029, inter-quartile range 0.038), the ratio of p-ERK positive cells in adenocarcinoma was close to negative (median ratio of p-ERK positive cells 0.01, inter-quartile range 0.03).

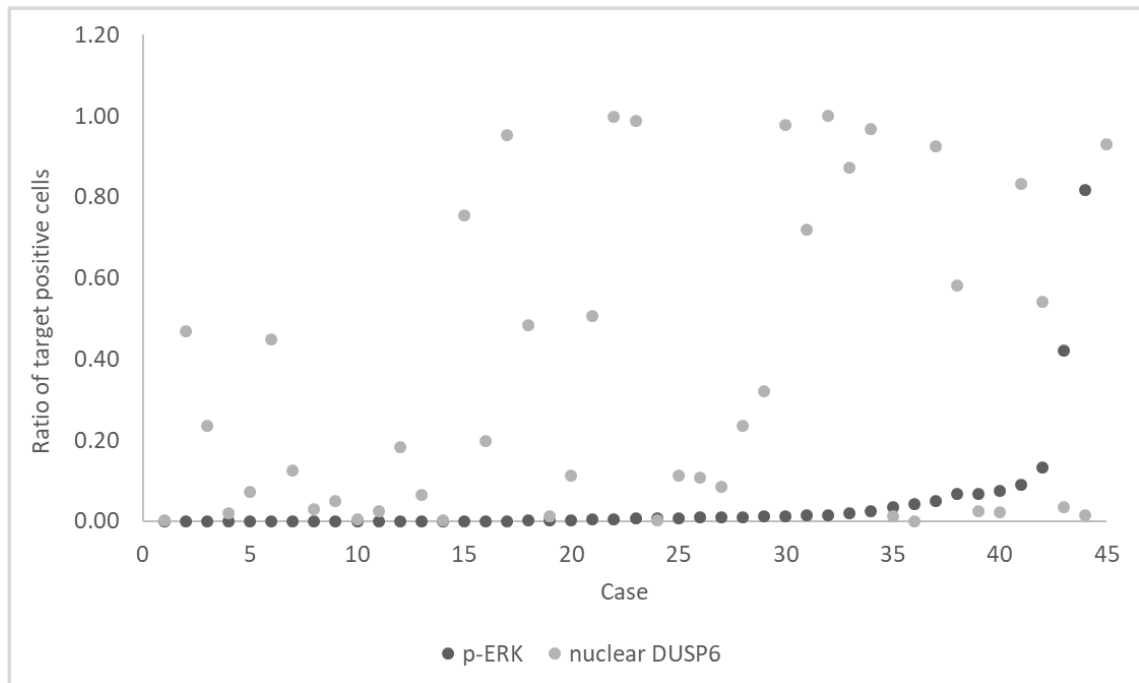


Figure 47. Ratio of p-ERK positive cells and DUSP6 positive cells in adenocarcinoma. The ratio of p-ERK positive cells was plotted against the ratio of nuclear DUSP6 positive cells per case for adenocarcinoma cases (n45). Statistical analysis with Wilcoxon-Signed Rank test identified a significantly higher ratio of nuclear DUSP6 positive cells than p-ERK in 88% of cases, $p < 0.001$, eta squared (η^2)=0.5 (large effect), (1-tailed Monte-Carlo significance).

4.2.9 Characterisation of DUSP6 function *in vitro*

4.2.9.1 *RG/C2/80 adenoma cells as a model cell line for DUSP6 transduction*

In order to investigate the functional relationship between DUSP6 and ERK *in vitro* the adenoma cell line RG/C2/80 was used, Fig. 48. RG/C2/80 originate from a sporadic colorectal adenoma with tubular histology¹¹⁵.

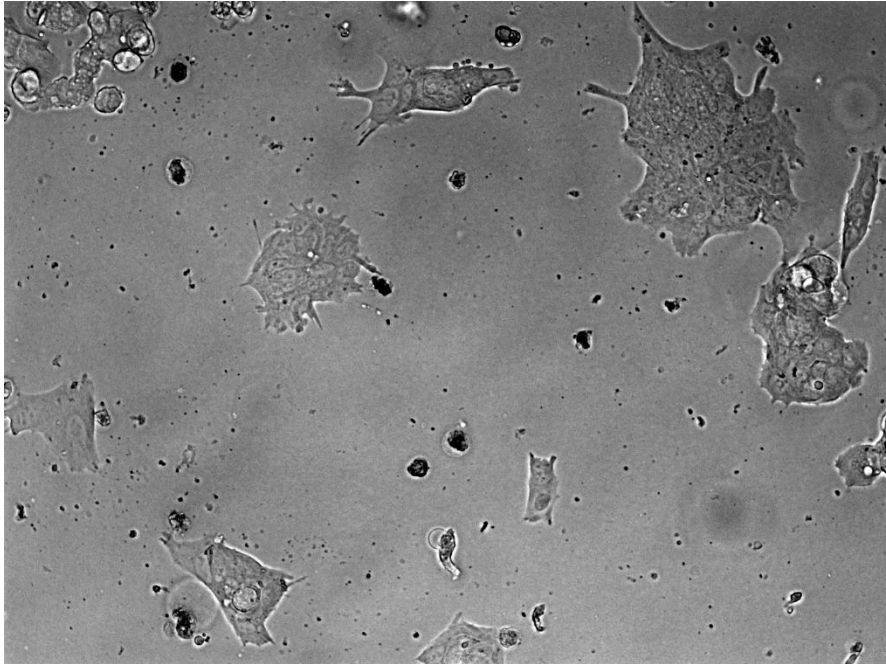


Figure 48. RG/C2/80 adenoma cells. Brightfield image of RG/C2/80 adenoma cells. x20 mag. Cells are derived from a sporadic colorectal adenoma with tubular histology.

4.2.9.2 Confirmation of DUSP6 transduction in RG/C2/80 cells

Initial screening of DUSP6 expression in RG/C2/80 by western blotting analysis cells demonstrated that undetectable levels of isoform A and high levels of isoform B were present in this cell line, making these cells an ideal candidate for overexpression of the canonical form of the protein.

RG/C2/80 cells were successfully transduced with a pLEX-MCS plasmid expressing full length DUSP6, Fig. 49. A 3-fold increase in DUSP6 isoform A expression was observed in RG/C2/80 cells containing the pLEX-MCS plasmid expressing full length DUSP6 in comparison to unmanipulated parental lines.

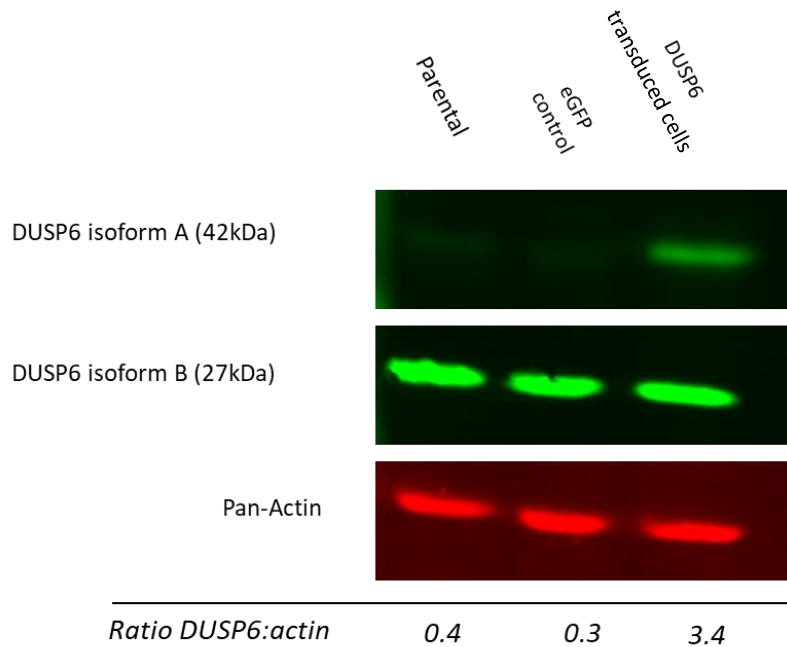


Figure 49. Western blot of DUSP6 expression in RG/C2/80 cells. Left to right: Parental, RG/C2/80 cells transfected with pLJM1-eGFP plasmid (enhanced green fluorescent protein (eGFP) control) and RG/C2/80 cells transfected with pLEX-MCS plasmid containing full-length DUSP6. Parental cells and eGFP control contained undetectable levels of isoform A whilst DUSP6 overexpression cells demonstrated a 3 fold increase in DUSP6 protein expression in comparison with parental line.

4.2.9.3 p-ERK protein expression is decreased in DUSP6 transduced RG/C2/80 cells

ERK protein expression was unchanged between the 3 cell lines assessed; RG/C2/80 (Parental), RG/C2/80 (enhanced green fluorescent protein (eGFP) transduced) and RG/C2/80 (DUSP6 transduced). A marked decrease in p-ERK expression was observed in RG/C2/80 (DUSP6 transduced) cells relative to parental (70% decrease) and eGFP controls (57% decrease), Fig. 50. Repeats presented in Appendix 8.1.3 and 8.1.4

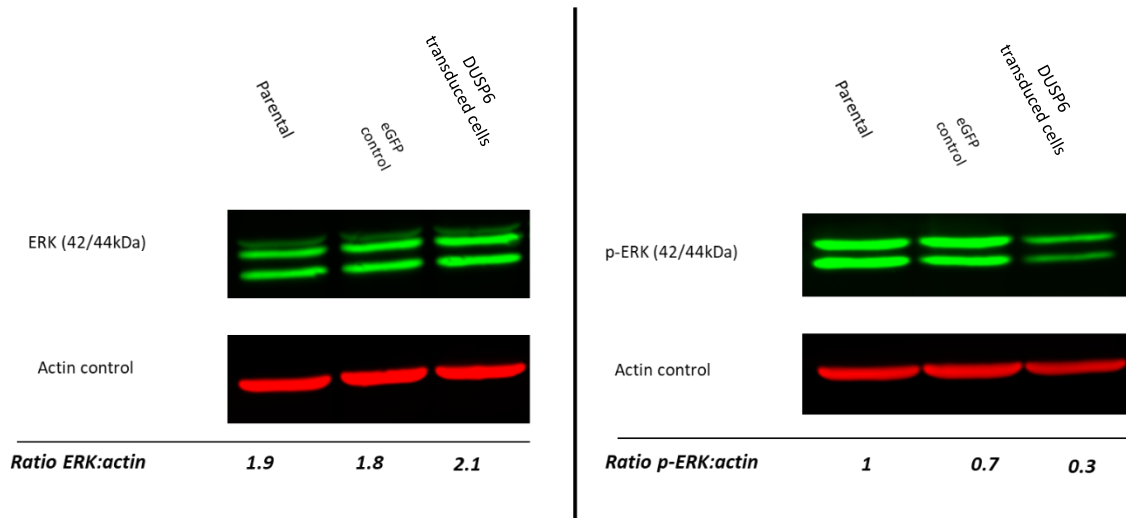


Figure 50. Western blot of ERK and p-ERK expression in RG/C2/80 cell lines. Relative protein expression of ERK and p-ERK to actin control was calculated using LICOR Odyssey imaging system. No significant difference in ERK expression was observed between cell lines. A marked decrease in p-ERK expression was identified in cells transfection with DUSP6 overexpression construct compared to parental (70% decrease) and enhanced fluorescent green protein (eGFP) transduced (57% decrease).

4.3 Discussion

The role of dual-specificity phosphatase 6 (DUSP6) has been extensively explored in a variety of cancers including pancreatic and lung. In these cancer types, a dichotomous expression profile of DUSP6 has been observed whereby higher DUSP6 expression was identified in cases with mild-severe dysplasia or in-situ carcinoma¹¹⁶. In contrast to this, in carcinoma, decreased DUSP6 expression has been associated with increasing histological grade and growth activity, poorer histological differentiation and epithelial-mesenchymal transition (EMT) like features^{61,62,65,66}. Despite the extensive investigation into the roles of DUSP6 in a number of cancer types, little is known whether DUSP6 plays a similar role in colorectal disease. In order to explore the role of DUSP6 in colorectal cancer development, a cohort of cases representative of stages in the adenoma-carcinoma sequence was identified¹¹.

DUSP6 is a phosphatase with exclusivity for ERK, for which it regulates both temporally and spatially^{55,56}. At its core function, DUSP6 de-phosphorylates active p-ERK preventing its transcriptional activation of factors responsible for cellular proliferation and survival. DUSP6 can also act as a cytoplasmic anchor of ERK, being able to bind irrespective of its phosphorylation state thus preventing its activation and translocation and subsequent proliferative and pro-survival activities. In addition to this, the presence of a nuclear export signal (NES) enables the shuttling of DUSP6 between the nucleus and cytoplasm, this allows for the sequestration of active p-ERK to the cytoplasm⁴⁴. In reflection of this body of research, DUSP6 appears to have a fundamental relationship with the control of ERK signalling. With ERK serving as an axis for the proliferative and pro-survival activity of cells and the frequent aberration of many of its upstream activators in cancer; investigation into its protein expression across the adenoma carcinoma sequence and its relationship to DUSP6 protein expression was explored.

In adenoma, analysis of DUSP6 identified that the ratio of DUSP6 positivity increased with increasing grade of dysplasia. This is consistent with observations in pancreatic cancer¹¹⁶. Additionally, when the ratio of nuclear DUSP6 positivity was explored in association with p-ERK positivity, the majority of cases with high grade dysplasia demonstrated a significantly higher ratio of nuclear DUSP6 positivity than p-ERK. It could be surmised that with an increased degree of dysplasia DUSP6 is upregulated in response to increased activation of the ERK pathway to address the imbalance in ERK activation. It may be proposed that in dysplasia, DUSP6 provides a protective mechanism to oppose excessive ERK signalling thus reducing proliferation and survival of dysregulated cells. It is important to note in this series of investigations target protein

analysis was carried out on different sections due to the limitations in immunohistochemical staining i.e. that only one marker can be assessed on one section of tissue at a time. Multiplexing of targets using immunofluorescence would enable assessment of co-expression of DUSP6 and p-ERK and may provide a more accurate measure of expression.

Despite the changes in DUSP6 positivity, no significant difference in ERK or p-ERK positivity was identified between grades of dysplasia. The disparity in DUSP6 positivity but not p-ERK between low and high grade dysplasia may be explained by the effect of *TP53* mutations on DUSP6 transcription. Piya et al⁵⁷ demonstrated the presence of two *TP53* binding sites upstream of the DUSP6 promotor. Additionally, Ets1 has been shown to upregulate DUSP6 transcription by binding to a specific Ets site in intron 1 of *DUSP6*^{55,117,118}. Furthermore, mutant *TP53* has been shown to interact with Ets1, increasing its transcriptional activity, the effects of this however on transcriptional upregulation of DUSP6 via this mechanism are yet to be investigated. A schematic of the proposed mechanism of action is detailed, Fig. 51.

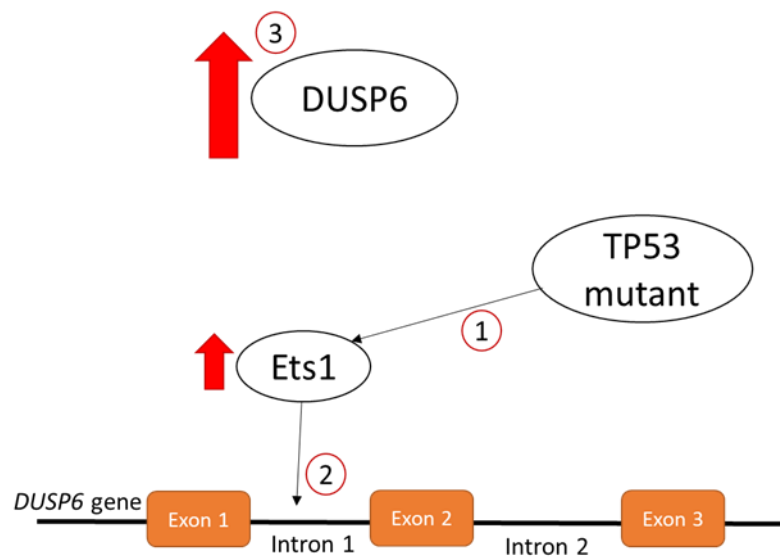


Figure 51. Proposed mechanism of *TP53* mediated upregulation of *DUSP6*. *TP53* binding sites have been identified upstream of the *DUSP6* promotor. In addition, Ets1 binding sites have been demonstrated in intron 1 of the *DUSP6* gene. Mutant *TP53* has been shown to interact and augment Ets1 expression. It is hypothesised that the increase in Ets1 results in increased binding to transcription sites in *DUSP6* and thus increased transcription of *DUSP6*.

An increasing frequency of *TP53* mutations have been observed with increasing grade of dysplasia¹¹⁹. A preliminary immunohistochemical assessment of mutant *TP53* demonstrated a higher incidence of mutations in high grade dysplasia cases in comparison with low grade

dysplasia. Thus the observed increase in the ratio of DUSP6 positivity but not ERK with increasing grade of dysplasia may be explained by *TP53* mutant mediated upregulation of DUSP6 in addition to upregulation of DUSP6 in response to increased ERK activation in these lesions. To further support the hypothesis of mutant *TP53* mediated upregulation of DUSP6 assessment of the ratio of DUSP6 positivity in adenocarcinoma identified a significantly higher ratio of DUSP6 positivity in *TP53* mutant cases compared to adenocarcinoma with *KRAS* mutation or dMMR. Further investigation into the relevance of *TP53* mutations to DUSP6 expression is warranted.

In addition to the above observations in adenoma it is important to highlight that serrated adenomas demonstrated a ratio of DUSP6 positivity akin to those of dMMR and *KRAS* mutant lesions i.e. low, compared to other adenoma subtypes. Similar to adenocarcinoma the ratio of p-ERK positivity was also low demonstrating a similar trend to adenocarcinoma cases than adenoma classes. Serrated adenomas (traditional and sessile) are a subtype of colorectal adenomas with a distinct histology. Adaptations of the traditional adenoma-carcinoma sequence identify adenocarcinomas derived from serrated adenomas as a distinct and molecularly different carcinogenic process, termed the serrated pathway^{32,120}. Serrated adenocarcinomas are likely to have a poorer prognosis than traditional adenocarcinoma and have also demonstrated differential responses to therapy in comparison to lesions arising from the traditional chromosomal instability (CIN) pathway²⁵. The isolation of serrated adenomas in this study with respect to marker expression aligns with the concept that malignant neoplasms derived from this pathway are distinctly different in both histology and molecular characteristics.

When assessing ERK and p-ERK expression, adenocarcinoma cases demonstrated significantly lower ERK and p-ERK positivity to levels below normal mucosa. Further analysis based upon molecular classification identified *KRAS* mutant cases having a higher ratio of ERK positivity compared to adenocarcinoma with dMMR or *TP53* mutation. This pattern of expression was expected as mutations in hotspot regions i.e. kinase domain of *RAS* confer constitutive activation of the protein thus increased activation of the downstream MAPK pathway and subsequent increase in ERK activation^{110,121}. Despite the elevated ERK expression in *KRAS* mutant cases, p-ERK expression remains significantly lower than adenoma groups and is expressed at a comparative level to normal mucosa. This suggests that despite ERK availability in *KRAS* mutant cases, limited phosphorylation occurs and thus reduced activation of proliferative and anti-apoptotic pathways in comparison to adenoma cases (excluding serrated). In order to

investigate this further, the adenocarcinoma group was re-classified based upon T stage of lesion. A trend in decreasing ERK and DUSP6 expression with increasing T stage was observed, consistent with previous findings in lung cancer studies. T stage 4 represents highly invasive lesions which have metastasised to another site and could be described as having higher propensity for EMT. The trend of decreased DUSP6 expression at this stage in comparison to lower staged lesions may be explained by DUSP6 dependent release of inhibition on factors involved in EMT. Further studies are required to elucidate additional pathways which are activated in response to the transition to adenocarcinoma such as those involved in the EMT. Reduced DUSP6 expression in adenocarcinoma would favour advancement into EMT as has been previously observed in this and other studies^{65,66}.

Additional analysis of the relative ratio of nuclear DUSP6 and p-ERK cells in adenocarcinoma identified a significantly lower ratio of DUSP6 and p-ERK positivity than adenomatous lesions. This may be implicit of complete abrogation of this pathway and loss of requirement for DUSP6 in adenocarcinoma. Consistent with patterns of expression observed in cases with high grade dysplasia, assessment of the ratio of nuclear DUSP6 positivity in association with p-ERK positivity in adenocarcinoma identified that the majority of cases demonstrated a significantly higher ratio of nuclear DUSP6 positivity than p-ERK. However, in adenocarcinoma, the ratio of p-ERK positivity in adenocarcinoma was close to negative. It may be inferred from these observations that in adenocarcinoma DUSP6 function becomes redundant with respect to its role in regulating ERK activation as additional aberrations acquired during malignant transformation may prove to be greater drivers of proliferation and differentiation (as inferred from near negative p-ERK expression in this cohort) thus overriding the inhibitory activity of DUSP6 on ERK.

With the involvement of *RAS* mutations in activation of the MAPK/ERK pathway assessment of the impact of *KRAS* mutation status on DUSP6 expression was carried out. Analysis of the ratio of DUSP6 positivity and *KRAS* mutation status in adenocarcinoma identified a significantly lower ratio of DUSP6 positivity in *KRAS* mutant lesions. It could be hypothesised that this decrease is associated with a decrease in the ratio of ERK positive cells in these lesions however the opposite was apparent; ERK positivity was significantly higher in *KRAS* mutant adenocarcinoma compared to wildtype lesions. Thus, it may be inferred that either in *KRAS* mutant individuals the relationship between DUSP6 and ERK regulation is lost due to pathway aberrance or the incorporation of *TP53* mutant individuals with very high DUSP6 positivity may introduce bias to analysis. Further investigation into the relationship between *KRAS* mutation status and DUSP6

expression in a cohort randomised for *RAS* mutation would be advised. Evidence of lack of association of DUSP6 with *KRAS* mutation status in adenoma may be suggestive that *KRAS* mutation status does not impact DUSP6 expression.

In addition to the relationship between DUSP6 and ERK, studies have demonstrated that DUSP6 phosphorylation and degradation can also be affected by the PI3K/mTOR pathway. Agonists of the mTOR pathway (excluding ERK activators) have been shown to induce DUSP6 phosphorylation and degradation¹²². A line of enquiry to explore the effect of DUSP6 transduction in adenocarcinoma cell lines on mTOR expression may assist elucidating mechanisms of DUSP6 downregulation. One would hypothesise mTOR expression would increase with increasing DUSP6 in order to counteract excess DUSP6 expression.

The transition from tissue to *in vitro* models enabled the investigation of DUSP6 function by its manipulation i.e. transduction into RG/C2/80 adenoma cells which do not express detectable levels of DUSP6 isoform A. The observation of unchanged ERK protein expression but markedly decreased p-ERK expression between cell lines demonstrates a functional relationship between DUSP6 and p-ERK. Further studies to assess the localisation of p-ERK in relation to DUSP6 transduction would assist in determining the impact of DUSP6 transfection on p-ERK shuttling from the nucleus. Cell fractionation studies on protein lysates from RG/C2/80 parental cells (undetectable isoform A, high isoform B) and DUSP6 transduced RG/C2/80 cells (high isoform A and B) to assess the nuclear versus cytoplasmic fraction of p-ERK with respect to DUSP6 expression may elucidate whether isoform B is able to act as a phosphatase or simply a cytoplasmic anchor of ERK. One would expect high nuclear p-ERK in parental cells compared to DUSP6 transduced, implicit of a lack of phosphatase function in isoform B.

In summary, the ratio of DUSP6 positivity increases with increasing grade of dysplasia. It is hypothesised that in the event of dysplasia DUSP6 acts to reduce proliferation of aberrant cells thus halting malignant transformation. This evidence is supportive of a tumour suppressor like role of DUSP6 in colorectal carcinogenesis. In adenocarcinoma, it is hypothesised that this role of DUSP6 is lost due to the acquisition of further mutations which negate the requirement/override the inhibitory activity of DUSP6 on ERK activation. Further investigation is required into the functional relationship between DUSP6 and ERK in adenoma.

5. Investigating the role of DUSP6 in treatment response and prognosis in colorectal adenocarcinoma

5.1 Introduction

DUSP6 protein expression has been demonstrated to be downregulated in a variety of malignant neoplasms compared to normal tissue. For example, in primary pancreatic cell lines DUSP6 expression was downregulated in adenocarcinoma compared to pre-neoplastic lesions with low to high grade dysplasia⁶⁴. Immunohistochemical assessment of DUSP6 protein expression in non-small cell lung cancer (NSCLC)⁶¹ and oesophageal squamous cell carcinoma (ESCC)¹²³ demonstrated that protein expression was significantly decreased in neoplastic lesions compared with adjacent normal tissue. Despite its extensive investigation in a variety of neoplasm types, little is known about the importance of DUSP6 in colorectal adenocarcinoma. DUSP6 protein expression as determined by immunohistochemistry has been shown to demonstrate a low but dynamic expression profile in colorectal adenocarcinoma tissue compared to adenoma and normal tissue as has been previously shown in Chapter 3.

To explore the importance of DUSP6 in colorectal adenocarcinoma further, the colorectal adenocarcinoma cohort comprising 525 primary resections previously described in Chapter 4 was explored for DUSP6 protein expression by semi-quantitative immunofluorescence and AQUA (Automated Quantitative Analysis). In addition to exploring the landscape of DUSP6 protein expression in a large and diverse cohort comprising a variety of stages of colorectal cancer, the association of DUSP6 with *RAS* mutation status was also investigated.

The MAPK pathway has been a focus due to its roles in the pathogenesis of colorectal cancer for many decades. With the advent of targeted therapies such as Cetuximab (Erbix) the importance of the *RAS* arm of the MAPK pathway has become ever more pertinent in determining treatment response and prognostic outcome. At present, *RAS* mutation is the only negative biomarker for response to the anti-epidermal growth factor receptor monoclonal antibody treatment, Cetuximab for the first line treatment of metastatic colorectal

adenocarcinoma. The direct association of DUSP6 with the MAPK pathway member, ERK makes it an ideal candidate to investigate its expression profile in response to *RAS* mutation. In addition, DUSP6 protein expression may prove a more sensitive surrogate of ERK activation and thus MAPK pathway addiction, better stratifying individuals for treatment with Cetuximab than *RAS* mutation status alone. Indeed DUSP6 has been shown to regulate drug sensitivity in a variety of tissues albeit presenting conflicting functions depending upon tissue context. For example, increased expression of DUSP6 in NSCLC and ovarian cancer increased sensitivity to Crizotinib and Cisplatin respectively^{89,90}. In contrast to this, increased expression in glioblastoma and ER positive breast cancer increase resistance to chemotherapy and Tamoxifen treatment respectively^{63,124}. DUSP6 appears to be a protein of two faces depending upon tissue context. In addition, DUSP6 function whether tumour suppressive or oncogenic may be dependent on the extent of oncogenic addiction of *RAS* as a driver of disease.

5.2 Results

5.2.1 Cohort clinicopathological dataset

Following study cohort identification as detailed in methods 2.2, clinicopathological datasets were obtained from clinical pathological reports. A summary of the cohort is detailed, Table 24.

Clinicopathological parameter	Sub-group	n	Clinicopathological parameter	Sub-group	n
Gender	Male	258	Tumour type	Adenocarcinoma	438
	Female	267		Mucinous adenocarcinoma	73
Age at diagnosis	21-30	3		Tumour differentiation	Medullary adenocarcinoma
	31-40	11	Good		22
	41-50	37	Moderate		398
	51-60	103	Poor	105	
	61-70	138	0-10	4	
	71-80	151	44136.00	47	
	81-90	75	21-30	102	
	91-100	7	31-40	124	
TNM staging	1	72	Tumour diameter (macroscopic measurements)	41-50	103
	2A	155		51-60	63
	2B	40		61-70	35
	3B	112		71-80	21
	3C	89		81-90	5
	4	56		91-100	8
Tumour site	Rectum	147		101-110	1
	Rectosigmoid	52		111-120	3
	Left	145		121-130	1
	Transverse	31		Venous	155
	Right	150		Lymphatic	24
				Venous and lymphatic	37
				No EMLVI	285

Table 24. Cohort summary of clinicopathological dataset. Clinicopathological dataset derived from original clinical pathology reports for each case. Tumour site, Left comprises sigmoid colon, descending colon and splenic flexure; Tumour site, Right comprises hepatic flexure, ascending colon, caecum and appendix. Extramural lymphovascular invasion (EMLVI) ascertained by pathologist by visual inspection of H & E slide.

5.2.2 DUSP6 expression in colorectal adenocarcinoma

Immunofluorescence was carried out on the cohort of 525 cases, Appendix 8.2.2. DUSP6 expression was semi-quantified using the Automated Semi-Quantitative (AQUA) system and subsequent analysis was carried out using X-Tile software¹⁰².

5.2.2.1 DUSP6 expression is dynamic in colorectal adenocarcinoma

DUSP6 expression was assessible in 498/525 (94.8%) cases (cases which were not assessed either failed immunofluorescence or tissue was no longer available for assessment). Median nuclear DUSP6 expression was higher than cytoplasmic DUSP6 (349.6 vs 318.7 arbitrary units AU), Fig. 52. A dynamic range of expression intensity was observed for both nuclear and cytoplasmic expression across cases (inter-quartile range 248.7 and 232.7 AU), Fig. 53.

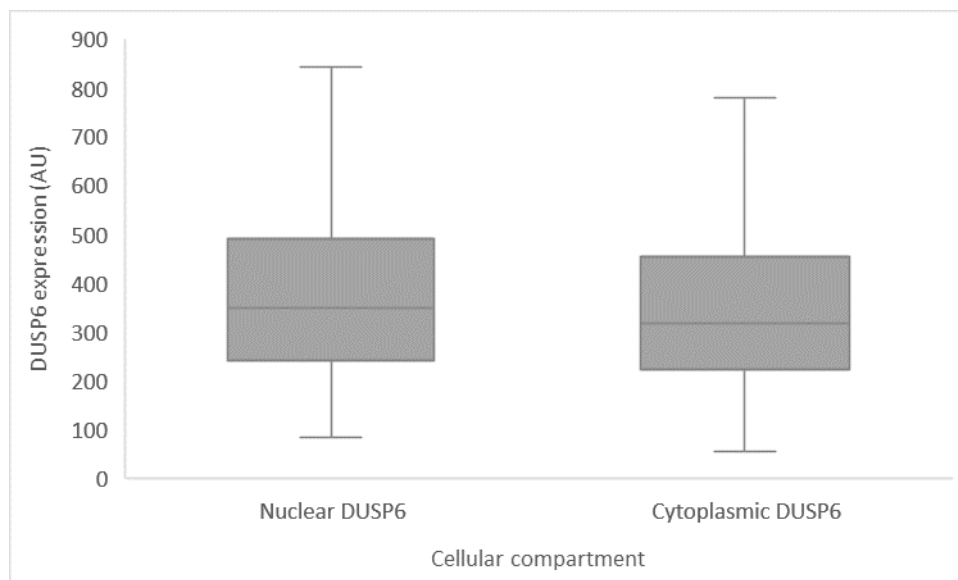


Figure 52. Range of nuclear and cytoplasmic DUSP6 protein expression values. DUSP6 expression was semi-quantified in both nuclear and cytoplasmic compartments in 498 cases (*Median* 349.6 and 318.7 arbitrary units (AU), inter-quartile range 248.7 and 232.7 AU respectively).

Normality tests were carried out on nuclear and cytoplasmic DUSP6 AQUA values. The data set was not normally distributed as determined by Kolmogorov-Simrnov test, $p < 0.001$ (both nuclear and cytoplasmic DUSP6 AQUA values). Non-parametric statistical analysis was employed.

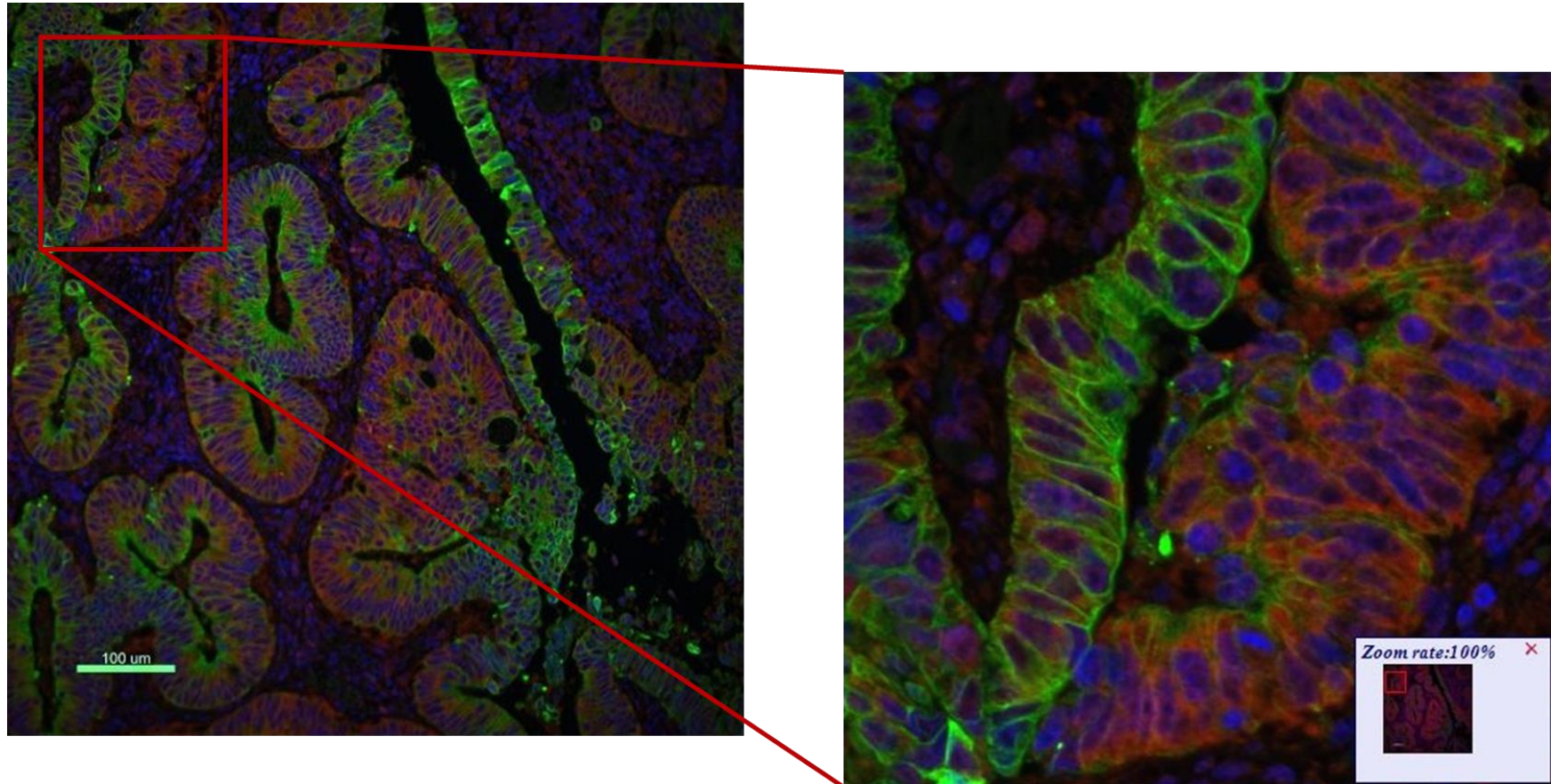


Figure 53. DUSP6 positivity in colorectal adenocarcinoma by immunofluorescence. Red: DUSP6, Green: Pan-cytokeratin and Blue: Hoechst. Immunofluorescence was carried out on 525 whole sections of primary colorectal adenocarcinoma using formalin fixed paraffin embedded sections. Automated semi-quantitative (AQUA) analysis was carried out to quantify DUSP6 intensity for each case. Protein expression of DUSP6 demonstrated a dynamic range across cases. Additionally, a degree of intra-tumour heterogeneity was observed.

5.2.3 DUSP6 expression and associations with clinicopathological dataset

A relationship between DUSP6 protein expression and histological grade in lung cancer has been previously demonstrated by Okudela et al., 2009⁶¹ whereby DUSP6 protein expression decreased with increasing histological grade. To explore whether this phenomenon is present in colorectal cancer associations between T stage and DUSP6 protein expression were assessed. In addition, associations with other clinicopathologic features were also explored, Appendix 8.2.3. Bonferroni corrections were applied for multiple testing.

5.2.3.1 DUSP6 expression is not associated with histological grade

A significant difference was identified in cytoplasmic DUSP6 expression and T stage as determined by Kruskal-Wallis test, $p=0.038$, Fig. 54. A significantly higher cytoplasmic DUSP6 expression in T stage III than T stage IV as determined by Mann-Whitney test (0.007, Median AQUA value 325.1, $n=289$ and 279.3 respectively, $n=132$). Despite this significant finding, analysis of effect size by eta-squared (η^2) identified that only a small effect ($\eta^2=0.18$) of the difference in DUSP6 expression could be explained by lesion size, additionally the large inter-quartile ranges of T stage III and IV lesions is implicit that the likelihood of these findings being biologically relevant is negligible. The findings observed in NSCLC that DUSP6 expression decreases with increasing histological grade are not evident in colorectal adenocarcinoma.

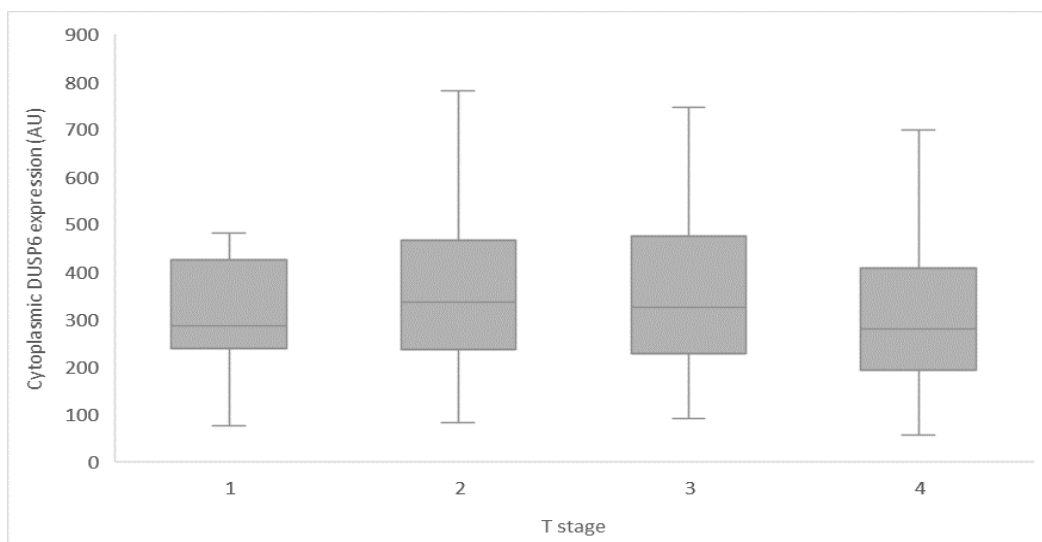


Figure 54. Cytoplasmic DUSP6 expression by T stage. Cytoplasmic DUSP6 expression was significantly different between T stage III and IV as determined by Wilcoxon Signed Rank test ($p=0.007$, Med 325.1 and 279.3 respectively). Lesions with T stage IV (primary lesion evident in other organs) demonstrated a lower DUSP6 protein expression than T stage III however analysis of effect size (measured by eta-squared) demonstrated a limited effect ($\eta^2=0.18$) of DUSP6 expression difference being attributed to stage of lesion implicit of limited biological relevance of this difference.

5.2.4 DUSP6 expression and association with RAS mutation status

Immunohistochemical analysis of adenocarcinoma cases in chapter 3 identified a difference in DUSP6 expression based upon *KRAS* mutation status however these observations were made in a cohort specifically chosen for molecular aberrations (*KRAS* mutant, *TP53* mutant (*KRAS* wildtype), deficient mis-match repair (*KRAS* wildtype) and thus may represent a bias population.

Identification of this study cohort was carried out in a non-selective manner with respect to *RAS* mutation status. Thus the previous observations of DUSP6 expression being lower in *KRAS* mutant individuals was tested in this cohort.

5.2.4.1 DUSP6 expression is not associated with RAS mutation status

The study cohort had been previously screened for the presence of *RAS* mutations at hotspot locations throughout *KRAS* and *NRAS* genes, Chapter 4. Statistical analysis was carried out with a number of *RAS* mutation classifications and nuclear or cytoplasmic DUSP6 expression.

RAS mutation classifications:

1. *RAS* mutation present or absent – any *RAS* mutation within either *KRAS* or *NRAS*.
2. *RAS* mutation type – mutant cases were grouped based upon the specific genomic characterisation of the mutation e.g. G12D.

Mann-Whitney and Kruskal-Wallis test were used to assess any significant relationships between the groups above and DUSP6 expression (nuclear and cytoplasmic), Table 25. No significant association between DUSP6 and *RAS* mutation status or mutation type was identified.

RAS classification group	Nuclear DUSP6	Cytoplasmic DUSP6
1	0.230	0.975
2	0.238	0.141

Table 25. Mann-Whitney and Kruskal-Wallis test results assessing relationship between *RAS* mutation status and DUSP6 expression.

In addition to *RAS* mutation status, the study cohort had been previously assessed for the presence of *BRAF* hotspot mutations. The relationship between DUSP6 and *BRAF* mutation status was also assessed by Mann-Whitney test. No significant difference was identified however a trend between nuclear DUSP6 expression and *BRAF* mutation status was identified, $p=0.063$.

5.2.4.2 Nuclear expression of DUSP6 is associated with RAS mutation status but with limited effect

The previous analysis assessed the association of RAS mutation status with DUSP6 expression by isolated compartment. The presence of a nuclear export signal enables the shuttling of DUSP6 between the nucleus and cytoplasm⁴⁴. To capture this dynamic flow of DUSP6 between sub-cellular compartments, DUSP6 protein expression in both nuclear and cytoplasmic compartments was assessed in combination per case, Fig. 55.

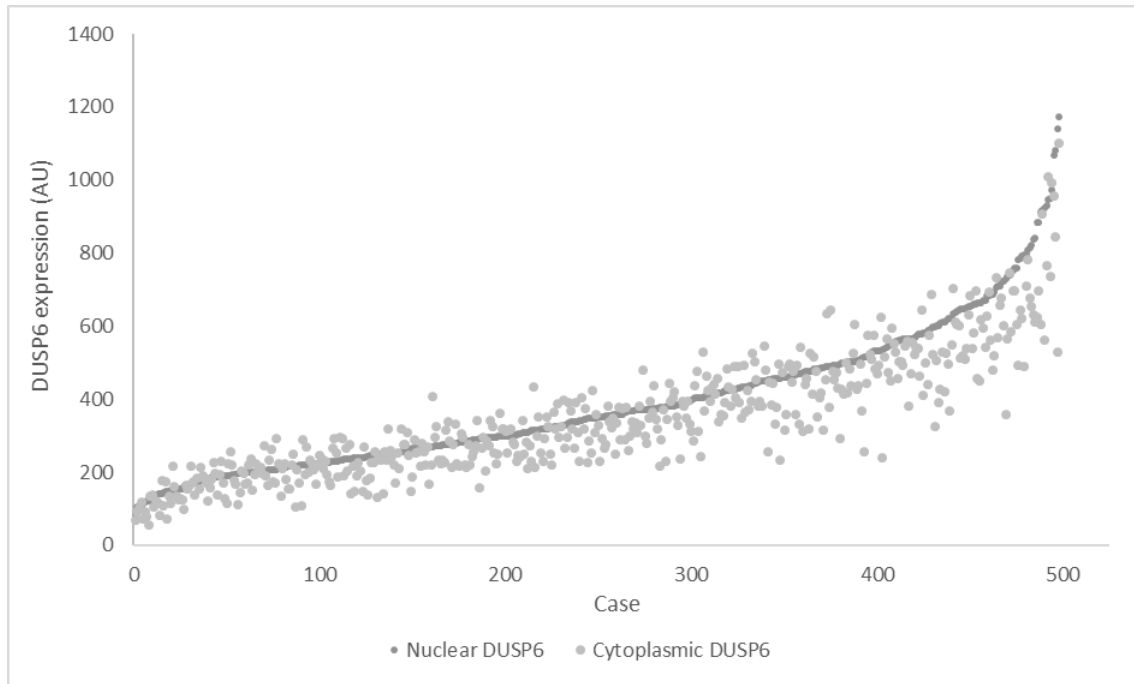


Figure 55. DUSP6 AQUA expression by sub-cellular compartment. DUSP6 expression was semi-quantified in both nuclear and cytoplasmic compartments using AQUA system. DUSP6 nuclear and cytoplasmic expression is plotted in ascending order of nuclear expression. Sixty-nine percent (342/498) of cases demonstrated significantly higher nuclear DUSP6 expression in comparison to cytoplasmic expression ($p < 0.001$, Median cytoplasmic AQUA value of cases with higher nuclear than cytoplasmic: 299.6 and higher cytoplasmic than nuclear: 338.5 respectively).

Overall, 69% (342/498) of lesions demonstrated significantly higher nuclear than cytoplasmic DUSP6 expression as determined by Wilcoxon-Signed rank test, $p < 0.001$. This may be indicative of an increased ERK activation within these cases irrespective of RAS mutation. The observation of a number of cases having higher nuclear vs cytoplasmic DUSP6 expression provided evidence to explore whether the sub-cellular localisation of DUSP6, a negative regulator of ERK is associated with RAS mutation status. When cases were categorised as either having higher nuclear DUSP6 than cytoplasmic or higher cytoplasmic than nuclear a larger proportion of RAS

mutant lesions (74.9% of *RAS* mutant lesions) demonstrated higher nuclear than cytoplasmic DUSP6 expression, than *RAS* wildtype (65% of *RAS* wildtype lesions) as determined by Chi² test for independence, p0.027 however the impact of *RAS* mutation status upon the degree of nuclear over cytoplasmic DUSP6 expression was considered little effect as determined by Phi test for effect size (Phi -0.1).

The evidence presented above suggests there is no significant overall relationship between *RAS* mutation status and DUSP6 protein expression.

5.2.5 Investigating DUSP6 expression and function *in vitro*

As previously mentioned, *RAS* mutation status is currently the only clinically implemented negative biomarker for the prediction of response of colorectal adenocarcinoma to the targeted treatment, Cetuximab and yet fails to sufficiently stratify individuals. Sixty-eight percent of cases in this study cohort demonstrated higher nuclear than cytoplasmic DUSP6 expression irrespective of *RAS* mutation status. This could be indicative that DUSP6 may be a useful surrogate of MAPK signalling pathway hyper-activation than *RAS* mutation status alone and in turn better identify individuals likely to respond to Cetuximab treatment.

A limited number of patients (n7) had been treated with Cetuximab from the cohort in which DUSP6 protein expression had been assessed, rendering investigation into the association between DUSP6 protein expression and cohort response to Cetuximab unattainable. Therefore, in order to A: explore the expression and function of DUSP6 in adenocarcinoma and B: investigate the effects of DUSP6 manipulation on Cetuximab response *in vitro*, identification of a suitable cell line was carried out.

5.2.5.1 DUSP6 protein expression across colorectal adenocarcinoma cell lines

To identify a suitable cell line to manipulate the expression of the canonical DUSP6 isoform A an initial screen of 10 colorectal adenocarcinoma cell lines available in the laboratory was carried out to identify DUSP6 protein expression, shown in Figure 56. HCT116, SK-CO-1, DLD1, HT29, SW626, LoVo, LS411N, SW480, NCI-H508 and C99.

As determined by western blot, DUSP6 isoform A and B protein expression differed between cell lines, Fig. 56a. DUSP6 protein expression was compared between cell lines, Fig. 56b. Relative DUSP6 isoform A protein expression ranged from 1620 – 2190. Relative DUSP6 isoform B protein expression demonstrated a larger degree of variation, range 1250 – 4520.

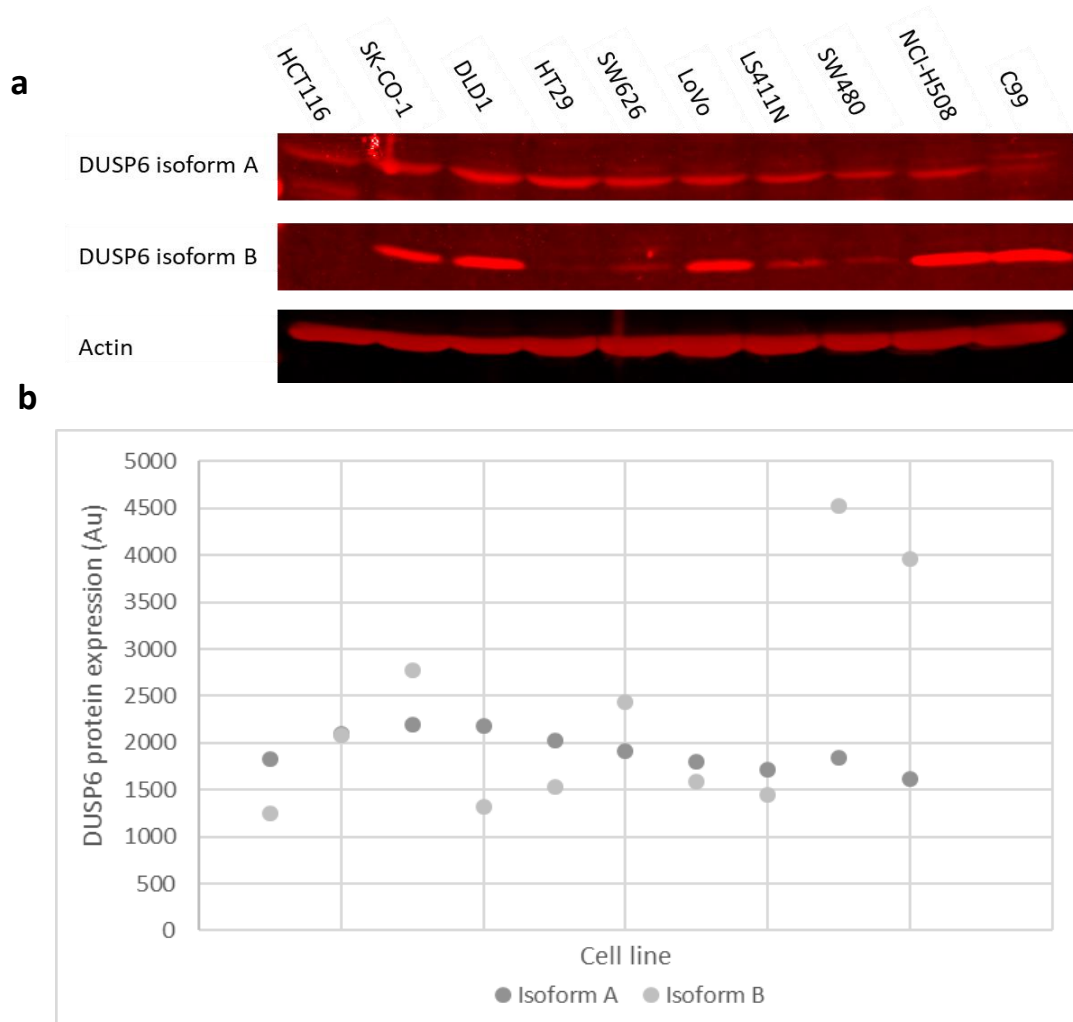


Figure 56. Screening of colorectal adenocarcinoma cell lines for DUSP6 protein expression. A. Western blot of 10 colorectal adenocarcinoma cell lines for DUSP6 isoform a and b. B. Relative protein expression of DUSP6 isoform A and B. Left to right: HCT116, SK-CO-1, DLD1, HT29, SW626, LoVo, LS411N, SW480, NCI-H508 and C99. Total protein loaded per sample = 30µg.

Molecular and response profiles of CRC cell lines to Cetuximab treatment had been previously carried out by Medico et al¹²⁵, Table 26. In addition to this, protein expression of DUSP6 isoforms A and B was expressed as a ratio for each of the 10 colorectal cell lines assessed.

Of the 10 cell lines assessed, NCI-H508 and C99 were the only reported to demonstrated sensitivity to Cetuximab treatment. These two cell lines also demonstrated the lowest ratio (0.4) of A:B isoforms, demonstrating markedly higher ratio of isoform B to isoform A. In addition to this observation, C99 and NCI-H508 cells were the only cell lines out of those assessed that were *RAS* and *BRAF* wildtype.

Cell line	Molecular analysis					Cetuximab ($\mu\text{g/mL}$)						DUSP6 protein expression		
	MSI/MS S status	KRAS	NRAS	BRAF	PIK3CA	0.001	0.01	0.1	1	10	100	A	B	Ratio A:B
HCT116	MSI	p.G13D	WT	WT	p.H1047R	-0.5	7.5	9.0	9.1	8.7	10.2	1830	1250	1.5
SK-CO-1	MSS	p.G12V	WT	WT	WT	-12.1	-10.7	-10.9	-7.4	-6.5	2.5	2100	2080	1.0
DLD1	MSI	p.G13D	WT	WT	p.E545K	0.5	1.5	5.9	0.4	-2.4	-0.3	2190	2770	0.8
HT29	MSS	WT	WT	p.V600E	WT	-0.5	-3.6	0.3	-0.8	-2.1	8.5	2180	1320	1.7
SW620	MSS	p.G12V	WT	WT	WT	0.8	-3.3	-6.9	-9.0	-9.8	2.9	2030	1530	1.3
LoVo	MSI	p.G13D	WT	WT	WT	-4.0	-2.1	-0.1	-2.5	1.6	11.6	1910	2440	0.8
LS411N	MSI	WT	WT	p.V600E	WT	-0.2	-2.2	1.1	-4.0	-2.3	-3.6	1800	1580	1.1
SW480	MSS	p.G12V	WT	WT	WT	-4.8	1.6	2.5	3.8	1.7	8.2	1710	1450	1.2
NCI-H508	MSS	WT	WT	WT	p.E545K	-2.4	4.1	40.1	80.6	85.4	84.3	1840	4520	0.4
C99	MSS	WT	WT	WT	WT	11.7	34.1	48.7	59.4	68.1	78.6	1620	3960	0.4

Table 26. Molecular and Cetuximab response data. Molecular analysis and Cetuximab response data was derived from work carried out by Medico et al, 2015. DUSP6 protein expression as determined by screening of protein lysates in this study is also detailed. The ratio of DUSP6 isoforms A:B was calculated. Of the 10 cell lines assessed, NCI-H508 and C99 demonstrated sensitivity to Cetuximab treatment, in addition these two lines demonstrated the lowest ratio of DUSP6 isoforms A:B and were the only all *RAS* and *BRAF* wildtype lines.

5.2.5.2 C99 cells as a model for investigation of DUSP6

The two cell lines, NCI-H508 and C99 were brought forward as potential candidates for the over-expression of DUSP6. The adenocarcinoma cell line C99 demonstrating the lowest levels of DUSP6 isoform A protein expression relative to the other cell lines assessed whilst also being an adherent cell line made this line an ideal candidate for transduction with DUSP6, Fig. 57.

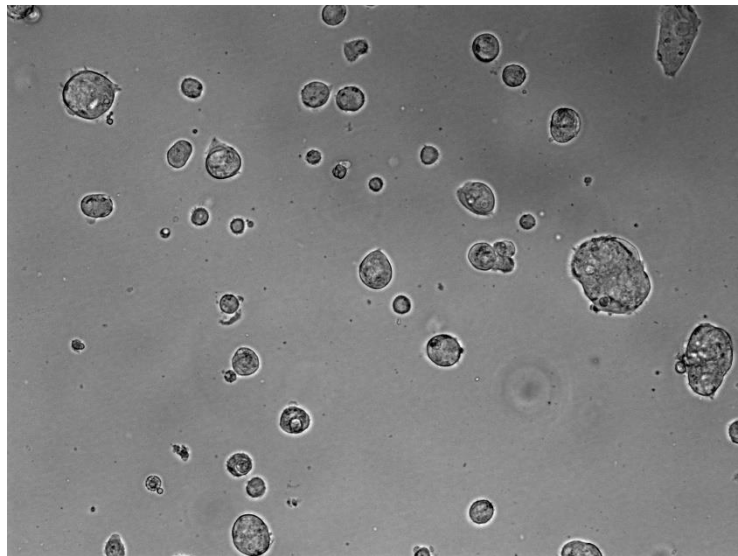


Figure 57. C99 adenocarcinoma cells. C99 cells originated from a moderately well differentiated colorectal adenocarcinoma classified as TNM stage III. Brightfield image of C99 adenocarcinoma cells. x20 magnification.

5.2.5.3 DUSP6 transduction in C99 cells

Following the identification of low levels of DUSP6 isoform A in C99 cells, a stable transduction of DUSP6 by lentiviral delivery was successfully achieved in these cells. Assessment of DUSP6 protein expression by western blot confirmed previous findings that C99 cells express minimal levels of isoform A and high levels of isoform B. In comparison to enhanced green fluorescent protein (eGFP) transfected control cells, a 40.5 fold change in DUSP6 isoform A expression was observed in DUSP6 transduced cells, Fig. 58. An 11.5 fold change in DUSP6 isoform A expression was observed in DUSP6 transduced cells compared to parental cells. No significant difference in DUSP6 isoform B was identified suggesting that over-expression was of isoform A alone.

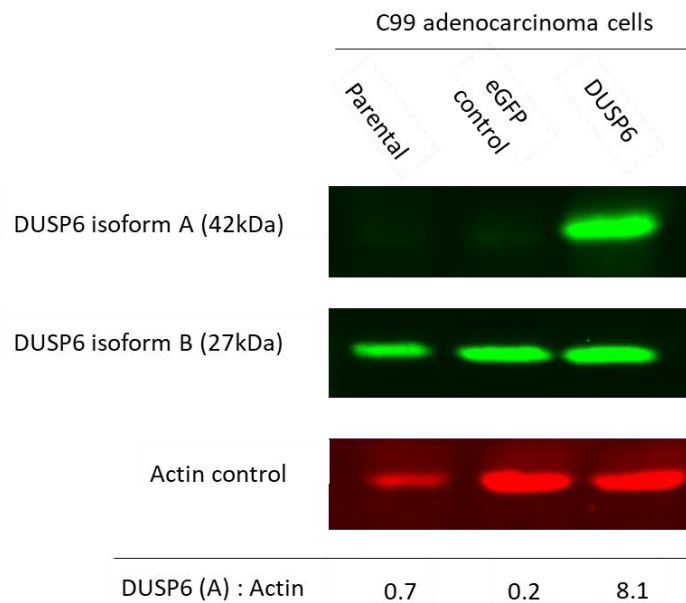


Figure 58. DUSP6 overexpression in C99 cells as determined by western blot. DUSP6 was successfully overexpressed in C99 cells. 40.5 fold change in DUSP6 isoform A was identified in DUSP6 transduced cells in comparison to enhanced green fluorescent protein (eGFP) transduced control. An 11.5 fold change in DUSP6 isoform A was identified in DUSP6 transduced cells in comparison to parental control. No significant difference in isoform B was identified confirming isolated over-expression of isoform A.

5.2.5.4 Investigating a functional relationship between DUSP6 and ERK in C99 cells

It has been previously demonstrated in Chapter 3 in the adenoma cell line RG/C2/80 that transduction of DUSP6 resulted in a marked decrease in p-ERK expression. This phenomenon was explored in C99 cells in which DUSP6 had been transduced, Fig. 59.

5.2.5.5 p-ERK expression is markedly decreased in DUSP6 transduced C99 cells

No significant difference in ERK protein expression was observed between C99 parental, C99 (eGFP transduced) and C99 (DUSP6 transduced) cell lines. Consistent with observations in RG/C2/80 adenoma cells transduced with DUSP6, a marked decrease in p-ERK protein expression was observed in C99 cells transduced with DUSP6. A 5 fold decrease in p-ERK protein expression between DUSP6 transduced and eGFP transduced C99 cells and a 25 fold decrease was observed between DUSP6 transduced and parental C99 cells. Repeats presented in Appendix 8.2.4 and 8.2.5.

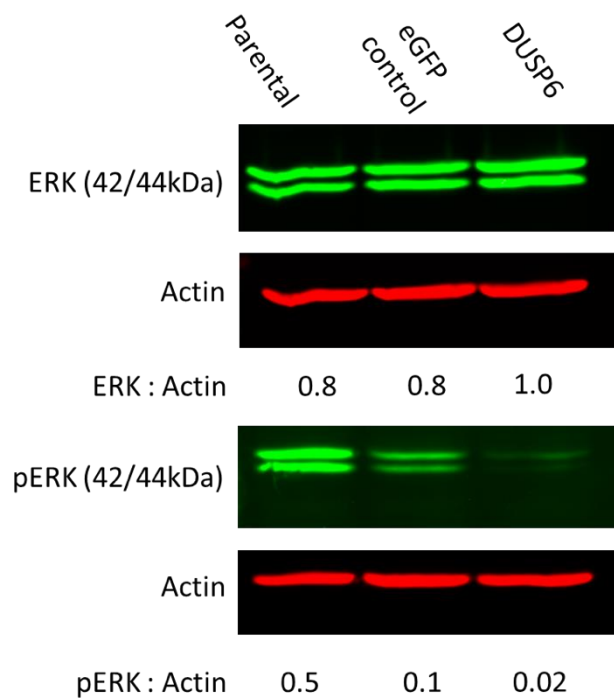


Figure 59. ERK and p-ERK protein expression in protein lysates of parental, eGFP transduced and DUSP6 transduced C99 cells. Protein expression of ERK and p-ERK was explored by western blot using protein lysates (repeated n3) from 3 C99 cell lines; parental, eGFP transduced control and DUSP6 transduced. No significant change in ERK protein expression was observed between cell lines. A marked decrease was observed in pERK expression. A 5 fold decrease in p-ERK protein expression was observed between eGFP transduced and DUSP6 transduced C99 cells. A 25 fold decrease in p-ERK protein expression was observed between parental and DUSP6 transduced C99 cells. Each sample was normalised to actin control.

5.2.5.6 Assessment of C99 response to Cetuximab treatment

Work previously carried out by Medico et al, 2015¹²⁵ identified that C99 cells were sensitive to Cetuximab treatment. In order to confirm this reported response profile the experimental procedure was recapitulated. Cells were treated at a range of concentrations 24hrs post seeding for a duration of 4 days following which the percentage change in growth in comparison to untreated cells was calculated, Fig 60. A decrease in the rate of growth of C99 parental cells with increasing Cetuximab concentration confirmed the response profile reported in previous studies by Medico et al., 2015; confirming sensitivity to Cetuximab treatment. Repeat measurements presented in Appendix 8.2.6.

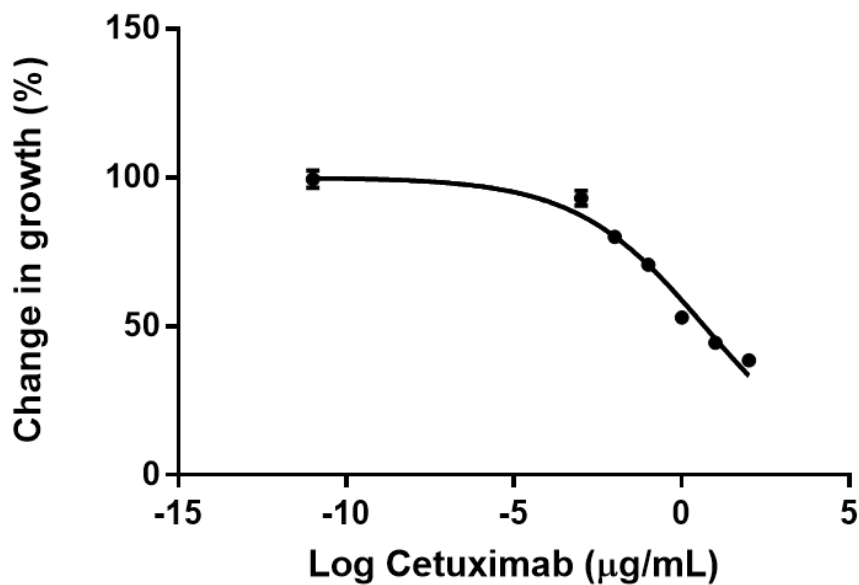


Figure 60. Dose-response for C99 parental lines. Parental C99 cells (repeated n6 for each treatment) were treated 24hrs post seeding (2×10^6 seeding density) with Cetuximab at increasing concentrations; 0.001, 0.01, 0.1, 1, 10 and 100µg/mL. Standard error of the mean represented.

5.2.5.7 DUSP6 transduction and Cetuximab response in C99 cells

DUSP6 has been implicated as a potential mechanism for resistance to a variety of chemotherapeutic and targeted treatments. In colorectal cancer *RAS* wildtype individuals are predicted to respond to targeted therapy Cetuximab however it remains that a proportion of these individuals are unresponsive.

Following transduction of C99 cells with DUSP6, a dose-response experiment following the same protocol as previously described was carried out for each cell line (C99 (parental), C99 (eGFP transduced and C99 (DUSP6 transduced), Fig 61. A control lane of media only was used as a comparator for change in growth for each cell line. Repeat measurements presented in Appendix 8.2.7.

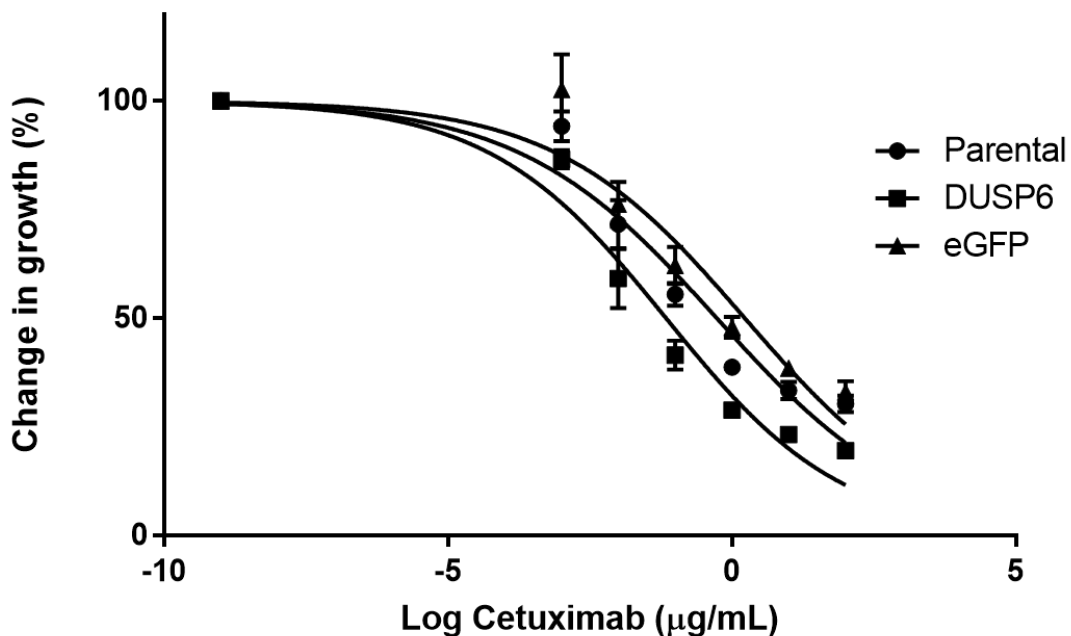


Figure 61. Dose-response for Cetuximab treated C99 cells. 3 cell lines; C99 parental, eGFP transduced C99 and DUSP6 transduced C99 cells were seeded $2.0 \times 10^6/200\mu\text{l}$ density in replicates of 6 for each treatment. Cells were treated with Cetuximab at varying concentrations (0, 0.001, 0.01, 0.1, 1, 10, 100µg/mL) 24 hours post seeding for a duration of 4 days. Untreated control was media only and used as a measure for change in growth compared to treatment for each cell line. Following treatment SRB assay was used to determine percentage change in growth in comparison to untreated control. DUSP6 transduced cells demonstrated an increased sensitivity to Cetuximab treatment than two other conditions. A 24.2 fold decrease in IC50 was observed between DUSP6 transduced cells (IC50 0.07) and eGFP transduced control (IC50 1.7). A 7.7 fold decrease in IC50 was observed between DUSP6 transduced C99 cells and parental (IC50 0.54).

Following treatment an SRB assay was used to determine percentage change in growth in comparison to untreated control for each cell line. DUSP6 transduced cells demonstrated an increased sensitivity to Cetuximab treatment compared to the changes in growth of the other two cell lines. A 24.2 fold decrease in IC50 was observed between DUSP6 transduced cells (IC50 0.07) and eGFP transduced control (IC50 1.7). A 7.7 fold decrease in IC50 was observed between DUSP6 transduced C99 cells and parental (IC50 0.54).

In this preliminary study, a significant decrease in growth was observed in C99 (DUSP6 transduced) cells compared to both C99 (Parental) and C99 (eGFP transduced) cell lines treated with Cetuximab (Significant differences in growth observed at Cetuximab concentrations 0.1, 1, 10 and 100µg/mL), Table 27.

Cell line	Cetuximab concentration (µg/mL)					
	0.001	0.01	0.1	1	10	100
Parental	ns	ns	0.015	0.002	0.009	0.002
eGFP	ns	ns	0.009	0.002	0.002	0.002

Table 27. Summary of statistical analysis of change in growth for C99 (DUSP6 transduced) cells and other conditions. Mann-Whitney test was carried out between C99 (DUSP6 transduced) cells and C99 (parental) and C99 (eGFP transduced) respectively. A significant difference between percentage change in growth upon Cetuximab treatment was observed between both condition compared at Cetuximab concentrations 0.1, 1, 10 and 100µg/mL.

No significant difference between percentage change in growth was identified between C99 (parental) and C99 (eGFP transduced) (excluding Cetuximab concentration 1µg/mL), suggestive that the effect observed in C99 (DUSP6 transduced) cells is attributable to the over-expression of DUSP6 and not merely stress exerted by the transduction process. The findings from this preliminary study is suggestive that overexpression of DUSP6 in C99 cell lines increases cell sensitivity to Cetuximab treatment however further repeats are required to confirm these findings.

5.2.6 DUSP6 as a biomarker of prognosis

In consideration of this, a hypothesis that low DUSP6 protein expression would confer a poorer prognostic outcome was hypothesised.

In order to investigate whether DUSP6 has a prognostic utility in colorectal cancer, 5 year survival data was obtained for 484/525 cases, Appendix 8.1.4. Of the 484 cases to which survival was assessable 189 patients died prior to the 5 year endpoint. Of the 189 deceased patients, colorectal specific death was attributable to 114/189 (60.3%) of cases.

To enable assessment of whether DUSP6 protein expression could prove a biomarker in colorectal prognosis X-tile software¹⁰² was used to generate thresholds for high and low DUSP6 expression with respect to overall survival (OS) outcome. Cases with cytoplasmic DUSP6 expression less than a <217.92 AQUA value were designated low (n110), cases with DUSP6 expression greater than this value were designated high (n388). These thresholds were applied to nuclear DUSP6 expression also; low (n87) and high (n411).

Cox-regression analysis was carried out for overall survival and disease specific survival using these thresholds for high and low DUSP6 expression for both nuclear and cytoplasmic DUSP6 expression. No significant association between nuclear expression and survival was identified. Significant associations between cytoplasmic expression are presented.

5.2.7 Low DUSP6 expression is associated with poor 5yr. overall survival and disease specific survival

Cox regression was carried out to assess 5yr. OS and low and high DUSP6 expression, Table 28. Individuals with low DUSP6 expression had a significantly lower 5yr. OS than individuals with high DUSP6 expression, Fig. 62. Bootstrapping analysis was performed using 1000 samples to validate findings. A significant association between low DUSP6 expression and poorer survival outcome was maintained in this analysis (p0.003).

Target	p-value	Hazard ratio	95% C.I.
^a Cytoplasmic DUSP6	0.002*	1.643	1.2 – 2.3
^b Cytoplasmic DUSP6 (bootstrapping)	0.003*	1.642	1.2 – 2.2

Table 28. Cox-regression analysis with categorical cytoplasmic DUSP6 expression and overall survival. The association of these cut offs with 5yr DSS were explored using Cox-regression analysis. ^aCases with low DUSP6 expression demonstrated a poorer survival outcome in comparison to those above the cut off of 217.92 AQUA score (p0.002). ^bBootstrapping analysis was performed using 1000 samples to validate findings. A significant association between low DUSP6 expression and poorer survival outcome was maintained in this analysis (p0.003).

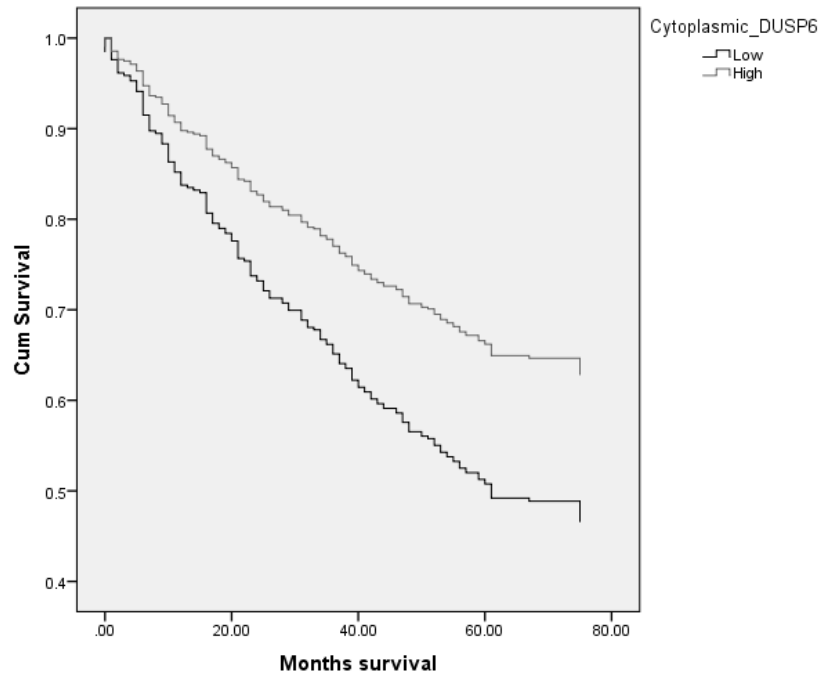


Figure 62. Kaplan-Meier plot for cytoplasmic DUSP6 expression and overall survival. Cytoplasmic DUSP6 expression was categorised by high and low expression as previously determined by X-tile software. Using Cox-regression analysis, individuals with low DUSP6 expression displayed a significantly shorter survival in comparison to those with high cytoplasmic DUSP6 expression, $p < 0.002$, HR 1.6, CI 1.2-2.3.

Cox-regression analysis for cytoplasmic DUSP6 and 5yr. disease specific survival DSS was carried out using the pre-defined cut offs for high and low expression, Table 29. Individuals with low DUSP6 expression had a significantly lower 5yr DSS than individuals with high DUSP6 expression, $p < 0.001$, Fig. 63. Bootstrapping was carried out to validate these observations. A significant association between low DUSP6 expression and poorer disease specific survival outcome was maintained in this analysis ($p < 0.001$).

Target	p-value	Hazard ratio	95% C.I.
^a Cytoplasmic DUSP6	<0.001	2.044	1.386-3.013
^b Cytoplasmic DUSP6 (bootstrapping)	0.001	2.044	1.404-3.031

Table 29. Cox-regression analysis with categorical cytoplasmic DUSP6 expression and disease specific survival. The association of high and low cytoplasmic DUSP6 expression with 5yr DSS was explored using Cox-regression analysis. ^aCases with low DUSP6 expression demonstrated a poorer survival outcome in comparison to those above the cut off of 217.92 AQUA score, $p < 0.001$, HR 2.04, CI 1.4 – 3.0. ^bBootstrapping analysis was performed using 1000 samples to validate these findings. A significant association between low DUSP6 expression and poorer survival outcome was maintained in this analysis, $p < 0.001$, HR 2.04, CI 1.4 – 3.0.

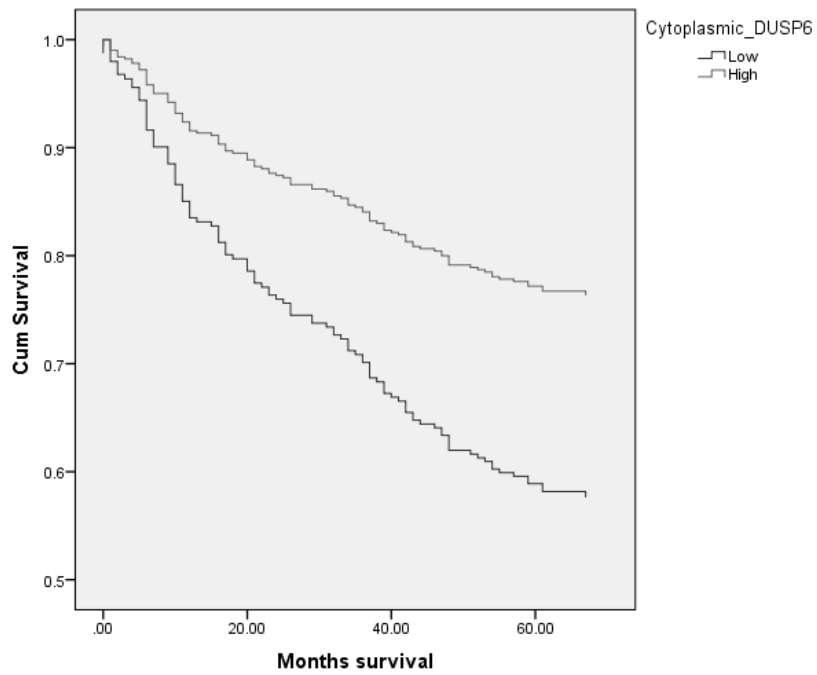


Figure 63. Kaplan-Meier plot for categorical cytoplasmic DUSP6 expression. Cytoplasmic DUSP6 expression was categorised by high and low expression as previously determined by X-tile software. Using Cox-regression analysis, individuals with low DUSP6 expression displayed a significantly shorter survival in comparison to those expression high cytoplasmic DUSP6, $p < 0.001$, HR 2.04, CI 1.4-3.0.

5.2.8 Low DUSP6 expression in Stage III colorectal cancer is associated with poor 5yr. overall survival and disease specific survival

In consideration of cytoplasmic DUSP6 expression being significantly associated with 5y. OS and DSS and previous trends observed in the association of DUSP6 expression with T stage, analysis into the impact of TNM staging and survival was carried out. Cox-regression analysis was carried with DUSP6 high and low expression and OS for each TNM stage (TNM stage I: high DUSP6 (n57) low DUSP6 (n10), stage II: high DUSP6 (n150) low DUSP6 (n36), stage III: high DUSP6 (n161) low DUSP6 (n60) and stage IV: high DUSP6 (n20) low DUSP6 (n4), Table 30. A significant association was identified between DUSP6 expression and OS in individuals with TNM stage III disease, Fig. 64.

TNM stage	p-value	Hazard ratio	95% C.I.
I	0.292	2.020	.546 – 7.467
II	0.311	1.362	.749 – 2.476
III	0.006*	1.791	1.183 – 2.713
IV	0.578	.704	.204 – 2.425

Table 30. Cox-regression analysis of TNM stage and relationship of cytoplasmic DUSP6 expression with 5yr overall survival. A significant relationship between DUSP6 expression and survival was identified in individuals with TNM stage 3 disease, p0.006 (HR 1.8, CI 1.18 – 2.71).

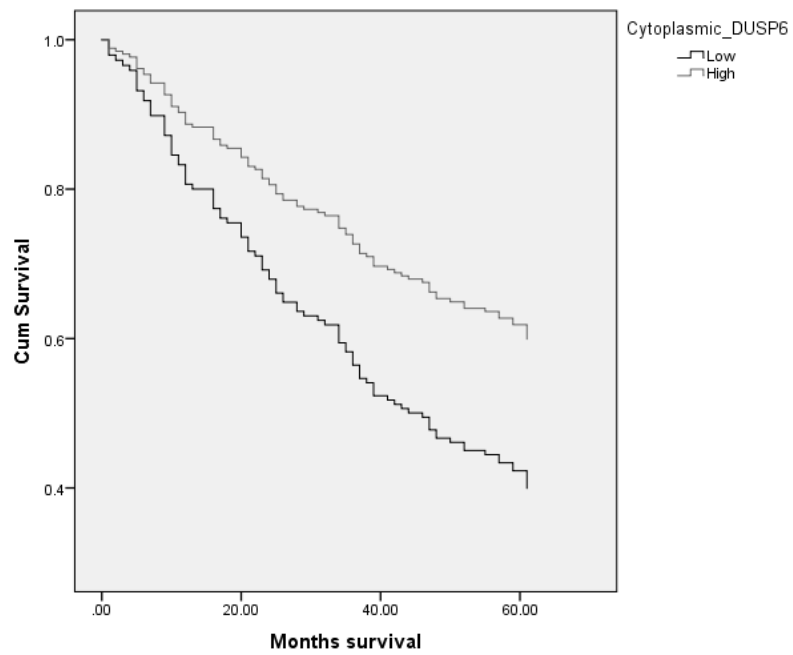


Figure 64. Kaplan-Meier for cytoplasmic DUSP6 expression in TNM stage 3 disease and overall survival. The relationship between cytoplasmic DUSP6 expression and 5yr OS was investigated with respect to TNM stage. In individuals with TNM stage 3 disease, a significant relationship was identified.

In overview, TNM stage III individuals demonstrated comparative median DSS to stage I and II (61, 67 and 65 months respectively). TNM stage specific investigation into the association between cytoplasmic DUSP6 and 5yr DSS was carried out using Cox-regression, Table 31. Analysis identified a significant association in individuals with TNM stage III disease and DUSP6 expression, Fig. 65. Individuals with low DUSP6 expression and TNM stage III disease demonstrated significantly poorer survival outcome than those with high DUSP6 expression. TNM stage III lesions with low DUSP6 expression demonstrated a median DSS of 44.5 months compared to those with high DUSP6 expression with a median of 61 months.

TNM stage	p-value	Hazard ratio	95% C.I.
I	0.527	2.077	.216-19.98
II	0.116	1.952	.848-4.490
III	0.004*	2.067	1.268-3.372
IV	0.699	.780	.221-2.753

Table 31. Cox-regression analysis of TNM stage and relationship of cytoplasmic DUSP6 expression with 5yr disease specific survival. The significant relationship between low DUSP6 expression and poorer survival outcome was limited to individuals with TNM stage 3 disease, $p=0.004$ (HR 2.067, CI 1.268-3.372). TNM stage III lesions with low DUSP6 expression demonstrated a median DSS of 44.5 months compared to those with high DUSP6 expression with a median of 61 months.

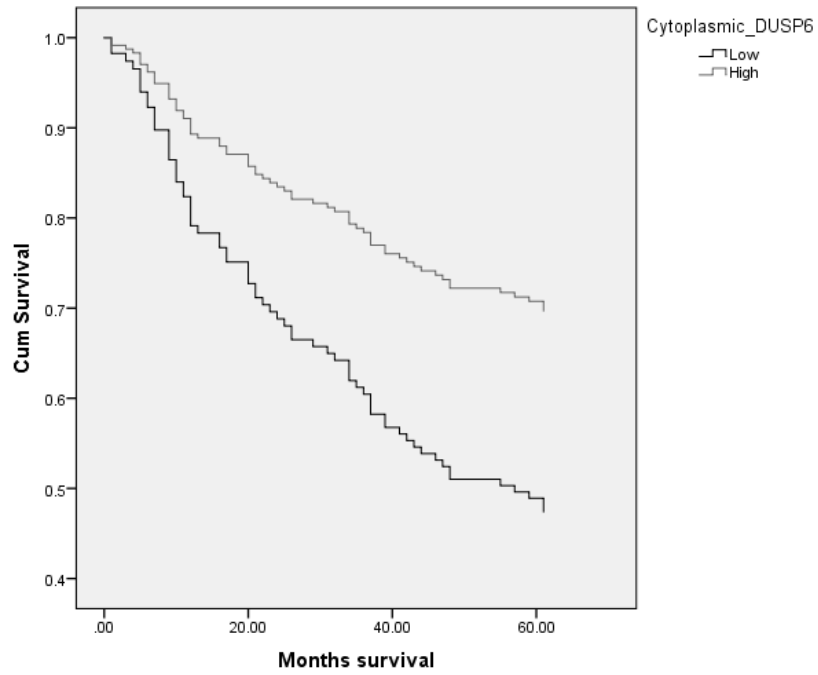


Figure 65. Kaplan-Meier for cytoplasmic DUSP6 expression in TNM stage 3 disease and disease specific survival. The relationship between cytoplasmic DUSP6 expression and 5yr DSS was investigated with respect to TNM stage. In individuals with TNM stage 3 disease, a significant relationship was identified.

5.3 Discussion

Previously in this series of studies it has been shown that DUSP6 demonstrates a dynamic range of protein expression intensity across colorectal adenocarcinoma FFPE tissue as determined by immunohistochemistry. In *KRAS* mutant lesions this was shown to associate with an increased ERK activation a finding consistent with observations in lung cancer⁵⁶. In addition to this observation, the association between DUSP6 protein expression and growth has been identified in a variety of tissues. In primary pancreatic cancer cell lines DUSP6 protein expression was downregulated in invasive carcinoma in comparison to pre-neoplastic lesions with mid-high grade dysplasia⁶⁴. In lung cancer studies, DUSP6 expression levels were significantly lower in lung cancer cell lines compared with normal bronchial epithelial cells⁶¹. Immunohistochemical analysis of DUSP6 expression in non-small cell lung cancer (NSCLC) resections identified a significant downregulation in DUSP6 expression in cancer compared with normal lung tissue¹²⁶. Similar findings were presented in a study by Ma et al¹²³, whereby a large proportion of oesophageal squamous cell carcinoma (ESCC) specimens demonstrated reduced DUSP6 protein expression with DUSP6 protein expression associating with pathological grade, this trend was also observed in assessment of cell lines.

Initial investigations into DUSP6 protein expression was carried out in this study cohort comprising 525 colorectal adenocarcinoma resections of varied histological stage by immunofluorescence. DUSP6 demonstrated a dynamic range of expression and upon assessment of associations with T stage and tumour size, trends were observed that both increasing histological stage i.e. T stage; however this effect analysis identified a small effect rendering these observations unlikely to transfer to being biologically relevant. This is contrasting with findings in other tissues.

DUSP6 is a direct regulator of activated ERK and it has been previously shown in this study that *RAS* mutations associate with increased ERK activation. In light of this, investigations into the associations between DUSP6 protein expression and *RAS* mutation status were carried out in this cohort. When assessed in isolation, neither nuclear or cytoplasmic DUSP6 expression was associated with *RAS* mutation status. When taken in the context relative to each other per case i.e. nuclear expression higher than cytoplasmic, a greater proportion of cases with *RAS* mutations demonstrated higher nuclear than cytoplasmic expression however the contribution of *RAS* mutation on the difference in DUSP6 expression was considered a small effect. This is suggestive that DUSP6 may provide a more sensitive measure of MAPK signalling pathway

hyperactivation i.e. by increased regulation of activated ERK, than *RAS* mutation status alone. Those lesions with hyper-activated MAPK pathway as a result of aberrations other than *RAS* mutation may be identified by DUSP6 protein expression.

The dynamic range of DUSP6 expression encouraged investigation into whether high DUSP6 expression; indicative of increased MAPK pathway activation, would confer resistance to Cetuximab treatment. Due to the limited number of patients treated with Cetuximab in the study cohort this was investigated using an *in vitro* model.

DUSP6 has two known isoforms; A: considered the canonical isoform with known ERK recognition, phosphatase function and nuclear export ability and B: a truncated form of the protein containing a coding sequence for ERK recognition and part of the phosphatase domain i.e. catalytic active site but lacks the nuclear export signal. Upon assessment of the DUSP6 protein expression of the 10 colorectal cell lines, a difference in the expression levels of isoform A and B expressed in each cell line became apparent. The greatest variation was observed in the expression levels of isoform B between cell lines; not all resistant cell lines demonstrated detectable levels of isoform B. This disparity in isoform B expression between cell lines may be suggestive of a novel function of this isoform.

To investigate the relationship between DUSP6 and ERK in adenocarcinoma, protein lysates from C99 cells transduced with DUSP6 were compared for ERK and p-ERK expression to parental lysates. Consistent with previous findings in the adenoma cell line RG/C2/80, over-expression of DUSP6 resulted in a marked decrease in p-ERK expression providing evidence of a functional relationship between DUSP6 and ERK in colorectal adenocarcinoma. The question however remains as to the contribution of isoform B to the regulation of ERK. The presence of high p-ERK in C99 parental cells compared to DUSP6 transduced cells fails to inform as to the sub-cellular location of p-ERK and whether isoform B has phosphatase ability. Lacking an NES and a portion of the phosphatase domain, can the DUSP6 B isoform dephosphorylate activated ERK?

Further investigations using cell fractionation techniques to assess the proportion of nuclear and cytoplasmic p-ERK with respect to the two DUSP6 isoforms is warranted. This could be achieved by assessing these features in protein lysates from C99 parental cells which have demonstrated undetectable levels of isoform A but high levels of isoform B and comparing with C99 DUSP6 transduced cells containing high levels of both isoforms. The presence of high nuclear p-ERK but low cytoplasmic p-ERK would suggest isoform B has the capacity to dephosphorylate p-ERK (cytoplasmic localised) but an inability to remove p-ERK from the nucleus due to the lack of NES.

All cell lines excluding NCI-H508 and C99 have previously been shown to be resistant to Cetuximab treatment¹²⁵ and demonstrated comparative levels of DUSP6 isoform A, supporting the hypothesis that high DUSP6 expression may confer resistance to Cetuximab. However, the only two cell lines responsive to Cetuximab treatment; NCI-H508 and C99, demonstrated the lowest ratio of DUSP6 isoforms A : B. In addition to this observation, these are the only two all *RAS* and *BRAF* wildtype cell lines. It could be hypothesised that the impact DUSP6 has on ERK regulation is dependent upon whether the pathway is constitutively activated by an aberration such as *RAS* mutation. Absence of constitutive activation of the MAPK pathway may relieve the requirement for a high degree of ERK regulation resulting in low expression of DUSP6 isoform A. To explore this further assessment of DUSP6 protein expression in all wildtype cell lines sensitive to cetuximab would be required, one would expect a low ratio of DUSP6 isoforms A : B in these cell lines.

The increase in sensitivity to Cetuximab treatment upon transduction of DUSP6 in C99 cells contradicted the initial hypothesis that low DUSP6 isoform A constitutes sensitivity to treatment. Cetuximab acts to bind the epidermal growth factor receptor (EGFR), targeting the receptor for degradation and reducing signal transduction through the MAPK signalling pathway which when activated results in transcription of genes responsible for cell proliferation and pro-survival. In the context of a *RAS* wildtype pathway considered responsive to Cetuximab due to absence of hyperactivation below the EGFR level, it is hypothesised that DUSP6 works to supplement Cetuximab activity by further reducing ERK dependent activation of proliferative and pro-survival genes, complementing Cetuximab activity thus increasing sensitivity to treatment.

To control for effects of transduction, C99 cells were transduced with eGFP. The protein expression levels of DUSP6 and ERK were comparable to the parental line. Despite protein expression of p-ERK being lower in the eGFP control in comparison to parental line, expression in DUSP6 transduced cells was markedly lower than the control suggesting that despite the potential impact transduction has on p-ERK expression, DUSP6 transduction markedly adds to this downregulation. However, the eGFP control involves the introduction of exogenous protein into the cell which most likely stresses the cell potentially impacting expression levels of proteins such as ERK, therefore to ensure the marked decrease in p-ERK can be attributed more conclusively to DUSP6 manipulation an empty vector control should be generated and protein expression analysis carried out again.

A study by Diaz-Garcia et al¹²⁶ highlighted the potential prognostic utility of DUSP6 in NSCLC. In this study high DUSP6 protein expression was associated with better survival outcome. The hypothesis that low DUSP6 protein expression confers poorer prognostic outcome was confirmed in this study cohort. For both overall survival and disease specific survival, individuals with low cytoplasmic DUSP6 expression demonstrated a poorer survival outcome. To understand the mechanisms underlying this observation further work is warranted to assess the effects DUSP6 overexpression has on phenotype using the C99 model generated from this body of work. One would hypothesise that over-expression of DUSP6 would slow the rate of proliferation as the increase in DUSP6 would increase negative regulation of ERK activation thus resulting in a decrease in transcription of proliferative and pro-survival genes in these individuals.

Interestingly, when DUSP6 protein expression was assessed based upon TNM stage specific overall and disease specific survival, a significant difference in patient outcome was restricted to TNM stage III patients. This expression profile remained significant within each assessable subgroup of the TNM stage III classification (i.e. IIIB and IIIC). The TNM stage III clinical classification characterises lesions which have invaded the surrounding muscularis mucosa and nodal positivity is present. It could be postulated that lesions of individuals with TNM stage III disease and low DUSP6 demonstrate more aggressive phenotypes than those expressing high levels of DUSP6. Identification of patients in this sub-group who progress to metastatic disease may provide insight into whether this hypothesis is supported.

In conclusion, herein this study, DUSP6 has been shown to exhibit tumour suppressive like role in colorectal adenocarcinoma. In contrast to findings in NSCLC, DUSP6 protein expression is not associated with histological grade in colorectal adenocarcinoma. Preliminary studies have demonstrated that increases in DUSP6 increase sensitivity to Cetuximab treatment, as has been observed in NSCLC and ovarian cancer however further investigation is required to confirm these findings. Possible mechanisms of increased sensitivity rely upon the complementary action of DUSP6 downregulation of activated ERK in addition to Cetuximab activity. It remains, however, the implication of DUSP6 isoform B on ERK regulation which warrants further investigation.

6. Discussion

Herein this body of work is presented a series of investigations into the role of DUSP6 in colorectal cancer. This study has investigated the role of DUSP6 in the pathogenic process of malignant transformation from normal – adenoma – adenocarcinoma, its function with respect to treatment response in adenocarcinoma and its prognostic utility.

Immunohistochemical analysis of colorectal tissue identified that DUSP6 protein expression increases in adenoma with increasing grade of dysplasia. In adenocarcinoma, DUSP6 expression is significantly decreased. This is consistent with findings in NSCLC and ESCC. ERK and p-ERK expression followed a similar pattern of expression^{61,62,116}. A proposed mechanism of action of DUSP6 upregulation in colorectal adenoma is detailed in Fig. 66.

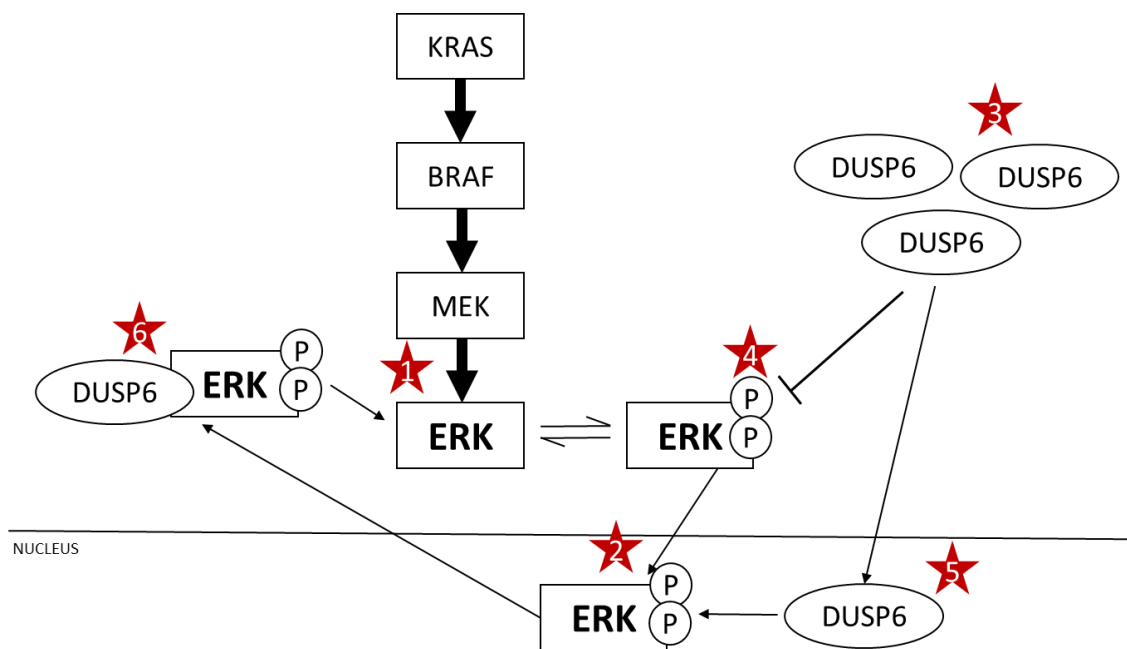


Figure 66. Schematic detailing activity of upregulated DUSP6 in colorectal adenoma. 1. ERK expression has been shown to be upregulated in colorectal adenoma in comparison to normal tissue. 2. Increased activated ERK (p-ERK) also shown in colorectal adenoma in comparison to normal tissue. 3. Upregulation of DUSP6 in response to increased ERK expression. 4. DUSP6 de-phosphorylates p-ERK and 5. Presence of NES in DUSP6 enables its shuttling to the nucleus to remove p-ERK to the cytoplasmic whereby 6. p-ERK is dephosphorylated to downregulate ERK activation of proliferative and pro-survival nuclear targets, ultimately acting to reduce cell growth and increase cell death.

DUSP6 regulates ERK activity in three ways: 1. De-phosphorylates ERK, inhibiting its activity, 2. DUSP6 anchors ERK irrespective of phosphorylation state in the cytoplasm, 3. DUSP6 shuttles to the nucleus to relocate activated ERK (p-ERK) to the cytoplasm. Evident from these activities, DUSP6 may act as a tumour suppressor by contributing to the reduction in p-ERK and subsequent p-ERK directed transcription of proliferative and pro-survival genes^{41-44,46}.

From the observations of DUSP6 expression in this study it is proposed that in adenoma, DUSP6 expression increases in response to increased ERK activity. For DUSP6 to successfully regulate overall ERK activity (thus reducing the proliferation and survival of dysregulated cells) DUSP6 expression must exceed ERK activation hence DUSP6 expression increases with increasing grade of dysplasia. Transduction of DUSP6 in adenoma RG/C2/80 cells resulted in a marked decrease in p-ERK protein expression, providing evidence for the regulation of ERK activity by DUSP6 in adenoma. Evidenced by the significant decrease in DUSP6 expression, in adenocarcinoma the regulation of ERK activity by DUSP6 appears to be lost perhaps due to the acquisition of additional aberrations which negate the reliance on ERK driving progression of dysregulated cells.

From the findings in this series of studies it is proposed that in colorectal adenoma the extent to which DUSP6 regulates ERK is dependent upon the balance of ERK and DUSP6 activation. Should the degree of ERK activation exceed the capacity of DUSP6 to de-phosphorylate and sequester p-ERK from the nucleus then cell proliferation and survival ensues. There is evidence to support a variety of mechanisms in the regulation of DUSP6 expression, Fig. 67.

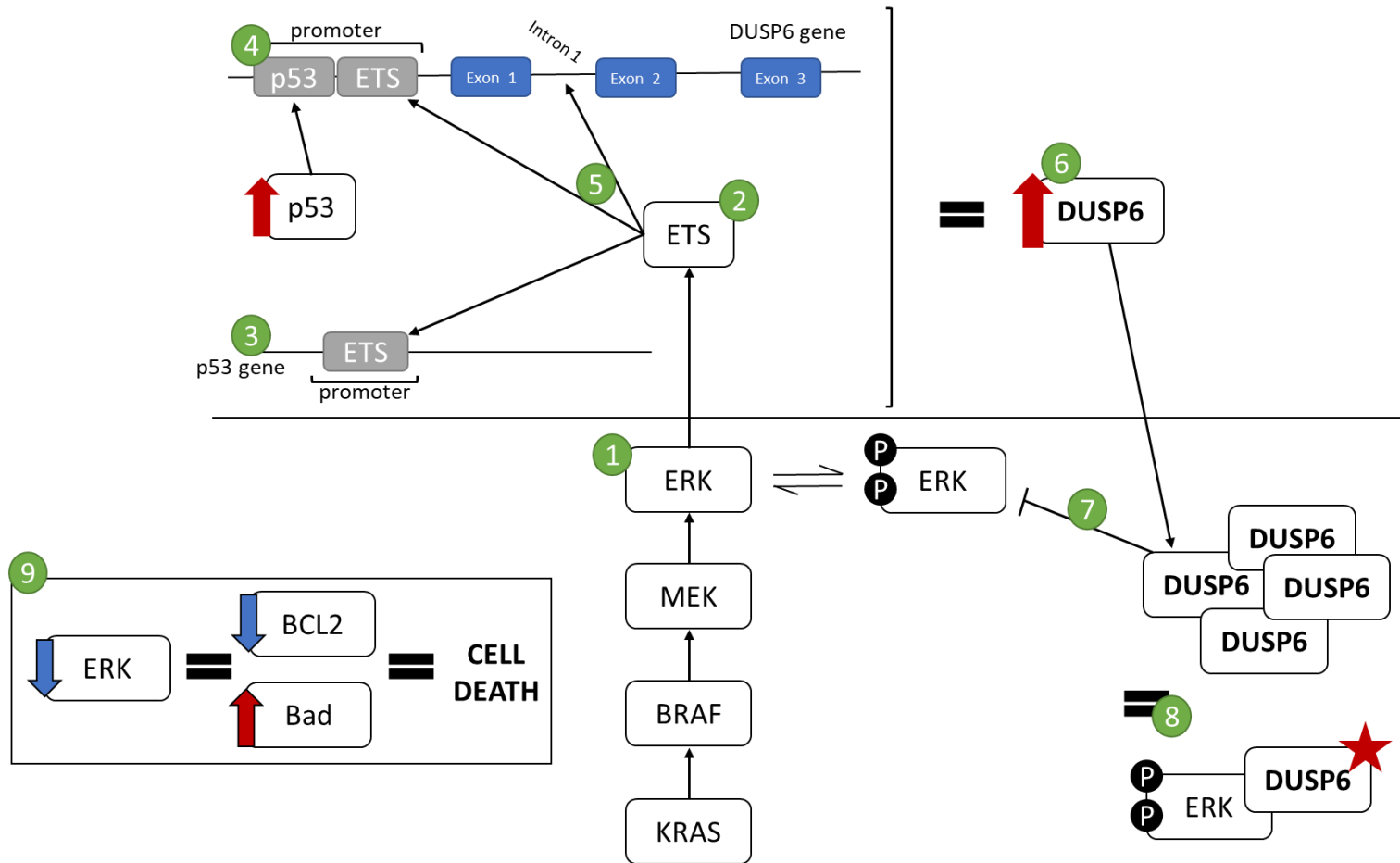


Figure 67. Schematic demonstrating the ERK-DUSP6 axis for regulation of cell proliferation and survival. 1. ERK activation promotes the upregulation of 2. ETS family transcription factors. ETS binding to motifs in the 3. p53 gene promoter resulting in upregulation of p53. 4. P53 binding to motifs in the DUSP6 gene promoter and intron 1 of the DUSP6 gene results in 6. upregulation of DUSP6. Increased DUSP6 is available to de-phosphorylate p-ERK which in turn 8. Targets DUSP6 for proteasomal degradation and 9. Reduced ERK expression and subsequently augments targets resulting in increased cell death. * Targeted for proteasomal degradation

It has been proposed by a variety of studies that DUSP6 expression is regulated by a variety of mechanisms including TP53 and ETS family of transcription factors. ETS factors have been shown to bind to ETS motifs in the p53 promoter resulting in upregulation of its transcription. TP53 has been shown to bind TP53 binding motifs in the DUSP6 promoter⁵⁷ thus providing evidence for a TP53 mediated upregulation of DUSP6. Further increasing DUSP6 upregulation, ETS motifs and binding has been confirmed in the promoter and intron 1 of the DUSP6 gene^{55,117,118}. With DUSP6 transcription increased by these mechanisms, a greater reserve of DUSP6 is available to dephosphorylate activated p-ERK. This reduces p-ERK mediated regulation of pro-survival factors such as BCL2 and increases pro-death factors such as Bad favouring cell death. Further investigation of the relevance of these mechanisms in the regulation of DUSP6 in colorectal tissue would be informative with aberrations in TP53 being pertinent to colorectal pathogenesis.

An additional mechanism for the regulation of DUSP6 expression is the association of DUSP6 with ERK which targets DUSP6 for proteasomal degradation⁵². Thus the greater degree of association between ERK and DUSP6 causes a greater degree of DUSP6 degradation. Degradation of DUSP6 to an extent below ERK activity enables the re-accumulation of activated ERK. In consideration of this, it is proposed that modulation of DUSP6/ERK activity occurs in an oscillatory manner, Fig. 68. Indeed, oscillatory patterns of signalling have been previously identified in the MAPK pathway by Hu et al., 2013¹²⁷.

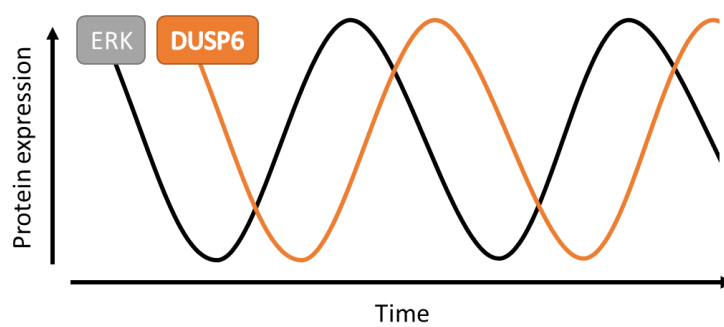


Figure 68. Proposed model of ERK – DUSP6 axis for regulation of MAPK signalling. It is hypothesised that DUSP6 expression oscillates with ERK activation. ERK activation results in upregulation of DUSP6 which negatively regulates and acts to reduce overall ERK activation. Association of DUSP6 with ERK targets DUSP6 for proteasomal degradation thus reducing DUSP6 levels. The reduction of DUSP6 allows for the recovery of activation ERK and thus the wave continues. When ERK exceeds DUSP6 expression, proliferation and survival is favoured.

Further work to elucidate the impact specific isoforms of DUSP6 have on ERK regulation in colorectal adenoma would prove informative to this series of studies. Indeed, the adenoma cell line RG/C2/80 demonstrated only detectable levels of isoform B. The immunohistochemical studies carried out herein failed to distinguish the expression of specific DUSP6 isoforms. Investigation into the isoforms present in the adenoma cohort of this study is warranted. Protein extraction and assessment of DUSP6 by western blot would enable identification of the ratios of both isoforms of DUSP6. Additionally, cell fractionation studies assessing the ratios of p-ERK in nuclear and cytoplasmic fractions of RG/C2/80 cell lines with and without DUSP6 transduction may elucidate whether isoform B can regulate ERK despite missing the nuclear export signal and part of the phosphatase domain.

In addition to a tumour suppressor role of DUSP6 in colorectal adenoma, a role of DUSP6 in response to Cetuximab treatment in colorectal adenocarcinoma is also identified. DUSP6 expression has been implicated in the response of a variety of cancers to a range of chemotherapeutic treatments⁸⁹⁻⁹².

Preliminary investigations identified that treatment of DUSP6 transduced cells resulted in an increased sensitivity to Cetuximab treatment. A proposed mechanism of action of DUSP6 over-expression in these cells is detailed in Fig. 69. In this model, it is inferred from these observations that over-expression of DUSP6 complements Cetuximab activity by further reducing ERK activation thus overall reducing activation of proliferative and pro-survival genes. Further studies are required to confirm these observations and investigation into the proposed mechanisms are advised. For example, forward-phase protein arrays to assess changes in members of ERK associated pathways such as apoptosis in C99 cells with and without DUSP6 transduction may prove informative as to the extent of additional ERK regulation by DUSP6 as a mechanism for increasing sensitivity to Cetuximab treatment.

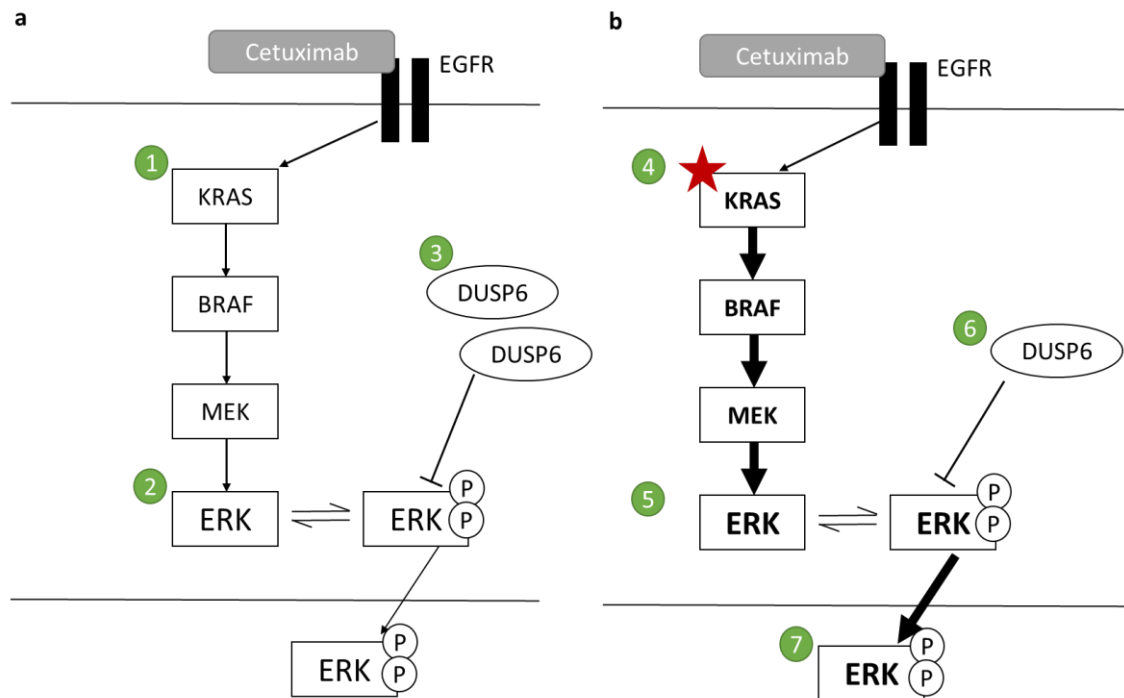


Figure 69. Schematic detailing complementary action of DUSP6 with Cetuximab treatment. A: 1. Wildtype *RAS* activates 2. ERK to a degree at which 3. DUSP6 activity can counteract ERK activation, thus reducing overall ERK signalling. B: 4. *RAS* mutation results in constitutive activation of the protein and subsequent hyperactivation of the MAPK signalling pathway including 5. ERK. The degree of ERK activation outweighs the activity of 6. DUSP6 regulation of ERK thus increasing 7. Activated p-ERK, thus conferring resistance to treatment. * *RAS* mutation

In conclusion, it is evident from this series of investigations that DUSP6 plays a tumour suppressive role in colorectal adenoma; to suppress proliferation and survival of dysregulated cell thus attempting to arrest malignant transformation. In adenocarcinoma, the impact of DUSP6 regulation over ERK activity appears lost which may be in part due to the acquisition of additional aberrations in signalling pathways driving tumorigenesis. Further studies are warranted to investigate the role of specific DUSP6 isoforms with respect to regulation of ERK in adenoma. Additional studies to confirm observations from preliminary studies identifying DUSP6 as a potential mediator of response to Cetuximab treatment are required. In addition to this investigation into the potential mechanisms underpinning the observed increases in sensitivity associated with DUSP6 expression in colorectal adenocarcinoma would prove informative to this series of investigations.

7. References

1. Cancer Research UK. Bowel Cancer Statistics. Available at: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer#heading-Two>. (Accessed: 12th June 2017)
2. Cancer Research UK. Introduction of the Faecal Immunochemical Test (FIT). (2018).
3. Cancer Research UK. Bowel Cancer Incidence Statistics. Available at: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/incidence#heading-Zero>. (Accessed: 18th September 2018)
4. Marley, A. R. & Nan, H. Epidemiology of colorectal cancer. *Int. J. Mol. Epidemiol. Genet.* **7**, 105–114 (2016).
5. Ponz de Leon, M. & Di Gregorio, C. Pathology of colorectal cancer. *Dig. Liver Dis.* **33**, 372–388 (2001).
6. Di Gregorio, C. *et al.* Histology of aberrant crypt foci in the human colon. *Histopathology* **30**, 328–334 (1997).
7. Fleming, M., Ravula, S., Tatishchev, S. F. & Wang, H. L. Colorectal carcinoma: Pathologic aspects. *J. Gastrointest. Oncol.* **3**, 153–173 (2012).
8. Edge, S., Byrd, D.R., Compton, C.C., Fritz, A.G., Greene, F., Trotti, A. AJCC Cancer Staging Handbook. **7th editio**, 173–206 (2010).
9. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
10. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
11. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
12. Grady, W. M. & Carethers, J. M. Genomic and Epigenetic Instability in Colorectal Cancer Pathogenesis. *Gastroenterology* **135**, 1079–1099 (2008).
13. Tariq, K. & Ghias, K. Colorectal cancer carcinogenesis: a review of mechanisms. *Cancer Biol. Med.* **13**, 120–135 (2016).

14. Pino, M. S. & Chung, D. C. The chromosomal instability pathway in colon cancer. *Gastroenterology* **138**, 2059–2072 (2010).
15. Vilar, E. & Gruber, S. B. Microsatellite instability in colorectal cancer-the stable evidence. *Nat. Rev. Clin. Oncol.* **7**, 153–162 (2010).
16. Grady, W. M. Genomic instability and colon cancer. *Cancer Metastasis Rev.* **23**, 11–27 (2004).
17. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).
18. Powell, S. M. *et al.* APC mutations occur early during colorectal tumorigenesis. *Nature* **359**, 235–237 (1992).
19. Kuipers, E. J. *et al.* COLORECTAL CANCER. *Nat. Rev. Dis. Prim.* **1**, 15065 (2015).
20. Shih, I.-M. *et al.* Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis. *Cancer Res.* **61**, 818–822 (2001).
21. Cahill, D. P., Kinzler, K. W., Vogelstein, B. & Lengauer, C. Genetic instability and darwinian selection in tumours. *Trends Cell Biol.* **9**, M57–M60 (1999).
22. Cross, W. *et al.* The evolutionary landscape of colorectal tumorigenesis. *Nat. Ecol. Evol.* **2**, 1661–1672 (2018).
23. Torlakovic, E., Skovlund, E., Snover, D. C., Torlakovic, G. & Nesland, J. M. Morphologic reappraisal of serrated colorectal polyps. *Am. J. Surg. Pathol.* **27**, 65–81 (2003).
24. Longacre, T. A. & Fenoglio-Preiser, C. M. Mixed hyperplastic adenomatous polyps/serrated adenomas. A distinct form of colorectal neoplasia. *Am. J. Surg. Pathol.* **14**, 524–537 (1990).
25. Patai, Á. V, Molnár, B., Tulassay, Z. & Sipos, F. Serrated pathway: Alternative route to colorectal cancer. *World J. Gastroenterol.* **19**, 607–615 (2013).
26. Thanki, K. 3967. *ri. et al.* Consensus Molecular Subtypes of Colorectal Cancer and their Clinical Implications. *Int. Biol. Biomed. J.* **3**, 105–111 (2017).
27. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350 (2015).

28. Müller, M. F., Ibrahim, A. E. K. & Arends, M. J. Molecular pathological classification of colorectal cancer. *Virchows Arch.* **469**, 125–134 (2016).
29. Isella, C. *et al.* Selective analysis of cancer-cell intrinsic transcriptional traits defines novel clinically relevant subtypes of colorectal cancer. *Nat. Commun.* **8**, 15107 (2017).
30. Liu, F., Yang, X., Geng, M. & Huang, M. Targeting ERK, an Achilles' Heel of the MAPK pathway, in cancer therapy. *Acta Pharm. Sin. B* **8**, 552–562 (2018).
31. Takayama, T. *et al.* Analysis of K-ras, APC, and beta-catenin in aberrant crypt foci in sporadic adenoma, cancer, and familial adenomatous polyposis. *Gastroenterology* **121**, 599–611 (2001).
32. Rosenberg, D. W. *et al.* Mutations in BRAF and KRAS differentially distinguish serrated versus non-serrated hyperplastic aberrant crypt foci in humans. *Cancer Res.* **67**, 3551–3554 (2007).
33. Malumbres, M. & Barbacid, M. RAS oncogenes: the first 30 years. *Nature reviews. Cancer* **3**, 459–465 (2003).
34. Patel, D. *et al.* Anti-epidermal growth factor receptor monoclonal antibody cetuximab inhibits EGFR/HER-2 heterodimerization and activation. *Int. J. Oncol.* **34**, 25–32 (2009).
35. Tzahar, E. *et al.* A hierarchical network of interreceptor interactions determines signal transduction by Neu differentiation factor/neuregulin and epidermal growth factor. *Mol. Cell. Biol.* **16**, 5276–5287 (1996).
36. Matrisian, L. M., Glaichenhaus, N., Gesnel, M. C. & Breathnach, R. Epidermal growth factor and oncogenes induce transcription of the same cellular mRNA in rat fibroblasts. *EMBO J.* **4**, 1435–1440 (1985).
37. Burotto, M., Chiou, V. L., Lee, J.-M. & Kohn, E. C. The MAPK pathway across different malignancies: a new perspective. *Cancer* **120**, 3446–3456 (2014).
38. Seger, R. & Krebs, E. G. The MAPK signaling cascade. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **9**, 726–735 (1995).
39. Wainstein, E. & Seger, R. The dynamic subcellular localization of ERK: Mechanisms of translocation and role in various organelles. *Current Opinion in Cell Biology* (2016). doi:10.1016/j.ceb.2016.01.007

40. Neuzillet, C., Hammel, P., Tijeras-Raballand, A., Couvelard, A. & Raymond, E. Targeting the Ras–ERK pathway in pancreatic adenocarcinoma. *Cancer Metastasis Rev.* **32**, 147–162 (2013).
41. Nichols, A. *et al.* Substrate recognition domains within extracellular signal-regulated kinase mediate binding and catalytic activation of mitogen-activated protein kinase phosphatase-3. *J. Biol. Chem.* **275**, 24613–24621 (2000).
42. Zhou, B. *et al.* Multiple regions of MAP kinase phosphatase 3 are involved in its recognition and activation by ERK2. *J. Biol. Chem.* **276**, 6506–6515 (2001).
43. Liu, S., Sun, J.-P., Zhou, B. & Zhang, Z.-Y. Structural basis of docking interactions between ERK2 and MAP kinase phosphatase 3. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5326–5331 (2006).
44. Karlsson, M., Mathers, J., Dickinson, R. J., Mandl, M. & Keyse, S. M. Both nuclear-cytoplasmic shuttling of the dual specificity phosphatase MKP-3 and its ability to anchor MAP kinase in the cytoplasm are mediated by a conserved nuclear export signal. *J. Biol. Chem.* (2004). doi:10.1074/jbc.M406720200
45. Furukawa, T. Impacts of Activation of the Mitogen-Activated Protein Kinase Pathway in Pancreatic Cancer. *Front. Oncol.* **5**, 23 (2015).
46. Bermudez, O., Pagès, G. & Gimond, C. The dual-specificity MAP kinase phosphatases: critical roles in development and cancer. *Am. J. Physiol. Physiol.* **299**, C189–C202 (2010).
47. Ahmad, M. K., Abdollah, N. A., Shafie, N. H., Yusof, N. M. & Razak, S. R. A. Dual-specificity phosphatase 6 (DUSP6): a review of its molecular characteristics and clinical relevance in cancer. *Cancer Biol. Med.* **15**, 14–28 (2018).
48. Stewart, A. E., Dowd, S., Keyse, S. M. & McDonald, N. Q. Crystal structure of the MAPK phosphatase Pyst1 catalytic domain and implications for regulated activation. *Nat. Struct. Biol.* **6**, 174–181 (1999).
49. Farooq, A. *et al.* Solution structure of ERK2 binding domain of MAPK phosphatase MKP-3: structural insights into MKP-3 activation by ERK2. *Mol. Cell* **7**, 387–399 (2001).
50. Furukawa, T. *et al.* Genomic analysis of DUSP6, a dual specificity MAP kinase phosphatase, in pancreatic cancer. *Cytogenet. Cell Genet.* **82**, 156–159 (1998).

51. UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* D158–D169 (2017).
52. Marchetti, S. *et al.* Extracellular signal-regulated kinases phosphorylate mitogen-activated protein kinase phosphatase 3/DUSP6 at serines 159 and 197, two sites critical for its proteasomal degradation. *Mol. Cell. Biol.* **25**, 854–864 (2005).
53. Ekerot, M. *et al.* Negative-feedback regulation of FGF signalling by DUSP6/MKP-3 is driven by ERK1/2 and mediated by Ets factor binding to a conserved site within the DUSP6/MKP-3 gene promoter. *Biochem. J.* **412**, 287–298 (2008).
54. Xu, S., Furukawa, T., Kanai, N., Sunamura, M. & Horii, A. Abrogation of DUSP6 by hypermethylation in human pancreatic cancer. *J. Hum. Genet.* **50**, 159–167 (2005).
55. Furukawa, T., Tanji, E., Xu, S. & Horii, A. Feedback regulation of DUSP6 transcription responding to MAPK1 via ETS2 in human cells. *Biochem. Biophys. Res. Commun.* **377**, 317–320 (2008).
56. Zhang, Z. *et al.* Dual specificity phosphatase 6 (DUSP6) is an ETS-regulated negative feedback mediator of oncogenic ERK signaling in lung cancer cells. *Carcinogenesis* (2010). doi:10.1093/carcin/bgq020
57. Piya, S. *et al.* DUSP6 is a novel transcriptional target of p53 and regulates p53-mediated apoptosis by modulating expression levels of Bcl-2 family proteins. *FEBS Lett.* (2012). doi:10.1016/j.febslet.2012.10.031
58. Zhang, H., Chi, Y., Gao, K., Zhang, X. & Yao, J. P53 protein-mediated Up-regulation of MAP kinase phosphatase 3 (MKP-3) contributes to the establishment of the cellular senescent phenotype through dephosphorylation of extracellular signal-regulated kinase 1/2 (ERK1/2). *J. Biol. Chem.* **290**, 1129–1140 (2015).
59. Bermudez, O., Marchetti, S., Pagès, G. & Gimond, C. Post-translational regulation of the ERK phosphatase DUSP6/MKP3 by the mTOR pathway. *Oncogene* **27**, 3685 (2008).
60. Kimura, M. *et al.* Identification of two common regions of allelic loss in chromosome arm 12q in human pancreatic cancer. *Cancer Res.* **58**, 2456–2460 (1998).
61. Okudela, K. *et al.* Down-Regulation of DUSP6 Expression in Lung Cancer. *Am. J. Pathol.* (2009). doi:10.2353/ajpath.2009.080489

62. Wong, V. C. L. *et al.* Tumor suppressor dual-specificity phosphatase 6 (DUSP6) impairs cell invasion and epithelial-mesenchymal transition (EMT)-associated phenotype. *Int. J. Cancer* (2012). doi:10.1002/ijc.25970
63. Messina, S. *et al.* Dual-specificity phosphatase DUSP6 has tumor-promoting properties in human glioblastomas. *Oncogene* **30**, 3813 (2011).
64. Furukawa, T., Sunamura, M., Motoi, F., Matsuno, S. & Horii, A. Potential Tumor Suppressive Pathway Involving DUSP6/MKP-3 in Pancreatic Cancer. *Am. J. Pathol.* (2003). doi:10.1016/S0002-9440(10)64315-5
65. Onder, T. T. *et al.* Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways. *Cancer Res.* **68**, 3645–3654 (2008).
66. Mendez, M. G., Kojima, S.-I. & Goldman, R. D. Vimentin induces changes in cell shape, motility, and adhesion during the epithelial to mesenchymal transition. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **24**, 1838–1851 (2010).
67. Miyamoto, Y., Suyama, K. & Baba, H. Recent Advances in Targeting the EGFR Signaling Pathway for the Treatment of Metastatic Colorectal Cancer. *Int. J. Mol. Sci.* **18**, (2017).
68. NHS Scotland, S. M. C. *cetuximab (Erbixux)*. (2015).
69. Li, S. *et al.* Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* **7**, 301–311 (2005).
70. Sunada, H., Magun, B. E., Mendelsohn, J. & MacLeod, C. L. Monoclonal antibody against epidermal growth factor receptor is internalized without stimulating receptor phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 3825–3829 (1986).
71. Goldstein, N. I., Prewett, M., Zuklys, K., Rockwell, P. & Mendelsohn, J. Biological efficacy of a chimeric antibody to the epidermal growth factor receptor in a human tumor xenograft model. *Clin. Cancer Res.* **1**, 1311–1318 (1995).
72. Wu, X. *et al.* Involvement of p27KIP1 in G1 arrest mediated by an anti-epidermal growth factor receptor monoclonal antibody. *Oncogene* **12**, 1397–1403 (1996).
73. Peng, D. *et al.* Anti-epidermal growth factor receptor monoclonal antibody 225 up-regulates p27KIP1 and induces G1 arrest in prostatic cancer cell line DU145. *Cancer Res.* **56**, 3666–3669 (1996).

74. Huang, S. M., Bock, J. M. & Harari, P. M. Epidermal growth factor receptor blockade with C225 modulates proliferation, apoptosis, and radiosensitivity in squamous cell carcinomas of the head and neck. *Cancer Res.* **59**, 1935–1940 (1999).
75. Bokemeyer, C. *et al.* Efficacy according to biomarker status of cetuximab plus FOLFOX-4 as first-line treatment for metastatic colorectal cancer: the OPUS study. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **22**, 1535–1546 (2011).
76. Bokemeyer, C. *et al.* Addition of cetuximab to chemotherapy as first-line treatment for KRAS wild-type metastatic colorectal cancer: pooled analysis of the CRYSTAL and OPUS randomised clinical trials. *Eur. J. Cancer* **48**, 1466–1475 (2012).
77. De Roock, W. *et al.* KRAS wild-type state predicts survival and is associated to early radiological response in metastatic colorectal cancer treated with cetuximab. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **19**, 508–515 (2008).
78. Karapetis, C. S. *et al.* K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N. Engl. J. Med.* **359**, 1757–1765 (2008).
79. Van Cutsem, E. *et al.* Cetuximab plus irinotecan, fluorouracil, and leucovorin as first-line treatment for metastatic colorectal cancer: updated analysis of overall survival according to tumor KRAS and BRAF mutation status. *J. Clin. Oncol.* **29**, 2011–2019 (2011).
80. Bokemeyer, S., *et al.* Treatment outcome according to tumor RAS mutation status in OPUS study patients with metastatic colorectal cancer (mCRC) randomized to FOLFOX4 with/without cetuximab. | 2014 ASCO Annual Meeting | Abstracts | Meeting Library. *J. Clin. Oncol.* **32**, 3506 (2014).
81. Ciardiello, Fortunato., Lenz, Heinz-Josef., Kohn, Claus-Henning., Heinemann, Volker., Tejpar, Sabine., Melezinek, I. Treatment outcome according to tumor RAS mutation status in CRYSTAL study patients with metastatic colorectal cancer (mCRC) randomized to FOLFIRI with/without Cetuximab. *ASCO* (2014).
82. Loupakis, F. *et al.* KRAS codon 61, 146 and BRAF mutations predict resistance to cetuximab plus irinotecan in KRAS codon 12 and 13 wild-type metastatic colorectal cancer. *Br. J. Cancer* **101**, 715–721 (2009).
83. Schirripa, M. *et al.* Role of NRAS mutations as prognostic and predictive markers in metastatic colorectal cancer. *Int. J. cancer* **136**, 83–90 (2015).

84. De Roock, W. *et al.* Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. *Lancet. Oncol.* **11**, 753–762 (2010).
85. Sorich, M. J. *et al.* Extended RAS mutations and anti-EGFR monoclonal antibody survival benefit in metastatic colorectal cancer: a meta-analysis of randomized, controlled trials. *Ann. Oncol.* **26**, 13–21 (2015).
86. Zhao, B. *et al.* Mechanisms of resistance to anti-EGFR therapy in colorectal cancer. *Oncotarget* **8**, 3980–4000 (2016).
87. Loboda, A. *et al.* A gene expression signature of RAS pathway dependence predicts response to PI3K and RAS pathway inhibitors and expands the population of RAS pathway activated tumors. *BMC Med. Genomics* **3**, 26 (2010).
88. Tian, S. *et al.* A combined oncogenic pathway signature of BRAF, KRAS and PI3KCA mutation improves colorectal cancer classification and cetuximab treatment prediction. *Gut* **62**, 540–549 (2013).
89. Hrustanovic, G. *et al.* RAS-MAPK dependence underlies a rational polytherapy strategy in EML4-ALK-positive lung cancer. *Nat. Med.* **21**, 1038–1047 (2015).
90. Chan, D. W. *et al.* Loss of MKP3 mediated by oxidative stress enhances tumorigenicity and chemoresistance of ovarian cancer cells. *Carcinogenesis* (2008). doi:10.1093/carcin/bgn167
91. Oliveras-Ferraros, C. *et al.* Stem cell property epithelial-to-mesenchymal transition is a core transcriptional network for predicting cetuximab (Erbitux™) efficacy in KRAS wild-type tumor cells. *J. Cell. Biochem.* **112**, 10–29 (2011).
92. Boeckx, C. *et al.* Overcoming cetuximab resistance in HNSCC: the role of AURKB and DUSP proteins. *Cancer Lett.* **354**, 365–377 (2014).
93. Williams, H.L., Walsh, K., Diamond, A., Oniscu, A., Deans, Z. C. Validation of the OncoPrint Focus Panel for Next Generation Sequencing of clinical tumour samples. *Virchows Arch.* (2018).
94. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotech* **29**, 24–26 (2011).
95. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV):

- high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
96. Camp, R. L., Chung, G. G. & Rimm, D. L. Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nat. Med.* **8**, 1323 (2002).
 97. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).
 98. Bagnyukova, T. V. *et al.* DUSP6 regulates drug sensitivity by modulating DNA damage response. *Br. J. Cancer* (2013). doi:10.1038/bjc.2013.353
 99. Sancak, Y. *et al.* The Rag GTPases bind raptor and mediate amino acid signaling to mTORC1. *Science* **320**, 1496–1501 (2008).
 100. NCBI. Basic Local Alignment Search Tool (Nucleotide).
 101. Handra-Luca, A., Bilal, H., Bertrand, J.-C. & Fouret, P. Extra-cellular signal-regulated ERK-1/ERK-2 pathway activation in human salivary gland mucoepidermoid carcinoma: association to aggressive tumor behavior and tumor cell proliferation. *Am. J. Pathol.* **163**, 957–967 (2003).
 102. Camp, R. L., Dolled-filhart, M. & Rimm, D. L. X-Tile : A New Bio-Informatics Tool for Biomarker Assessment and Outcome-Based Cut-Point Optimization. **10**, 7252–7259 (2004).
 103. Forbes, S. A. *et al.* The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr. Protoc. Hum. Genet.* **CHAPTER**, Unit-10.11 (2008).
 104. Hall, T. BioEdit. (2005).
 105. Alberts, B., Johnson, A., Lewis, J., *et al.* *Molecular Biology of the Cell.* (2002).
 106. Mazumdar, M. & Glassman, J. R. Categorizing a prognostic variable: review of methods, code for easy implementation and applications to decision-making about cancer treatments. *Stat. Med.* **19**, 113–132 (2000).
 107. Greer, C. E., Peterson, S. L., Kiviat, N. B. & Manos, M. M. PCR Amplification from Paraffin-Embedded Tissues: Effects of Fixative and Fixation Time. *Am. J. Clin. Pathol.* **95**, 117–124 (1991).

108. Srinivasan, M., Sedmak, D. & Jewell, S. Effect of Fixatives and Tissue Processing on the Content and Integrity of Nucleic Acids. *Am. J. Pathol.* **161**, 1961–1971 (2002).
109. Forbes, S. A. *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–D783 (2017).
110. Schubbert, S., Shannon, K. & Bollag, G. Hyperactive Ras in developmental disorders and cancer. **7**, 295–308 (2007).
111. Prior, I. A., Lewis, P. D. & Mattos, C. A comprehensive survey of Ras mutations in cancer. *Cancer Res.* **72**, 2457–2467 (2012).
112. Janakiraman, M. *et al.* Genomic and biological characterization of exon 4 KRAS mutations in human cancer. *Cancer Res.* **70**, 5901–5911 (2010).
113. Vogelstein, B. *et al.* Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* **319**, 525–532 (1988).
114. Walther, A. *et al.* Genetic prognostic and predictive markers in colorectal cancer. *Nat. Rev. Cancer* **9**, 489 (2009).
115. Paraskeva, C., Finerty, S., Mountford, R. A. & Powell, S. C. Specific cytogenetic abnormalities in two new human colorectal adenoma-derived epithelial cell lines. *Cancer Res.* **49**, 1282–1286 (1989).
116. Furukawa, T., Sunamura, M., Motoi, F., Matsuno, S. & Horii, A. Potential tumor suppressive pathway involving DUSP6/MKP-3 in pancreatic cancer. *Am. J. Pathol.* **162**, 1807–1815 (2003).
117. Nunes-Xavier, C. E. *et al.* Differential up-regulation of MAP kinase phosphatases MKP3/DUSP6 and DUSP5 by Ets2 and c-Jun converge in the control of the growth arrest versus proliferation response of MCF-7 breast cancer cells to phorbol ester. *J. Biol. Chem.* **285**, 26417–26430 (2010).
118. Venanzoni, M. C., Robinson, L. R., Hodge, D. R., Kola, I. & Seth, A. ETS1 and ETS2 in p53 regulation: spatial separation of ETS binding sites (EBS) modulate protein: DNA interaction. *Oncogene* **12**, 1199–1204 (1996).
119. Hao, X. P. *et al.* The spectrum of p53 mutations in colorectal adenomas differs from that in colorectal carcinomas. *Gut* **50**, 834 LP-839 (2002).

120. Yamane, L., Scapulatempo-Neto, C., Reis, R. M. & Guimarães, D. P. Serrated pathway in colorectal carcinogenesis. *World J. Gastroenterol.* **20**, 2634–2640 (2014).
121. Prior, I. A., Lewis, P. D. & Mattos, C. A comprehensive survey of ras mutations in cancer. *Cancer Research* (2012). doi:10.1158/0008-5472.CAN-11-2612
122. Bermudez, O. *et al.* Post-transcriptional regulation of the DUSP6/MKP-3 phosphatase by MEK/ERK signaling and hypoxia. *J. Cell. Physiol.* **226**, 276–284 (2011).
123. Ma, J., Yu, X., Guo, L. & Lu, S. H. DUSP6, a tumor suppressor, is involved in differentiation and apoptosis in esophageal squamous cell carcinoma. *Oncol. Lett.* (2013). doi:10.3892/ol.2013.1605
124. Cui, Y. *et al.* Elevated expression of mitogen-activated protein kinase phosphatase 3 in breast tumors: a mechanism of tamoxifen resistance. *Cancer Res.* **66**, 5950–5959 (2006).
125. Medico, E. *et al.* The molecular landscape of colorectal cancer cell lines unveils clinically actionable kinase targets. *Nat. Commun.* **6**, 7002 (2015).
126. Diaz-Garcia, C. V. *et al.* Prognostic value of dual-specificity phosphatase 6 expression in non-small cell lung cancer. *Tumour Biol.* **36**, 1199–1206 (2015).
127. Hu, H. *et al.* Feedforward and feedback regulation of the MAPK and PI3K oscillatory circuit in breast cancer. *Cell. Signal.* **25**, 26–32 (2013).

8. Appendix

8.1 Cohort 1

8.1.1 Genotyping of cohort by pyrosequencing

Case ID	Dysplasia group	Histological group	KRAS genotype
1	High	Tubulovillous	c.34G>T
2	High	Tubulovillous	WT
3	High	Tubulovillous	WT*
4	High	Tubulovillous	c.35G>A
5	High	Tubular	c.182T>C
6	High	Tubular	WT
7	High	Tubular	WT
8	High	Tubulovillous	WT
9	High	Tubulovillous	~
10	High	Tubulovillous	WT
11	High	Tubulovillous	WT**
12	High	Tubulovillous	WT
13	High	Villous	WT*
14	High	Tubulovillous	c.34G>A
15	High	Tubulovillous	WT
16	Low	Tubular	^
17	Low	Tubulovillous	c.38G>A
18	Low	Tubular	WT
19	Low	Tubulovillous	WT
20	Low	Tubular	WT
21	Low	Tubulovillous	WT
22	Low	Tubulovillous	c.34G>A
23	Low	Villous	c.34G>T, c.38G>A
24	Low	Villous	c.35G>A
25	Low	Villous	WT
26	Low	Villous	c.38G>A
27	Low	Tubular	WT
28	Low	Tubulovillous	c.35G>A
29	Low	Tubular	c.35G>A
30	Low	Tubular	WT
53	Low	Serrated	WT
54	Low	Serrated	WT*
55	Low	Serrated	WT*
56	Low	Tubular	WT
57	Low	Serrated	WT
58	Low	Tubulovillous	WT

59	Low	Tubular	WT
60	Low	Tubular	WT
61	Low	Tubular	WT
62	Low	Tubular	WT*
63	Low	Tubular	WT
64	Low	Tubular	WT
65	Low	Tubular	^
66	Low	Tubulovillous	c.182T>A
67	Low	Tubular	WT*
68	Low	Tubular	WT*
69	Low	Tubular	WT
70	Low	Tubular	WT*
71	Low	Tubular	WT
72	Low	Tubular	WT*
73	Low	Tubulovillous	WT
74	Low	Tubulovillous	WT*
75	Low	Tubulovillous	WT
76	Low	Tubulovillous	WT
77	Low	Tubulovillous	c.35G>A
78	Low	Tubulovillous	c.35G>A
79	Low	Tubulovillous	WT
80	Low	Tubulovillous	c.183T>G
81	Low	Tubulovillous	WT*
82	Low	Tubulovillous	WT
83	Low	Tubulovillous	c.35G>C
84	Low	Tubulovillous	WT*
85	Low	Tubulovillous	c.38G>A
86	Low	Tubulovillous	WT
87	High	Tubulovillous	WT*
88	Low	Villous	c.34G>T
89	Low	Villous	c.35G>A
90	Low	Villous	WT
91	Low	Tubular	WT*
92	Low	Tubular	^
93	Low	Villous	WT
94	Low	Villous	c.35G>T
95	High	Villous	c.35G>T
96	Low	Villous	c.38G>A
97	Low	Villous	WT
98	High	Villous	c.35G>A
99	Low	Villous	WT*
100	Low	Villous	^
101	Low	Villous	c.183T>G
102	Low	Villous	c.35G>T
46	Low	Serrated	WT*

47	Low	Villous	WT
48	Low	Serrated	^
49	Low	Serrated	^
50	Low	Serrated	c.35G>A
51	Low	Serrated	WT*
52	Low	Serrated	WT*

Table 32. Genotyping of cohort by pyrosequencing. Pyrosequencing was carried out for each adenoma in cohort 1 (comprising 147 colorectal lesions ranging from normal mucosa, adenoma to adenocarcinoma). Variant in *KRAS* codon 12, 13, 59 and 61 were assessed. Six cases failed sequencing. One case was not tested due to tissue availability. **KRAS* exon 1 only. ***KRAS* exon 2 only. ^Failed sequencing. ~Not tested.

8.1.2 DUSP6 protein expression: QuPath analysis data

Case	Classification			Ratio of target positive cells			
	Dysplasia class	Adenoma-carcinoma sequence classification	Histology classification	ERK	pERK	Nuclear DUSP6	Cytoplasmic DUSP6
1	High	Adenoma	Tubulovillous	0.85	0.10	0.15	0.11
2	High	Adenoma	Tubulovillous	0.97	0.22	0.69	0.54
3	High	Adenoma	Tubulovillous	0.87	0.09	0.01	0.00
4	High	Adenoma	Tubulovillous	0.82	0.31	0.03	0.02
5	High	Adenoma	Tubular	0.74	0.82	0.48	0.35
6	High	Adenoma	Tubular	0.93	0.26	0.63	0.49
7	High	Adenoma	Tubular	0.89	0.11	0.86	0.76
8	High	Adenoma	Tubulovillous	0.40	0.28	0.34	0.16
9	High	Adenoma	Tubulovillous	0.91	0.12	0.30	0.15
10	High	Adenoma	Tubulovillous	0.67	0.80	0.27	0.08
11	High	Adenoma	Tubulovillous	0.60	0.14	0.48	0.44
12	High	Adenoma	Tubulovillous	0.44	0.67	0.72	0.66
13	High	Adenoma	Villous	0.71	0.32	0.80	0.69
14	High	Adenoma	Tubulovillous	0.70	0.25	0.77	0.65
15	High	Adenoma	Tubulovillous	0.96	0.41	0.72	0.52
16	Low	Adenoma	Tubular	0.86	0.39	0.14	0.09
17	Low	Adenoma	Tubulovillous	0.73	0.29	0.55	0.32
18	Low	Adenoma	Tubular	0.80	0.33	0.22	0.11
19	Low	Adenoma	Tubulovillous	0.88	0.14	0.35	0.22
20	Low	Adenoma	Tubular	0.93	0.07	0.53	0.35
21	Low	Adenoma	Tubulovillous	0.87	0.68	0.32	0.08
22	Low	Adenoma	Tubulovillous	0.52	0.54	0.44	0.18
23	Low	Adenoma	Villous	0.90	0.25	0.19	0.08

24	Low	Adenoma	Villous	0.94	0.03	0.28	0.16
25	Low	Adenoma	Villous	0.93	0.14	0.27	0.12
26	Low	Adenoma	Villous	0.91	0.58	0.40	0.17
27	Low	Adenoma	Tubular	0.96	0.21	0.01	0.00
28	Low	Adenoma	Tubulovillous	0.97	0.33	0.57	0.35
29	Low	Adenoma	Tubular	0.89	0.42	0.64	0.39
30	Low	Adenoma	Tubular	0.94	0.37	0.76	0.41
31	Normal	Normal	Normal	0.99	0.00	0.26	0.07
32	Normal	Normal	Normal	0.94	0.01	0.05	0.00
33	Normal	Normal	Normal	0.37	0.01	0.41	0.04
34	Normal	Normal	Normal	0.74	0.16	0.62	0.35
35	Normal	Normal	Normal	0.40	0.51	0.08	0.00
36	Normal	Normal	Normal	0.11	0.73	0.12	0.00
37	Normal	Normal	Normal	0.82	0.03	0.06	0.00
38	Normal	Normal	Normal	0.31	0.28	0.28	0.01
39	Normal	Normal	Normal	0.13	0.02	0.56	0.12
40	Normal	Normal	Normal	0.46	0.10	0.48	0.20
41	Normal	Normal	Normal	0.93	0.00	0.38	0.11
42	Normal	Normal	Normal	0.53	0.01	0.41	0.05
43	Normal	Normal	Normal	0.64	0.57	0.46	0.11
44	Normal	Normal	Normal	0.57	0.01	0.37	0.04
45	Normal	Normal	Normal	0.47	0.02	0.01	0.20
46	Low	Adenoma	Serrated	0.67	0.03	0.06	0.00
47	Low	Adenoma	Villous	0.50	0.01	0.54	0.35
48	Low	Adenoma	Serrated	0.62	0.10	0.32	0.08
49	Low	Adenoma	Serrated	0.42	0.00	0.01	0.31
50	Low	Adenoma	Serrated	0.79	0.18	0.04	0.00
51	Low	Adenoma	Serrated	0.42	0.22	0.17	0.01

52	Low	Adenoma	Serrated	0.95	0.12	0.36	0.01
53	Low	Adenoma	Serrated	0.76	0.00	0.37	0.06
54	Low	Adenoma	Serrated	0.85	0.36	0.19	0.01
55	Low	Adenoma	Serrated	0.96	0.02	0.21	0.00
56	Low	Adenoma	Tubular	0.99	0.26	0.30	0.06
57	Low	Adenoma	Serrated	0.17	0.00	0.05	0.00
58	Low	Adenoma	Tubulovillous	0.71	0.33	0.10	0.03
59	Low	Adenoma	Tubular	0.87	0.27	0.22	0.07
60	Low	Adenoma	Tubular	0.93	0.24	0.09	0.01
61	Low	Adenoma	Tubular	0.81	0.57	0.13	0.05
62	Low	Adenoma	Tubular	0.94	0.29	0.10	0.03
63	Low	Adenoma	Tubular	0.97	0.09	0.53	0.22
64	Low	Adenoma	Tubular	0.92	0.30	0.62	0.38
65	Low	Adenoma	Tubular	0.78	0.72	0.30	0.08
66	Low	Adenoma	Tubulovillous	0.91	0.44	0.53	0.16
67	Low	Adenoma	Tubular	0.80	0.50	0.03	0.08
68	Low	Adenoma	Tubular	0.35	0.02	0.14	0.03
69	Low	Adenoma	Tubular	0.94	0.49	0.82	0.45
70	Low	Adenoma	Tubular	0.64	0.60	0.08	0.03
71	Low	Adenoma	Tubular	0.95	0.60	0.06	0.02
72	Low	Adenoma	Tubular	0.73	0.13	0.54	0.17
73	Low	Adenoma	Tubulovillous	0.63	0.46	0.26	0.10
74	Low	Adenoma	Tubulovillous	0.81	0.80	0.12	0.03
75	Low	Adenoma	Tubulovillous	0.68	0.82	0.10	0.04
76	Low	Adenoma	Tubulovillous	0.56	0.10	0.71	0.38
77	Low	Adenoma	Tubulovillous	0.56	0.15	0.19	0.13

78	Low	Adenoma	Tubulovillous	0.78	0.25	0.14	0.03
79	Low	Adenoma	Tubulovillous	0.85	0.14	0.16	0.07
80	Low	Adenoma	Tubulovillous	0.83	0.25	0.28	0.14
81	Low	Adenoma	Tubulovillous	0.88	0.20	0.20	0.07
82	Low	Adenoma	Tubulovillous	0.94	0.29	0.18	0.07
83	Low	Adenoma	Tubulovillous	0.96	0.81	0.62	0.30
84	Low	Adenoma	Tubulovillous	0.80	0.59	0.05	0.02
85	Low	Adenoma	Tubulovillous	0.94	0.50	0.38	0.17
86	Low	Adenoma	Tubulovillous	0.92	0.30	0.19	0.10
87	High	Adenoma	Tubulovillous	0.98	0.52	0.59	0.35
88	Low	Adenoma	Villous	0.89	0.71	0.75	0.42
89	Low	Adenoma	Villous	0.97	0.55	0.66	0.14
90	Low	Adenoma	Villous	0.44	0.05	0.71	0.44
91	Low	Adenoma	Tubular	0.21	0.26	0.90	0.34
92	Low	Adenoma	Tubular	0.66	0.50	0.26	0.03
93	Low	Adenoma	Villous	0.89	0.18	0.22	0.09
94	Low	Adenoma	Villous	0.97	0.94	0.79	0.41
95	High	Adenoma	Villous	0.97	0.82	0.84	0.42
96	Low	Adenoma	Villous	0.86	0.09	0.51	0.22
97	Low	Adenoma	Villous	0.79	0.57	0.32	0.15
98	High	Adenoma	Villous	0.50	0.21	0.69	0.12
99	Low	Adenoma	Villous	0.96	0.82	0.22	0.02
100	Low	Adenoma	Villous	0.93	0.25	0.52	0.14
101	Low	Adenoma	Villous	0.80	0.11	0.17	0.03
102	Low	Adenoma	Villous	0.89	0.67	0.17	0.05
103	Adenocarcinoma	Adenocarcinoma	dMMR	0.83	0.01	0.32	0.26
104	Adenocarcinoma	Adenocarcinoma	dMMR	0.01	0.00	0.03	0.00
105	Adenocarcinoma	Adenocarcinoma	dMMR	0.29	0.00	0.11	0.01

106	Adenocarcinoma	Adenocarcinoma	dMMR	0.18	0.00	0.07	0.01
107	Adenocarcinoma	Adenocarcinoma	dMMR	0.02	0.13	0.54	0.15
108	Adenocarcinoma	Adenocarcinoma	dMMR	0.03	0.00	0.51	0.02
109	Adenocarcinoma	Adenocarcinoma	dMMR	0.01	0.01	0.11	0.00
110	Adenocarcinoma	Adenocarcinoma	dMMR	0.07	0.00	0.47	0.21
111	Adenocarcinoma	Adenocarcinoma	dMMR	0.01	0.01	0.00	0.00
112	Adenocarcinoma	Adenocarcinoma	dMMR	0.04	0.00	0.13	0.08
113	Adenocarcinoma	Adenocarcinoma	dMMR	0.09	0.00	0.03	0.00
114	Adenocarcinoma	Adenocarcinoma	dMMR	0.01	0.00	0.05	0.01
115	Adenocarcinoma	Adenocarcinoma	dMMR	0.12	0.00	0.00	0.00
116	Adenocarcinoma	Adenocarcinoma	dMMR	0.01	0.00	0.00	0.00
117	Adenocarcinoma	Adenocarcinoma	dMMR	0.01	0.00	0.02	0.00
118	Adenocarcinoma	Adenocarcinoma	<i>KRAS</i> mutant	0.05	0.00	0.01	0.00
119	Adenocarcinoma	Adenocarcinoma	<i>KRAS</i> mutant	0.51	0.07	0.02	0.02
120	Adenocarcinoma	Adenocarcinoma	<i>KRAS</i> mutant	0.75	0.00	0.01	0.00
121	Adenocarcinoma	Adenocarcinoma	<i>KRAS</i> mutant	0.14	0.00	0.07	0.05
122	Adenocarcinoma	Adenocarcinoma	<i>KRAS</i> mutant	0.60	0.82	0.02	0.01
123	Adenocarcinoma	Adenocarcinoma	<i>KRAS</i> mutant	0.17	0.01	0.08	0.18
124	Adenocarcinoma	Adenocarcinoma	<i>KRAS</i> mutant	0.31	0.00	0.18	0.16
125	Adenocarcinoma	Adenocarcinoma	<i>KRAS</i> mutant	0.49	0.04	0.00	0.00
126	Adenocarcinoma	Adenocarcinoma	<i>KRAS</i> mutant	0.23	0.07	0.02	0.02
127	Adenocarcinoma	Adenocarcinoma	<i>KRAS</i> mutant	0.97	0.04	0.01	0.01
128	Adenocarcinoma	Adenocarcinoma	<i>KRAS</i> mutant	0.05	0.00	0.20	0.17
129	Adenocarcinoma	Adenocarcinoma	<i>KRAS</i> mutant	0.83	0.01	0.11	0.00
130	Adenocarcinoma	Adenocarcinoma	<i>KRAS</i> mutant	0.49	0.00	0.24	0.27
131	Adenocarcinoma	Adenocarcinoma	<i>KRAS</i> mutant	0.35	0.01	0.98	0.85
132	Adenocarcinoma	Adenocarcinoma	<i>KRAS</i> mutant	0.25	0.01	0.99	0.78
133	Adenocarcinoma	Adenocarcinoma	<i>TP53</i> mutant	0.01	0.00	1.00	0.88

134	Adenocarcinoma	Adenocarcinoma	<i>TP53</i> mutant	0.54	0.00	0.76	0.26
135	Adenocarcinoma	Adenocarcinoma	<i>TP53</i> mutant	0.01	0.00	0.45	0.26
136	Adenocarcinoma	Adenocarcinoma	<i>TP53</i> mutant	0.24	0.00	0.95	0.74
137	Adenocarcinoma	Adenocarcinoma	<i>TP53</i> mutant	0.98	0.02	1.00	
138	Adenocarcinoma	Adenocarcinoma	<i>TP53</i> mutant	0.28		0.93	0.76
139	Adenocarcinoma	Adenocarcinoma	<i>TP53</i> mutant	0.13	0.02	0.87	0.69
140	Adenocarcinoma	Adenocarcinoma	<i>TP53</i> mutant	0.91	0.03	0.97	0.92
141	Adenocarcinoma	Adenocarcinoma	<i>TP53</i> mutant	0.01	0.01	0.23	0.04
142	Adenocarcinoma	Adenocarcinoma	<i>TP53</i> mutant	0.61	0.00	0.48	0.21
143	Adenocarcinoma	Adenocarcinoma	<i>TP53</i> mutant	0.02	0.07	0.58	0.42
144	Adenocarcinoma	Adenocarcinoma	<i>TP53</i> mutant	0.30	0.42	0.03	0.01
145	Adenocarcinoma	Adenocarcinoma	<i>TP53</i> mutant	0.01	0.05	0.93	0.93
146	Adenocarcinoma	Adenocarcinoma	<i>TP53</i> mutant	0.04	0.01	0.72	0.26
147	Adenocarcinoma	Adenocarcinoma	<i>TP53</i> mutant	0.27	0.09	0.83	0.46

Table 33. Ratio of target positive cells. Immunohistochemistry was carried out for each case of cohort 1. Cases were classified based upon lesions type (normal mucosa, adenoma, adenocarcinoma), degree of dysplasia (low or high grade dysplasia – adenoma only) and histological and molecular classification (normal mucosa, tubular adenoma, tubulovillous adenoma, villous adenoma, serrated adenoma, adenocarcinoma with deficient mis-match repair (dMMR), adenocarcinoma with KRAS mutation or adenocarcinoma with TP53 mutation. The ratio of target positive cells out of total cells assessed for each case is presented. For each classification the median ratio of target positive cells was calculated. Targets included DUSP6 (nuclear and cytoplasmic compartment), ERK (cytoplasmic compartment) and p-ERK (nuclear compartment).

8.1.3 Assessment of ERK expression in RG/C2/80 cell lines by western blot (repeats)

Western blots on protein lysates from RG/C2/80 cell lines (Parental, eGFP and DUSP6) for ERK expression were repeated, Fig. 70. Confirming initial findings, no marked difference in ERK protein expression was observed.

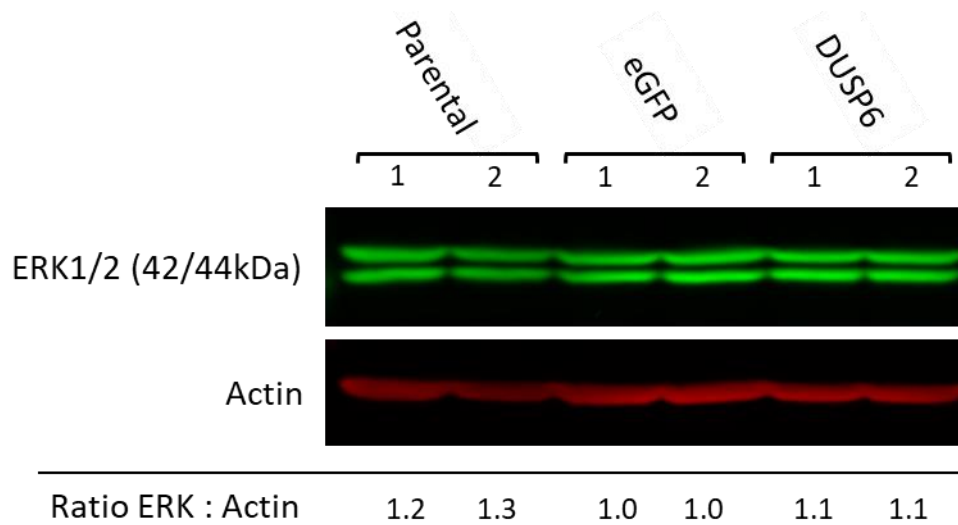


Figure 70. Western blot repeats of ERK protein expression in RG/C2/80 cell lines. ERK protein expression assessment was repeated in RG/C2/80 cells line (parental, enhanced green fluorescent protein (eGFP) transduced and DUSP6 transduced. No marked difference in ERK protein expression between the three cell lines was observed.

8.1.4 Assessment of p-ERK expression in RG/C2/80 cell lines by western blot (repeats)

Western blots on protein lysates from RG/C2/80 cell lines (Parental, eGFP and DUSP6) for p-ERK expression were repeated, Fig. 71. Confirming initial findings, a marked decrease in p-ERK protein expression was observed in RG/C2/80 (DUSP6 transduced) cells compared to parental (75% decrease) and eGFP transduced (72% decrease) lines.

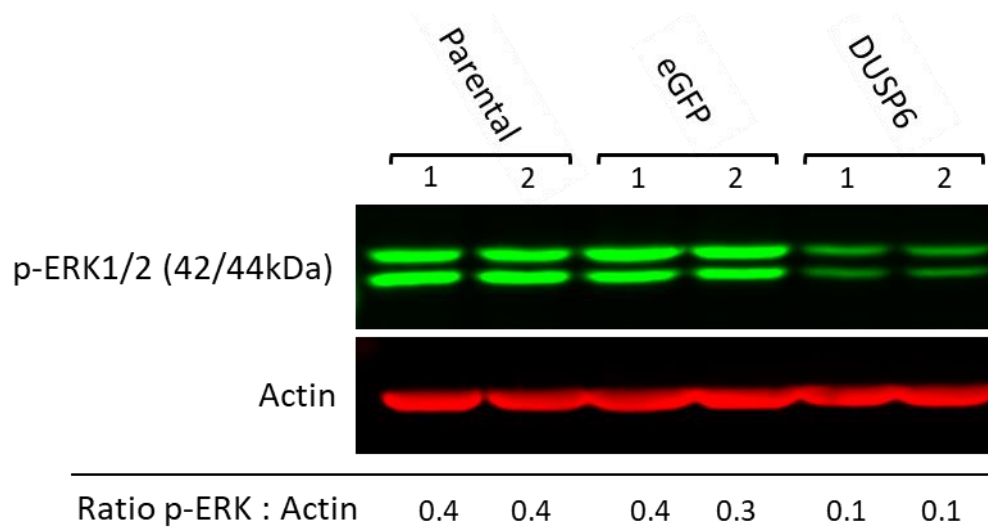


Figure 71. Western blot repeats of p-ERK protein expression in RG/C2/80 cell lines. p-ERK protein expression assessment was repeated in RG/C2/80 cells line (parental, enhanced green fluorescent protein (eGFP) transduced and DUSP6 transduced). A marked decrease in p-ERK protein expression was observed between RG/C2/80 (DUSP6 transduced) compared to parental (75% decrease) and eGFP transduced (72% decrease) cells.

8.2 Cohort 2

8.2.1 Genotyping of cohort by pyrosequencing

Case ID	KRAS codon 12, 13	KRAS codon 59, 61	KRAS codon 117	KRAS codon 146	NRAS codon 12, 13	NRAS codon 59	BRAF*
1	M (G12D)	WT	WT	WT	WT	WT	WT
3	M (G12S)	WT	WT	WT	WT	WT	WT
4	WT	WT	WT	WT	WT	WT	WT
6	WT	WT	WT	^	WT	WT	WT
7	WT	WT	WT	WT	WT	^	WT
10	M (G12C)	WT	^	^	WT	^	WT
11	M (G12V)	WT	WT	WT	WT	WT	WT
12	WT	WT	WT	WT	WT	WT	WT
13	WT	M (Q61R)	WT	WT	WT	WT	WT
14	WT	WT	WT	WT	WT	WT	WT
15	WT	WT	WT	WT	WT	WT	WT
16	WT	WT	WT	WT	WT	WT	WT
17	WT	WT	WT	WT	WT	WT	WT
18	WT	WT	WT	WT	WT	WT	WT
19	M (G12C)	WT	WT	WT	WT	WT	WT
20	WT	WT	WT	WT	WT	WT	WT
22	WT	WT	WT	WT	WT	WT	M (V600E)
23	WT	WT	WT	WT	WT	WT	WT
24	WT	WT	WT	WT	WT	WT	WT
25	WT	WT	WT	WT	WT	WT	M (V600E)
26	WT	M (Q61H)	WT	WT	WT	WT	WT
27	M (G12V)	WT	WT	WT	WT	WT	WT
28	WT	WT	WT	WT	WT	WT	WT
29	WT	WT	WT	WT	WT	WT	M (V600E)
30	WT	WT	WT	WT	WT	WT	M (V600E)
32	WT	WT	WT	WT	WT	WT	WT
33	WT	WT	WT	WT	WT	WT	M (V600E)
34	WT	WT	WT	WT	WT	^	WT
35	M (G12D)	WT	^	WT	WT	WT	WT
36	M (G12A)	WT	WT	WT	WT	WT	WT
37	M (G12D)	WT	WT	WT	WT	WT	WT
38	WT	WT	WT	WT	WT	WT	WT
39	M (G12D)	WT	WT	WT	WT	WT	WT
40	WT	WT	WT	WT	WT	WT	M (V600E)
41	WT	WT	WT	WT	WT	WT	M (V600E)
42	M (G13D)	WT	WT	WT	WT	WT	WT
43	M (G12D)	WT	WT	WT	WT	WT	WT
44	WT	WT	WT	WT	WT	WT	M (V600E)

46	WT	WT	WT	WT	WT	^	WT
47	WT	WT	WT	WT	WT	WT	M (V600E)
48	M (G12A)	WT	WT	WT	WT	WT	WT
49	M (G12D)	WT	WT	WT	WT	WT	WT
50	WT	WT	WT	WT	WT	WT	WT
51	M (G12V)	WT	WT	WT	WT	WT	WT
52	WT	WT	WT	WT	WT	WT	WT
54	WT	WT	WT	WT	WT	WT	M (V600E)
55	WT	WT	WT	WT	WT	WT	WT
56	WT	WT	WT	WT	WT	WT	WT
57	WT	WT	WT	WT	WT	WT	WT
58	WT	WT	WT	WT	WT	WT	WT
59	M (G12V)	WT	WT	WT	WT	WT	WT
60	M (G12V)	WT	WT	WT	WT	WT	WT
61	WT	WT	WT	WT	WT	WT	M (V600E)
62	M (G13D)	WT	WT	WT	WT	WT	WT
63	M (G12C)	WT	WT	WT	WT	WT	WT
65	WT	WT	WT	WT	WT	WT	M (V600E)
66	WT	WT	WT	WT	WT	WT	WT
67	WT	WT	WT	WT	WT	WT	WT
69	WT	WT	WT	WT	WT	WT	WT
70	WT	WT	WT	WT	WT	M (Q61L)	WT
72	WT	WT	^	^	WT	WT	WT
73	WT	WT	WT	WT	WT	WT	WT
74	WT	M (Q61L)	WT	WT	WT	WT	WT
75	WT	WT	WT	WT	WT	WT	WT
77	M (G13D)	WT	WT	WT	WT	WT	WT
78	WT	WT	WT	WT	WT	WT	WT
79	WT	WT	WT	WT	WT	WT	WT
80	WT	WT	WT	M (A146V)	WT	WT	WT
81	M (G12D)	WT	WT	WT	WT	WT	WT
82	WT	WT	WT	WT	WT	WT	WT
83	WT	WT	WT	WT	WT	WT	WT
84	WT	M (Q61L)	WT	WT	WT	WT	WT
85	WT	WT	WT	WT	WT	WT	WT
86	WT	WT	WT	WT	WT	WT	WT
87	WT	WT	WT	WT	WT	WT	M (V600E)
88	WT	WT	WT	WT	WT	WT	WT
91	M (G12D)	WT	WT	WT	WT	WT	WT
92	WT	WT	WT	WT	WT	WT	M (V600E)
93	WT	WT	WT	WT	WT	WT	WT
94	WT	WT	WT	WT	WT	WT	WT
95	M (G12C)	WT	WT	WT	WT	WT	WT
96	WT	WT	WT	WT	WT	WT	WT
97	WT	WT	WT	WT	WT	WT	WT

98	WT	WT	WT	WT	WT	M (Q61L)	WT
99	M (G12D)	WT	WT	WT	WT	WT	WT
100	WT	WT	WT	WT	WT	WT	WT
101	M (G12D)	WT	^	WT	WT	WT	WT
102	WT	WT	WT	WT	WT	WT	M (V600E)
103	WT	WT	WT	WT	WT	WT	WT
104	WT	WT	WT	WT	WT	WT	WT
105	WT	WT	WT	WT	WT	WT	WT
106	WT	WT	^	^	WT	^	WT
107	M (G12A)	WT	WT	WT	WT	WT	WT
108	WT	WT	WT	WT	WT	WT	WT
109	M (G12D)	WT	WT	WT	WT	WT	WT
110	WT	WT	WT	WT	WT	WT	WT
111	WT	WT	WT	WT	WT	M (Q61R)	WT
112	WT	WT	WT	WT	WT	WT	WT
113	WT	WT	WT	WT	WT	WT	WT
114	WT	WT	WT	WT	WT	WT	WT
115	WT	WT	WT	WT	WT	WT	WT
116	WT	WT	^	^	WT	WT	WT
117	WT	WT	WT	WT	WT	WT	WT
119	WT	WT	WT	WT	WT	WT	WT
120	WT	WT	WT	WT	WT	WT	M (V600E)
121	WT	WT	WT	WT	WT	WT	WT
122	WT	WT	WT	WT	WT	WT	WT
123	WT	WT	WT	WT	WT	WT	WT
124	WT	WT	WT	^	WT	WT	WT
125	WT	WT	WT	WT	WT	WT	WT
126	M (G12S)	WT	WT	WT	WT	WT	WT
127	M (G12D)	WT	WT	WT	WT	WT	WT
128	WT	WT	WT	WT	WT	WT	WT
129	WT	WT	WT	WT	WT	WT	WT
130	WT	WT	WT	WT	WT	WT	WT
131	WT	WT	WT	WT	WT	WT	M (V600E)
132	WT	WT	WT	WT	WT	WT	WT
133	M (G12D)	WT	WT	WT	WT	WT	WT
134	WT	WT	WT	WT	WT	WT	WT
135	WT	WT	WT	WT	WT	WT	WT
136	WT	WT	WT	WT	WT	WT	WT
137	WT	WT	WT	WT	WT	WT	WT
138	M (G12V)	WT	WT	WT	WT	WT	WT
140	M (G12C)	WT	WT	WT	WT	WT	WT
141	WT	WT	WT	WT	WT	WT	M (V600E)
142	WT	WT	WT	WT	WT	WT	M (V600E)
143	WT	WT	WT	WT	WT	WT	WT
144	M (G12A)	WT	^	^	WT	WT	WT

145	M (G12S)	WT	WT	WT	WT	WT	WT
146	WT	WT	WT	WT	WT	WT	WT
147	M (G13D)	WT	WT	WT	WT	WT	WT
148	M (G12D)	WT	WT	WT	WT	WT	WT
149	M (G12D)	WT	WT	WT	WT	WT	WT
150	M (G12D)	WT	WT	WT	WT	WT	WT
152	WT	WT	WT	WT	WT	WT	WT
154	WT	WT	WT	WT	WT	WT	WT
155	WT	WT	WT	WT	WT	WT	M (V600E)
156	M (G12A)	WT	WT	WT	WT	WT	WT
157	WT	WT	WT	WT	WT	WT	WT
158	WT	WT	^	^	WT	WT	WT
160	WT	WT	WT	WT	WT	WT	WT
161	WT	WT	WT	WT	WT	WT	M (V600E)
162	WT	WT	WT	WT	WT	WT	WT
163	WT	WT	WT	WT	WT	WT	WT
164	WT	WT	WT	WT	WT	WT	M (V600E)
165	WT	WT	WT	WT	WT	WT	M (V600E)
166	WT	WT	WT	WT	WT	WT	WT
168	WT	WT	WT	WT	WT	WT	M (V600E)
169	WT	WT	WT	WT	WT	WT	WT
170	WT	WT	WT	WT	WT	WT	WT
171	M (G12V)	WT	WT	WT	WT	WT	WT
172	WT	WT	WT	WT	WT	WT	M (V600E)
173	M (G12V)	WT	WT	WT	WT	WT	WT
174	WT	WT	WT	WT	WT	WT	WT
175	WT	WT	WT	WT	WT	WT	WT
176	WT	WT	WT	WT	WT	WT	M (V600E)
177	WT	WT	WT	WT	WT	WT	WT
178	M (G12D)	WT	WT	WT	WT	WT	WT
179	WT	WT	WT	WT	WT	WT	M (V600E)
180	WT	WT	WT	WT	WT	WT	M (V600E)
181	M (G12V)	WT	WT	WT	WT	WT	WT
182	WT	WT	WT	WT	WT	WT	WT
184	M (G12V)	WT	WT	WT	WT	WT	WT
186	WT	WT	WT	WT	WT	WT	WT
187	M (G12C)	WT	WT	WT	WT	WT	WT
188	M (G12C)	WT	WT	WT	WT	WT	WT
189	WT	WT	WT	WT	WT	WT	WT
190	WT	WT	WT	WT	WT	WT	WT
191	M (G12A)	WT	WT	WT	WT	WT	WT
192	WT	WT	WT	WT	WT	WT	WT
194	WT	WT	WT	WT	WT	WT	WT
195	WT	WT	WT	WT	WT	WT	WT
196	WT	WT	WT	WT	WT	WT	WT

197	WT	WT	WT	WT	WT	WT	WT
198	WT	WT	WT	WT	WT	WT	WT
200	WT	WT	^	^	WT	^	WT
201	WT	M (G60V)	WT	WT	WT	WT	WT
202	WT	WT	WT	WT	WT	WT	WT
203	WT	WT	^	WT	WT	WT	WT
205	WT	WT	WT	WT	WT	WT	WT
206	WT	WT	WT	WT	WT	WT	WT
208	M (G12D)	WT	WT	WT	WT	WT	WT
209	WT	WT	WT	WT	WT	WT	WT
210	M (G12C)	WT	WT	WT	WT	WT	WT
211	M (G13D)	WT	WT	WT	WT	WT	WT
212	WT	WT	WT	WT	WT	WT	WT
213	WT	WT	WT	WT	M (G12D)	WT	WT
214	WT	WT	WT	WT	WT	WT	WT
216	WT	WT	WT	WT	WT	WT	M (V600E)
217	WT	WT	WT	WT	WT	WT	M (V600E)
220	M (G12S)	WT	WT	WT	WT	WT	WT
221	WT	WT	WT	WT	WT	WT	WT
223	WT	WT	WT	WT	WT	WT	M (V600E)
224	M (G12D)	WT	WT	WT	WT	WT	M (V600E)
225	WT	WT	WT	WT	WT	WT	WT
226	M (G12A)	WT	WT	WT	WT	WT	WT
227	M (G12V)	WT	WT	WT	WT	WT	WT
229	WT	WT	WT	WT	WT	WT	WT
230	WT	M (Q61K)	WT	WT	WT	WT	WT
231	WT	WT	WT	WT	WT	WT	M (V600E)
232	WT	WT	WT	WT	WT	WT	WT
233	M (G13D)	WT	WT	WT	WT	WT	WT
234	WT	WT	WT	WT	WT	WT	WT
235	WT	WT	WT	WT	WT	WT	M (V600E)
236	M (G13D)	WT	WT	WT	WT	WT	WT
237	WT	WT	WT	WT	WT	WT	WT
238	WT	WT	WT	WT	WT	WT	WT
239	M (G12D)	WT	^	WT	WT	WT	WT
240	M (G13D)	WT	WT	WT	WT	WT	WT
241	M (G12V)	WT	WT	WT	WT	WT	WT
242	WT	WT	WT	WT	WT	WT	WT
243	WT	WT	WT	WT	WT	WT	M (V600E)
244	WT	WT	WT	WT	WT	WT	WT
245	WT	WT	WT	WT	WT	WT	WT
246	WT	WT	WT	WT	WT	WT	WT
247	WT	WT	WT	WT	WT	M (Q61K)	WT
248	WT	WT	WT	WT	WT	WT	WT
249	WT	WT	WT	WT	WT	WT	WT

250	WT	WT	WT	WT	WT	WT	WT
252	WT	M (Q61R)	WT	WT	WT	WT	WT
253	M (G12C)	WT	WT	WT	WT	WT	WT
254	WT	WT	WT	WT	WT	WT	M (V600E)
255	M (G12C)	WT	WT	WT	WT	WT	WT
256	WT	WT	WT	WT	M (G12D)	WT	WT
258	WT	WT	WT	WT	WT	WT	WT
259	WT	WT	WT	WT	WT	WT	M (V600E)
260	M (G12D)	WT	WT	WT	WT	WT	WT
261	WT	WT	WT	WT	WT	WT	WT
262	M (G12V)	WT	WT	WT	WT	WT	WT
263	WT	WT	WT	WT	WT	WT	WT
264	WT	WT	WT	WT	WT	WT	WT
265	FAIL	FAIL	WT	WT	FAIL	FAIL	M (V600E)
266	WT	WT	WT	WT	WT	WT	WT
267	WT	WT	WT	WT	WT	WT	WT
268	WT	WT	WT	WT	WT	WT	M (V600E)
269	M (G12V)	WT	WT	WT	WT	WT	WT
270	WT	WT	WT	WT	WT	WT	M (V600E)
271	M (G12V)	WT	WT	WT	WT	WT	WT
272	M (G12D)	WT	WT	WT	WT	WT	WT
273	M (G13D)	WT	WT	WT	WT	WT	WT
274	WT	WT	WT	WT	WT	WT	WT
275	M (G13D)	WT	WT	WT	WT	WT	WT
276	WT	WT	WT	WT	WT	WT	WT
277	M (G12D)	WT	WT	WT	WT	WT	WT
278	WT	WT	WT	WT	WT	WT	WT
279	WT	WT	WT	WT	WT	WT	M (V600E)
280	M (G12V)	WT	WT	WT	WT	WT	WT
281	WT	WT	WT	WT	WT	WT	WT
282	WT	WT	WT	WT	WT	WT	WT
284	WT	WT	WT	WT	WT	WT	WT
285	M (G12V)	WT	WT	WT	M (G12V)	WT	WT
286	WT	WT	WT	WT	WT	WT	WT
287	M (G12V)	WT	WT	WT	WT	WT	WT
288	WT	WT	WT	WT	WT	WT	WT
289	WT	WT	WT	WT	WT	WT	WT
290	WT	WT	^	^	WT	WT	WT
291	M (G12D)	WT	WT	WT	WT	WT	WT
293	M (G12D)	WT	^	^	WT	^	WT
294	M (G12V)	WT	WT	WT	WT	WT	WT
296	M (G12V)	WT	WT	WT	WT	WT	WT
297	WT	WT	WT	WT	WT	WT	WT
298	M (G12S)	WT	^	^	WT	WT	WT
299	WT	WT	WT	WT	WT	WT	WT

300	WT	WT	WT	WT	WT	WT	WT
301	WT	WT	WT	WT	WT	WT	WT
302	WT	WT	WT	WT	WT	WT	WT
303	WT	WT	WT	WT	WT	WT	WT
305	M (G12C)	WT	WT	WT	WT	WT	WT
306	M (G12V)	WT	WT	WT	WT	WT	WT
308	WT	WT	WT	WT	WT	WT	WT
309	M (G12D)	WT	WT	WT	WT	WT	WT
311	WT	WT	WT	WT	WT	WT	WT
312	M (G12D)	WT	^	^	WT	^	WT
314	WT	WT	WT	WT	WT	WT	WT
315	WT	WT	WT	WT	WT	WT	M (V600E)
316	WT	WT	WT	WT	WT	WT	WT
317	WT	M (Q61K)	WT	WT	WT	WT	WT
318	M (G13D)	WT	WT	WT	WT	WT	WT
319	WT	WT	^	^	WT	WT	WT
321	WT	WT	WT	WT	WT	WT	WT
322	WT	WT	WT	WT	WT	WT	WT
323	WT	WT	WT	WT	WT	WT	WT
324	WT	WT	WT	WT	WT	WT	M (V600E)
325	WT	WT	WT	WT	WT	WT	WT
326	M (G12D)	WT	WT	WT	WT	WT	WT
327	WT	WT	^	WT	WT	WT	FAIL
329	WT	WT	WT	WT	WT	WT	WT
330	WT	WT	WT	WT	WT	WT	M (V600E)
331	M (G13D)	WT	WT	WT	WT	WT	WT
332	WT	WT	WT	WT	WT	WT	M (V600E)
333	M (G12V)	WT	^	WT	WT	WT	WT
334	M (G12D)	WT	^	^	WT	WT	M (V600E)
335	M (G12S)	WT	WT	WT	WT	WT	WT
336	M (G12D)	WT	WT	WT	WT	WT	WT
337	WT	WT	WT	WT	WT	WT	WT
338	WT	M (Q61H)	WT	WT	WT	WT	WT
339	M (G12V)	WT	WT	WT	WT	WT	WT
340	M (G13D)	WT	WT	WT	WT	WT	WT
341	WT	WT	WT	WT	WT	WT	WT
342	M (G13D)	WT	WT	WT	WT	WT	WT
343	WT	WT	WT	WT	WT	WT	WT
344	WT	WT	WT	WT	WT	WT	WT
345	WT	WT	WT	WT	WT	WT	WT
346	M (G12S)	WT	WT	WT	WT	WT	WT
347	WT	WT	WT	WT	WT	WT	WT
348	M (G12D)	WT	WT	WT	WT	WT	WT
349	WT	WT	WT	WT	WT	WT	M (V600E)
350	M (G12A)	WT	WT	WT	WT	WT	WT

351	WT	WT	WT	WT	WT	WT	WT
352	WT	WT	WT	WT	WT	WT	WT
353	WT	WT	WT	WT	WT	WT	WT
354	WT	WT	WT	WT	WT	WT	M (V600E)
355	WT	WT	WT	WT	WT	WT	WT
356	WT	WT	WT	WT	WT	WT	WT
357	WT	WT	WT	WT	WT	WT	WT
358	WT	M (Q61L)	WT	WT	WT	WT	WT
359	WT	WT	WT	WT	WT	WT	WT
360	WT	WT	WT	WT	WT	WT	WT
361	WT	WT	^	WT	WT	WT	WT
362	WT	M (Q61H)	WT	WT	M (G12A)	WT	WT
363	WT	WT	WT	WT	WT	WT	M (V600E)
364	WT	WT	WT	WT	WT	WT	WT
366	WT	WT	WT	WT	WT	WT	WT
367	WT	WT	WT	^	WT	WT	WT
368	WT	WT	WT	WT	WT	WT	M (V600E)
369	WT	WT	WT	WT	WT	WT	WT
371	WT	WT	WT	WT	WT	WT	M (V600E)
372	M (G13D)	WT	WT	WT	WT	WT	WT
373	M (G12D)	WT	WT	WT	WT	WT	WT
374	WT	WT	WT	WT	WT	WT	WT
375	WT	WT	WT	M (A146T)	WT	WT	WT
376	WT	WT	WT	WT	WT	WT	M
377	WT	WT	WT	WT	WT	WT	WT
378	WT	WT	WT	M (A146S)	WT	WT	WT
379	WT	WT	WT	WT	WT	WT	WT
380	WT	WT	WT	WT	WT	WT	WT
381	WT	WT	M (K117N)	WT	WT	WT	WT
382	M (G13D)	WT	WT	WT	WT	WT	WT
383	M (G13D)	WT	WT	WT	WT	WT	WT
384	WT	WT	WT	WT	WT	WT	WT
385	WT	WT	WT	WT	WT	WT	WT
386	WT	WT	WT	WT	WT	WT	WT
387	WT	WT	WT	^	WT	WT	WT
388	WT	WT	WT	WT	WT	WT	M (V600E)
389	WT	WT	WT	WT	WT	WT	WT
390	WT	WT	WT	WT	WT	WT	WT
391	M (G12S)	WT	WT	WT	WT	WT	WT
392	WT	WT	^	WT	WT	WT	WT
393	M (G12A)	WT	WT	WT	WT	WT	WT
394	WT	WT	WT	WT	WT	WT	WT
395	M (G12V)	WT	WT	WT	WT	WT	WT
396	WT	WT	WT	WT	WT	WT	WT
397	WT	WT	WT	WT	WT	WT	WT

398	WT	WT	WT	WT	WT	WT	WT
399	WT	M (Q61L)	WT	WT	WT	WT	WT
401	WT	WT	WT	WT	WT	WT	M (V600E)
402	M (G12D)	WT	WT	WT	WT	WT	WT
403	WT	WT	WT	WT	WT	WT	WT
404	WT	WT	WT	WT	WT	M (Q61K)	WT
405	WT	WT	WT	WT	WT	WT	M (V600E)
406	M (G12D)	WT	WT	WT	WT	WT	WT
407	WT	WT	WT	WT	WT	WT	WT
408	M (G13D)	WT	WT	WT	WT	WT	WT
409	WT	WT	WT	WT	WT	WT	WT
410	WT	WT	WT	WT	WT	WT	WT
412	WT	WT	WT	WT	WT	WT	WT
413	M (G12C)	WT	WT	WT	WT	WT	WT
414	M (G12D)	WT	WT	WT	WT	WT	WT
415	WT	WT	WT	WT	WT	WT	WT
416	M (G12C)	WT	WT	WT	WT	WT	WT
417	WT	WT	WT	WT	WT	WT	WT
418	WT	WT	WT	WT	WT	WT	WT
419	M (G12V)	WT	WT	WT	WT	WT	WT
421	WT	WT	WT	^	WT	WT	WT
422	WT	WT	WT	WT	WT	WT	WT
423	M (G12V)	WT	WT	WT	WT	WT	WT
424	M (G12D)	WT	WT	WT	WT	WT	WT
425	M (G12D)	WT	WT	WT	WT	WT	WT
426	M (G12D)	WT	WT	WT	WT	WT	WT
428	WT	WT	^	WT	WT	WT	WT
429	M (G12D)	WT	WT	WT	WT	WT	WT
430	WT	WT	WT	WT	WT	WT	WT
431	WT	WT	WT	WT	WT	WT	M (V600E)
433	WT	WT	WT	WT	WT	M (Q61R)	WT
434	WT	WT	WT	WT	WT	WT	M (V600E)
435	WT	WT	WT	WT	WT	WT	M (V600E)
436	M (G12D)	WT	WT	WT	WT	WT	WT
437	WT	WT	WT	WT	WT	WT	WT
438	WT	WT	WT	WT	WT	WT	WT
439	WT	WT	WT	WT	WT	WT	WT
440	M (G12D)	WT	WT	WT	WT	WT	WT
441	WT	WT	WT	WT	WT	WT	WT
442	WT	WT	^	^	WT	WT	WT
443	M (G12V)	WT	WT	WT	WT	WT	WT
444	WT	WT	WT	WT	WT	WT	WT
445	M (G12D)	WT	WT	WT	WT	WT	WT
446	WT	WT	WT	WT	WT	M (Q61L)	WT
447	WT	WT	WT	WT	WT	WT	WT

448	WT	WT	WT	WT	WT	WT	M (V600E)
449	WT	WT	WT	WT		WT	WT
450	WT	WT	WT	WT	WT	WT	WT
451	M (G12V)	WT	WT	WT	WT	WT	WT
452	WT	WT	WT	WT	WT	WT	WT
453	WT	WT	WT	WT	M (G12D)	WT	WT
454	WT	WT	WT	WT	WT	WT	WT
455	M (G12S)	WT	WT	WT	WT	WT	WT
456	WT	WT	WT	WT	WT	WT	WT
457	WT	WT	WT	WT	WT	WT	WT
458	M (G12V)	WT	WT	WT	WT	WT	WT
459	M (G12D)	WT	^	^	WT	^	WT
460	WT	WT	WT	WT	WT	WT	M (V600E)
461	M (G12C)	WT	WT	WT	WT	WT	WT
462	WT	WT	WT	WT	WT	WT	WT
463	WT	WT	^	^	WT	WT	WT
465	M (G12D)	WT	WT	WT	WT	WT	WT
466	WT	WT	WT	WT	WT	WT	WT
468	WT	WT	WT	WT	WT	WT	WT
469	WT	WT	WT	WT	WT	WT	M (V600E)
470	WT	WT	WT	WT	WT	WT	WT
471	M (G12V)	WT	^	^	WT	WT	WT
472	WT	WT	WT	WT	WT	WT	WT
473	M (G13D)	WT	WT	WT	WT	WT	WT
474	WT	WT	WT	WT	WT	WT	WT
475	WT	WT	^	^	WT	^	M (V600E)
476	WT	WT	WT	WT	WT	WT	WT
477	WT	WT	WT	WT	WT	WT	WT
478	WT	WT	WT	WT	WT	WT	WT
479	WT	WT	WT	WT	WT	WT	WT
480	WT	WT	WT	WT	WT	WT	WT
481	WT	WT	WT	WT	WT	WT	WT
482	WT	WT	WT	WT	WT	WT	WT
483	M (G12S)	WT	WT	WT	WT	WT	WT
484	WT	WT	WT	WT	M (G12D)	WT	WT
486	WT	WT	WT	WT	WT	WT	WT
487	M (G12D)	WT	WT	WT	WT	WT	WT
488	M (G12V)	WT	WT	WT	WT	WT	WT
489	WT	WT	WT	WT	WT	WT	WT
490	WT	WT	WT	WT	WT	WT	WT
491	M (G12V)	WT	WT	WT	WT	WT	WT
492	WT	WT	WT	WT	WT	WT	WT
493	WT	WT	WT	WT	WT	WT	M (V600E)
494	WT	WT	WT	WT	WT	WT	WT
495	WT	WT	WT	WT	WT	WT	WT

496	WT	WT	WT	WT	WT	WT	WT
498	M (G12D)	WT	WT	WT	WT	WT	WT
499	WT	WT	WT	WT	WT	WT	WT
500	WT	WT	WT	WT	WT	WT	WT
502	WT	WT	WT	WT	WT	WT	WT
503	WT	WT	WT	WT	WT	M (Q61R)	WT
504	WT	WT	WT	WT	WT	WT	WT
505	WT	WT	WT	M (A146V)	WT	WT	WT
506	WT	WT	WT	WT	WT	WT	WT
507	WT	WT	WT	WT	WT	WT	WT
508	WT	M (Q61R)	WT	WT	WT	WT	WT
509	WT	WT	WT	WT	WT	WT	WT
510	M (G12V)	WT	WT	WT	WT	WT	WT
511	M (G12D)	WT	WT	WT	WT	WT	WT
512	WT	WT	WT	WT	WT	WT	WT
513	WT	WT	WT	WT	WT	WT	M (V600E)
514	WT	WT	WT	WT	WT	WT	WT
515	WT	WT	WT	WT	WT	WT	WT
516	WT	WT	WT	WT	M (G12D)	WT	WT
517	WT	WT	WT	WT	WT	WT	WT
518	M (G12D)	WT	WT	WT	WT	WT	WT
519	WT	WT	WT	WT	WT	WT	WT
520	WT	WT	WT	WT	WT	WT	WT
521	WT	WT	WT	WT	WT	WT	WT
522	WT	WT	WT	WT	WT	WT	WT
523	M (G13D)	WT	WT	WT	WT	WT	WT
524	WT	M (Q61H)	WT	WT	WT	WT	WT
526	M (G12D)	WT	WT	WT	WT	WT	WT
527	WT	WT	^	WT	WT	WT	WT
528	M (G12C)	WT	WT	WT	WT	WT	WT
529	M (G12V)	WT	WT	WT	WT	WT	WT
531	WT	WT	WT	WT	WT	WT	WT
532	WT	WT	WT	WT	WT	WT	WT
533	M (G12D)	WT	WT	WT	WT	WT	WT
534	WT	WT	WT	WT	WT	WT	WT
535	WT	WT	^	^	WT	^	WT
536	WT	WT	WT	WT	WT	WT	WT
537	M (G12D)	WT	WT	WT	WT	WT	WT
538	M (G12V))	WT	WT	WT	WT	WT	WT
539	WT	WT	WT	WT	WT	WT	WT
540	WT	WT	WT	WT	WT	WT	WT
541	WT	WT	WT	WT	WT	WT	WT
542	M (G12D)	WT	WT	WT	WT	WT	WT
543	M (G12D)	WT	WT	WT	WT	WT	WT
544	WT	WT	WT	WT	WT	M (Q61L)	WT

546	WT	WT	WT	WT	WT	WT	M (V600E)
548	WT	WT	WT	WT	WT	WT	WT
549	WT	WT	WT	WT	WT	WT	WT
550	WT	WT	WT	WT	WT	WT	WT
551	WT	WT	^	WT	WT	WT	WT
552	M (G12D)	WT	WT	WT	WT	WT	WT
553	WT	WT	WT	WT	WT	WT	WT
554	WT	WT	WT	WT	WT	WT	WT
555	M (G12D)	WT	^	WT	WT	WT	WT
556	WT	WT	WT	WT	WT	WT	M (V600E)
557	WT	WT	WT	WT	WT	WT	WT
558	M (G12V)	WT	WT	WT	WT	WT	WT
560	M (G12V)	WT	WT	WT	WT	WT	WT
561	WT	WT	WT	WT	WT	WT	WT
563	WT	WT	WT	WT	WT	WT	WT
564	M (G12D)	WT	WT	WT	WT	WT	WT
565	WT	WT	WT	WT	WT	WT	WT
566	WT	WT	WT	WT	WT	WT	WT
567	WT	WT	WT	WT	WT	WT	WT
568	M (G12S)	WT	WT	WT	WT	WT	WT
569	WT	WT	WT	WT	WT	WT	WT
570	M (G13D)	WT	WT	WT	WT	WT	WT
571	M (G12D)	WT	WT	WT	WT	WT	WT
572	M (G12C)	WT	WT	WT	WT	WT	WT
573	WT	M (Q61L)	WT	WT	WT	WT	WT
574	WT	WT	WT	WT	WT	M (Q61R)	WT
575	WT	WT	WT	WT	WT	WT	WT
576	M (G12S)	WT	WT	WT	WT	WT	WT
577	WT	WT	WT	WT	WT	WT	M (V600E)
578	M (G12C)	WT	WT	^	WT	WT	WT
579	M (G12V)	WT	WT	WT	WT	WT	WT
580	WT	WT	^	^	M (G12A)	WT	WT
581	WT	WT	WT	WT	WT	WT	WT
582	WT	WT	WT	WT	WT	WT	WT
583	WT	WT	WT	WT	WT	WT	WT
584	WT	WT	WT	WT	WT	WT	WT
585	M (G12V)	WT	WT	WT	WT	WT	WT

Table 34. Genotyping of adenocarcinoma cohort by pyrosequencing. ^Failed sequencing. *Only *BRAF* V600E mutation was assessed. Results in black had been previously reported by molecular pathology at the Royal Infirmary of Edinburgh. Results in red had been detected during pyrosequencing of the cohort during this study.

8.2.2 DUSP6 protein expression: Survival and DUSP6 AQUA data

Case ID	Survival data			Average DUSP6 AQUA value	
	Survival (Months)	5yr. OS*	5yr. DSS*	Nuclear	Cytoplasmic
1	23.0	1		593.7	686.8
3	21.0	1		519.8	575.6
4	70.0	0	0	338.2	389.8
6	67.0	0	0	230.0	290.9
7	61.0	1		217.0	248.1
10	72.0	0	0	324.8	388.4
11	11.0	1	1	502.0	438.3
12	2.0	1		255.5	275.4
13	78.0	0	0	246.7	182.4
14	70.0	0	0		
15	12.0	1		524.0	522.3
16	69.0	0	0	381.0	420.8
17	60.0	1		301.6	317.0
18	7.0	1		329.7	397.0
19	71.0	0	0	141.9	176.4
20	60.0	0	0	414.8	412.8
22	64.0	0	0	280.7	222.7
23	40.0	1	1	692.7	519.8
24	60.0	0	0	307.9	341.6
25	37.0	1	1		
26	74.0	0	0	395.3	392.8
27	28.0	1	1		
28	53.0	1	1	380.1	394.9
29	72.0	0	0		
30	74.0	0	0	458.2	234.4
32	66.0	0	0	337.5	319.9
33	72.0	0	0	641.5	606.2
34	63.0	0	0	354.1	380.1
35	73.0	0	0	285.3	220.7
36	64.0	0	0	325.3	295.2
37	9.0	1	1	564.3	543.9
38	64.0	0	0	489.9	633.0
39	67.0	0	0	227.8	175.3
40	66.0	0	0		
41	23.0	1		534.0	238.2
42	71.0	0	0	175.7	213.6
43	69.0	0	0	401.3	284.2
44	16.0	1		564.6	559.6
46	64.0	0	0	268.8	254.9

47	79.0	0	0	349.0	359.6
48	64.0	0	0	200.7	206.4
49	78.0	0	0		
50	31.0	1	1	277.0	275.4
51	64.0	0	0	374.4	393.5
52	73.0	0	0	229.6	265.5
54	66.0	0	0	344.8	373.1
55	53.0	1		383.7	351.7
56	70.0	0	0	916.2	907.3
57	66.0	0	0	267.1	216.6
58	17.0	1	1	611.7	530.6
59	73.0	0	0		
60	74.0	0	0	290.7	266.9
61	54.0	1	1	469.3	330.4
62	63.0	0	0	648.3	510.1
63	0.0	1		276.6	215.0
65	54.0	1	1	619.8	497.4
66	11.0	1	1	589.7	575.4
67	77.0	0	0	395.3	331.0
69	61.0	0	0	270.6	406.1
70	60.0	0	0	273.3	224.8
72	70.0	0	0	164.6	154.6
73	60.0	0	0	662.7	451.7
74	67.0	0	0	541.2	564.1
75	70.0	0	0	222.3	222.4
77	66.0	0	0	377.4	371.4
78	66.0	0	0	432.6	348.8
79	75.0	0	0	199.2	169.9
80	61.0	1		325.5	325.5
81	69.0	0	0	329.7	265.0
82	6.0	1	1	146.4	174.4
83	10.0	1	1	597.3	521.2
84	7.0	1	1	189.7	230.7
85	58.0			265.0	253.5
86	75.0	0	0	319.7	351.3
87	72.0	0	0	369.8	382.3
88	74.0	0	0	298.4	319.4
91	11.0	1	1	659.3	581.3
92	76.0	0	0	422.8	392.5
93	74.0	0	0	262.6	148.2
94	73.0	0	0	622.7	547.4
95	65.0	0	0	200.0	235.5
96	79.0	0	0	533.4	493.1
97	65.0	0	0	267.0	255.0
98	39.0	1	1	287.4	242.4

99	36.0	1		165.9	163.4
100	5.0	1	1	424.6	380.6
101	7.0	1	1	359.6	375.4
102	20.0	1	1	279.7	332.0
103	64.0	0	0	538.4	515.3
104	72.0	0	0	299.3	311.6
105	62.0	0	0	739.3	747.5
106	61.0	0	0	490.5	644.2
107	68.0	0	0	634.6	703.5
108	75.0	1		477.2	526.3
109	5.0	1	1	280.8	217.9
110	18.0	1		322.0	248.1
111	3.0	1	1	705.6	731.4
112	70.0	0	0	254.1	209.1
113	60.0	0	0	356.5	348.9
114	47.0	1	1	430.8	391.8
115	71.0	0	0	171.2	175.2
116	17.0	1	1	564.3	494.3
117	65.0	0	0	492.8	454.9
119	57.0			972.7	991.1
120	69.0	0	0	369.7	336.0
121	69.0	0	0		
122	63.0	0	0	179.9	185.4
123	1.0	1	1	154.6	128.9
124	73.0	0	0	234.5	291.4
125	16.0	1	1	179.7	190.1
126	37.0	1	1	407.9	424.8
127	72.0	0	0	492.3	475.2
128	69.0	0	0	371.3	326.9
129	79.0	0	0	451.2	480.0
130	77.0	0	0	216.8	220.2
131	12.0	1	1	201.2	195.6
132	79.0	0	0	272.7	233.8
133	7.0	1		251.5	257.3
134	25.0	1		191.1	256.4
135	65.0	0	0	586.7	439.8
136	69.0	0	0	533.5	623.3
137	77.0	0	0		
138	71.0	0	0	260.4	309.3
140	61.0	0	0	655.6	682.2
141	78.0	0	0	236.9	264.4
142	61.0	0	0	252.4	259.0
143	62.0	0	0	292.0	296.2
144	79.0	0	0	284.0	295.8
145	72.0	0	0	438.0	392.9

146	69.0	0	0	500.0	414.0
147	74.0	0	0	275.8	227.6
148	61.0	0	0	565.2	470.8
149	69.0	0	0	477.4	414.3
150	74.0	0	0	194.9	165.7
152	36.0	1		240.9	268.1
154	62.0	0	0	255.7	277.3
155	69.0	0	0	759.5	605.5
156	38.0	1	1	225.3	237.0
157	76.0	0	0	200.4	192.8
158	26.0	1	1	433.5	494.1
160	77.0	0	0	190.5	169.3
161	74.0	0	0	553.0	596.0
162	66.0	0	0	247.8	254.8
163	68.0	0	0	252.4	243.2
164	1.0	1	1	324.7	366.0
165	64.0	0	0	347.5	287.3
166	48.0	1	1	299.3	253.3
168	40.0	1		421.3	355.5
169	67.0	0	0	337.0	369.4
170	68.0	0	0	615.3	526.5
171	65.0	0	0	494.2	428.9
172	64.0	0	0	484.7	417.4
173	48.0	1	1	452.9	384.0
174	47.0	1	1	343.1	404.4
175	79.0	0	0	252.7	252.6
176	56.0	1		294.9	340.2
177	61.0	0	0	380.6	402.5
178	67.0	0	0	560.2	503.9
179	42.0	1	1	560.2	503.9
180	60.0	0	0	792.5	621.0
181	74.0	0	0	348.9	309.1
182	59.0	0	0	294.7	285.7
184	60.0	0	0	301.6	281.4
186	66.0	0	0	206.3	261.9
187	63.0	0	0	557.3	530.5
188	9.0	1	1	339.2	307.9
189	75.0	0	0	483.4	351.5
190	70.0	0	0	506.8	436.7
191	64.0	0	0	230.3	253.3
192	77.0	0	0	255.2	318.9
194	72.0	0	0		
195	39.0	1	1	375.3	438.0
196	70.0	0	0	583.0	500.0
197	71.0	0	0	484.6	404.2

198	66.0	0	0	133.3	104.0
200	72.0	0	0	644.6	599.8
201	73.0	0	0	758.6	696.7
202	58.0			490.5	378.9
203	53.0	1		189.1	193.4
205	6.0	1	1	85.0	69.0
206	63.0	0	0	148.0	72.2
208	8.0	1		477.2	318.4
209	65.0	0	0	426.1	489.4
210	21.0	1	1	250.4	139.4
211	26.0	1	1	308.4	250.0
212	10.0	1	1	380.0	443.4
213	65.0	0	0	414.8	438.9
214	37.0	1	1	154.8	123.4
216	6.0	1	1	1139.5	528.3
217	23.0	1	1	434.8	364.0
220	33.0	1	1	231.4	187.6
221	59.0			456.2	442.8
223	70.0	0	0	369.0	318.9
224	21.0	1	1	369.7	308.9
225	71.0	0	0		
226	65.0	0	0		
227	70.0	0	0	671.0	543.3
229	40.0	1	1		
230	77.0	0	0	211.9	222.3
231	6.0	1	1	621.9	368.0
232	72.0	0	0	331.1	296.3
233	23.0	1	1	238.9	146.3
234	68.0	0	0	224.6	220.7
235	73.0	0	0	375.6	346.3
236	76.0	0	0	389.8	394.1
237	73.0	0	0	438.7	527.1
238	16.0	1		290.0	342.9
239	10.0	1	1	121.8	56.4
240	77.0	0	0	324.7	318.8
241	60.0	0	0	451.9	256.2
242	67.0	0	0	276.5	284.5
243	69.0	0	0	396.1	310.4
244	70.0	0	0	532.4	470.3
245	16.0	1		721.7	677.7
246	76.0	0	0		
247	44.0	1		214.2	181.1
248	77.0	0	0	450.9	546.6
249	71.0	0	0	271.9	296.8
250	34.0	1		196.2	203.4

252	67.0	0	0	296.7	253.3
253	69.0	0	0	662.4	458.3
254	56.0	1		214.1	222.0
255	65.0	0	0	358.2	334.3
256	70.0	0	0	809.1	781.0
258	77.0	0	0		
259	62.0	0	0	670.8	594.3
260	10.0	1		183.7	156.1
261	57.0	1		197.0	196.3
262	21.0	1	1	685.7	561.3
263	77.0	0	0	103.7	90.9
264	64.0	0	0	239.7	203.8
265	7.0	1	1	378.1	229.8
266	75.0	0	0		
267	41.0	1	1	252.0	233.2
268	11.0	1	1	285.9	214.4
269	72.0	0	0	531.5	574.8
270	7.0	1	1	223.1	205.1
271	32.0	1	1	348.1	251.4
272	64.0	0	0	198.9	166.6
273	70.0	0	0	494.5	292.1
274	57.0			161.0	98.8
275	61.0	0	0	366.3	259.5
276	9.0	1	1	925.0	561.5
277	79.0	0	0	426.5	402.7
278	67.0	0	0	354.7	321.4
279	60.0	0	0	351.5	228.1
280	55.0	1		786.9	642.8
281	69.0	0	0	191.0	114.7
282	65.0	0	0		
284	60.0	0	0	473.8	311.2
285	74.0	0	0	224.8	246.2
286	77.0	0	0	271.9	324.3
287	6.0	1		195.2	110.0
288	80.0	0	0	220.5	194.8
289	63.0	0	0	205.9	234.6
290	16.0	1	1	238.0	217.7
291	10.0	1	1	294.5	203.2
293	72.0	0	0	885.0	624.8
294	23.0	1		220.1	109.6
296	66.0	0	0	234.2	295.0
297	67.0	0	0	276.0	313.7
298	64.0	0	0	504.3	605.3
299	72.0	0	0	475.1	439.7
300	31.0	1		351.7	267.4

301	18.0	1	1	751.4	698.2
302	34.0	1	1	119.1	92.2
303	10.0	1		684.2	692.0
305	22.0	1	1	708.9	656.4
306	74.0	0	0	376.8	290.3
308	60.0	0	0	195.9	142.8
309	17.0	1		200.1	150.3
311	72.0	0	0	152.2	162.0
312	70.0	0	0	203.2	190.3
314	73.0	0	0	1174.5	1102.0
315	13.0	1	1	244.1	138.3
316	31.0	1			
317	78.0	0	0	290.0	273.1
318	79.0	0	0	213.8	210.0
319	71.0	0	0	683.4	627.5
321	46.0	1	1	171.3	137.5
322	79.0	0	0	236.9	202.6
323	72.0	0	0		
324	70.0	0	0	951.1	735.6
325	73.0	0	0	458.8	315.9
326	6.0	1		460.1	359.3
327	67.0	0	0	282.5	304.9
329	76.0	0	0	137.3	121.8
330	67.0	0	0	464.8	447.0
331	4.0	1	1	598.0	324.6
332	0.0	1		795.3	491.1
333	35.0	1		367.4	299.2
334	57.0			270.2	167.2
335	67.0	0	0	239.2	190.5
336	6.0	1	1	559.2	456.8
337	75.0	0	0	107.6	101.2
338	1.0	1		735.3	566.5
339	21.0	1		393.2	348.4
340	75.0	0	0	506.7	426.7
341	74.0	0	0	310.0	210.0
342	26.0	1	1	200.7	213.3
343	69.0	0	0	515.1	442.5
344	2.0	1	1	286.8	274.6
345	61.0	1	1	287.0	220.3
346	75.0	0	0	708.8	567.2
347	59.0	1		581.2	410.4
348	71.0	0	0	528.4	509.6
349	12.0	1	1	514.8	255.5
350	63.0	0	0	313.4	272.0
351	72.0	0	0	741.4	584.6

352	2.0	1	1	354.3	339.4
353	69.0	0	0		
354	72.0	0	0	358.7	378.2
355	0.0	1	1	722.3	601.8
356	64.0	0	0	202.7	271.5
357	71.0	0	0	636.8	610.9
358	60.0	0	0	780.8	494.4
359	76.0	0	0	407.3	464.8
360	39.0	1		448.1	380.8
361	50.0	1		661.3	696.1
362	73.0	0	0	180.3	120.2
363	19.0	1		501.1	418.7
364	57.0			229.5	162.5
366	76.0	0	0	500.4	499.7
367	60.0	0	0		
368	2.0	1	1	449.3	394.0
369	78.0	0	0	242.1	177.3
371	65.0	0	0	354.1	276.6
372	36.0	1	1	237.7	142.6
373	72.0	0	0	409.4	394.8
374	21.0	1		402.7	478.0
375	73.0	0	0	403.3	311.4
376	77.0	0	0		
377	69.0	0	0	687.4	481.3
378	77.0	0	0	796.7	711.4
379	69.0	0	0	842.7	612.7
380	66.0	0	0	371.6	478.7
381	75.0	1		565.6	555.6
382	39.0	1	1	912.8	604.8
383	76.0	0	0	607.6	427.7
384	52.0	1		611.9	421.5
385	15.0	1	1	117.3	72.8
386	67.0	0	0	343.6	317.4
387	31.0	1		424.2	483.2
388	68.0	0	0	358.1	249.8
389	61.0	0	0	469.7	543.2
390	57.0			342.0	230.8
391	16.0	1	1	655.2	630.6
392	31.0	1		579.7	644.4
393	75.0	0	0	351.6	291.8
394	76.0	0	0	267.9	278.7
395	67.0	0	0	578.0	537.4
396	74.0	0	0	405.1	529.4
397	0.0	1	1	247.8	225.4
398	16.0	1		297.9	249.6

399	61.0	0	0	578.1	463.3
401	4.0	1	1	274.8	233.5
402	60.0	1		289.2	255.7
403	69.0	0	0	494.1	471.9
404	65.0	0	0	315.5	298.7
405	59.0	1	1	223.4	168.8
406	20.0	1	1	646.8	520.1
407	61.0	0	0	550.1	450.3
408	37.0	1	1	367.4	336.7
409	25.0	1		208.1	171.0
410	78.0	0	0	139.5	115.0
412	79.0	0	0	513.1	368.4
413	42.0	1	1	150.3	114.6
414	43.0	1	1	179.9	172.2
415	71.0	0	0	320.0	324.8
416	24.0	1	1	316.8	351.5
417	75.0	0	0	220.3	287.8
418	68.0	0	0	665.3	618.2
419	72.0	0	0	322.2	221.5
421	69.0	0	0	452.6	448.2
422	68.0	0	0	531.6	483.1
423	61.0	1		504.2	524.6
424	72.0	0	0	465.8	476.7
425	79.0	0	0	139.9	82.6
426	77.0	0	0	646.1	512.5
428	79.0	0	0	237.2	274.4
429	12.0	1	1	215.6	152.8
430	0.0	1	1	444.8	453.2
431	70.0	0	0	350.3	328.1
433	66.0	0	0	458.7	471.8
434	69.0	0	0	815.1	677.9
435	10.0	1	1	212.8	133.4
436	65.0	0	0	297.1	223.7
437	68.0	0	0	264.9	290.6
438	65.0	0	0	373.0	347.6
439	29.0	1	1	296.3	324.3
440	58.0			501.2	481.5
441	80.0	0	0	187.3	195.9
442	67.0	0	0	416.5	456.5
443	57.0	1	1	376.2	215.1
444	71.0	0	0	152.1	215.5
445	68.0	0	0	435.3	423.7
446	32.0	1		314.3	229.1
447	73.0	0	0	162.2	165.6
448	72.0	0	0	456.5	378.6

449	6.0	1	1	372.3	299.9
450	60.0	0	0	555.5	553.6
451	50.0	1		200.4	222.7
452	67.0	0	0	303.5	231.2
453	61.0	0	0	544.9	472.5
454	59.0			385.1	329.1
455	76.0	0	0	239.4	225.5
456	66.0	0	0	348.5	425.2
457	77.0	0	0	462.2	497.0
458	65.0	0	0	461.1	489.9
459	29.0	1	1	220.6	268.8
460	39.0	1		457.4	496.3
461	64.0	0	0	570.9	541.3
462	69.0	0	0	430.5	490.9
463	67.0	0	0	347.4	227.1
465	72.0	0	0	364.7	377.5
466	79.0	0	0	144.0	108.4
468	60.0	0	0	172.2	182.7
469	47.0	1		270.3	219.6
470	63.0	0	0	276.3	336.7
471	52.0	1		599.8	506.7
472	71.0	0	0	335.4	389.6
473	35.0	1	1	386.3	235.6
474	6.0	1	1	187.5	138.3
475	48.0	1	1	655.6	540.3
476	80.0	0	0	306.3	243.8
477	47.0	1		599.9	391.1
478	61.0	1	1	216.3	153.3
479	78.0	0	0	193.0	221.2
480	51.0	1	1	315.0	212.6
481	64.0	0	0	299.9	316.4
482	76.0	0	0	225.9	228.5
483	24.0	1		199.8	220.4
484	65.0	0	0	298.2	361.0
486	68.0	0	0	206.7	173.4
487	67.0	0	0	337.5	267.2
488	69.0	0	0	191.0	224.1
489	74.0	0	0	249.9	239.8
490	28.0	1		264.6	185.7
491	66.0	0	0	418.7	358.8
492	70.0	0	0	568.2	523.9
493	43.0	1	1	236.1	189.1
494	48.0	1	1	498.5	428.2
495	77.0	0	0	314.3	433.4
496	79.0	0	0	476.8	455.1

498	5.0	1		930.1	766.9
499	28.0	1		391.8	381.5
500	64.0	0	0	208.3	293.2
502	32.0	1	1	425.3	428.1
503	76.0	0	0		
504	20.0	1	1		
505	41.0	1		456.2	330.7
506	72.0	0	0	190.9	126.3
507	61.0	1		306.2	247.7
508	60.0	0	0	225.9	208.5
509	64.0	0	0	406.2	368.1
510	79.0	0	0	443.1	502.1
511	29.0	1		488.8	453.7
512	65.0	0	0	290.7	158.1
513	2.0	1		128.8	138.1
514	62.0	0	0	252.5	218.8
515	60.0	0	0	252.2	319.7
516	46.0	1		885.0	696.9
517	38.0	1		728.7	357.5
518	77.0	0	0	375.9	318.6
519	71.0	0	0	379.7	345.7
520	78.0	0	0	483.2	475.1
521	65.0	0	0	363.6	290.5
522	60.0	0	0	359.4	287.6
523	23.0	1		309.0	283.8
524	64.0	0	0	468.7	493.0
526	59.0	1	1	414.0	429.1
527	67.0	1	1	424.5	435.0
528	74.0	0	0	107.9	117.5
529	78.0	0	0	319.2	228.4
531	25.0	1	1	221.4	200.8
532	73.0	0	0	186.9	228.1
533	20.0	1	1	217.3	104.5
534	66.0	0	0	205.6	165.1
535	57.0			481.5	517.3
536	57.0			375.2	365.0
537	69.0	0	0	319.5	292.8
538	12.0	1	1	240.1	147.7
539	9.0	1	1	299.0	319.0
540	25.0	1		173.6	155.3
541	74.0	0	0	252.7	171.5
542	44.0	1	1	565.0	379.7
543	65.0	0	0	186.1	180.7
544	72.0	0	0	372.6	279.3
546	67.0	0	0	247.3	226.0

548	56.0	1		250.3	233.7
549	19.0	1		153.2	133.0
550	69.0	0	0	836.7	630.6
551	52.0	1	1	193.5	176.5
552	67.0	0	0	269.6	220.9
553	36.0	1	1	294.8	273.2
554	75.0	0	0	302.8	277.1
555	65.0	0	0		
556	73.0	0	0	262.0	247.2
557	59.0			418.3	338.8
558	21.0	1		507.8	496.6
560	69.0	0	0	823.0	653.2
561	17.0	1	1	125.3	135.9
563	69.0	0	0	649.9	539.7
564	34.0	1	1	248.7	129.8
565	40.0	1		219.5	169.8
566	6.0	1	1		
567	59.0			484.9	316.0
568	69.0	0	0	318.7	254.2
569	14.0	1		222.1	233.1
570	69.0	0	0	1068.8	957.4
571	78.0	0	0	404.2	242.8
572	61.0	0	0	218.2	206.2
573	65.0	0	0	1080.7	843.2
574	67.0	0	0	402.4	438.0
575	5.0	1	1	270.1	270.3
576	70.0	0	0	120.6	77.2
577	16.0	1	1	450.5	381.4
578	69.0	0	0	169.4	216.9
579	22.0	1	1		
580	37.0	1	1	466.7	356.7
581	55.0	1	1	437.9	372.0
582	42.0	1	1	154.0	124.8
583	69.0	0	0	948.1	1008.0
584	64.0	0	0	226.0	193.2
585	34.0	1	1	148.5	133.2

Table 35. Survival and Automated semi-quantitative (AQUA) analysis of DUSP6 protein expression in cohort 2. Semi-quantitation of DUSP6 protein expression was carried out per case on whole sections of cohort 2 by immunofluorescence. DUSP6 expression was assessed by AQUA system. Associations between DUSP6 protein expression and 5yr survival outcome were explored using Cox regression. *Event occurred = 1.

8.2.3 Statistical analysis of clinicopathologic features with DUSP6 protein expression

Statistical analysis was carried out to investigate associations between DUSP6 protein expression and clinicopathologic parameters, Table 42. A significant association between DUSP6 protein expression (nuclear and cytoplasmic) was identified with tumour size however this was rejected due to the limited sample size of lesions 101-150 macroscopic measurements (mm).

Clinicopathologic parameter	p-value	
	Nuclear DUSP6	Cytoplasmic DUSP6
Age at diagnosis	0.494	0.164
Gender	0.438	0.726
Tumour site	0.084	0.123
TNM stage	0.768	0.211
Tumour size	0.011*	0.008*
Tumour type	<0.001	0.24
Tumour differentiation	0.908	0.29
EMLVI	0.525	0.033

Table 36. Summary of statistical analysis of clinicopathologic parameters and DUSP6 expression. Kruskal-Wallis test was used to assess associations between each clinicopathologic parameter and DUSP6 protein expression (continuous scale). *A significant association between tumour size and DUSP6 protein expression (nuclear and cytoplasmic) was identified however rejected due to limited sample size of lesions sized 101-150mm.

8.2.4 Assessment of ERK expression in C99 cell lines by western blot (repeats)

Western blots on protein lysates from C99 cell lines (Parental, eGFP and DUSP6) for ERK expression were repeated, Fig. 72. Confirming initial findings, no marked difference in ERK protein expression was observed.

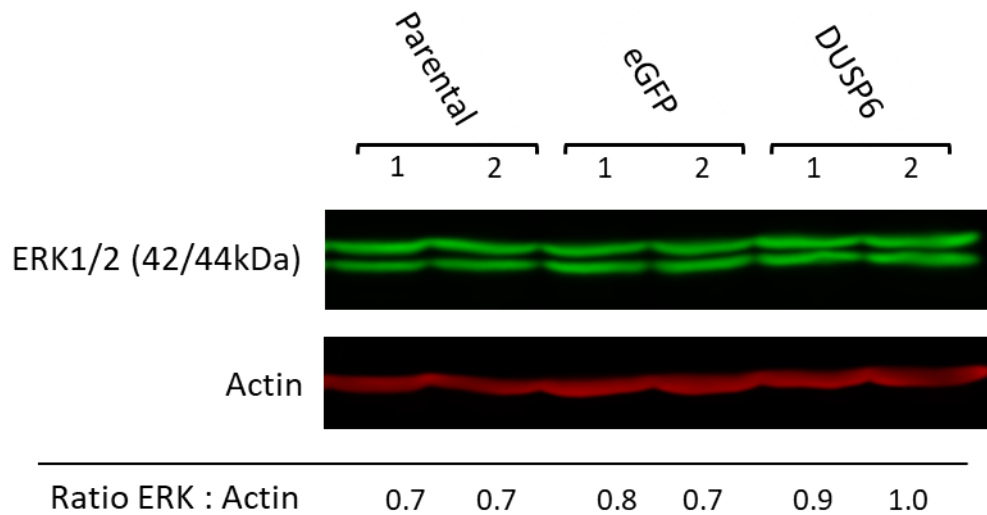


Figure 72. Western blot repeats of ERK protein expression in C99 cell lines. ERK protein expression assessment was repeated in C99 cells line (parental, enhanced green fluorescent protein (eGFP) transduced and DUSP6 transduced. No marked difference in ERK protein expression between the three cell lines was observed.

8.2.5 Assessment of p-ERK expression in C99 cell lines by western blot (repeats)

Western blots on protein lysates from C99 cell lines (Parental, eGFP and DUSP6) for p-ERK expression were repeated, Fig. 73. Confirming initial findings, a marked decrease in p-ERK protein expression was observed in C99 (DUSP6 transduced) cells compared to parental (92% decrease) and eGFP transduced lines (72% decrease).

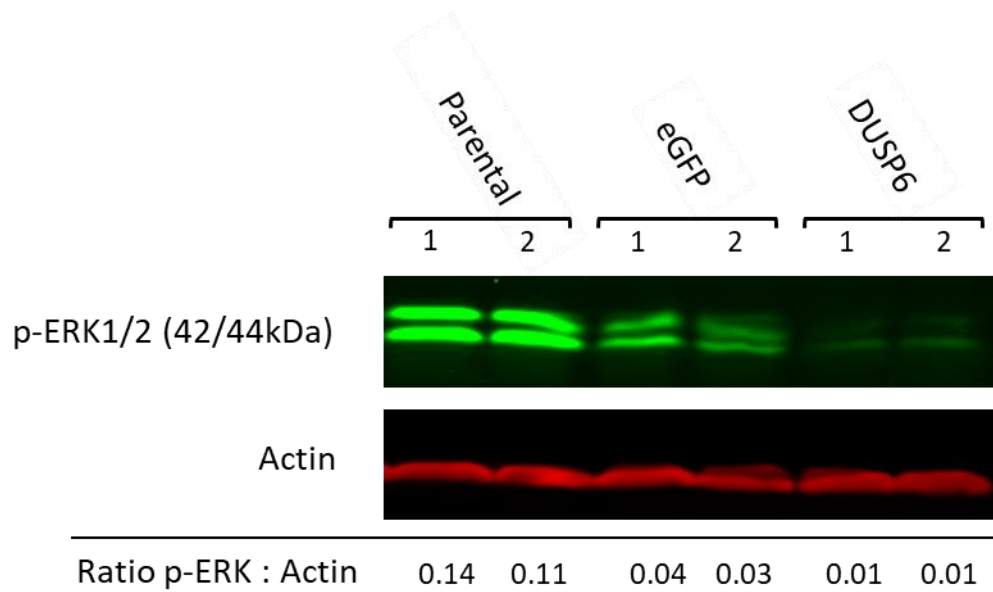


Figure 73. Western blot repeats of p-ERK protein expression in C99 cell lines. p-ERK protein expression assessment was repeated in C99 cells line (parental, enhanced green fluorescent protein (eGFP) transduced and DUSP6 transduced. A marked decrease in p-ERK protein expression was observed between RG/C2/80 (DUSP6 transduced) compared to parental (92%) and eGFP transduced (72%) cells.

8.2.6 Growth response to Cetuximab treatment – parental cell line

To confirm sensitivity of C99 cells to Cetuximab treatment as reported by Medico et al¹²⁵ cells were plated in six-plicate for each treatment condition (0 (media only control), 0.001, 0.01, 0.1, 1, 10, 100µg/mL). Cells were treated for four days, 24hrs post seeding. At the end of treatment an SRB assay was carried out to determine changes in cell growth, Table 43.

Replicate	Cetuximab concentration (µg/mL)						
	Control	0.001	0.01	0.1	1	10	100
1	98.0	109.7	87.1	70.7	53.2	33.2	26.2
2	119.2	89.5	80.0	66.5	46.2	42.2	32.6
3	115.7	82.5	75.2	69.9	45.7	41.2	31.7
4	94.7	83.0	70.0	57.9	43.9	37.2	30.0
5	79.1	86.4	76.5	61.5	39.2	34.5	31.3
6	93.2	92.3	68.6	68.2	47.4	31.3	26.1
Average change in growth (%)	100	90.6	76.2	65.8	46.0	36.6	29.7

Table 37. Change in growth in C99 adenocarcinoma cells treated with Cetuximab. For the parental C99 cell line sixplicates were set up for each concentration of Cetuximab (0 (control), 0.001, 0.01, 0.1, 1, 10 and 100µg/ml to confirm treatment response reported in Medico et al., 2015. At the end of the 4 day treatment an SRB assay was used to determine the change in growth compared to control lane. Using the optical density values, each well treated with Cetuximab was expressed as a percentage of the average of the control lane. An average of each sixplicate for each cell line was calculated.

8.2.7 Growth response to Cetuximab treatment – all cell lines

To assess the impact of DUSP6 transduction on Cetuximab sensitivity in C99 cell lines, each cell line (C99 (Parental), C99 (eGFP transduced) and C99 (DUSP6 transduced) were plated in six-plicate for each treatment condition (0 (control), 0.001, 0.01, 0.1, 1, 10, 100µg/mL). An SRB assay was used to determine change in growth compared to control lane for each cell line, Table 44.

A	Replicate	Cetuximab concentration (µg/mL)						
		Control	0.001	0.01	0.1	1	10	100
	1	106.6	97.7	93.5	62.6	43.4	38.9	36.1
	2	121.1	101.1	80.7	58.6	37.9	35.0	28.4
	3	88.5	84.9	73.1	49.4	40.3	36.3	31.6
	4	75.7	99.6	61.8	50.2	36.1	31.9	32.6
	5	118.7	100.1	62.3	62.3	34.5	24.8	31.1
	6	89.3	82.0	58.6	50.0	40.5	33.7	21.9
	<i>Average change in growth (%)</i>	<i>100</i>	<i>94.2</i>	<i>71.7</i>	<i>55.5</i>	<i>38.8</i>	<i>33.4</i>	<i>30.3</i>

B	Replicate	Cetuximab concentration (µg/mL)						
		Control	0.001	0.01	0.1	1	10	100
	1	119.1	131.8	92.1	68.1	52.7	39.7	42.7
	2	97.5	116.2	88.3	72.9	47.8	37.8	26.2
	3	76.7	107.5	58.4	53.8	57.0	44.3	36.5
	4	104.5	87.0	70.2	60.5	41.9	37.6	32.7
	5	105.9	78.6	78.9	46.7	44.6	36.2	27.8
	6	96.2	94.8	69.7	71.1	43.5	35.7	32.7
	<i>Average change in growth (%)</i>	<i>100</i>	<i>102.7</i>	<i>76.3</i>	<i>62.2</i>	<i>47.9</i>	<i>38.6</i>	<i>33.1</i>

C	Replicate	Cetuximab concentration (µg/mL)						
		Control	0.001	0.01	0.1	1	10	100
	1	120.2	93.7	81.2	52.7	33.8	26.8	20.8
	2	121.8	88.2	73.7	47.8	27.7	21.1	21.2
	3	80.3	87.1	55.2	35.5	26.6	19.6	18.3
	4	98.4	82.0	49.8	30.9	30.3	22.7	17.0
	5	100.1	79.3	35.0	38.5	30.6	24.2	19.8
	6	79.1	89.5	60.0	43.8	24.5	25.2	20.1
	<i>Average change in growth (%)</i>	<i>100</i>	<i>86.6</i>	<i>59.2</i>	<i>41.5</i>	<i>28.9</i>	<i>23.3</i>	<i>19.5</i>

Table 38. Change in growth of C99 cell lines treated with Cetuximab. A: C99 (Parental). B: C99 (eGFP transduced). C: C99 (DUSP6 transduced). For each cell line sixplicates were set up for each concentration of Cetuximab (0 (control), 0.001, 0.01, 0.1, 1, 10 and 100µg/ml). At the end of the 4 day treatment an SRB assay was used to determine the change in growth compared to control lane. Using the optical density values, each well treated with Cetuximab was expressed as a percentage of the average of the control lane. An average of each sixplicate for each cell line was calculated.

8.3 Ethics approval



University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

15th January 2018

Professor David Harrison
School of Medicine

Dear Professor Harrison

Thank you for submitting your amendment application which comprised the following documents:

1. Ethical Amendment Application Form

The School of Medicine Ethics Committee is delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has approved this ethical amendment application. The particulars of this approval are as follows –

Original Approval Code:	MD9202	Approved on:	16/10/2012
Amendment 5 Approval Date:	15/01/18	Approval Expiry Date:	16/10/2022
Term of Approval	10 YEARS		
Project Title:	Systems pathology of disease		
Researcher(s):	David Harrison, Fiona McKissock, Jennifer Bre, Peter Cate, In Hwa Um, Oliver Read, Ines Nearchou, Christos Gavriel, Nsofytos Dimitriou, Hannah Williams, Mustafa Elshani, Raffaele De Filippis, Mary Kudsy, Mark Dates, Romina Brilla, Nourjahan Khafaga, Sophie Rao, Tsz Chan, Matthew Scoll, Beth Gwyther	Supervisor(s):	David Harrison

Ethical amendment approval does not extend the originally granted approval period, rather it validates the changes you have made to the originally approved ethical application. If you are unable to complete your research within the original validation period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

Any serious adverse events or significant change which occurs in connection with this study and/or which may alter its ethical consideration, must be reported immediately to the School Ethics Committee, and an Ethical Amendment Form submitted where appropriate.

Approval is given on the understanding that you adhere to the 'Guidelines for Ethical Research Practice' (<http://www.st-andrews.ac.uk/media/UTREC/guidelines%20Feb%2008.pdf>).

Yours sincerely

Dr Morven Shearer, Convener School of Medicine Ethics Committee

School of Medicine Ethics Committee

Medical and Biological Sciences Building, North Haugh, St Andrews, Fife, KY16 9TE, Scotland, UK
Email: medelao@st-andrews.ac.uk Tel No: 01334 463385
The University of St Andrews is a charity registered in Scotland: No SC013532

8.4 Tissue Governance approval

8.4.1 Cohort 1 tissue governance approval

Document Name	QF-TGU-A-SAMREQA	VERSION 1.1	Page	1 of 1	Review date	25-Sep-2019
---------------	------------------	-------------	------	--------	-------------	-------------

LOTHIAN NRS BIORESOURCE SAMPLE REQUEST ANSWER FORM

Sample Request number:	SR963
Name of Researcher:	Professor David Harrison Hannah Williams
Address of Researcher:	Royal Infirmary Edinburgh Little France
Study Title:	The role of Oxidative stress and DUSP6 in colorectal carcinogenesis.
Ethical status:	15/ES/0094
Material Requested	<p>Release and use of the following is approved by Tissue Governance for the purposes of the above named study:</p> <p>Approx 200 archival FFPE samples from 2012. Survival data for the cases.</p> <p>All samples and data must be used in a de-identified manner for the purposes of the research so that patient confidentiality is maintained at all times.</p>

REQUEST AUTHORISED

Date:	10-Nov-2017
Authorised by:	Frances Rae

REQUEST REJECTED

Date:	
Authorised by:	
Reason	

Author	: Frances Rae	Date	: 25-Sep-2015
Authority for issue	: Craig Marshall	Date	: 25-Sep-2015
Quality Checked	: Craig Marshall	Date	: 25-Sep-2015

8.4.2 Cohort 2 tissue governance approval

Document Name	QF-TGU-A-SAMREQA	VERSION 1.0	Page	1 of 1	Review date	15-Jul-2014
---------------	------------------	-------------	------	--------	-------------	-------------

NHS Lothian SAHSC Bioresource Sample Request Answer Form

Sample Request number:	SR386
Name of Researcher:	Dr Anca Oniscu / Prof David Harrison
Address of Researcher:	Pathology Dept Laboratory Medicine Royal Infirmary Edinburgh
Study Title:	Molecular stratification of colorectal cancer
Ethical status:	13/ES/0126
Material Requested	Use of the following is approved for the purposes of the above named study. Anonymised CRC samples (H&Es and paraffin blocks) which were part of the Phase I CRUK Stratified Medicine programme. Archival anonymised CRC samples tested locally within Molecular Pathology, RIE. This will be approximately 400 cases.

REQUEST AUTHORISED

Date:	05-Aug-2014
Authorised by:	Frances Rae

REQUEST REJECTED

Date:	
Authorised by:	
Reason	

Author	: Frances Rae	Date	: 15-Jul-2010
Authority for Issue	: Craig Marshall	Date	: 15-Jul-2010
Quality Checked	: Craig Marshall	Date	: 15-Jul-2010

8.5 Standard operating procedures



Document Number:	SASoM/EQUIP/101.v1
Title:	Use of Leica SCN 400 Scanner
Version:	v1
Author:	In Hwa Um

Effective from:	01/03/17
Valid to:	28/02/22

SOP History		
Number	Date	Reason for Change
v1		Original

1.0 Purpose –

The purpose of this SOP is to outline the principles of the routine use and maintenance of Leica SCN400 scanner in Laboratory 248 at the St Andrews School of Medicine (SASoM).

2.0 Scope –

This SOP applies to routine use and maintenance of Leica SCN400 scanner within the SASoM.

3.0 Responsibilities –

It is the responsibility of all users of Leica SCN400 scanner within the SASoM to comply with this SOP.

4.0 Procedure –

4.1 Switching on Scanner and Computer

4.1.1 Order for switching on scanner and computer:

- 1.QSNAP Memory Buffer: Should be on and left on, if not, turn this on first - will take some time to stabilise,
- 2.Scanbox: Press and hold button on front of Scanbox – green light will come on.
- 3.Scanner PC: Turn on the PC and Monitor



Equipment Operation Procedure

4.Scanner: Initialises by itself after Scanbox is turned on (sometimes requires software to be opened before initialisation. Light on scanner turns green. While initialising the second light on scanner will flash orange. When the scanner is ready both lights will shine green.

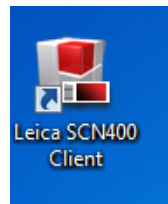
4.1.2 Username and Password to log onto computer:

Username: dih_slidepath

Password: Svc11023.spDIH

4.1.3 Open Software:

Double click on the SCN400 Client icon (will take a few moments to open)



Log on details for software:

User name: dih_slidepath

Password: Svc11023.spDIH

4.2 Loading and pre-scanning

4.2.1 Loading slides:

Insert the slides into the slide tray and then the slide tray into the scanner.

If the tray is inside the machine and not visible:

Go to "Setup" menu in software and press Eject:

Prior to scanning; naming and exporting images

Click on "Configuration" Drop down menu and select "Options"

4.2.2 Auto Naming:

Click on the "Auto Naming" tab and in "Entity Selection" window select "ImageCollection (User)" from the dropdown menu.

Click "Edit" in the "Entity data" window and enter the name you want to call the slide collection in the "Stem" dialogue box.

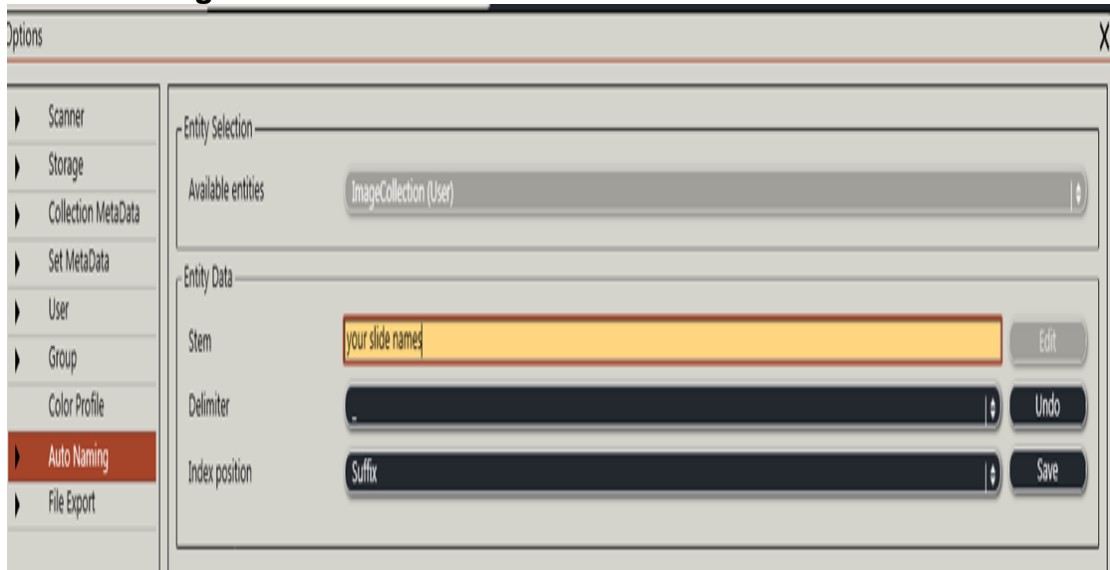
You can only use one stem for each scan run and a suffix of 1,2,3 or 4 will appear after the stem to identify individual slide images.

Click "Save" when complete.



Equipment Operation Procedure

Auto Naming window:



File Export:

Click on the “File Export” tab in Options and press “Edit” in the “File Export Data” window

Check the box “Export after Overview and Scan”

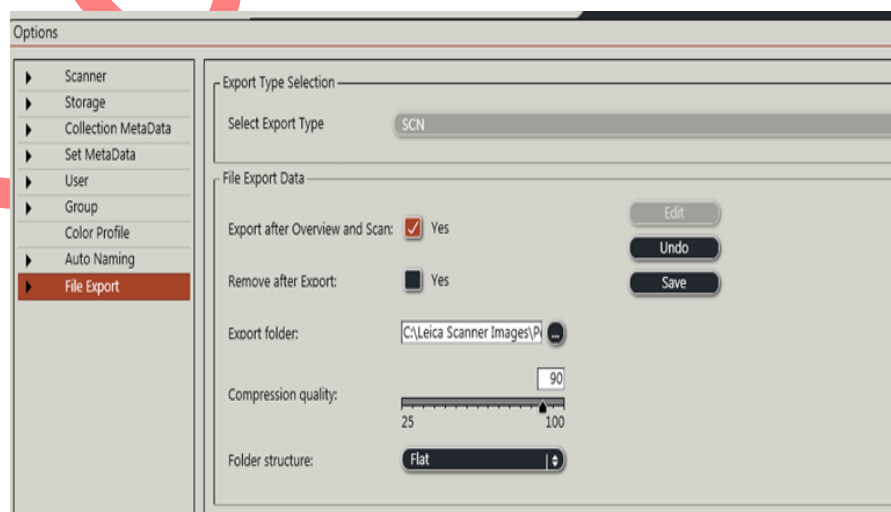
In the Export folder browser box click the three dots “...” and browse to the folder you want the images to be exported to.

Compression Quality should be between 85-100%

Press Save

Press Close (bottom right of screen)

File Export window:



Equipment Operation Procedure

4.3 Manual Scanning BF overview:

Within the “Setup” Menu of the software and under “Scanner” window ensure the full window is displayed by pressing the arrow at the right so it is facing downward.

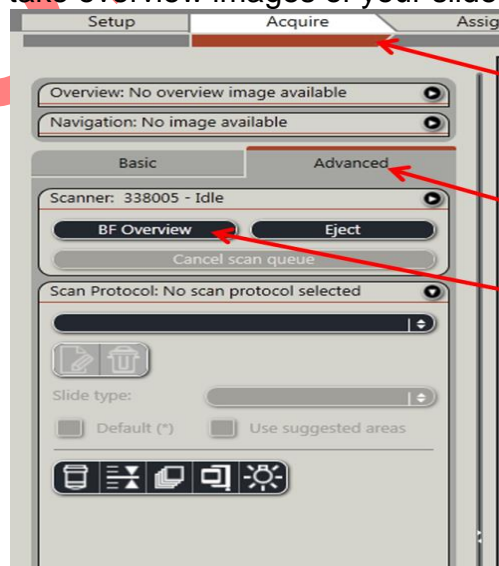


Next ensure the “Attended” box is checked

Now switch to the “Acquire” Menu and the “Advanced” window

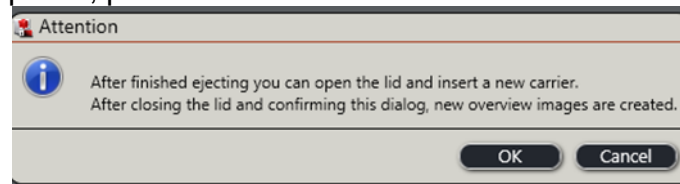
Press “BF Overview”

This will load the slides and take overview images of your slides.



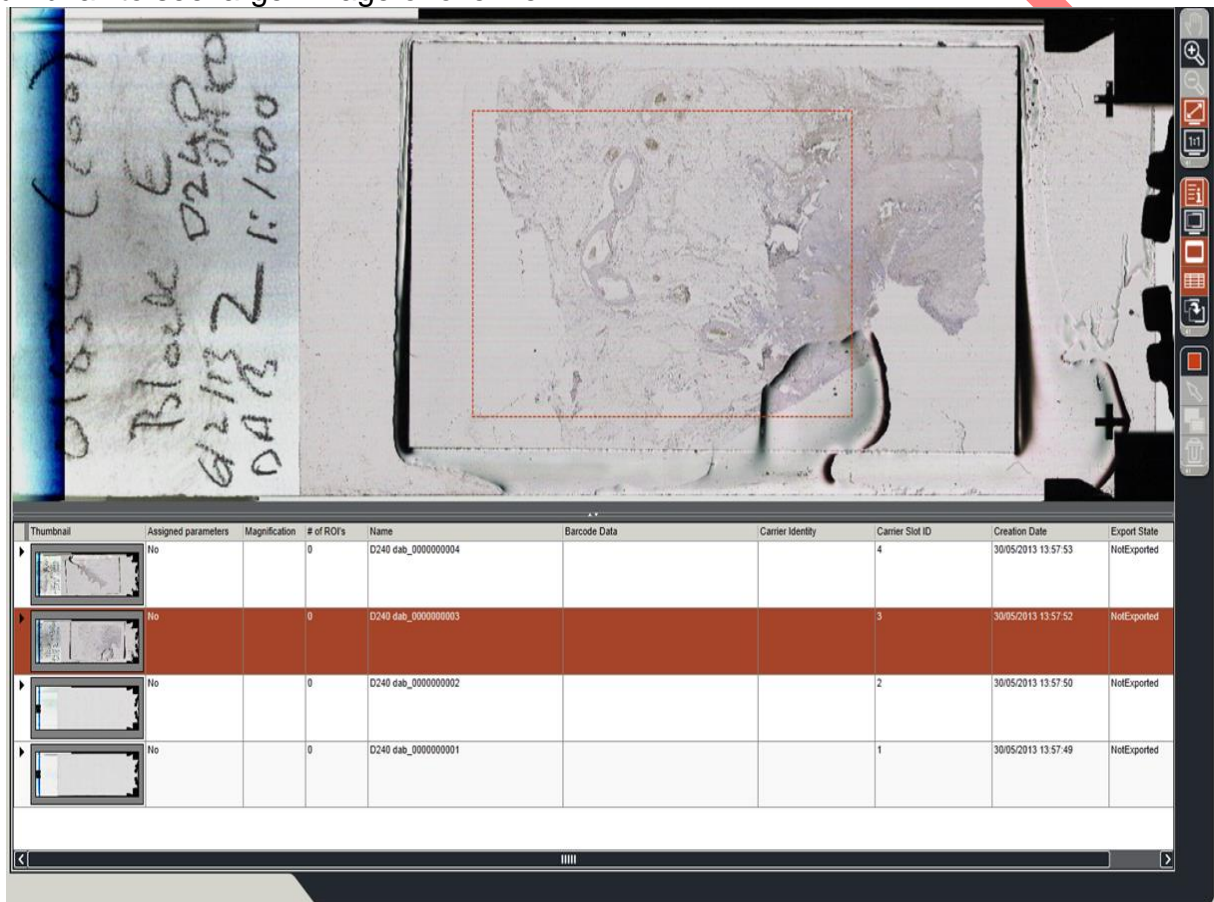
Equipment Operation Procedure

If this Attention box appears, press OK



The orange light on the SCN400 will flash indicating the platform is scanning

The low magnification overview slides will now appear in the thumbnail window, click on thumbnail to see larger image of overview:



4.4 Manual Scanning High magnification

4.4.1 Left Click on your 1st overview thumbnail you wish to image:

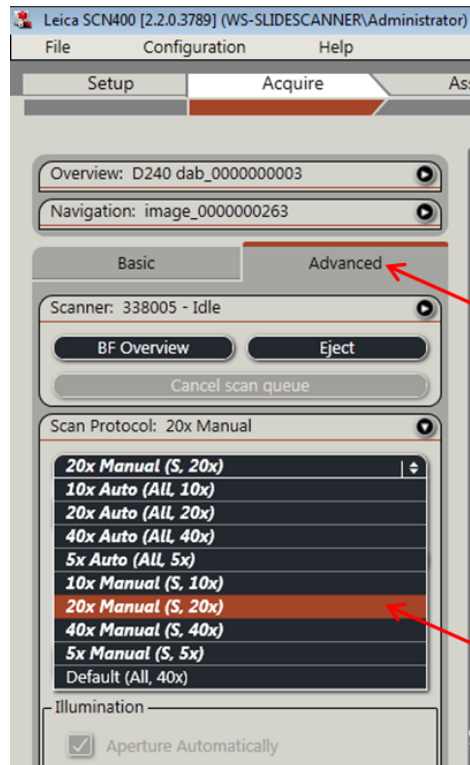
Select the "Advanced" tab

In the "Scan protocol" window, select the magnification you would like to scan at from the dropdown menu:

5x, 10x, 20x, 40x – ensure the it is the Manual version NOT the Advanced

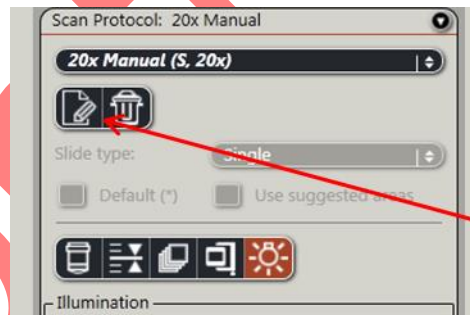


Equipment Operation Procedure



If the red dotted rectangle on your overview slide image is not over the area you wish to scan then:

Click the edit button in Scan Protocol window:

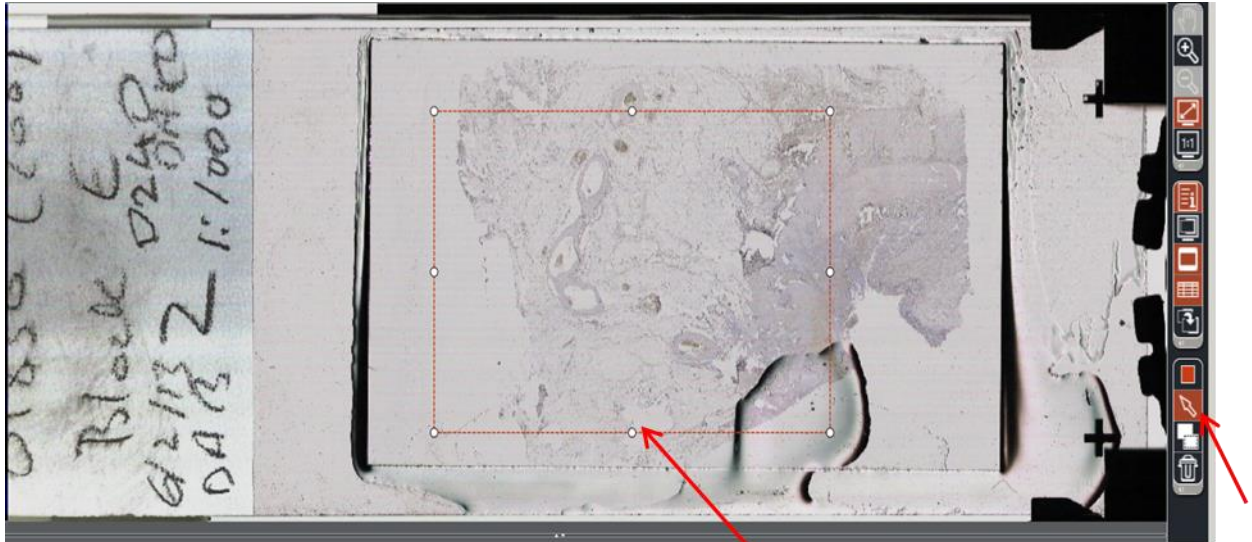


On the far right of the image window press the arrow icon

Now move the arrow with the mouse onto the red dotted rectangle so that the white circles appear on it

Reshape the dotted outline using the mouse so that it encompasses the desired scan area.

Equipment Operation Procedure

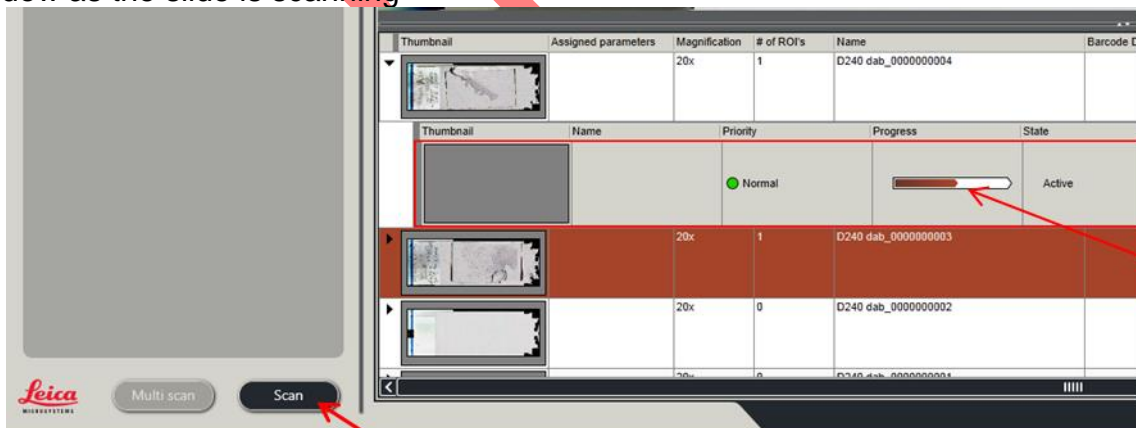


To add a new scan area to the same slide press and draw the new scan area.



Press Save in the Scan Protocol window

Press Scan at the bottom left of the screen; a progress bar will appear in the thumbnail window as the slide is scanning



The Orange scan light will flash at the front of the instrument to show that the scanner is operating

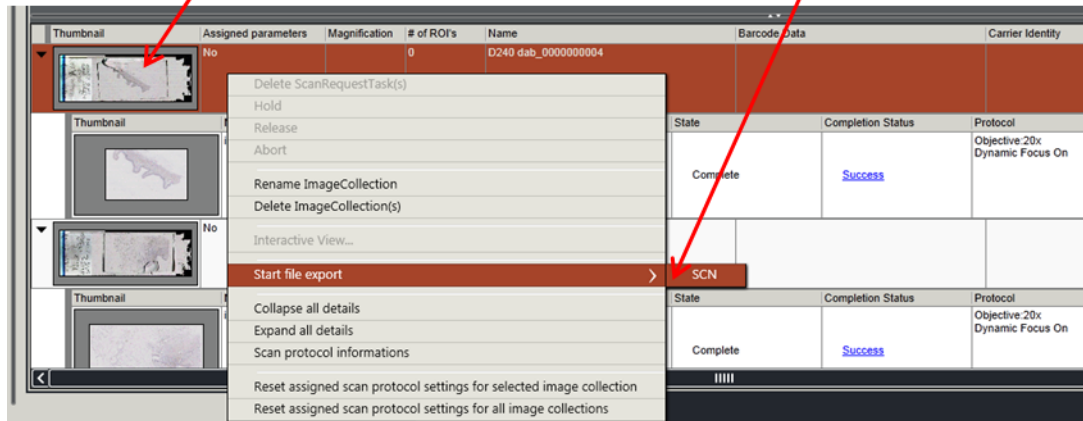
4.4.2 Click on your 2nd overview thumbnail and repeat process D.



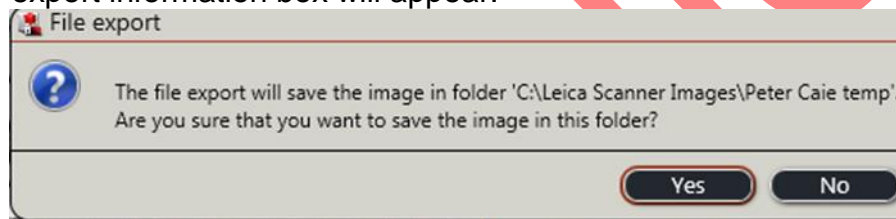
Equipment Operation Procedure

Manual Exporting of images

Right click on the overview thumbnail (not the hi-resolution thumbnail) of the image you wish to export and highlight “Start file export” then highlight “SCN”



A File export information box will appear:



Ensure that the export directory is correct and press OK.

4.5 Switching off Scanner and Computer

After scanning is finished the scanner may be shut down in the software.

The Scanner may also be shutdown by pressing and holding the ON/OFF button (with green light in centre) on the front of the Scan Box – NOT the scanner itself.

Once the scanner is shutdown the PC and monitor can be switched off

Leave the QSNAP Memory Buffer on.

5.0 Personal protection –

- Howie coat must be worn at all times.

6.0 Training –

All users have to be trained before using the Instrument by a designated person.

7.0 Related documents –

7.1 Risk assessments – RA/MH/002, COSHH/008



8.0 Approval and sign off –

Author:

Name: In Hwa Um

Position: Research technician

Signature:

Date:

Management Approval:

Name: Mary Wilson

Position: Laboratory Manger

Signature:

Date:

QA release by:

Name: Alex MacLellan

Position: QA Manager

Signature:

Date:

Confidential



Document Number:	SASoM/EQUIP/025.v3
Title:	Image Acquisition using HistoRx® PM-2000, AQUA (<u>A</u>utomated <u>Q</u>uantitative <u>A</u>nalysis)
Version:	v3
Author:	Peter Mullen

Effective from:	02/07/2018
Valid to:	01/07/2023

SOP History		
Number	Date	Reason for Change
v1	01/01/1013	Original
V2	02/07/2013	Minor Amendments
V3	02/07/2018	Update

1.0 Purpose –

The purpose of this SOP is to outline the principles of image acquisition using HistoRx ® PM-2000, AQUA in Laboratory 248 at the St Andrews School of Medicine (SASoM).

2.0 Scope –

This SOP applies to routine image acquisition using AQUA within the SASoM.

3.0 Responsibilities –

It is the responsibility of all users of the AQUA within the SASoM to comply with this SOP.

4.0 Procedure –

Hardware components:

External light source - EXPO X-cite 120XL contains 120W pre-aligned lamp. Up to 1500 hours will be used. This lamp is a mercury lamp. All the users must read risk assessment and be aware of dealing with breakage of mercury lamp.

Enclosed microscope - The microscope consists of five Olympus UIS2 objective lenses (x4, x10, x20, x40, x60) and Prior precision stage with high speed PCI controller and Numerik Jena linear encoders.

Digital camera – 2048 x 2048 pixels



Dell workstation – Two dual core, hyper threaded intel processors, 2GB RAM, Windows XP Professional SP2 operating system are featured.

Joystick – used to control stage movements directly in all three axes.

Software componets:

AQUAsition used for image acquisition

AQUAnalysis used for image analysis

Turn Unit On:

The AQUA station consists of four main power lines plugged into one extension hub located on the rear of the monitor and EXFO lamp box. Turn on by moving switch to ON position in the extension hub.

Turn on the light source using the rocker switch at the front of the EXFO lamp unit (allow the lamp to warm up for a minimum of 20 minutes in order to stabilize the illuminator's light output.).

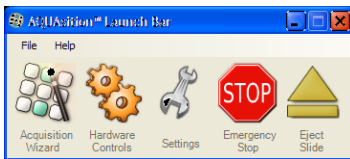
Image Acquisition(AQUAsition):

On the **Desktop**, double-click the **AQUAsition** icon to start the program.



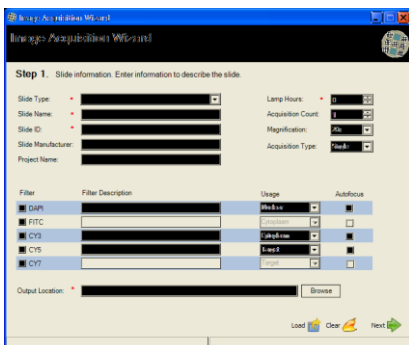
Open the PM-2000 access doors and place a slide in the stage insert.

The specimen should face up and the label end of the slide should be towards the doors. Ensure that the slide is well secured with the retaining clip, and then close the access doors.



On the **AQUAsition Launch Bar** window, click the **Acquisition Wizard** icon to begin the slide acquisition process.

STEP1. Slide information:



In the **Slide Type** box, select the kind of slide to be scanned.

The two available options are either a Tissue Microarray (TMA) or Whole Tissue Section (WTS). This choice will only affect the way in which the fields-of-view to be acquired are selected and will have no effect on the analysis.



Equipment Operation Procedure

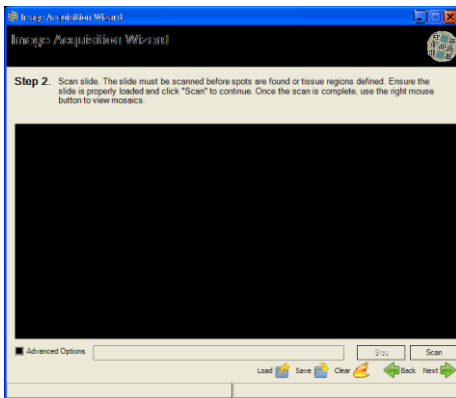
On the **Slide information** window, select or fill all other relevant text boxes.

This information will be permanently attached to the image file. It can subsequently be used to assist in tracking extensive projects and can aid in the integration of the output data into a personal experiment database.

NOTE: All fields marked with a red asterisk must be filled out to continue to the next step of the slide acquisition process.

When all required fields have been completed, click the **Next** arrow to proceed to the following step.

STEP2. Scan slide:



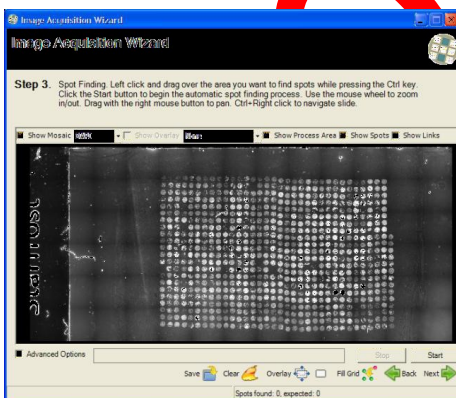
On the **Scan slide** window, click the **Scan** button to initiate the low-resolution composite acquisition of the DAPI filter.

The low-resolution composite(s) are used in the spot finding or region selection steps only and will not be used in the quantification.

TIP: Before clicking the **Scan** button, select the **Advanced Options** checkbox to reveal additional selections allowing the opportunity to select some or all filters in use for low-resolution scanning. The composites for all filters often prove a valuable tool in a quick evaluation of the quality of the tissue sample and of the staining.

When the scan is completed, click on the **Next** arrow to proceed to the next step.

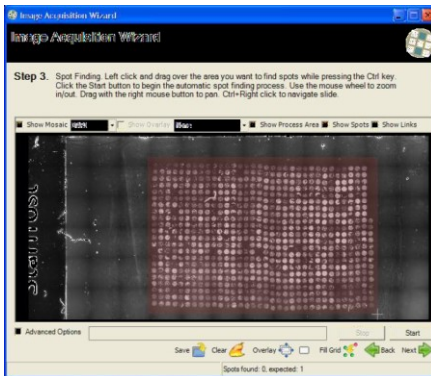
STEP3. Spot finding:



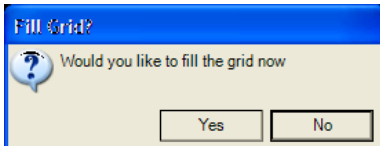
On the **Spot Finding** window, hold down **CTRL** and drag the mouse pointer to select the region of the slide containing the tissue microarray samples to be acquired.

The selected area will now be overlaid in a translucent red rectangle.

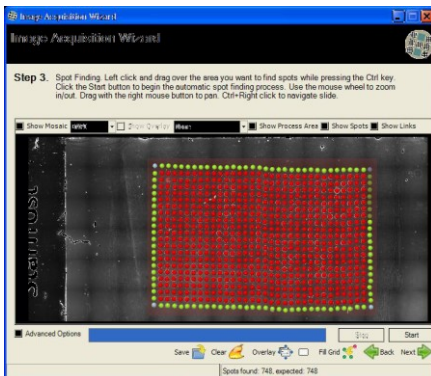
Equipment Operation Procedure



Click the **Start** button to initiate the *AQUAsition* spot finding algorithm, which will attempt to identify all tissue microarray cores within the selected region.



On the **Fill Grid** dialog box, click **Yes** to fill in missing spots based on the intersection of rows and columns.



Once a core is discovered by the algorithm, a colour-coded circle is overlaid on the composite image to mark its location. Initially, the overlaid spots will be orange, and then turn to blue, green, or red. These colours represent the spot's current position and row and column links within the array as assigned by the program. Depending on the quality of the tissue microarray and other factors, the algorithm may not be able to identify all spot array links correctly. Therefore, the colour-coding can prove very helpful for making the final adjustments to ensure proper row and column linking.

Red spot: Inner spots; directly linked to four other immediately adjacent spots (top, bottom, left, right).

Orange spot: Found spots, array links not yet determined.

Blue spot: Corner spots; directly linked to only two other spots.

Green spot: Peripheral spots; directly linked to three other spots.

On the composite image, click an area to add a missing spot or remove an incorrectly placed one.

Found spots may also be repositioned by dragging. For detailed adjustments, the low-resolution composite image may be zoomed into by rotating the mouse wheel.

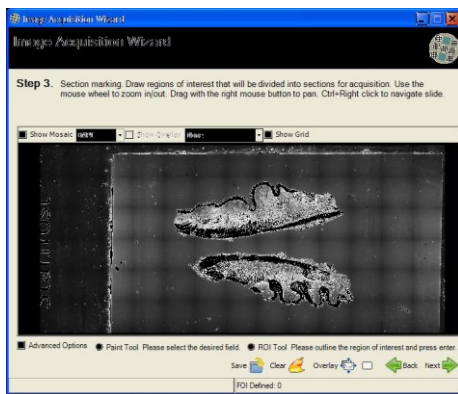
Equipment Operation Procedure

When all existing array cores are marked with a red, green, or blue spot, ensure that the array is rectangular by adding spots to missing row column intersections that were missed by the **Fill Grid** function.

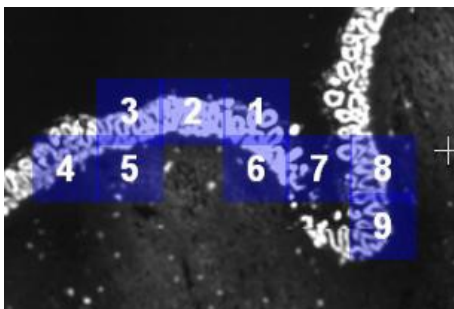
TIP: Attempting to focus on a spot location that contains no tissue can cause auto-focus failure for the offending spot as well as all those following it during high-resolution acquisition. It is recommended that spots on the array that act as placeholders for missing cores are designated as *Virtual Spots*, which will be skipped entirely during acquisition yet the spot grid integrity will be preserved. To do so, hold down **SHIFT** and click each placeholder spot.

When the spot finding is completed, click on the **Next** arrow to proceed to the next step.

STEP4. Section marking:



On the **Section marking** window, click the composite image in an area to be acquired.



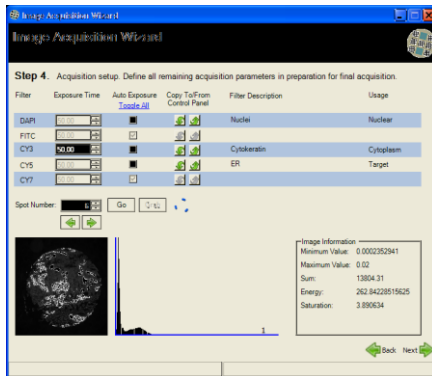
A blue, translucent field-of-view marker, its size corresponding to the magnification selected on the **Slide information** window, will now mark the area to be acquired. To select larger areas to be acquired, drag the pointer over the areas to be acquired.

NOTE: *Fields-of-view to be acquired need not be contiguous.*

For increased precision, the low-resolution composite image may be zoomed into by rotating the mouse wheel and then panned around by holding down the right mouse button and dragging.

When the section marking is completed, click on the **Next** arrow to proceed to the next step.

STEP5. Image Acquisition setup:



On the **Acquisition setup** window, clear the **Auto Exposure** checkboxes corresponding to the filters for which exposure times are to be manually set or else click **Toggle All** to clear the checkboxes for all filters.

NOTE: The **Auto Exposure** function greatly enhances quantification accuracy and the dynamic range of scores by automatically adapting to the pixel intensity levels of each spot and filter. Using the manual exposure setting is not recommended for most acquisitions.

Click on the **Next** arrow to proceed to the next step.

STEP6. Image acquire:



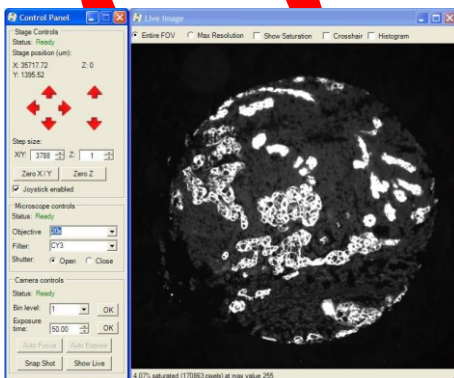
On the **Acquire** window, click **Set Focus** to set the initial focal point.

The **Live View** and **Control Panel** windows will now open and the stage will navigate to the first microarray spot or the first field of view of a whole tissue section.

NOTE: The shutter does not open automatically to avoid bleaching of the slide.

On the **Control Panel** window, click **Open** to open the shutter.

Referring to the live image displayed in the **Live View** window, bring the tissue into focus



On the **Control Panel** window, click **Open** to open the shutter.



Equipment Operation Procedure

Referring to the live image displayed in the **Live View** window, bring the tissue into focus by rotating the mouse wheel or else using the joystick.

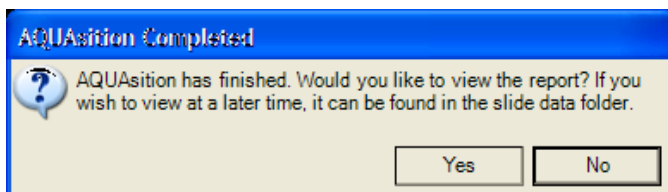
*NOTE: The mouse wheel can be used for focusing only when the **Live View** window is active.*

On the **Acquire** window, click **Go**.

The high-resolution acquisition will now proceed automatically. The **Approximate Time Remaining** counter displays an estimate of the time left to completion of acquisition. It may take up to three spots before an estimate appears and its accuracy improves with each additional spot.

Close down the viewer window in order to conserve memory on the PC. Failure to close the viewer may result in acquisition failing midway through the run. Similarly close the Settings window

When image acquisition is completed, the shutter will close, the stage will return to its home position at the upper left corner of the slide and a notification dialog will be displayed.

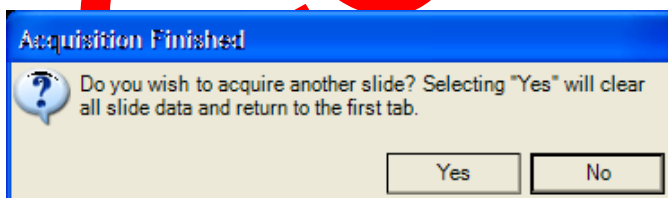


Click **Yes** to view the **AQUAsition™ Results** report.

The report file is in the PDF format and will be displayed in the computer's default PDF viewer application.

Close the PDF viewer application when finished with the report.

On the **Acquire** windows, click **Done**.



On the **Acquisition Finished** dialog, click **No**.

On the **AQUAsition™ Launch Bar** window, click **Exit** on the **File** menu or click **Close** button to exit the program.



5.0 Personal protection -

Howie coat must be worn at all times.

6.0 Training

Under no circumstances should this machine be operated by anyone who is not acquainted with it or has not been trained to use it.

7.0 Related documents –

- 7.1 Equipment manual
- 7.2 Equipment Maintenance Log
- 7.3 Equipment Maintenance Information sheet
- 7.4 Risk assessments – RA/GEN/038, RA/MH/002 and COSHH/008
- 7.5 SOP SASoM/EQUIP/026
Image analysis using HistoRx® PM-2000, AQUA
- 7.6 SOP SASoM/EQUIP/024
Maintenance and replacement of EXFO lamp of the HistoRx® PM-2000, AQUA

8.0 Approval and sign off –

Author:

Name: Peter Mullen

Position: Research Fellow

Signature: _____ Date: _____

Management Approval:

Name: Mary Wilson

Position: Laboratory Manager

Signature: _____ Date: _____

QA release by:

Name: Alex MacLellan

Position: QA Manager



Equipment Operation Procedure

Signature:	Date:
------------	-------

Controlled



Document Number:	SASoM/EQUIP/026.v2
Title:	Image Analysis using HistoRx® PM-2000, AQUA (<u>A</u>utomated <u>Q</u>uantitative <u>A</u>nalysis)
Version:	v2
Author:	Peter Mullen

Effective from:	01/01/2018
Valid to:	31/12/2022

SOP History		
Number	Date	Reason for Change
v1	01/01/2013	Original
V2	01/01/2018	Update

1.0 Purpose –

The purpose of this SOP is to outline the principles of image analysis using HistoRx® PM-2000, AQUA in Laboratory 248 at the St Andrews School of Medicine (SASoM).

2.0 Scope –

This SOP applies to routine image analysis using HistoRx® PM-2000, AQUA in the SASoM.

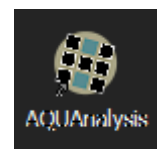
3.0 Responsibilities –

It is the responsibility of all users of the AQUA within the SASoM to comply with this SOP.

4.0 Procedure –

AQUAnalysis:

Double-click the **AQUAnalysis** icon on the desktop to start the program.

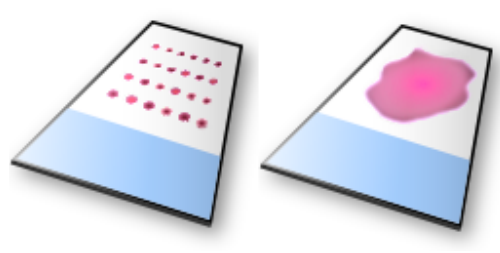


On the **File** menu, click **Load TMA**.

Browse for the TMA file created in the previous section, and then click **OK**.



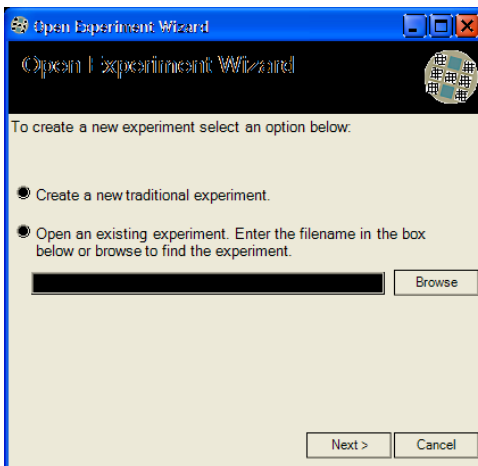
Equipment Operation Procedure



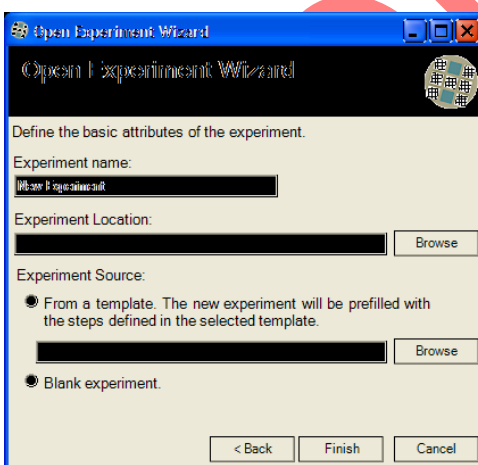
The selected TMA file appears in the **Slide Information** pane on the *AQUAnalysis Main* window.

NOTE: Multiple TMA files can be loaded simultaneously into the AQUAnalysis program if they contain images from an identical set of filters. Only one TMA file is active at any one time. To activate a different loaded TMA file, click on the corresponding slide icon. The eye icon marks the active TMA file.

Creating a New Experiment:



On the **File** menu, click **Open Experiment Wizard**.



Select **Create a new traditional experiment**, then click **Next**.

In the **Experiment name** text box type in a name for the experiment.

Click the **Browse** button to the right of the **Experiment Location** text box then select a directory for experiment data to be stored.



Equipment Operation Procedure

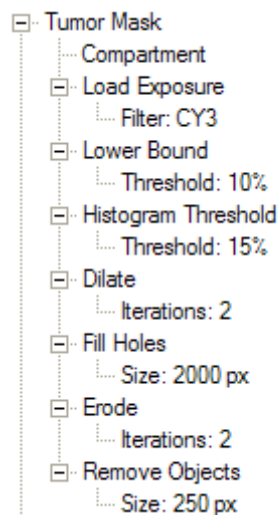
Select From template then open the browse to find previously set experimental template located in Experimental template in AQUA folder in Y: drive as an experiment source. Then click **Finish** to complete the wizard.

TIP: Frequently used experimental setups can be saved as AQUA templates files and then loaded in this step.

Experiment Steps and Procedures:

The combination of steps and procedures detailed below is a standard experimental setup for epithelial tissue analysis and is contained in the *Standard HARP AQUA template* file. Values in the following AQUA experimental setup steps are recommendations. The user should define variable values for procedures such as **Histogram Threshold**, **Fill Holes**, **Remove Objects**, etc., to reflect the staining and tissue properties of the TMA file being analyzed.

a. Tumour mask



Load Cy3 exposure then run Lower bound, Histogram threshold, dilate, fill holes, erode, remove objects step by step.

The **Lower Bound** procedure removes the lowest X percentage of the pixel range recorded in the image, effectively setting them to zero. This can be useful for removing background fluorescence haze. The recommended value for the **Threshold** variable is 10%.

The **Histogram Threshold** procedure sets all pixels with CY3 signal above the chosen threshold to ON\1\white and all other pixels to OFF\0\black, effectively binerizing the image. The recommended value for the **Threshold** variable is 15%.

The **Dilate** procedure blooms all non-negative pixel areas by the number of pixels specified in the **Iterations** variable. The recommended value for the **Iterations** variable is 2.

The **Fill Holes** procedure fills the nuclei holes in the cytokeratin stain. The recommended value for the **Size** variable is 2000px.

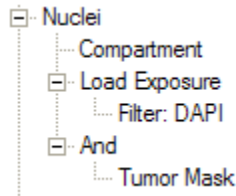


Equipment Operation Procedure

The **Erode** procedure contracts all pixels bloomed by the **Dilate** procedure in Step 16 above, with the exception of those pixels now part of a filled hole. The recommended value for the **Iterations** variable is 2.

The **Remove Objects** procedure cleans up the mask by removing small, stand-alone clusters of positive pixels. The recommended value for the **Size** variable is 250 px.

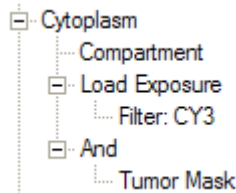
b. Nuclei



After loading Dapi exposure run And tumour mask.

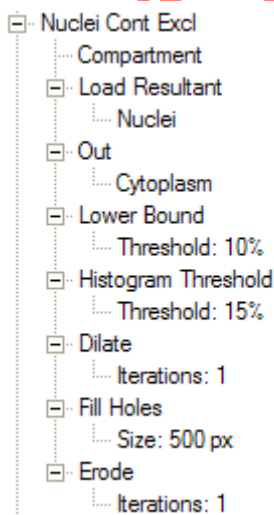
The **And** procedure output is a multiplication of the matching pixels from the exposure loaded at the beginning of the step and those of the variable resultant image.

c. Cytoplasm



d. Nuclei Cont Excl

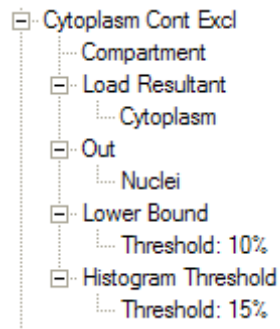
Run the series of steps.





e. Cytoplasmic Cont Excl

Run the series of steps.

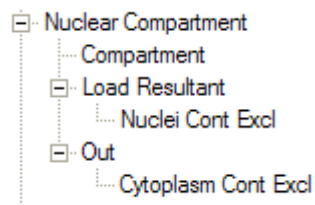


f. Nuclear compartment

Run the series of steps.

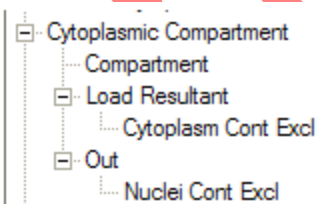
The **Out** procedure retains only those pixels from the current image that do not co-localize with pixels in the user selected **Resultant**. In this case, nuclear compartment pixels that co-localize with those of the Cytoplasmic compartment will be eliminated.

This is the area where we detect target protein pixel intensity in tumour nuclei.

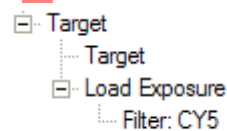


g. Cytoplasmic compartment

Run the series of steps. This is the area where we detect target protein pixel intensity in tumour cytoplasm.



h. Target



NOTE: All step titles are suggestions only and can be changed. For example, it would be perfectly acceptable to name the target step Estrogen Receptor or after any other biomarker being used.

This completes the experiment setup.



AQUA® Scoring:

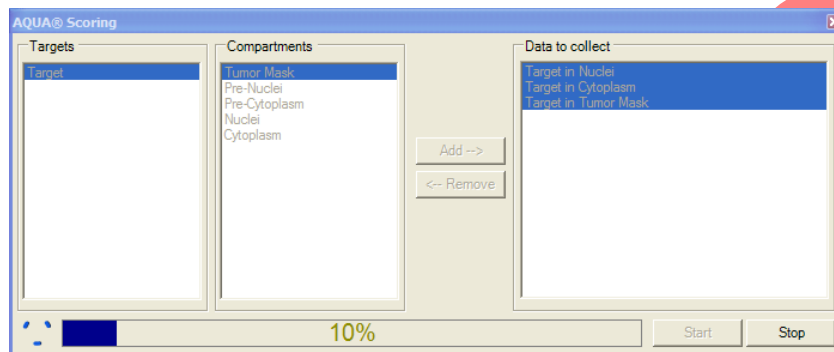
On the **Tools** menu, click **AQUA® Scoring**.

In the **Targets** pane of the **AQUA® Scoring** window, select Target.

In the **Compartments** pane of the **AQUA® Scoring** window, select the compartments in which the target will be quantified.

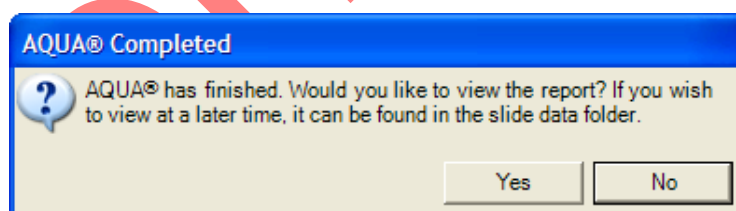
TIP: Hold down the **CTRL** button while clicking the target or compartment names to select multiple items at the same time.

Click the **Add-->** button to confirm the selection. To remove erroneous target\compartment combinations from the **Data to collect** pane, highlight the unwanted entries and click the **--Remove** button



Click **Start** to begin the scoring process.

Although the analysis time varies depending on the number of steps as well as their complexity, an average experiment will take up to 30 seconds per spot. A progress bar shows the approximate completion level.



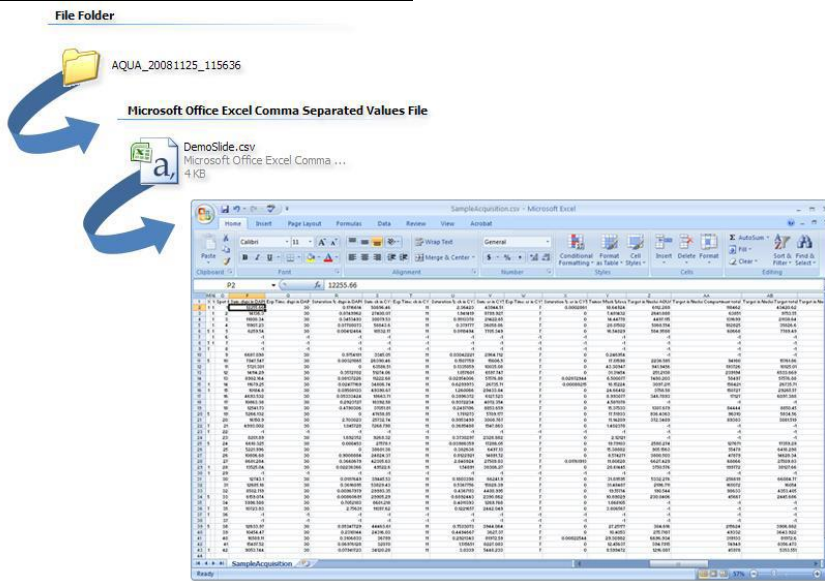
On the **AQUA® Completed** dialog, click **Yes** to view the **AQUAnalysis™ Scoring Results** report.

The report file is in the PDF format and will be displayed in the computer's default PDF viewer application.

Close the PDF viewer application when finished with the report.

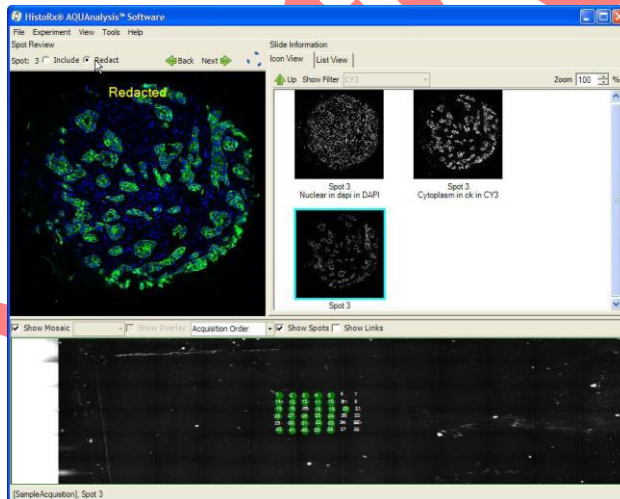
Equipment Operation Procedure

Retrieve Analysis Output File:



The analysis output data is automatically saved in the same directory as the .tma file, in a subfolder named according to **AQUA_YYYYMMDD_HHMMSS**.

The results file is in **CSV** (comma separated values) text format which can be opened by MS Excel as well as most statistical programs.



AQUAnalysis will now enter into the optional **Spot Review Mode**. In **Spot Review Mode**, the user can cycle through all analyzed images with the goal of flagging defective or faulty spots/fields-of-view.

5.0 Personal protection -
Howie coat must be worn at all times.



6.0 Training

Under no circumstances should this machine be operated by anyone who is not acquainted with it or has not been trained to use it.

7.0 Related documents –

- 7.1 Equipment manual
- 7.2 Equipment Maintenance Log
- 7.3 Equipment Maintenance Information sheet
- 7.4 Risk assessments – RA/GEN/038, RA/MH/002 and COSHH08
- 7.5 SOP - SASoM/EQUIP/024
Maintenance and replacement of EXFO lamp of the HistoRx[®] PM-2000, AQUA
- 7.6 SOP - SASoM/EQUIP/025
Image acquisition using HistoRx[®] PM-2000, AQUA

8.0 Approval and sign off –

Author:

Name: Peter Mullen
Position: Research Fellow
Signature: _____ Date: _____

Management Approval:

Name: Mary Wilson
Position: Laboratory Manager
Signature: _____ Date: _____

QA release by:

Name: Alex MacLellan
Position: QA Manager
Signature: _____ Date: _____



Document Number:	SASoM/METHOD/034.v4
Title:	Western Blot Antibody Detection Using Licor Odyssey Scanner.
Version:	V4
Author:	Peter Mullen

Effective from:	01/02/2019
Valid to:	31/01/2021

SOP History		
Number	Date	Reason for Change
V1	01/02/2013	Biennial review. Changes to sections 3, 4, 5, 6 & 8.
V2	01/02/2015	Update
V3	01/02/2017	Update
V4	01/02/2019	Update

1.0 Purpose –

This SOP describes the current procedure for performing Western Blot antibody detection using LiCor Odyssey in Laboratory 248 at the St Andrews School of Medicine (SASoM).

2.0 Scope –

This SOP applies to all staff in the SASoM involved in visualising western Blots with the Licor Odyssey scanner.

3.0 Responsibilities –

All staff involved in performing Western Blot antibody detection using LiCor Odyssey are responsible for ensuring that the methods are followed in accordance with this SOP.

All staff must have read and signed the relevant risk assessment documents before performing this procedure.

4.0 Procedure –**SDS-PAGE and Transfer onto nitrocellulose membranes.**

Prepare samples for western blot and run down an SDS-PAGE gel as per usual protocols.



Transfer samples onto nitrocellulose membrane in the usual manner (30V) overnight at 4°C.

Primary Antibody Detection

After transfer, block the membrane in Li-Cor Odyssey Blocking Buffer (diluted 50:50 in PBS) for 1 hour at room temperature.

Dilute the primary antibodies in Li-Cor Odyssey Blocking Buffer (diluted 50:50 in PBS) using dilutions comparable to those used for western blotting.

Incubate membrane in primary antibodies overnight at 4°C as per standard Western Blotting protocol.

Secondary Antibody Detection using LiCor Scanner

Make up 0.1% PBS-Tween20 (PBS-T; 1ml Tween20 /1L PBS).

Wash membrane in PBS-T at room temperature for 5 mins (x3).

Dilute fluorescently-labelled secondary antibodies at 1:10,000 dilution (1.5µL/15ml) in Odyssey Blocking Buffer (diluted 50:50 in PBS) containing 0.01% SDS in the first instance - this may require optimisation between 1:5,000 and 1:25,000 depending on the antibodies being used. Mouse-derived primary antibodies are detected using an anti-mouse fluorescently-labelled secondary antibody (either 680nm or 800nm wavelength) whilst rabbit-derived primary antibodies are detected using an anti-rabbit fluorescently-labelled secondary antibody (again of either 680nm or 800nm wavelength). By combining a mouse primary and a rabbit primary along with their respective secondary antibodies (one of 680nm and the other of 800nm), dual-labelled Westerns can be obtained.

Incubate membrane in secondary antibodies at room temperature for 45mins with gentle shaking – it is important to protect the membrane from the light until such time as it has been finally scanned.

Wash membrane in PBS-T at room temperature for 5 mins (x3), keeping the membrane in the dark.

Wash membrane in PBS at room temperature for 5 mins (x3) to remove residual Tween20, again keeping the membrane in the dark.

Lie membrane flat on a piece of filter paper in the dark and allow to air dry – allowing the membrane to dry may enhance the signal but render it useless for stripping and re-probing.

Scan the membrane on the LiCor Odyssey scanner. Keep the membrane in the dark until such time as it has been scanned.



5.0 Personal protection –

A Howie coat must be worn at all times. Gloves as specified in the appropriate COSHH RA.

6.0 Spillages –

Always clean up any spills immediately after use, only you know what you have spilt and are aware of its hazard.

Spillages should be mopped up with paper towel, disinfected with 70% ethanol and finally washed with detergent.

7.0 Training –

All staff should under go training in this technique before performing the procedure.

8.0 Related documents –

- 8.1 Risk assessments COSHH/013
General RA 49
- 8.2 SOP SASoM/METHOD/027
Polyacrylamide Gel Electrophoresis.
- 8.3 SOP SASoM/METHOD /050
Western Blotting.
- 8.4 SOP SASoM/EQUIP/037
Use and Maintenance of the Licor Odyssey.



9.0 Approval and sign off –

Author:

Name: Peter Mullen

Position: Research Fellow

Signature:

Date:

Management Approval:

Name:

Position: Laboratory Manager

Signature:

Date:

QA release by:

Name: Alex MacLellan

Position: QA Manager

Signature:

Date:

Control



Document Number:	SASoM/EQUIP/037.v2
Title:	Use and Maintenance of the Licor Odyssey Scanner
Version:	v2
Author:	Peter Mullen

Effective from:	01/01/2018
Valid to:	31/12/2022

SOP History		
Number	Date	Reason for Change
v1	01/01/2013	Original
V2	01/01/2018	Update

1.0 Purpose –

The purpose of this SOP is to outline the principles of the routine use and maintenance of the Odyssey in Laboratory 248 at the St Andrews School of Medicine (SASoM).

2.0 Scope –

This SOP applies to routine use and maintenance of the Odyssey within the SASoM.

3.0 Responsibilities –

It is the responsibility of all users of the Odyssey within the SASoM to comply with this SOP.

4.0 Procedure –

Principles of Operation:

The Li-Cor Odyssey is used to scan for immuno-reactivity by infrared imaging.

For Western blot scanning:

Turn on the instrument power switch, and turn on computer.

Open the lid of the Odyssey scanner by lifting up the front edge of the lid.

Wipe the scanning surface with soft tissue with 70% ethanol followed by distilled water.



Equipment Operation Procedure

Put the membrane on the glass plate with sample surface down and the top of the membrane facing the front of the scanner.

Flatten the membrane by placing the soft square pad on it.

Remember the area on which the membrane is according to the alignment guide at the left and bottom edge of the scanning surface.

Close the lid by pulling down gently.

UV light can harm eyes & skin. Do not try to take measurements with the sample chamber open.

Log in to the linked computer with the user's name.

Double click on the 'Odyssey V3.0' icon on the desktop to open the software.

Choose File>New to create a new project.

Enter the directory pathway and project name in the New Project dialog box.

Click the blue arrow icon on toolbar to start a scan. Both the user name and the password are 'Breakthrough'. Click 'OK'.

The scanner console window opens after log in.

Parameter settings.

Group: Breakthrough

Preset: membrane

Resolution: 169µm

Quality: medium

Focus offset: 0.0 mm

Selection of the channel(s) depends on which fluorophore is conjugated to the secondary antibodies. If IRDye 680 is used, tick the channel of 700. Tick 800 for IRDye 800CW. Tick both channels if both secondary antibodies are used. Scan area (cm): it can be defined by setting the values of 'X cord' & 'Y cord' in origin and 'width' & 'height' in Size. Or press the left mouse button and drag the cursor on the area definition grids. The settings of 'Intensity' are flexible. The bigger the values, the higher the intensity of both the bands and background. Try different values and click preview to see what values can cause best signal-to-noise ratio. Click 'start scan' once the optimal intensity values are decided.

Click 'save...' to save the scan after the scan is complete.

Enter a scan name or leave the default scan name as it is.

Enter a name for the first analysis of the images and click 'OK'. If you need to flip or rotate the images before saving them, Click 'Advanced'.



DO NOT do any modification to the folders containing the scan. If users want to analyze the images with other softwares (e.g. Microvigeine), please copy and paste the folders to another drive.

For Reverse phase Protein array (RPPA) scanning:

The operation is the same as Western Blot scanning except the parameter settings.

Parameter settings for RPPA:

Group: Breakthrough

Preset: membrane

Resolution: 21 μ m

Quality: medium

Focus offset: 0.0 mm

Selection of the channel(s) depends on which fluorophore is conjugated to the secondary antibodies. If IRDye 680 is used, tick the channel of 700. Tick 800 for IRDye 800CW. Tick both channels if both secondary antibodies are used. Scan area (cm): it can be defined by setting the values of 'X cord' & 'Y cord' in origin and 'width' & 'height' in Size. Or press the left mouse button and drag the cursor on the area definition grids. The settings of 'Intensity' are flexible. The bigger the values, the higher the intensity of both the bands and background. Try different values and click preview to see what values can cause best signal-to-noise ratio. Click 'start scan' once the optimal intensity values are decided.

For In-cell Western scanning:

The operation is the same as Western Blot scanning except on how to place the black/clear 96-well plates in step 3 and the parameter settings.

Place the 96-well plates on the scanning surface with bottom down and the edge with column number '1,2,3.....12' facing the rear of the scanner. There is no need to use the soft flattening pad.

Parameter settings for RPPA:

Group: Breakthrough

Preset: membrane

Resolution: 169 μ m

Quality: medium

Focus offset: 4.0 mm

Selection of the channel(s) depends on which fluorophore is conjugated to the secondary antibodies. If IRDye 680 is used, tick the channel of 700. Tick 800 for IRDye 800CW. Tick both channels if both secondary antibodies are used. Scan area (cm): it can be defined by setting the values of 'X cord' & 'Y cord' in origin and 'width' & 'height' in Size. Or press the left mouse button and drag the cursor on the area definition grids. The settings of 'Intensity' are flexible. The bigger the values, the higher the intensity of both the bands and background. Try different values and click preview to see what values can cause best signal-to-noise ratio. Click 'start scan' once the optimal intensity values are decided.



5.0 Personal protection -

Howie coat must be worn at all times.

6.0 General maintenance -

Clean surfaces of the apparatus with soft cloth and mild detergent.

Glass surface should be cleaned after use.

7.0 Training -

All users have to be trained before using the Instrument by a designated person.

8.0 Related documents –

- 8.1 Equipment manual
- 8.2 Equipment Maintenance Information sheet
- 8.3 Risk assessments – RA/GEN/030, COSHH/013

9.0 Approval and sign off –

Author:

Name: Peter Mullen
Position: Research Fellow
Signature: _____ Date: _____

Management Approval:

Name: Mary Wilson
Position: Laboratory Manager
Signature: _____ Date: _____

QA release by:

Name: Alex MacLellan
Position: QA Manager
Signature: _____ Date: _____

9. Publications



Validation of the Oncomine[™] focus panel for next-generation sequencing of clinical tumour samples

Hannah L. Williams^{1,2,3} · Kathy Walsh² · Austin Diamond⁴ · Anca Oniscu² · Zandra C. Deans¹

Received: 5 February 2018 / Revised: 25 May 2018 / Accepted: 9 July 2018
© The Author(s) 2018

Abstract

The clinical utility of next-generation sequencing (NGS) for a diverse range of targets is expanding, increasing the need for multiplexed analysis of both DNA and RNA. However, translation into daily use requires a rigorous and comprehensive validation strategy. The aim of this clinical validation was to assess the performance of the Ion Torrent Personal Genome Machine (IonPGM[™]) and validate the Oncomine[™] Focus DNA and RNA Fusion panels for clinical application in solid tumour testing of formalin-fixed, paraffin-embedded (FFPE) tissue. Using a mixture of routine FFPE and reference material across a variety of tissue and specimen types, we sequenced 86 and 31 samples on the Oncomine[™] Focus DNA and RNA Fusion assays, respectively. This validation considered a number of parameters including the clinical robustness of the bioinformatics pipeline for variant detection and interpretation. The Oncomine[™] Focus DNA assay had a sample and variant-based sensitivity of 99.1 and 97.1%, respectively, and an assay specificity of 100%. The Oncomine[™] Focus Fusion panel had a good sensitivity and specificity based upon the samples assessed, however requires further validation to confirm findings due to limited sample numbers. We observed a good sequencing performance based upon amplicon, gene (hotspot variants within gene) and sample specific analysis with 92% of clinical samples obtaining an average amplicon coverage above 500X. Detection of some indels was challenging for the routine IonReporter[™] workflow; however, the addition of NextGENe® software improved indel identification demonstrating the importance of both bench and bioinformatic validation. With an increasing number of clinically actionable targets requiring a variety of methodologies, NGS provides a cost-effective and time-saving methodology to assess multiple targets across different modalities. We suggest the use of multiple analysis software to ensure identification of clinically applicable variants.

Keywords Next-generation sequencing · Molecular pathology · Clinical validation · FFPE

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00428-018-2411-4>) contains supplementary material, which is available to authorized users.

✉ Hannah L. Williams
hlw37@st-andrews.ac.uk

¹ UKNEQAS for Molecular Genetics, Royal Infirmary of Edinburgh, Edinburgh, UK

² NHS Lothian, Department of Molecular Pathology, Royal Infirmary of Edinburgh, Edinburgh, UK

³ University of St Andrews, School of Medicine, North Haugh, St Andrews, Fife KY16 9TF, UK

⁴ NHS Lothian, Molecular Genetics, Western General Hospital, Edinburgh, UK

Introduction

Personalised medicine for the treatment of cancer provides directed therapy for patients based upon the genetic and epigenetic alterations of their disease. This requires laboratories to provide rapid assessment of the molecular landscape of the tumour to enable informed treatment decisions. Many of the current clinical testing algorithms are laborious, with multiple tests performed separately for a single patient. Next-generation sequencing (NGS) enables the testing of multiple genes, from multiple patient samples across dual modalities in one assay. The Ion Torrent NGS system is compatible with FFPE tissue which is currently routine processing of pathology specimens [1] and requires minimal nucleic acid input

(10 ng total of DNA or RNA) which is beneficial for testing frequently very small, diagnostic samples.

With increasing affordability enabling the implementation of NGS into clinical laboratories, validation of both assay and bioinformatic analytical pipelines are the primary challenges clinical laboratories face. To comply with ISO15189 accreditation [2], validation of the detection of somatic variants must, as a minimum, assess limit of detection, analytical sensitivity and specificity, repeatability and reproducibility and set appropriate thresholds and quality control parameters for reliable analysis of clinical specimens. Validation must also include the handling of large amounts of data produced by multi-gene panels [3].

This study presents the validation of the OncoPrint™ Focus DNA and RNA panel on the Ion Torrent Personal Genome Machine (IonPGM™, Thermo Fisher Scientific) for joint implementation within the Department of Molecular Pathology and United Kingdom National External Quality Assessment Service (UK NEQAS) for Molecular Genetics, Royal Infirmary of Edinburgh, UK. The OncoPrint™ Focus DNA and RNA assay comprises two separate panels (DNA and RNA) which were designed to interrogate hotspot mutations (35 genes), copy number variations (19 genes) and gene fusions in 23 genes. Combined, these two panels can identify current actionable genetic variants and potential future targets for personalised therapy.

Materials and methods

Sample selection

Seventy-eight anonymised FFPE tissues comprising of melanoma ($n = 18$), colorectal cancer (CRC) ($n = 28$), non-small-cell lung cancer (NSCLC) ($n = 22$) and gastrointestinal stromal tumours (GIST) ($n = 10$) were processed from a range of specimen types: resections ($n = 57$), biopsies ($n = 13$), cell blocks ($n = 6$), fine needle aspirate (FNA) ($n = 1$) and polyps ($n = 1$). Neoplastic content was assessed and ranged from 20 to 90% as determined by a pathologist. In addition, nine reference samples were tested including four commercially available standards AcroMetrix™ Oncology Hotspot Control catalog no. 969056, AcroMetrix™ Frequency Ladder (six variant allele frequencies: 2.8, 5.4, 11, 18.4, 29.5 and 47.9%), Horizon Structural Multiplex Reference Standard catalog no. HD753 and Horizon EGFR Gene-Specific Multiplex Reference Standard catalog no. HD300 and six in-house reference standards: REF 2 (68 variants), REF 3 (6 variants), REF 4 (133 variants), REF 5 (131 variants), REF 6 (9 variants) and REF 7 (9 variants) (S1 file). The limit of detection was calculated using data from the AcroMetrix™ Hotspot Frequency Ladder. RNA assay specificity and sensitivity was assessed using clinical samples, four in-house reference

standards and the ALK-RET-ROS1 Fusion FFPE RNA Reference standard (Horizon Diagnostics catalog no. HD784, RNA REF 1–4). The Human Brain Total RNA (Thermo Fisher Scientific, catalog no. AM7962) was used to assess RNA reproducibility.

Nucleic acid extraction and quantification

DNA was extracted from melanoma, CRC and GIST samples using QIAmp DNA FFPE Tissue Kit (Qiagen) following manufacturer's protocol (excluding de-paraffinisation). Dual DNA and RNA isolation was performed from NSCLC tissues using the RecoverAll Total Nucleic Acid Isolation Kit for FFPE (Thermo Fisher Scientific). DNA and RNA concentrations were determined by fluorometric quantitation using Qubit 2.0 Fluorimeter with Qubit DNA dsDNA BR Assay Kit and Qubit RNA BR Assay Kit (Qiagen) as appropriate.

Next-generation sequencing

Complementary DNA (cDNA) synthesis prior to library preparation for RNA panel was carried out using SuperScript™ VILO™ cDNA Synthesis Kit (Thermo Fisher Scientific, 11754050). Library preparation was carried out using the OncoPrint™ Assay™ (comprising the DNA OncoPrint™ Focus Assay (Thermo Fisher Scientific) and RNA OncoPrint™ Fusions assay (Thermo Fisher Scientific)) following manufacturer's instructions using a total of 10 ng input DNA and/or RNA per sample (minimum 0.83 ng/μl sample DNA concentration). A maximum of seven DNA samples were prepared per run (six samples if both DNA and RNA analyses were required) on an Ion 318™ v2 chip (Thermo Fisher Scientific, catalog no. 4488150). The DNA panel can identify hotspot mutations in the following genes: *AKT1*, *ALK*, *AR*, *BRAF*, *CDK4*, *CTNNB1*, *DDR2*, *EGFR*, *ERBB2*, *ERBB3*, *ERBB4*, *ESR1*, *FGFR2*, *FGFR3*, *GNA11*, *GNAQ*, *HRAS*, *IDH1*, *IDH2*, *JAK1*, *JAK2*, *JAK3*, *KIT*, *KRAS*, *MAP2K1*, *MAP2K2*, *MET*, *MTOR*, *NRAS*, *PDGFRA*, *PIK3CA*, *RAF1*, *RET*, *ROS1* and *SMO*; however, not all genes were assessed for the purposes of this validation. The RNA panel can identify rearrangements in *ALK*, *RET*, *ROS1*, *NTRK1*, *NTRK2*, *NTRK3*, *FGFR1*, *FGFR2*, *FGFR3*, *MET*, *BRAF*, *RAF1*, *ERG*, *ETV1*, *ETV4*, *ETV5*, *ABL1*, *AKT3*, *AXL*, *EGFR*, *ERBB2*, *PDGFRA* and *PPARG*, not all fusions were assessed for this validation. Nineteen copy number variant (CNV) targets are also included in the OncoPrint™ Focus Panel; however, these were not validated in this study. Template preparation was performed on the Ion Chef System (Thermo Fisher Scientific) using the Ion PGM Hi-Q Chef Kit and/or the Ion One Touch™ 2 System using the Ion PGM Template OT2 200 Kit. Sequencing was performed using the Ion PGM Hi-Q Sequencing Kit on the Ion Torrent Personal Genome Machine (Ion PGM).

Data analysis

Analysis was carried out using Ion Torrent Suite™ Browser version 5.0 and Ion Reporter™ version 5.0. The Torrent Suite™ Browser was used to perform initial quality control including chip loading density, median read length and number of mapped reads. The Coverage Analysis plugin was applied to all data and used to assess amplicon coverage for regions of interest. Variants were identified by Ion Reporter filter chain 5% OncoPrint™ Variants (5.0). A cut off of 500X coverage was applied to all analyses. All identified variants were checked for correct nomenclature using Alamut Visual v.2.7.1 (Interactive Biosoftware). Any discrepancies in variant identification, between Ion Reporter and Alamut, were validated manually using the Integrative Genomics Viewer [4, 5] and NextGENe® v2.4.2 (SoftGenetics®). For the purposes of this validation, amplicons covering clinically actionable regions with known mutation status (termed target amplicons; Table 1) were assessed as a subset of all amplicons (amplicons which target hot spot variants, i.e. SNVs and indels) covered in the OncoPrint™ Focus hot spot BED file.

Results

OncoPrint™ focus DNA panel

Sequencing performance

Of the 78 FFPE samples, for 4 samples (exclusively NSCLC), no amplicons were covered to 500X (minimum criteria for this validation) and were considered failed samples. All failed samples had an input DNA concentration below 2.34 ng/μl; however, not all samples below this DNA input failed sequencing. No relationship between DNA concentration and amplicon coverage was identified.

The overall panel performance was assessed by average amplicon (n110) coverage across all cases (n78). The majority (99%) of all amplicons were covered on average to a minimum of 500X. The *PIK3CA* amplicon, CHP2_PIK3CA_6, which covers nucleotides in the later portion of exon 8 was the only amplicon with an average coverage below 500X across all cases (Fig. 1a). A high variability in amplicon coverage within and between gene variants (n35) was observed across the combined study cohort. For example, intra-gene variation in *EGFR* amplicon coverage across eight amplicons ranged from *Median (Md)* 686–4853, inter-gene variation in *PIK3CA* exon 8 (*Md* 327) and *KIT* exon 11 (*Md* 4008) (Fig. 1b). A trend was observed between median amplicon coverage and amplicon length (Spearman's rho; $p = 0.072$) and between mean amplicon length and amplicon GC content (Spearman's rho; $p = 0.071$).

The average amplicon coverage per sample was also assessed; 89.7% (70/78) of all samples had an average amplicon coverage above 500X. A large proportion of samples (62/78, 79.4%) had an average amplicon coverage for the OncoPrint Focus assay between 500 and 3000X (Fig. 2). No significant association between DNA concentration, sample type or tissue type could be identified in the seven samples exceeding an average amplicon coverage of 3000X.

Sample performance using the Focus panel was assessed based upon the proportion of amplicons reaching a minimum of 500X coverage for all amplicons and target amplicons, respectively. All amplicons were covered to a minimum of 500X in 12.8% of samples, and all had an input DNA concentration above the recommended 10-ng total input (range 1.97–125.5 ng/μl). Sixty-four samples (82%) had a proportion of amplicons covered to a minimum of $\times 500$ (range 4.5–99.1%, *Md* 90%). Of the target amplicons (Table 1), 19 (24.4%) samples had all target amplicons covered to a minimum of 500X, all cases of which had an input DNA concentration above the recommended 10-ng input (range 1.29–125 ng/μl). Seventy four (94.9%) had a proportion of target amplicons covered to

Table 1 Target amplicons

Number of amplicons	Gene	Tissue type
8	<i>EGFR</i>	Lung
3	<i>KRAS</i>	
10	<i>PIK3CA</i>	
3	<i>NRAS</i>	CRC
2	<i>BRAF</i>	
5	<i>KIT</i>	Melanoma
3	<i>PDGFRA</i>	
		GIST

Table details 'target amplicons' per tissue type and number of amplicons covering each gene of interest based upon current clinical and EQA requirements within UKNEQAS and Molecular Pathology at the Royal Infirmary of Edinburgh

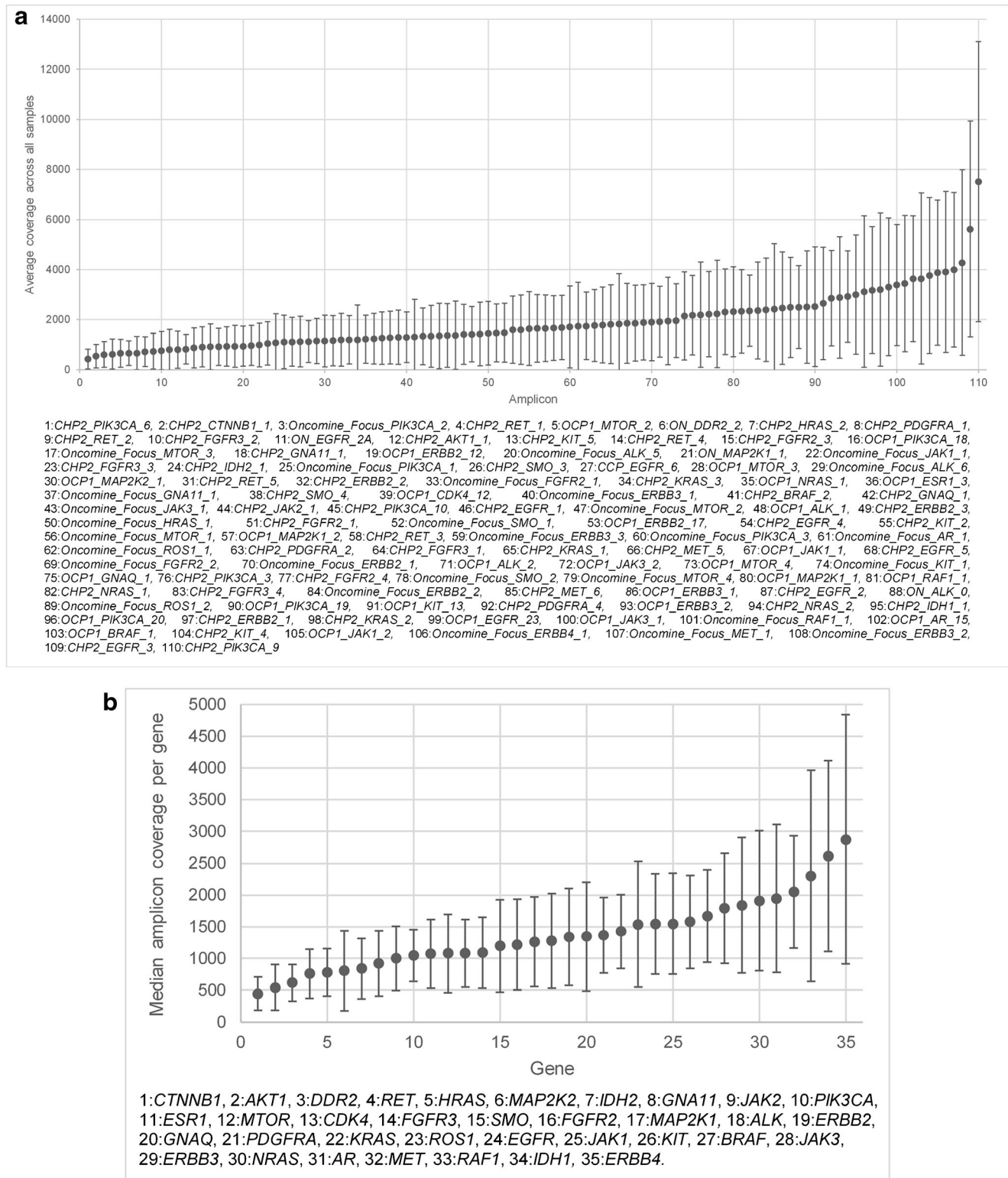


Fig. 1 Assessment of amplicon- and gene-based sequencing performance by average amplicon coverage. **a** Average amplicon coverage across all clinical samples tested (n78). Ninety-nine percent of amplicons were covered on average to a minimum of 500X¹. Average amplicon coverage was assessed for all hotspot amplicons in the OncomineTM Focus assay. **b**

Median amplicon coverage across all genes. Median coverage per gene (n35) comprising of a number of hotspot variants across exons per gene. A high variability in amplicon coverage was observed within and between genes. Intra-gene variability is depicted by interquartile range

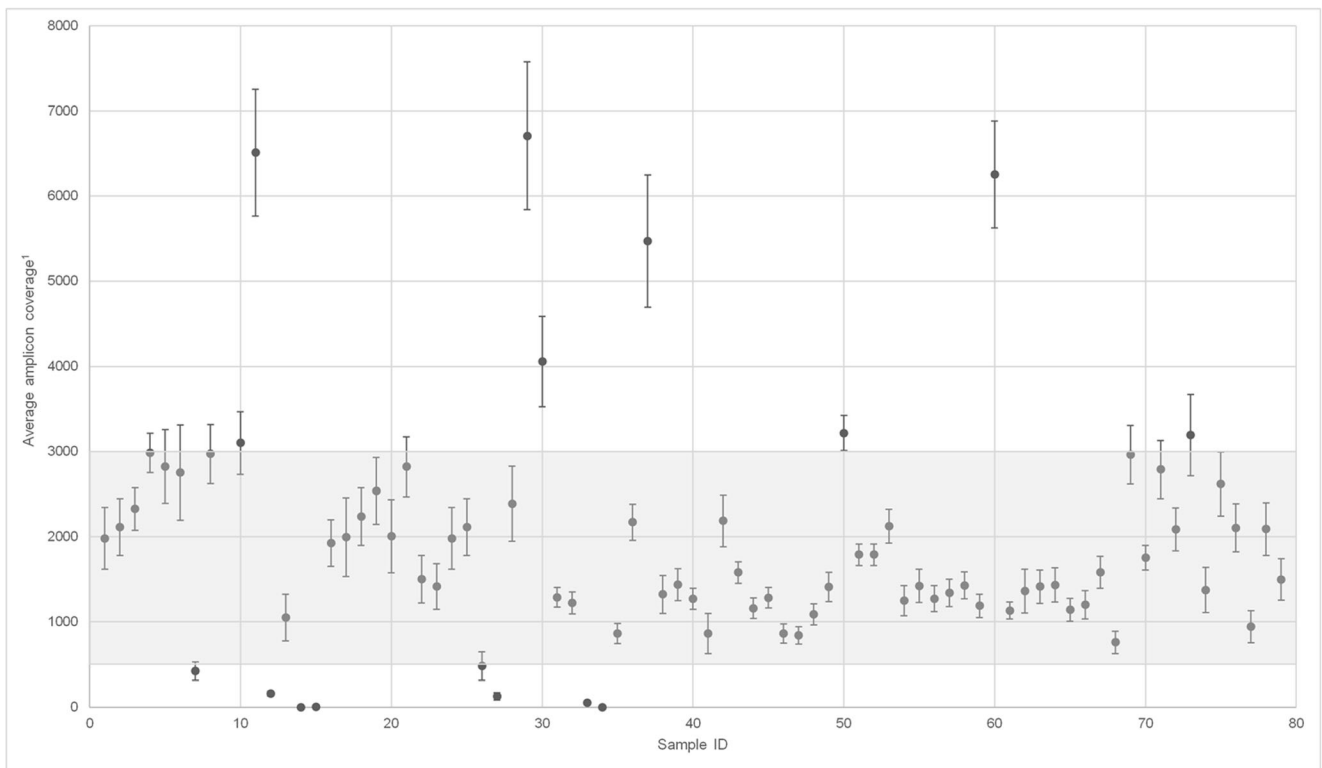


Fig. 2 Sample-based sequencing performance. Sample-based sequencing performance was assessed by the average amplicon coverage (all hotspot amplicons) for all genes (n35) for each sample. A large proportion of samples had an average amplicon coverage for all genes between 500

and 3000X. Seven samples exceeded 3000X coverage with a maximum 6707X coverage. Eight samples had average amplicon coverage below 500X with two samples failing to sequence any genes

Table 2 Sequencing performance metrics

	Total mapped reads (CV)	Average amplicon coverage depth (CV)	Percent of all amplicons \geq 500	Percent of target amplicons \geq 500
Total (n78)				
Tissue type				
Lung (n22)	390,884 (0.89)	1,545 (0.83)	61	64.3
GIST (n10)	507,137 (0.35)	1,861 (0.35)	81	88.8
Melanoma (n18)	396,993 (0.32)	1,447 (0.33)	82.6	85.6
CRC (n28)	441,153 (0.72)	1,613 (0.73)	73.7	81.7
<i>p</i> value	0.120	0.175	0.186	0.033 ^a
Specimen type				
Cell block (n6)	469,270 (1.06)	2,166 (0.8)	61.2	62.8
Biopsy (n13)	374,194 (0.79)	1,371 (0.8)	59	69.2
Resection (n57)	430,195 (0.59)	1,569 (0.59)	77.3	81.9
<i>p</i> value	0.840	0.613	0.664	0.692

Sequencing performance metrics (total mapped reads, average amplicon coverage, percentage of all amplicons \geq 500X, percentage of target amplicons \geq 500X). Metrics are presented by tissue type and specimen type. Mean values are reported. A significant difference in % target amplicons at 500X between tissue type was identified; however, adjustment for false discovery rate (FDR) using Bonferoni correction deemed this not significant (Kruskal-Wallis $p = 0.033$; $n78$). There was however a trend in NSCLC samples having a lower percentage of amplicons at \geq 500X than the other tissue types

^a*p* value significant at 0.05. For specimen type analysis 2 samples, one fine needle aspirate and one polyp were excluded from statistical analysis due to limited numbers of samples of this type. Coefficient of variation stated for total mapped reads and average amplicon coverage

a minimum of 500X (range 2.9–97.1%, *Md* 88.2%). Based upon mean sequencing performance metrics, there was no significant difference in average amplicon coverage, total mapped reads or percentage of amplicons at 500X between samples based upon tissue type or sample type (Table 2).

Tissue-specific sequencing performance was assessed by the percentage of target amplicons (specific to tissue type; Table 1) with minimum 500X depth of coverage per sample; 35 samples (44.8%) had all target amplicons covered to minimum 500X. GIST and melanoma samples had a greater proportion of samples achieving minimum 500X for all target amplicons (80 and 61.1%, respectively). Melanoma and GIST had the highest average percentage of tissue-specific target amplicons at 500X (per sample) (91.3 and 90.3%, respectively) whilst CRC and NSCLC had 78.8 and 68.4%, respectively.

Limits of detection

The limit of detection (LOD) were ascertained by repeated sequencing of the AcroMetrix™ Frequency Ladder which was analysed at the three lowest expected allele frequencies (EAFs) (2.8, 5.4 and 11%) for the presence of variants across all three repeats (S2 file).

Of the target genes assessed (Table 1), all (7) *EGFR* exon 21 variants failed to be detected at any of the three frequencies. These variants were reliably identified at 18.4% EAF; however, the observed variant allele frequency (VAF) was much lower than expected (average observed VAF 6.6%). Due to the pertinence of *EGFR* mutations in NSCLC, this was deemed an unsatisfactory LOD and required further investigation. Using the Horizon Discovery *EGFR* Gene specific multiplex reference standard (HD300) containing exon 21 variants, c.2582 T>A and c.2573 T>G at an EAF of 5%, all five variants (in addition to those above, c.2369C>T, c.2236_2250del and c.2155G>A) were detected at VAF between 4.5 and 6.2%, confirming a minimum LOD of 5% for the *EGFR* gene for these variants. LODs for *BRAF*, *EGFR*, *KRAS* and *NRAS* were 5.6% across all exons assessed and 11% for *PDGFRA* whilst *KIT* and *PIK3CA* demonstrated varying LODs depending upon exon assessed (Table 3).

Of the variants assessable in the AcroMetrix™ Frequency Ladder, all variants in only one gene (*GNA11*) could be detected at the 2.8% EAF for all variants. Ten genes (40% of total genes identified) had all variants detected across triplicates at 5.4% EAF (*ALK*, *BRAF*, *CTNNB1*, *EGFR*, *FGFR1*, *FGFR2*, *IDH1*, *KRAS*, *MET*, *NRAS* and *SMO*). In addition to these, a further eight genes (76% of total genes identified) had all variants detected across all three runs at 11% EAF (*AKT1*, *ERBB2*, *GNA11*, *HRAS*, *IDH2*, *JAK2*, *MAP2K1* and *PDGFRA*). LODs for all genes included in the AcroMetrix™ Frequency Ladder are detailed in S2 file.

In addition, during LOD analysis, we identified observed allelic frequencies with an element of positive or negative bias across repeats; e.g. observed VAFs were consistently higher than

expected in some genes (*BRAF*, *EGFR*, *KRAS*, *NRAS* and *PIK3CA*) and consistently lower than expected in others (*KIT* and *PDGFRA*).

Robustness

Library preparations and sequencing runs were performed four times using the AcroMetrix™ Oncology Hotspot Control, which contained 146 targeted variants across 25 genes (S3 table), to determine assay inter-run reproducibility. One-hundred and forty-three variants were detected across all runs.

Using Ion Reporter™ (IR™) routine workflow, three false negatives were identified (*FGFR3* c.1928A>G, *PDGFRA* c.1698_1712del15 and *IDH2* c.474A>G) being absent from 2, 2 and 1 repeat, respectively. Conferring a routine workflow reproducibility for the variants assessed of 97.9%. The FASTQ files from the four repeats were further analysed by NextGENe® software (SoftGenetics®). Using this analysis, the variants that comprised the initial three false negatives from the IR routine workflow were identified in all four repeats; however, three different false negatives were produced using this analysis (*MET* c.3757 T>G, *MET* c.3778G>T and *RET* c.1894_1906>AGCT) being absent from 1, 1 and 4 repeats, respectively (S3 file).

Intra-run repeatability was assessed by duplicate analysis of the 5.4% EAF (11/25) and 11% EAF (20/25) Acro-Metrix Hotspot Frequency Ladder samples containing 25 genes (hotspot variants within gene) comprising 140 variants; 44% (mean VAF 6.9%, CV 0.22) and 80% (mean VAF 11%, CV 0.22) repeatability was observed for all genes (hot spot variants within gene) at EAF levels 5.4 and 11%. On a variant basis for duplicate analysis of the 5.4 and 11% EAF, a repeatability of 96/140 (68.5%, mean VAF 6.6%, CV 0.22) and 132/140 (94.2%, mean VAF 10.6%, CV 0.23), respectively, was identified (S4 file).

For some variants, there was a large difference between the expected and observed VAFs demonstrating a positive or negative bias, thus preventing accurate VAFs to be derived. This was observed at both inter- and intra-gene level (Figs. 3 and 4).

Analytical sensitivity and specificity

Sensitivity assessment of the OncoPrint™ Focus DNA Panel was determined at both variant (reference samples and clinical material) and sample level (clinical material only). For the purposes of sensitivity analysis, only pre-validated variants were assessed. Any additional identified variants not previously validated were not included in analysis. A total of 588 variants (across 86 samples) were sequenced, 6 of these failed due to quality and were removed from further analysis. At the variant level, 571 of 576 known variants were correctly identified conferring an analytical variant-based sensitivity of 99.13% (95% CI 99.1–99.15%). Five false negative results were identified (Table 4).

Table 3 Limits of detection determination for target genes

Gene	Mutation CDS	Exon	Bias	Expected allele frequency			LOD (average percent of allele frequency)*	
				2.75	5.4	11		
<i>BRAF</i>	c.1799 T > A	15	1.4		•	•	5.4 (4.3–6.5)	
	c.1790 T > G				•	•		
	c.1781A > G				•	•		
	c.1742A > G				•	•		
	c.1391G > T	11	1.7		•	•		
	c.1380A > G				•	•		
	c.1359 T > C				•	•		
<i>EGFR</i>	c.323G > A	3	1.3	•	•	•	5.4 (5–5.8)	
	c.340G > A				•	•		
	c.866C > T	7	2.7		•	•		
	c.874G > T				•	•		
	c.1793G > T	15	2.4		•	•		
	c.2092G > A	18	5.4		•	•		
	c.2156G > C				•	•		
	c.2170G > A				•	•		
	c.2197C > T	19	2.9		•	•		
	c.2203G > A				•	•		
	c.2222C > T				•	•		
	c.2235_2249del15				•	•		
	c.2293G > A	20	3		•	•		
	c.2375 T > C				•	•		
	c.2485G > A	21 ^a	0.3					5
	c.2497 T > G							
	c.2504A > T							
c.2515G > A								
	c.2573 T > G							

Table 3 (continued)

	c.2582 T > A						
	c.2588G > A						
<i>KIT</i>	c.1509_1510insGCC TAT	9	-0.4	•	•	•	5.4 (5.3–5.5)
	c.1516 T > C				•	•	
	c.1526A > T				•	•	
	c.1535A > G				•	•	
	c.1588G > A	10	0.2		•	•	11 (10.5–11.5)
	c.1698C > T	1	-1.9		•	•	
	c.1727 T > C				•	•	
	c.1755C > T					•	
	c.1924A > G	13	-0.7		•	•	5.4 (5.2–5.6)
	c.1961 T > C				•	•	
c.2410C > T	17	-2.2			•	11 (10.9–11.1)	
<i>KRAS</i>	c.351A > C	4	0.7		•	•	5.4 (5.3–5.5)
	c.183A > C	3	0.5		•	•	
	c.175G > A				•	•	
	c.111 + 1C > T				•	•	
	c.104C > T				•	•	
	c.35G > A				•	•	
	c.24A > G				•	•	
<i>NRAS</i>	c.182A > G	3	3.2		•	•	5.4 (4.7–6.1)
	c.174A > G				•	•	
	c.52G > A	2	0.8		•	•	
	c.35G > A				•	•	
	c.29G > A				•	•	
<i>PDGFRA</i>	c.1743 T > C	14	-1.4			•	11 (10.9–11.1)
	c.1977C > A				•	•	
	c.2001A > G					•	
	c.2517G > T	18	-1.6			•	
	c.2525A > T				•	•	
	c.2544C > A					•	

Table 3 (continued)

	c.93A > G	2	1		•	•	5.4 (5.2–5.6)		
	c.180A > G				•	•			
	c.971C > T	5	0.8		•	•			
	c.1002C > T				•	•			
	c.1035 T > A				•	•			
	c.1258 T > C	8	2.1		•	•		11 (9.5–12.5)	
	c.1370A > G					•			
	<i>PIK3CA</i>	c.1616C > G	10	1.5		•		•	5.4 (5.2–5.6)
		c.1624G > A				•		•	
		c.1633G > A				•		•	
c.1640A > G					•	•			
c.2102A > C		14	1.2		•	•			
c.2702G > T		19	0.6	•	•	•			
c.2725 T > C					•	•			
c.3110A > G		21	1.3		•	•			
c.3140A > G				•	•				
Limit of detection analysis for target genes. <i>BRAF</i> , <i>KRAS</i> and <i>NRAS</i> were successfully identified across three repeats at 5.4% EAF. <i>PDGFRA</i> was identified across three repeats at 11% EAF. Varying exons of <i>KIT</i> and <i>PIK3CA</i> demonstrated different LODs within the same gene (5.4 and 11%). Bullet indicates variant detected at expected allele frequency. All variants listed are included in the AcroMetrix™ Frequency Ladder									
*95% confidence intervals at LOD stated									
^a EGFR exon 21 variants (c.2582 T > A and c.2573 T > G, highlighted in dark grey) were assessed using Horizon Discovery <i>EGFR</i> gene-specific multiplex reference standard (HD300). Variants listed in <i>EGFR</i> exon 21 other than those assessed in the Horizon Discovery gene specific multiplex reference standard (highlighted in light grey) were successfully repeated across three repeats at 18% EAF									

At the sample level, 71 out of 74 samples were successfully sequenced (69 gave concordant genotypes) giving an analytical sensitivity of 97.1% (95% CI 97.06–97.3%).

Variant-based assay specificity was assessed on 157 alternate variant negative genes (hot spot regions previously assessed) and conferred a 100% sensitivity with no false positives identified in hotspot regions assessed. Analytical assay specificity was determined based on the analysis of 77 FFPE samples; no false positives were detected conferring 100% analytical specificity. Overall, we identified a positive predictive value of 100% and a negative predictive value of 97.5%.

Bioinformatic performance

Of the false negatives identified in sensitivity analysis, sample 6 was previously validated using Sanger sequencing and contained a *KIT* c.1652_1663del p.Pro551_Val555delinsLeu mutation. This variant was absent in the initial output from Ion Reporter™; adjustment of analysis parameters and re-analysis resulted in successful identification of the variant, indicating that this mutation had been successfully sequenced but filtered out by the bioinformatics pipeline. The presence of the variant was confirmed by analysis of the FASTQ file with NextGENe® software (SoftGenetics®). Sample 71 that contained a second kit mutation c.1676_1694delinsA

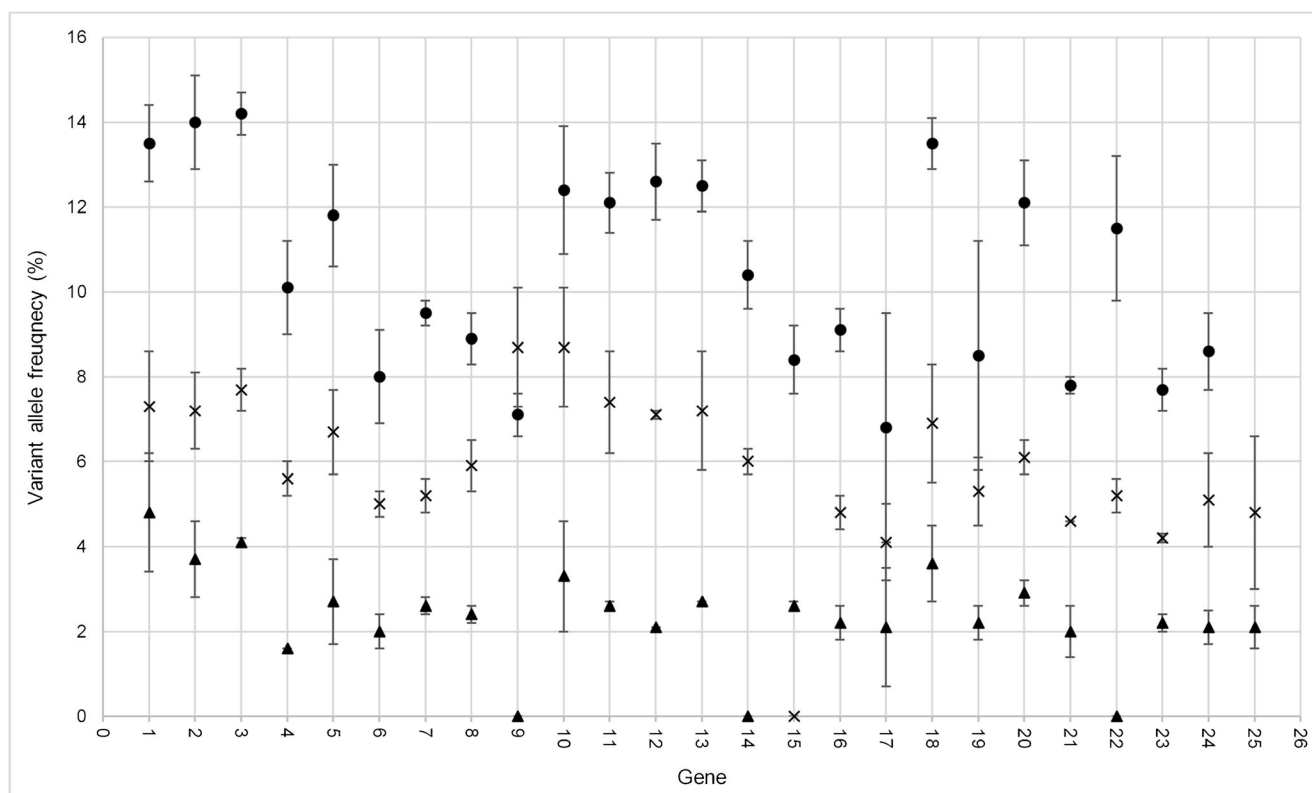


Fig. 3 Inter-gene variation in expected allele frequency (EAF). Average variant allele frequency for 25 genes represented in the AcroMatrix™ hotspot frequency ladder. Standard deviation of represented variants within each gene is depicted. 1 *NRAS*, 2 *ALK*, 3 *IDH1*, 4 *CTNNB1*, 5

6 *PIK3CA*, 6 *FGFR3*, 7 *PDGFRA*, 8 *KIT*, 9 *APC*, 10 *EGFR*, 11 *MET*, 12 *SMO*, 13 *BRAF*, 14 *FGFR1*, 15 *JAK2*, 16 *GNAQ*, 17 *RET*, 18 *FGFR2*, 19 *HRAS*, 20 *KRAS*, 21 *AKT1*, 22 *MAP2K1*, 23 *IDH2*, 24 *ERBB2*, 25 *GNAI1*. Black triangle 2.8% EAF, cross 5.4% EAF, black circle 11% EAF

p.(Val559_565delinsGlu) was not identified by either the original Ion Reporter™ algorithm or the adapted pipeline. The presence of this variant was also identified by NextGENe® software (SoftGenetics®).

Due to a number of clinical targets residing within *KIT* and *PDGFRA* genes involving indels, we sequenced a further five cases with known indels. Using the normal Ion Reporter™ workflow for OncoPrint DNA single sample analysis, 1/5 (20%) variants were identified. One additional variant was identified using a modified Ion Reporter™ workflow (soft-clipping parameters were decreased to enable greater sensitivity at ends of reads), three remained unidentified. The FASTQ files from these samples were analysed via NextGENe® software (SoftGenetics®), whereby 3/5 (60%) variants were successfully identified (Table 5). No additional false positives were identified using these workflows. Basic detection parameters, i.e. minimum SNV coverage and SNV allele frequency, were comparable between the two analysis software. No single workflow successfully identified all five variants.

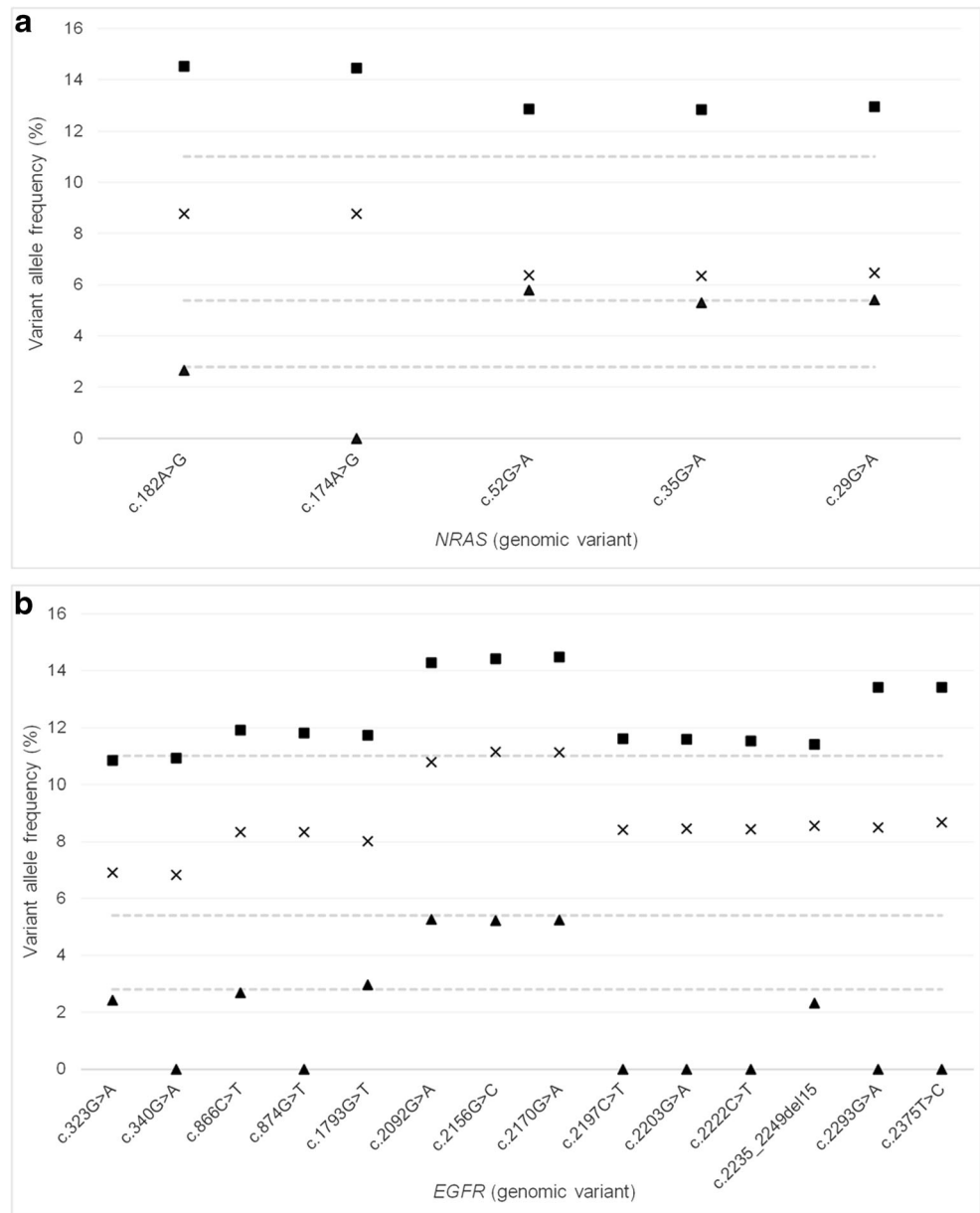
Six variants were identified with incorrect nomenclature by the Ion Report algorithm 5.0. Validation and correction of these calls was carried out using NextGENe® v2.4.2 (SoftGenetics®) and Alamut Visual v.2.7.1 (Interactive Biosoftware). A large proportion of nomenclature

inconsistencies were limited to deletions, insertions and duplications (Table 6).

OncoPrint™ RNA fusion panel

Thirty-one FFPE samples (6 biopsies, 8 cell blocks, 9 resections and 8 reference samples) previously validated by FISH were tested using the OncoPrint™ Focus RNA fusions panel. A higher sequencing failure rate was observed with the RNA panel than the DNA; eight (25.8%) cases failed sequencing. These failures were not associated with age of sample. All failed samples had a DNA concentration below 8 ng/μl; however, not all samples below this concentration failed fusion analysis. Of the 23 samples which passed quality control, all fusion positive cases ($n=6$) were correctly identified conferring an assay sensitivity of 100%. At the fusion level, all 31 fusions were identified across 23 samples conferring 100% specificity. One sample was identified as having an additional variant MET(8)–MET(9), which would not have been identified by current testing methodologies as this is not part of current clinical testing algorithms. Intra- and inter-run reproducibility was assessed using an EML4(6)–ALK(10) positive sample and demonstrated 100% concordance between and within runs. In addition, repeated analysis of the

Fig. 4 Intra-gene variation in expected allele frequency (EAF). **a** *NRAS* variants at 2.8, 5.4 and 11% EAF. **b** *EGFR* variants at 2.8, 5.4 and 11% EAF excluding exon 21 variants. A large variance is observed within genes at the 2.8% EAF; this variance decreases with increasing EAFs. A large proportion of genes demonstrate positive or negative bias from the EAF. Variation patterns observed between exons of same gene



FirstChoice Human Brain Reference Total RNA showed 100% concordance between expression control presence, imbalance assay and fusion assay calls. Further confirmation of validation parameters is required prior to the clinical implementation of the RNA Fusion assay.

Discussion

With an increasing requirement of clinical laboratories to perform multiple gene testing in both DNA and RNA, NGS panels designed for FFPE material provide a solution. Many commercially

Table 4 False negative results for variant and sample-based sensitivity

False negatives	Sample type	Sample number	Gene	Expected variant (genomic nomenclature)
Variant sensitivity	Reference	REF1	<i>PDGFRA</i>	c.1698_1712del
	Reference	REF1	<i>IDH2</i>	c.474A > G
	Reference	REF 2	<i>FGFR3</i>	c.1928A > G
Variant and sample sensitivity	Clinical	6	<i>KIT</i>	c.1652_1663del
	Clinical	71	<i>KIT</i>	c.1676_1694delinsA

Table 5 Challenging variant identification

Gene	Expected variant (genomic nomenclature)	IR™ normal workflow	IR™ deletion workflow	NextGENe®
<i>KIT</i>	c.1655_1660del		•	
	c.1728_1766dup			
	c.1726_1731dup	•	•	•
	c.1656_1676del			•
<i>PDGFRA</i>	c.2526_2537del			•

One of five variants was identified by IR™ routine workflow; an additional variant was identified by the IR™ deletion workflow. Three variants were identified by NextGENe® (SoftGenetics®) analysis, two of which had not previously been identified by either IR™ workflows

available NGS panels include >400 targets, making the task of validating a solid tumour panel for implementation in a clinical setting very challenging. In this study, we validated the Oncomine™ Focus Panel for DNA and the Oncomine™ Fusion panel for RNA application using a diverse cohort of validation material consisting of both clinical and reference material enabling comprehensive validation of both sequencing platform and bioinformatics performance.

We validated the assay on both wet bench and bioinformatics processes across a broad spectrum of validation parameters including sequencing performance, analytical sensitivity and specificity, reproducibility, repeatability, robustness and limit of detection. A number of guidelines for the application of NGS sequencing and analysis clinical testing are available; however, a general consensus as to validation size, its application in the somatic setting and how parameters are assessed remains difficult to elucidate [6, 7, 11]. This validation was conducted using samples representative of the clinical samples routinely processed on site at the Royal Infirmary of Edinburgh including NSCLC, CRC, Melanoma and GIST. The main aim of the validation was to determine the overall applicability of the Ion PGM and Oncomine™ Focus Panel for DNA and Oncomine™ Fusion Panel for RNA to routine clinical testing. The appeal of the Oncomine™ Focus panel for clinical application is the requirement of a small starting input of DNA, which is particularly applicable when tissue availability is limited for example with NSCLC specimens.

Overall, we observed good amplicon coverage and sequencing performance for the Focus panel; however, sub-optimal performance for specific cases could not be associated with either sample or tissue type and may be derived from pre-processing procedures, prior to reaching the testing facility. Inadequate sample fixation and the type of fixative used have been shown to have an impact on DNA/RNA quality and the performance of downstream molecular procedures [12, 13]. For our FFPE sample cohort, we identified a clinically suitable analytical sample sensitivity and specificity of 100%, which is comparable with studies of similar panels and platforms [14, 15]. We did identify differences in the performance of variant identification between SNVs and indels, with the detection of indels presenting a greater challenge for routine bioinformatic workflows in comparison to SNV detection; this has been previously identified in a number of studies. In order to commence integration of NGS into routine clinical testing algorithms, parallel testing using both single-gene methods and NGS for prospective cases may be implemented to further cement the findings from this initial validation.

The applicability of the Oncomine™ Focus panel was easier to assess due to the wide range of clinical material available for validation. Two of the 78 samples assessed for this validation were below the minimum input requirements for the assay of which one sample failed to sequence any amplicons. The DNA quantification kit used in this study is known to lack precision below 5 ng/μl, which comprised 32 (41%) of the total samples assessed by the DNA panel. For future assessment of DNA

Table 6 Nomenclature inconsistencies by Ion Reporter™

Gene	Expected variant	Ion Reporter™ variant
<i>KIT</i>	c.1672_1674dupAAG p.(Lys558dup)	c.1670_1671insGAA p.([Lys558dup])
	c.1679_1681delTTG p.(Val560del)	c.1675_1677delGTT p.(Val559del)
	c.1735_1737 p.(Asp579del)	c.1733_1735delATG p.(Asp579del)
	c.1730_1738del p.(Pro577_Asp579del)	c.1728_1736del p.(Pro577_Asp579del)
<i>EGFR</i>	c.2303_2311dup p.(Ser768_Asp770dup)	c.2300_2301insCAGCGTGGA p.(Ala767_Ser768insSerValAsp)
<i>BRAF</i>	c.1798_1799delGTinsAA p.(Val600Lys)	c.1798_1799delGTinsAA p.(Val600Lys) plus c.1798G > A p.(Val600Met)

Five variants were identified by the IonReporter™ analysis workflow with the incorrect nomenclature and checked by Alamut Visual v2.7.1 (Interactive Biosoftware). A large proportion of nomenclature errors were deletions, insertions and duplications

quantification, the Qubit dsDNA HS Kit (Thermo Fisher Scientific), which has been shown to be sufficient in determining sample DNA concentration for NGS, would be a more sensitive method to better assess sequencing performance in relation to DNA input [8, 16]. To explore the clinical impact minimum DNA input requirements of the assay has on the number of cases which would be applicable to this method, an audit was carried out on 865 clinical samples for which DNA concentration had been quantified previously prior to testing using current routine methods. Ninety percent of cases would be considered applicable to sequencing by the Ion PGM platform with DNA concentrations greater than the 10 ng total requirement of the assay. Of those that would be below this threshold, 65% are derived from NSCLC samples.

We successfully sequenced 94.8% of samples in our FFPE validation cohort, which is comparable with other studies [10, 17]. A number of previous studies validating NGS platforms for solid tumour application have used 500X coverage as a minimum coverage criterion, which theoretically provides sufficient coverage to detect a 2% MAF, although coverage below this can be informative when variant alleles are at a higher frequency [3, 11, 18]. A broad range in average amplicon coverage was observed at both the inter- and intra-gene level. We hypothesise that amplicons with lower average amplicon coverage could be more affected by amplification-associated issues such as comprising highly repetitive sequences resulting in reduced PCR proficiency and quality for subsequent sequencing [11].

Non-small-cell lung cancer comprises a large proportion of the clinical workload within molecular pathology at the Royal Infirmary of Edinburgh, with the requirement of the assessment of multiple targets via a variety of methods supporting the application of multi-gene panels on NGS systems. From this validation however, NSCLC would appear to be the most challenging cases within the validation cohort with respect to meeting the initial demands of DNA input through to sequencing performance. Samples from this tissue type had the lowest percentage of all amplicons and target amplicons covered to a minimum coverage of 500X with the four samples failing to sequence any amplicons being exclusive to NSCLC. Studies involving assessment of NGS application for NSCLC have reported similar findings [9]. Despite that coverage below 500X can be informative when variant alleles are at a higher frequency, the findings from this validation demonstrate a higher degree of difficulty in the identification of lower-frequency variants in those amplicons which are below 500X coverage. For example using a minimum variant coverage of 10X in order for a variant call to be made by IonReporter™, coverage of 200X would be sufficient (assuming good quality reads) of detecting an alternate variant at approximately 5% frequency. The challenges presented by this sample group in terms of meeting input requirements and deriving quality sequencing data put pressure on meeting the clinical demands of returning timely results if repeat testing is required due to failed samples.

Current methods for NSCLC testing (qPCR, pyrosequencing and FISH) enable results to be published to clinicians from the site of this validation within 5 days. To achieve a balance between cost per sample and time to results, turnaround times would be required to be increased from 5 to a minimum of 10 days. In order to meet testing demands, urgent requests will continue to be assessed using single-gene tests. Based upon the reduced sequencing performance of NSCLC in this validation, further validation to identify an accurate optimal threshold of sample quality prior to input will be required in order to triage samples more likely to fail and test these using single-gene test methods. NSCLC samples showed a high failure rate when assessed using the Fusions RNA panel which we hypothesise may be attributed to a higher proportion of larger samples in our validation cohort. In addition, this may also be as a result of sample processing methods prior to molecular testing such as length and extent of fixation of sample [19]. Obtaining clinically relevant ALK-positive material for validation is challenging; for example, out of 82 NSCLC cases tested for ALK rearrangements within Molecular Pathology in 2017, approximately 1% would test positive for ALK rearrangements. In addition to this, ROS testing is not currently carried out within Molecular Pathology making the identification of ROS positive cases for validation even more difficult. The limitations in the availability of clinical FFPE material to validate the panel further demonstrate the challenges in validating the RNA fusions panels for clinical application. In comparison to NGS fusion analysis, current methods for detecting ALK rearrangements in NSCLC, i.e. fluorescent in situ hybridisation (FISH), are relatively quick and cost-effective; based upon this, the RNA Fusion panel is not currently implementable as a clinical assay for assessment of ALK rearrangements.

In addition to the issues identified with sequencing performance of NSCLC specimens, we also identified issues in determining LODs for exon 21 of *EGFR* when using the AcroMetrix™ Oncology Hotspot Control. This exon is pertinent to NSCLC as it is required for the assessment of patient suitability for treatment with tyrosine kinase inhibitors such as erlotinib (Tarceva®). Detection of *EGFR* exon 21 variants using the AcroMetrix™ Hotspot Frequency Ladder gave surprisingly high LODs, comparable to those produced by Sanger sequencing. To further explore the LODs of this exon using an additional reference standard, we confirmed that the clinically required variants were detectable at 5% VAF. The challenges faced with ascertaining LODs in this validation study highlight the importance of using multiple types of reference material to gauge LODs on a per exon basis.

The use of the AcroMetrix™ Hotspot Frequency Ladder enabled us to assess limits of detection across a broad range of genes and variants in one sample which would otherwise be a costly and time-consuming process. Our assessment of LODs using this reference standard demonstrated inter- and intra-gene variability from the EAFs. This highlights the unsuitability of this

platform for the accurate reporting of VAFs and the importance of validating LODs on a wide range of variants if clinical targets span a number of amplicons as LODs may differ by exon [20, 21]. Variation in the observed VAF and EAF was identified, which varied depending upon the gene and exon assessed. In addition, during LOD analysis, we identified observed allelic frequencies with an element of positive or negative bias across repeats; e.g. observed VAFs were consistently higher than expected in some genes and consistently lower than expected in others. We hypothesise that this could be caused by a number of factors including library preparation and sequencing, location of targets within the gene or the surrounding context of the content to be sequenced, i.e. large homopolymer regions. The quality of DNA input into the assay may have a large impact on the allelic frequencies observed following sequencing due to the nature of AmpliSeq technology. We are unable to determine the extent of duplicate reads in our final product prior to sequencing using our current protocols and are therefore unable to deduce whether this has impacted the observed 'bias'. Improvements in DNA quantitation using more sensitive methods as mentioned and assessment of DNA quality by methods such as the ProNex® DNA QC Assay (Promega, NG1002) prior to library preparation would enable control of DNA input quality and to triage samples most applicable to this analytical procedure. In addition, more recent improvements in AmpliSeq panels have resulted in the incorporation of tag-based sequencing in which DNA is barcoded prior to PCR, enabling the identification of duplicate reads. The incorporation of this into the current workflow may negate the current inaccuracies in extrapolating VAFs from this assay.

During the validation process, the need to validate bioinformatic pipelines using multiple software providers became apparent. Despite a large proportion of indels being identified by the IonReporter™ routine analysis algorithm, we did identify issues in the routine Ion Reporter™ analysis algorithm for the detection of some indels, a type of variant known to present a challenge for NGS analysis [22]. Both false negatives identified in sample sensitivity analysis were indels in *KIT*, which failed to be identified by the routine Ion Reporter™ analysis workflow. Adjustment of analysis parameters, namely soft-clipping (the indel was located at the end of amplicon), enabled the successful detection of one indel by the Ion Reporter analysis software. FASTQ files from both indels were further analysed using NextGENe® software (SoftGenetics®) and were successfully identified. Further investigation into indel identification demonstrated that neither Ion Reporter™ nor NextGENe® was 100% successful in identification of five indels we ran. Further validation of this bioinformatics workflow, i.e. IonReporter™ routine workflow followed by NextGENe® workflow with the same detection parameters on prospective samples, will be required to ensure suitability of the workflow in identification of clinically applicable variants. We believe that this demonstrates the

importance of robust and appropriate validation of the bioinformatics pipeline for clinical application and the use of multiple analysis software to ensure detection of all types of variants. We also noted a number of incorrect nomenclature calls on identified variants. We suggest that reporting of NGS-derived results should be made by individuals experienced in the platform, bioinformatics and clinical application of data derived.

In conclusion, with an increasing number of clinically actionable targets requiring a variety of methodologies, an NGS test becomes the more viable option in terms of cost, time and availability of material. For example, within our clinical setting, NSCLC samples now require a plethora of testing across multiple modalities: *ALK* IHC, *ALK* FISH, *ROS1* IHC, *ROS1* FISH, *PDL1* IHC and PCR for *EGFR* and *KRAS* from normally small biopsies with limited material available. NGS enables the assessment of multiple targets using limited input material. The challenge clinical laboratories face is in how much future proofing is appropriate. Here, we have demonstrated that a balance is required between testing current clinically relevant targets and ensuring additional targets which do not currently have clinical utility are a suitable trade-off for sequencing space. It is important to take into account the costly and time-consuming validation/verification process following assay changes when deciding on the size of panel to be implemented into clinical practice.

Acknowledgements The authors would like to thank Dr. Jennifer Fairley for her support in delivering this validation study.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Wilkins B (2015) The retention and storage of pathological records and specimens, 5th edn
2. International Organisation for Standardization (2012) ISO 15189: 2012 Medical Laboratories – Requirements for quality and competence
3. Deans ZC, Costa JL, Cree I, Dequeker E, Edsjo A, Henderson S et al (2017) Integration of next-generation sequencing in clinical diagnostic molecular pathology laboratories for analysis of solid tumours; an expert opinion on behalf of IQN path ASBL. *Virchows Arch* 470(1):5–20

4. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178–192. <https://doi.org/10.1093/bib/bbs017>
5. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G et al (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24–26. <https://doi.org/10.1038/nbt.1754>
6. Deans ZC, Watson CC, Charlton R, Ellard S, Wallis Y, Mattocks C, Abbs S (2015) Practice guidelines for targeted next generation sequencing analysis and interpretation
7. Hall A (2014) Guidelines for diagnostic next generation sequencing. 1–5
8. Robin JD, Ludlow AT, LaRanger R, Wright WE, Shay JW (2016) Comparison of DNA quantification methods for next generation sequencing. *Sci Rep* 6:24067. <https://doi.org/10.1038/srep24067>
9. Hagemann IS, Devarakonda S, Lockwood CM, Spencer DH (2015) Clinical next-generation sequencing in patients with non – small cell lung cancer. 631–9
10. Hadd AG, Houghton J, Choudhary A, Sah S, Chen L, Marko AC et al (2013) Targeted , high-depth , next-generation sequencing of Cancer genes in formalin-fixed , Paraffin-embedded and fine-needle aspiration tumor specimens. *J Mol Diagn* 15(2):234–247. <https://doi.org/10.1016/j.jmoldx.2012.11.006>
11. Jennings LJ, Arcila ME, Corless C, Kamel-reid S, Lubin IM, Pfeifer J et al (2017) Guidelines for validation of next-generation sequencing based oncology panels a joint consensus recommendation of the Association for Molecular Pathology and College of American pathologists. *J Mol Diagn* 19(3):341–365. <https://doi.org/10.1016/j.jmoldx.2017.01.011>
12. Srinivasan M, Sedmak D, Jewell S (2002) Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am J Pathol* 161(6):1961–1971 Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1850907/>
13. Do H, Dobrovic A (2015) Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. *Clin Chem* 61(1):64 LP–64 71 Available from: <http://clinchem.aaccjnl.org/content/61/1/64.abstract>
14. Lih C-J, Harrington RD, Sims DJ, Harper KN, Bouk CH, Datta V et al (2017) Analytical validation of the next-generation sequencing assay for a Nationwide signal-finding clinical trial: molecular analysis for therapy choice clinical trial. *J Mol Diagn* 19(2):313–327
15. Tumors S, Hovelson DH, Mcdaniel AS, Cani AK, Johnson B, Rhodes K (2015) Development and validation of a scalable next-generation sequencing system for assessing relevant somatic variants in solid tumors. *Neoplasia* 17(4):385–399. <https://doi.org/10.1016/j.neo.2015.03.004>
16. Simbolo M, Gottardi M, Corbo V, Fassan M, Mafficini A, Malpeli G et al (2013) DNA qualification workflow for next generation sequencing of histopathological samples. *PLoS One* 8(6):e62692. <https://doi.org/10.1371/journal.pone.0062692>
17. Reiman A, Kikuchi H, Scocchia D, Smith P, Tsang YW, Snead D, et al. (2017) Validation of an NGS mutation detection panel for melanoma. 1–7
18. D’Haene N, Le Mercier M, De Neve N, Blanchard O, Delaunoy M, El Housni H et al (2015) Clinical validation of targeted next generation sequencing for Colon and Lung cancers. *PLoS One* 10(9):e0138245
19. Greer CE, Peterson SL, Kiviat NB, Manos MM (1991) PCR amplification from paraffin-embedded tissues: effects of fixative and fixation time. *Am J Clin Pathol* 95(2):117–124. <https://doi.org/10.1093/ajcp/95.2.117>
20. Bragg LM, Stone G, Butler MK, Hugenholtz P, Tyson GW (2013) Shining a light on dark sequencing: characterising errors in ion torrent PGM data. *PLoS Comput Biol* 9(4):e1003031
21. Yeo ZX, Wong JCL, Rozen SG, Lee ASG (2014) Evaluation and optimisation of indel detection workflows for ion torrent sequencing of the BRCA1 and BRCA2 genes. *BMC Genomics* 15(1):516. <https://doi.org/10.1186/1471-2164-15-516>
22. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J et al (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 30(5):434–439

Review of the implementation of plasma ctDNA testing on behalf of IQN Path ASBL: a perspective from an EQA providers' survey

Zandra C. Deans¹ · Hannah Williams¹ · Elisabeth M. C. Dequeker² · Cleo Keppens³ · Nicola Normanno⁴ · Ed Schuurin⁵ · Simon J. Patton⁶ · Melanie Cheetham⁶ · Rachel Butler⁷ · Jacqueline A. Hall^{8,9} · On behalf of IQN Path ASBL

Received: 18 April 2017 / Revised: 6 July 2017 / Accepted: 14 August 2017 / Published online: 25 August 2017
© The Author(s) 2017. This article is an open access publication

Introduction

Molecular biomarker analysis for the personalised treatment of non-small cell lung cancer (NSCLC) and colorectal cancer (CRC) is becoming more common, due to the number and availability of molecular targets for predictive biomarker testing increasing [1]. Clinical laboratories must implement accurate test procedures and provide timely and reliable test results, to ensure that appropriate therapies are administered to patients [2]. The challenge for laboratories is to keep pace with molecular biomarker developments while maintaining excellence in service standards.

Plasma circulating tumour DNA (ctDNA) may be found in the blood of cancer patients, alongside a larger fraction of circulating free DNA (cfDNA). Plasma ctDNA testing is becoming more common in the management of cancer patients [3]. It has

several advantages: in the absence of suitable or sufficient tissue biopsy, it yields material for molecular analysis, can demonstrate molecular resistance to targeted treatment and is an alternative to invasive tissue sampling [4]. Plasma ctDNA analysis may also prove useful in cases of intra- and inter-tumour heterogeneity [5]. With formal approval from the European Medicine Agency (EMA), several clinical applications for plasma ctDNA testing are now being considered, including the detection of Epidermal Growth Factor Receptor (*EGFR*) mutations in the plasma of patients with advanced NSCLC [6].

The implementation of new methods such as plasma ctDNA testing can be challenging for diagnostic laboratories. Indeed, it has been shown that inexperience in specialised and complex techniques can compromise the result quality [2, 7]. To address these issues, four EQA providers came together under the umbrella organisation the International Quality

This article is part of the Topical Collection on *Quality in Pathology*

Electronic supplementary material The online version of this article (<http://doi.org/10.1007/s00428-017-2222-z>) contains supplementary material, which is available to authorized users.

✉ Zandra C. Deans
sandi.deans@ed.ac.uk

¹ UK NEQAS for Molecular Genetics, Department of Laboratory Medicine, Royal Infirmary of Edinburgh, Little France Crescent, Edinburgh EH16 4SA, UK

² Biomedical Quality Assurance Research Unit, Department of Public Health and Primary Care, KU Leuven / University Hospital of Leuven, Leuven, Belgium

³ Biomedical Quality Assurance Research Unit, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

⁴ Cell Biology and Biotherapy Unit, Istituto Nazionale Tumori “Fondazione Giovanni Pascale”, IRCCS, Naples, Italy

⁵ Department of Pathology, University of Groningen, University Medical Center of Groningen, Groningen, The Netherlands

⁶ European Molecular Genetics Quality Network, Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester M13 9WL, UK

⁷ All Wales Genetic Laboratory, Institute of Medical Genetics, University Hospital of Wales, Heath Park, Cardiff CF14 4XW, UK

⁸ International Quality Network for Pathology (IQN Path ASBL), 3A Sentier de l'Esperance, L-1474 Luxembourg, Luxembourg

⁹ Division of Cancer, Department of Surgery and Cancer, Imperial College London, London, UK

Network for Pathology IQN Path (IQN Path): Association Italiana di Oncologia Medica (AIOM), European Molecular Genetics Quality Network (EMQN), European Society of Pathology (ESP) EQA and the United Kingdom National External Quality Assessment Service (UK NEQAS) for Molecular Genetics. Their aim was to survey testing methods currently in use and to pilot an EQA which assessed the standards of plasma ctDNA testing. This article summarises the results of the survey, which evaluated current laboratory practices in this field and which will subsequently inform the design of a pilot EQA scheme for plasma ctDNA testing.

Methods

An online survey of plasma ctDNA testing practice was designed by the IQN Path collaborative group. The survey was circulated by the four EQA members to their global network of participants, EMQN (1480), UK NEQAS (500), AIOM (47) and ESP testing schemes (568). The survey comprised six sections which included questions about laboratory participation in EQA for solid tumour testing in NSCLC and CRC, current experience and technologies used in plasma ctDNA testing and any analytical limitations of current test methodologies. The survey opened for completion between February 2016 and the middle of March 2016. The responses were analysed to understand current practices in the field of *EGFR* and *RAS* mutation testing using ctDNA and will inform design of future pilot EQA scheme.

Results

Completed surveys were received from 167 laboratories. The submitted data was collated and summarised.

The survey showed that some form of ctDNA plasma testing for *EGFR*, *KRAS* and *NRAS* was used in the majority of responding laboratories (151/167, 90%) but that only 62 (37%) laboratories currently perform diagnostic plasma ctDNA testing (Fig. 1). A further 56 laboratories (34%) have plasma ctDNA test methodologies in the development phase (Fig. 1). During 2015, 46 diagnostic laboratories tested fewer than 100 samples, while 9 tested more than 101 samples (4 did not respond).

The most frequently used method for plasma ctDNA testing was next-generation sequencing (NGS), used by 27% of laboratories. The most commonly used testing platform was Ion PGM System®/Thermo Fisher Scientific (Fig. 2a). After NGS, the most frequently used methods were Roche cobas®, Qiagen therascreen® and ddPCR (Fig. 2a). Of the ddPCR assays, BioRad's QX 200 ddPCR assays were the most commonly used (Fig. 2a).

Of the laboratories using more than one plasma ctDNA testing method, 90 (54%) employ a single method, 51 (31%) use

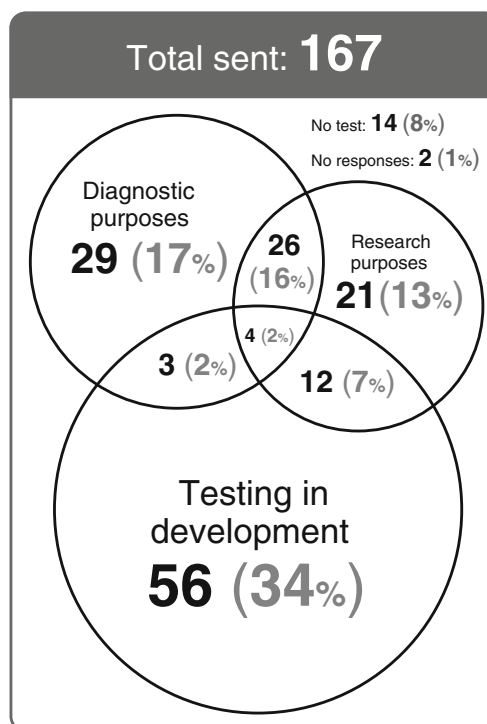


Fig. 1 Summary of the survey results of *EGFR*, *KRAS* and *NRAS* mutation testing in plasma samples. Figures reported are based on the number of laboratories offering testing for either research or diagnostic use or have testing in the development phase. Note: some laboratories use testing for more than one purpose as may be seen by the intersections in the Venn diagram

two and 8 (5%) use three different methods. The remaining 18 did not provide methodology information. Figure 2b illustrates the diversity of methods currently employed in plasma ctDNA testing.

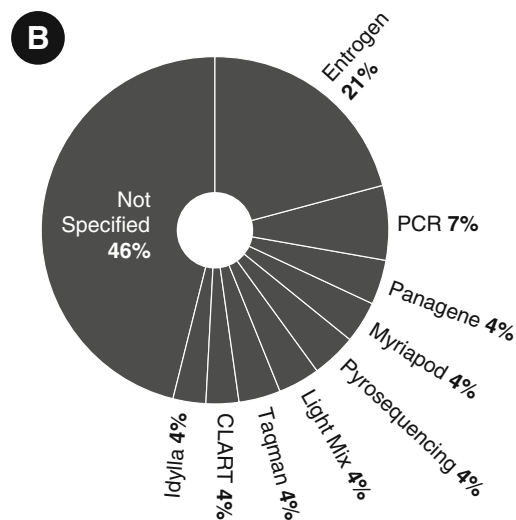
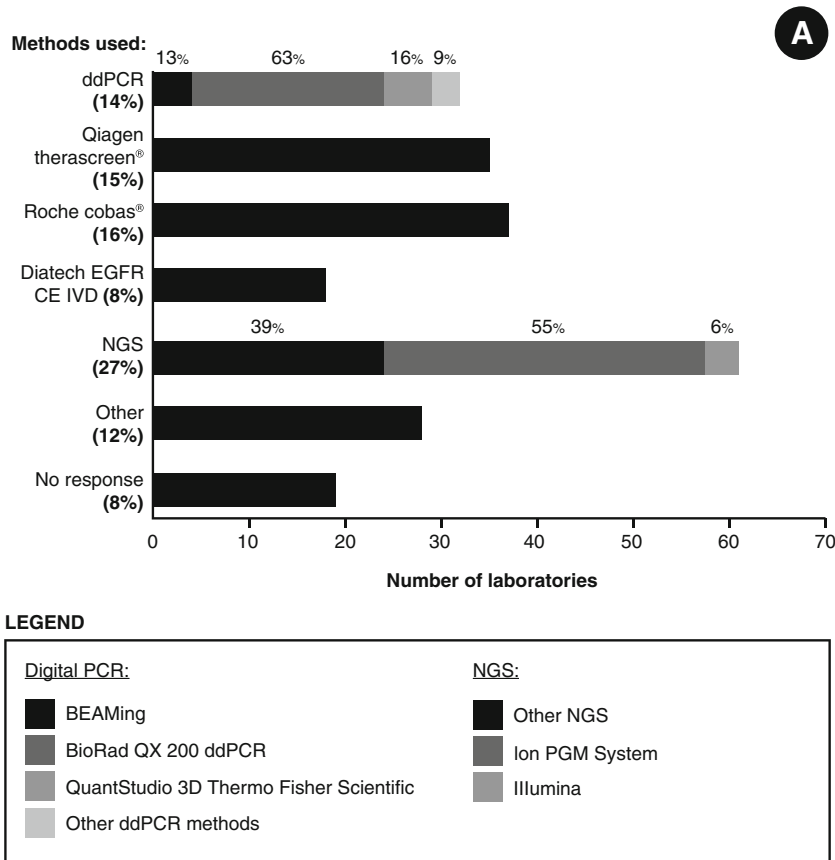
The stated limit of detection (LoD) of all allele frequencies for all laboratories was below 20%. The LoD was 5–20% in 13 laboratories (7.8%), < 5% in 56 laboratories (33.5%) and < 1% in 62 laboratories (37.1%). The remaining laboratories did not provide any LoD data (21.6%). Most laboratories performing NGS (97.8%) provided an LoD level of < 5% (44.2%), < 1% (53.8%) and 1% (< 5%). The other laboratories stated an LoD level of > 20% or provided no data. For ddPCR, 65.6% of laboratories gave an LoD of < 1 and 21.8% stated an LoD between 1 and < 5%. A single laboratory stated an LoD of < 10% and the remaining 3 offered no data.

The *EGFR*, *KRAS* and *NRAS* mutations targeted for analysis in plasma samples were collected. The three most common targets for each gene are outlined in Table 1.

Discussion

Current practice for plasma ctDNA testing in CRC and NSCLC tumour diagnostic testing was examined by the survey. The laboratories offering plasma ctDNA testing were

Fig. 2 The breakdown of methodologies used for mutation testing in plasma samples. Values represent the number of laboratories running a specific platform. Some laboratories use more than one testing method. **a** The current methods used for plasma ctDNA mutation testing by number of laboratories (% included). **b** A detailed breakdown of methods specified as “other testing methods”



those which already provide tissue-based molecular pathology services. Most survey participants (86.8%) offered diagnostic testing of *EGFR*, *KRAS* and *NRAS* from formalin-fixed paraffin-embedded (FFPE) tissue and participated in EQA for solid tumours. Of these, 78% participated in EQA assessment for NSCLC and 68% for CRC, so these laboratories already have experience in molecular technologies and understand how to interpret and report results.

The data suggest that no single, definitive technology for the analysis of plasma ctDNA has yet emerged. The methods currently used are a mixture of commercial and locally developed assays. These assays must be optimised and validated: they must also support adequate test sensitivity and available starting material and must also cover the range of molecular targets that require analysis.

Table 1 Tabulated frequency of the genes and variants analysed by laboratories for ctDNA testing

Gene	Target	Number of laboratory responses
<i>EGFR</i>	All variants within specified exons	61
	p.(G719A), p.(G719C), p.(G719S), p.(S768I), p.(T790M), p.(L858R), p.(L861Q), deletions in exon 19 and insertions in exon 20	28
	p.(T790M), p.(L858R) and deletions in exon 19	27
<i>KRAS</i>	All variants within specified exons	62
	p.(G12D), p.(G12R), p.(G12A), p.(G12C), p.(G12S), p.(G12V), p.(G13D) and p.(Q61H)	13
	p.(G12D), p.(G12R), p.(G12A), p.(G12C), p.(G12S), p.(G12V) and p.(G13D)	11
<i>NRAS</i>	All variants within specified exons	57
	p.(G12D), p.(Q61K), p.(Q61R), p.(Q61L) and p.(Q61H)	10
	All variants in codons 12, 13, 59, 117 and 146	4

Plasma ctDNA is present at low quantities, mixed within circulating free DNA (cfDNA) in the blood. Therefore, in order to have confidence in the results of plasma ctDNA testing, attention must be paid to the assay sensitivity. However, the optimal sensitivity for ctDNA testing is not yet clear.

An example of a specific clinical application of plasma ctDNA is the phase IV *EGFR* tyrosine kinase inhibitor Gefitinib ‘Follow Up Measure’ trial that facilitated the approval of plasma ctDNA testing for *EGFR* in NSCLC patients (6). The trial showed that although the Qiagen therascreen® kit had a low sensitivity (65.7%), it had a good correlation with the response of patients to first-line treatment with Gefitinib [4].

In the future, more sensitive techniques are likely to detect more patients with *EGFR* mutations and may also identify *EGFR* variants in patients with heterogeneous expression [8].

The use of highly sensitive ctDNA testing methods has permitted new insights into heterogeneity, e.g. *p.T790M* mutations in the ctDNA of patients with a tumour mass that tested negative for the resistance mutation [9]. Patients with positive plasma ctDNA tests and negative tissue results had shorter progression-free survival compared to patients with *EGFR p.T790M* detected in both their tumour tissue and plasma ctDNA [9]. In the future, accurate measurements of the ratio of resistant *EGFR* mutations to sensitising mutations might help select patients who are more likely to benefit from treatment with drugs targeting *p.T790M* [10].

Similarly, NGS panels may be used to determine the relative abundance of a tumour variant to support individually tailored therapy. Where NGS is used, a broad range of

molecular targets may be detected simultaneously; however, this may be at the cost of lower test LoD. Non-NGS-based methods may provide greater sensitivity but have the limitation of assaying fewer molecular targets.

There is a significant interest in the development of plasma ctDNA services. However, despite the 2014 EMA approval for plasma ctDNA biopsies which determine the suitability of first-line treatment of NSCLC with Gefitinib, few laboratories currently deliver NSCLC or CRC clinical diagnostic services [11]. For plasma ctDNA testing to become integrated into routine practice, those offering clinical services must be educated on its applications. However, until local services can validate and embed testing in patient pathways, laboratory and clinical uptake of plasma ctDNA may be hindered.

Current clinical applications for ctDNA are largely confined to NSCLC and CRC, although there is potential for its use in many other areas of oncology. Laboratories must provide high-quality testing services in which clinical teams and patients have confidence. The delivery of National and International EQA schemes is essential to maintain quality through the standardisation of sample logistics, molecular assays and result interpretation, as well as playing an important role in supporting education [4].

As many laboratories plan to implement the testing of plasma ctDNA, it is clear that support from well-designed EQA schemes is needed. Without this support, laboratories may be slower to offer plasma ctDNA clinical services or may encounter issues. Surveying current practices and collecting data to inform EQA design is a task that may be harmonised between several EQA providers, all with the aim of increasing efficiencies and supporting best practice standards in quality assessment [12].

Acknowledgements We would like to thank the sponsors who supported the plasma ctDNA pilot of IQN Path, the initial step of which is presented here by the survey results. Sponsors include Amgen, AstraZeneca, Merck KGA, Sysmex Inostics, Boehringer Ingelheim, Roche and Thermo Fisher Scientific.

Funding IQN Path provided administrative support for the project.

Compliance with ethical standards This research does not contain any studies involving human participants or animals performed by any of the authors. Human subjects were not used; therefore, formal consent was not required.

Conflict of interest Zandra C Deans, Hannah Williams, Cleo Keppens, Simon J Patton, Melanie Cheetham and Rachel Butler declare that they have no conflict of interest. Elisabeth MC Dequeker receives grants from Roche and Pfizer. Nicola Normanno receives grants and personal fees from AstraZeneca, Qiagen, Roche, Merck Serono, Sysmex. Ed Schuurings is on the Advisory Board and receives personal fees and travel expenses from Novartis, Pfizer and AstraZeneca and receives travel fees only for participation in an Advisory Board for Amgen. He is also on an advisory board for QCMD and receives a grant for laboratory costs. Jacqueline Hall owns stock in Vivactiv Ltd.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References


1. Cree IA, Deans Z, Ligtenberg MJ, Normanno N, Edsjo A, Rouleau E, Sole F, Thunnissen E, Timens W, Schuurin E, Dequeker E, Murray S, Dietel M, Groenen P, Van Krieken JH, European Society of Pathology Task Force on Quality Assurance in Molecular Pathology, Royal College of Pathologists (2014) Guidance for laboratories performing molecular pathology for cancer patients. *J Clin Pathol* 67:923–931. <https://doi.org/10.1136/jclinpath-2014-202404>
2. Dubbink HJ, Deans ZC, Tops BB, van Kemenade FJ, Koljenovic S, van Krieken HJ, Blokx WA, Dinjens WN, Groenen PJ (2014) Next generation diagnostic molecular pathology: critical appraisal of quality assurance in Europe. *Mol Oncol* 8:830–839. <https://doi.org/10.1016/j.molonc.2014.03.00410.1016/j.molonc.2014.03.004>
3. Rolfo C, Castiglia M, Hong D, Alessandro R, Mertens I, Baggerman G, Zwaenepoel K, Gil-Bazo I, Passiglia F, Carreca AP, Taverna S, Vento R, Peeters M, Russo A, Pauwels P (2014) Liquid biopsies in lung cancer: the new ambrosia of researchers. *Biochim Biophys Acta* 1846:539–546. <https://doi.org/10.1016/j.bbcan.2014.10.001>
4. Douillard JY, Ostoros G, Cobo M, Ciuleanu T, Cole R, McWalter G, Walker J, Dearden S, Webster A, Milenkova T, McCormack R (2014) Gefitinib treatment in EGFR mutated caucasian NSCLC: circulating-free tumour DNA as a surrogate for determination of EGFR status. *J Thorac Oncol* 9:1345–1353. <https://doi.org/10.1097/JTO.0000000000000263>
5. Normanno N, Rachiglio AM, Roma C, Fenizia F, Esposito C, Pasquale R, La Porta ML, Iannaccone A, Micheli F, Santangelo M, Bergantino F, Costantini S, De Luca A (2013) Molecular diagnostics and personalised medicine in oncology: challenges and opportunities. *J Cell Biochem* 114:514–524. <https://doi.org/10.1002/jcb.24401>
6. Douillard JY, Ostoros G, Cobo M, Ciuleanu T, McCormack R, Webster A, Milenkova T (2014) First-line gefitinib in Caucasian EGFR mutation-positive NSCLC patients: a phase-IV, open-label, single-arm study. *Br J Cancer* 110:55–62. <https://doi.org/10.1038/bjc.2013.721>
7. Tack V, Ligtenberg MJ, Tembuysen L, Normanno N, Vander Borgh S, Han van Krieken J, Dequeker EM (2015) External quality assessment unravels interlaboratory differences in quality of RAS testing for anti-EGFR therapy in colorectal cancer. *Oncologist* 20:257–262. <https://doi.org/10.1634/theoncologist.2014-0382>
8. Rachiglio AM, Abate RE, Sacco A, Pasquale R, Fenizia F, Lambiase M, Morabito A, Montanino A, Rocco G, Romano C, Nappi A, Laffaioli RV, Tatangelo F, Botti G, Ciardiello F, Maiello MR, De Luca A, Normanno N (2016) Limits and potential of targeted sequencing analysis of liquid biopsy in patients with lung and colon carcinoma. *Oncotarget* 7:66595–66605. [10.18632/oncotarget.10704](https://doi.org/10.18632/oncotarget.10704)
9. Oxnard GR, Thress KS, Alden RS, Lawrence R, Paweletz CP, Cantarini M, Yang JC, Barrett JC, Jänne PA (2016) Association between plasma genotyping and outcomes of treatment with osimertinib (AZD9291) in advanced non-small-cell lung cancer. *J Clin Oncol* 34:3375–3382. <https://doi.org/10.1200/JCO.2016.66.7162>
10. Normanno N, Maiello MR, Chicchinelli N, Iannaccone A, Esposito C, De Cecio R, D'alesio A, De Luca A (2017) Targeting the EGFR T790M mutation in non-small-cell lung cancer. *Expert Opin Ther Targets* 21:159–165. <https://doi.org/10.1080/14728222.2017.1272582>
11. Annex 1. Summary of product characteristics. European Medicines Agency http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/001016/WC500036358.pdf. (last accessed on 27/12/2016)
12. van Krieken JH, Normanno N, Blackhall F et al (2013) Guideline on the requirements of external quality assessment programs in molecular pathology. *Virchows Arch* 462:27–37. <https://doi.org/10.1007/s00428-012-1354-4>

RESEARCH ARTICLE

Open Access



International pilot external quality assessment scheme for analysis and reporting of circulating tumour DNA

Cleo Keppens^{1,2*} , Elisabeth M. C. Dequeker^{1,2}, Simon J. Patton³, Nicola Normanno⁴, Francesca Fenizia⁴, Rachel Butler⁵, Melanie Cheetham³, Jennifer A. Fairley⁶, Hannah Williams⁶, Jacqueline A. Hall^{7,8}, Ed Schuurin^{2,9}, Zandra C. Deans⁶ and On behalf of IQN Path ASBL

Abstract

Background: Molecular analysis of circulating tumour DNA (ctDNA) is becoming increasingly important in clinical treatment decisions. A pilot External Quality Assessment (EQA) scheme for ctDNA analysis was organized by four European EQA providers under the umbrella organization IQN Path, in order to investigate the feasibility of delivering an EQA to assess the detection of clinically relevant variants in plasma circulating cell-free DNA (cfDNA) and to analyze reporting formats.

Methods: Thirty-two experienced laboratories received 5 samples for *EGFR* mutation analysis and/or 5 samples for *KRAS* and *NRAS* mutation analysis. Samples were artificially manufactured to contain 3 mL of human plasma with 20 ng/mL of fragmented ctDNA and variants at allelic frequencies of 1 and 5%.

Results: The scheme error rate was 20.1%. Higher error rates were observed for *RAS* testing when compared to *EGFR* analysis, for allelic frequencies of 1% compared to 5%, and for cases including 2 different variants. The reports over-interpreted wild-type results and frequently failed to comment on the amount of cfDNA extracted.

Conclusions: The pilot scheme demonstrated the feasibility of delivering a ctDNA EQA scheme and the need for such a scheme due to high error rates in detecting low frequency clinically relevant variants. Recommendations to improve reporting of cfDNA are provided.

Keywords: *KRAS*, *NRAS*, *EGFR*, Mutation testing, ctDNA, cfDNA, Lung cancer, Colorectal cancer

Background

In the last decade, the analysis of predictive biomarkers has become an essential step in the optimisation of therapy for cancer patients [1, 2]. In routine practice, tumour-specific mutation testing entails the analysis of DNA extracted from tumour tissue which is harvested from resections or biopsies. However, tumour tissue sampling is often difficult, especially in patients with advanced disease. In some cases, the tumour sample can yield insufficient DNA for molecular analysis. This is particularly evident in non-small-cell lung cancer

(NSCLC) patients, where in approximately 30% of patients a tissue sample is not available for epidermal growth factor receptor (*EGFR*) mutation analysis, either at diagnosis or as the disease progresses [3]. In these cases, the analysis of circulating cell-free (cfDNA) derived from plasma has been proposed as an alternative method for mutation testing [4, 5].

Plasma-derived cfDNA contains both circulating tumour DNA (ctDNA) and nucleic acids released by normal dividing cells. The mechanism by which tumour cells release ctDNA into the blood is not fully known. It is thought to involve mechanisms such as apoptosis and necrosis, as suggested by the specific fragmentation pattern of ctDNA (+/- 160 base pairs) which in turn is suggestive of a nuclease-dependent degradation [6, 7]. It

* Correspondence: cleo.keppens@kuleuven.be

¹Department of Public Health and Primary Care, Biomedical Quality Assurance Research Unit, University of Leuven, Kapucijnenvoer 35d, 3000 Leuven, Belgium

²European Society of Pathology (ESP), Anderlecht, Belgium

Full list of author information is available at the end of the article



has also been proposed that tumour cells may secrete DNA fragments through vesicles [3].

The advantage of cfDNA testing is that it is minimally invasive and avoids incomplete or variable results arising from tumour heterogeneity [8]. It may also be used to monitor tumour progression [4, 5]. Many studies have demonstrated the effectiveness of assessing tumour-specific alterations by testing plasma cfDNA. This evidence led the European Medicine Agency to approve the use of plasma to detect *EGFR* mutations in the plasma of patients with advanced NSCLC, when adequate tissue is not available [9–11].

In patients with metastasized colorectal cancer (CRC), cfDNA testing for Kirsten rat sarcoma viral oncogene homolog (*KRAS*) and neuroblastoma rat sarcoma viral oncogene homolog (*NRAS*) mutations also holds prognostic value [12]. Consequently, numerous diagnostic tools for the detection of *EGFR*, *KRAS*, *NRAS* and *BRAF* mutations in cfDNA have recently become available. Subsequently the role of cfDNA has moved from use in diagnostic research to becoming a relevant testing matrix in patients with solid tumours [13]. However, the introduction of this novel methodology into clinical practice can be challenging for many laboratories. For instance, the standardization of testing procedures is complex, ranging from plasma collection, cfDNA extraction and cfDNA mutation analysis, to result interpretation. In addition, the analysis must be sufficiently sensitive to identify rare mutant molecules in a background of wild-type DNA at range of 0.1–1%. Currently, clinical applications of cfDNA are focused on the identification of primary mutations in pretreatment samples and the subsequent detection of resistant mutations upon progression in longitudinal samples, which inform treatment decisions. However, the potential uses are numerous and could include tumour monitoring and early tumour diagnosis [4].

The objectives of this External Quality Assessment (EQA) pilot scheme were to (i) investigate the feasibility of designing and delivering a technically challenging EQA (ii) evaluate and compare the ability of laboratories to detect cfDNA in plasma samples (iii) evaluate which extraction methodologies and testing method strategies were used and (iv) to assess the reporting of ctDNA testing results. For this purpose, four European EQA providers (Associazione Italiana di Oncologia Medica – AIOM, European Molecular Genetics Quality Network – EMQN, European Society of Pathology – ESP, United Kingdom National External Quality Assessment Service (UK NEQAS) for Molecular Genetics under the umbrella organization the International Quality Network for Pathology (IQN Path) [14], organised a pilot ctDNA EQA scheme. In this paper, we present the results of this scheme for the analysis of cfDNA for clinically relevant

mutations as well as provide recommendations for reporting.

Methods

EQA scheme design

The pilot EQA was developed in 2016 and delivered to participants during 2017 as a collaboration between the four EQA providers. It was co-ordinated under the banner of an IQN Path working group, with additional expertise provided by scientific advisors. The pilot was carried out according to the requirements of the International Standard for Conformity assessment of proficiency testing ISO 17043 [15] to ensure a robust audit trail was associated with its design, development and implementation.

Thirty-two participant laboratories (eight from each EQA provider) were chosen from a pool of 167 potential candidates who completed a selection survey [13]. Selection criteria included technology available (to ensure material suitability for a range of different technologies), clinical diagnostic workload (to ensure inclusion of laboratories delivering a clinical ctDNA testing service), global location (to assess sample stability during transportation) and testing for *EGFR* and/or *RAS* genes (to ensure relevance to current clinical practice).

The pilot EQA scheme consisted of a set of eight samples containing mutations in the *EGFR*, *KRAS* or *NRAS* genes, in addition to two wild-type samples. The samples were shipped on dry ice (BioCair, Cambridge, United Kingdom) to each participant laboratory and the transit temperatures were monitored. Participants were asked to test the samples for the isolation of cfDNA and subsequent genotyping according to their established routine procedures. A central system for electronic result collection was set up in accordance with ISO 17043 [15] to which the validating laboratories as well as the participants were able to submit their genotyping results and background information on the testing process.

Participating laboratories were asked to submit interpretative diagnostic reports for assessment via their EQA provider. All results provided within the submitted reports were assessed independently by at least two IQN Path working group members against the same pre-defined scoring criteria, harmonized between the four EQA providers (Table 1). Samples A-E versus samples F-J were scored, for *RAS* and *EGFR* testing, respectively. For every case, a maximum of 2 points was awarded and points were deducted depending on the type of error made (Table 1). This yielded a total genotyping score on 20 points for participants to both *RAS* and *EGFR* analysis, and a total score on 10 points for participants to one of both sample sets. For every case, an average genotyping score was calculated on the maximum of 2 points across all participants. Each participant laboratory received an individual feedback report (Additional file 1), as well as a

Table 1 Average genotyping score, assigned score criteria and error rates for every case analysed by the pilot scheme participants

Sample	A	B	C	D	E	F	G	H	I	J	A-J
Variant status	5% KRAS c.35G>C; p.(G12A)	1% KRAS c.35G>C; p.(G12A)	5% KRAS c.182A>G; p.(Q61R)	1% NRAS c.182A>G; p.(Q61R)	1% NRAS c.182A>G; p.(Q61R)	5% EGFR c.2235_2249del15; p.E746_A750del	1% EGFR c.2235_2249del15; p.E746_A750del	5% EGFR c.2573T>G; p.(L858R) and c.2369C>T; p.(T790M)	1% EGFR c.2573T>G; p.(L858R) and c.2369C>T; p.(T790M)	EGFR Wild-type	Overall
Average score per sample on 2 points	1.5	1.3	1.7	1.5	2.0	1.7	1.7	1.9	1.6	1.9	1.7
Score criteria	# obtained scores (%)										
Correct (2 points)	11 (35.5%)	7 (22.6%)	12 (38.7%)	8 (25.8%)	23 (74.2%)	5 (16.1%)	4 (12.9%)	29 (93.5%)	18 (58.1%)	30 (96.8%)	147 (47.4%)
Correct but unspecified (2 points)	5 (16.1%)	3 (9.7%)	4 (12.9%)	2 (6.5%)	0 (0.0%)	16 (51.6%)	16 (51.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	46 (14.8%)
False-negative ≤ LOD (2 points) ^a	0 (0.0%)	5 (16.1%)	0 (0.0%)	4 (12.9%)	0 (0.0%)	0 (0.0%)	2 (6.5%)	0 (0.0%)	6 (19.4%)	0 (0.0%)	17 (5.5%)
False-negative but variant not tested for (2 points) ^a	0 (0.0%)	0 (0.0%)	3 (9.7%)	3 (9.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	6 (1.9%)
False-negative > /unknown LOD (0 points) ^a	1 (3.2%)	5 (16.1%)	3 (9.7%)	4 (12.9%)	0 (0.0%)	0 (0.0%)	1 (3.2%)	0 (0.0%)	3 (9.7%)	0 (0.0%)	17 (5.5%)
False-positive (0 points) ^a	2 (6.5%)	2 (6.5%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (3.2%)	0 (0.0%)	0 (0.0%)	1 (3.2%)	6 (1.9%)
Incorrect mutation detected with therapeutic implications (0 points) ^a	0 (0.0%)	0 (0.0%)	1 (3.2%)	1 (3.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (0.6%)
Incorrect mutation detected without therapeutic implications (2 points)	3 (9.7%)	1 (3.2%)	0 (0.0%)	1 (3.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	5 (1.6%)
One variant missed in double mutation sample (0 points) ^a	N/A	N/A	N/A	N/A	N/A	N/A	N/A	2 (6.5%)	4 (12.9%)	N/A	6 (1.9%)
Mutation described incorrectly (1 point)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	10 (32.3%)	6 (19.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	16 (5.2%)
Technical failure (not scored)	1 (3.2%)	0 (0.0%)	0 (0.0%) [*]	0 (0.0%) [*]	0 (0.0%)	0 (0.0%)	1 (3.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (0.6%)
Not tested (not scored)	8 (25.8%)	8 (25.8%)	8 (25.8%)	8 (25.8%)	8 (25.8%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	40 (12.9%)
Total scored (n = 31)	22 (71.0%)	23 (74.2%)	23 (74.2%)	23 (74.2%)	23 (74.2%)	31 (100.0%)	30 (96.8%)	31 (100.0%)	31 (100.0%)	31 (100.0%)	268 (86.5%)

Table 1 Average genotyping score, assigned score criteria and error rates for every case analysed by the pilot scheme participants (Continued)

Sample	A	B	C	D	E	F	G	H	I	J	A-J	
Variant status	5% KRAS c.35G>C; p.(G12A)	1% KRAS c.35G>C; p.(G12A)	5% KRAS c.182A>G; p.(Q61R)	1% NRAS c.182A>G; p.(Q61R)	1% NRAS Wild-type	5% EGFR c.2235_2249del15; p.E746_A750del	1% EGFR c.2235_2249del15; p.E746_A750del	5% EGFR c.2573T>G; p.(L858R) and c.2369C>T; p.(T790M)	1% EGFR c.2573T>G; p.(L858R) and c.2369C>T; p.(T790M)	EGFR Wild-type	EGFR c.2573T>G; p.(L858R) and c.2369C>T; p.(T790M)	Overall
Error rate	# obtained scores (%)											
Total with implication on therapy decision ^a	3 (13.6%)	12 (52.2%)	7 (30.4%)	12 (52.2%)	0 (0.0%)	0 (0.0%)	4 (13.3%)	2 (6.5%)	13 (41.9%)	1 (3.2%)	54 (20.1%)	
Gene not tested/No method info given	0	0	3	3	0	0	0	0	0	0	6	
NGS	2	7	3	7	0	0	1	1	7	1	29	
Commercial Kit	0	3	1	1	0	0	0	0	1	0	6	
LDT	1	1	0	0	0	0	1	0	1	0	4	
BEAMing	0	0	0	0	0	0	0	0	0	0	0	
ddPCR	0	1	0	1	0	0	2	1	4	0	9	

N/A Not applicable

^aValues were used to calculate the error rates. ^bOne partial technical failure for NRAS, only correct KRAS WT status was assessed on these cases. Reference sequence at time of scoring: EGFR: NM_005228.4 or LRG_304t1; KRAS: NM_033360.3 or NM_004985.4; NRAS: NM_002524.4 or LRG_92t1. Methods are ranked from least to more sensitive techniques as reported in literature: NGS 1–3%, commercial kit 0.1%, BEAMing 0.01%, ddPCR 0.001% [28]. For the LDT, a LOD of 0.1% was reported by the laboratory. Abbreviations: BEAMing Beads, emulsification, amplification, and magnetics, ddPCR Droplet digital polymerase chain reaction, LDT Laboratory-developed test, NGS Next-generation sequencing

general report summarizing the expected results, scheme statistics and final results.

EQA sample preparation and validation

A panel of 10 artificial samples, 5 samples each for colorectal (Cases A-E) and lung (Cases F-J) cancer testing, were manufactured by and purchased from Horizon Discovery Ltd. (Cambridge, United Kingdom) according to a specification provided by the IQN Path working group. These included common, clinically relevant mutations in the *KRAS*, *NRAS* and *EGFR* genes with variant allelic frequencies of 1% or 5%, and also incorporated two wild-type samples (Table 1). Each sample comprised 3 mL human plasma containing 20 ng/mL ctDNA, fragmented to 150 base pairs in length.

Samples were created by reviving and expanding characterised cell lines of which gDNA pellets were created. DNA was extracted from the pellets, fragmented to 150 base pairs (+/- 10%), and diluted to the target concentration. The obtained cfDNA was spiked into normal human donor plasma, for which a copy detection analysis was performed on the background genes. The DNA was extracted once more and a final quality check was performed by estimating the fractional abundance.

Prior to their use in the pilot EQA scheme, each sample was characterised and validated by five reference laboratories, using a range of methodologies (Table 2) to verify sample performance in the pre-analytical and analytical processes, as well as to confirm that the expected genotype

met the material specification provided by the IQN Path working group, and to ensure that the material reflected routine clinical samples in the hands of multiple laboratories. Extraction and analysis methods were selected based on the available methodologies that were validated for *EGFR* and/or *RAS* analysis in the reference laboratories, with the purpose of reflecting at least one method for every technique type, namely next-generation sequencing (NGS), droplet digital PCR (ddPCR), commercial kit, and beads, emulsification, amplification, and magnetics (BEAMing). Optionally, a second laboratory validated the samples using the same methodology if available. The analysed results from the validation trial were collectively reviewed by the IQN Path working group before the materials were released for use in the pilot EQA scheme.

Computational and statistical analysis

EQA participant and validation data from the pilot EQA scheme were analyzed using Microsoft Excel 2013 (Microsoft, Redmond, WA, United States of America). The overall error rate was calculated by dividing the total number of false-positive and false-negative results over the total number of genotypes reported by the participants. False-positive or false-negative results for which the treatment outcome would be affected were considered as critical errors when calculating the rate. Incorrect variants at the same codon were not classified as critical genotyping errors. False-negative results for which the sample genotype was not included in the

Table 2 Overview of error rates per case for different methods for cfDNA extraction and variant analysis during validation

cfDNA extraction method	Cobas cfDNA sample preparation kit (Roche)	QIAamp Circulating Nucleic Acid Kit (Qiagen)					Reference laboratory code
		Cobas® EGFR Mutation Test v2 (Roche)	Capture SureSelect (Agilent), MiSeq (Illumina)	QX200 Droplet Digital PCR System (Bio-rad)	Ampliseq 50 gene hotspot panel, Ion Proton (LifeTechnologies)	Therascreen® EGFR Plasma RGQ PCR Kit (Qiagen)	
	1, 2	2	3*, 4	4	5	5	1-5
Sample	# errors/# genotypes analyzed (error rate in %)						
A	/	1/1 (100.0%)	0/2 (0.0%)	0/1 (0.0%)	/	0/1 (0.0%)	1/5 (20.0%)
B	/	1/1 (100.0%)	0/2 (0.0%)	1/1 (100.0%)	/	0/1 (0.0%)	2/5 (40.0%)
C	/	1/1 (100.0%)	0/1 (0.0%)	0/1 (0.0%)	/	0/1 (0.0%)	1/4 (25.0%)
D	/	1/1 (100.0%)	1/1 (100.0%)	1/1 (100.0%)	/	0/1 (0.0%)	3/4 (75.0%)
E	/	0/1 (0.0%)	0/2 (0.0%)	0/1 (0.0%)	/	0/1 (0.0%)	0/5 (0.0%)
F	0/2 (0.0%)	1/1 (100.0%)	0/2 (0.0%)	0/1 (0.0%)	0/1 (0.0%)	/	1/7 (14.3%)
G	0/2 (0.0%)	1/1 (100.0%)	0/2 (0.0%)	0/1 (0.0%)	0/1 (0.0%)	/	1/7 (14.3%)
H	0/2 (0.0%)	1/1 (100.0%)	0/2 (0.0%)	1/1 (100.0%)	0/1 (0.0%)	/	2/7 (28.6%)
I	1/2 (50.0%)	1/1 (100.0%)	1/2 (50.0%)	1/1 (100.0%)	0/1 (0.0%)	/	4/7 (57.1%)
J	0/2 (0.0%)	0/1 (0.0%)	0/2 (0.0%)	0/1 (0.0%)	0/1 (0.0%)	/	0/7 (0.0%)
A-J	1/10 (10.0%)	8/10 (80.0%)	2/18 (11.1%)	4/10 (40.0%)	0/5 (0.0%)	0/5 (0.0%)	15/58 (25.9%)

/, Sample not tested because gene not included in validated methodology. *Reference laboratory n°3 did not test *NRAS* status. Reference sequence at time of scoring: *EGFR*: NM_005228.4 or LRG_304t1; *KRAS*: NM_033360.3 or NM_004985.4; *NRAS*: NM_002524.4 or LRG_92t1

methodology, or where it was below the stated limit of detection (LOD), were included in error rates because laboratories offering diagnostic mutational analysis on cfDNA should test for the selected clinically relevant variants. Technical failures were excluded from the total number of genotypes. Participants that did not subscribe and thus did not receive either the 5 *EGFR* or 5 *RAS* samples were also not scored for those samples. Statistical difference between reported variant allele frequencies (VAFs) were compared between techniques using a Mann Whitney U (MWU) test, for both the 1% and 5% variants, with a significance level of $\alpha = 0.05$.

Results

No technical failures were observed by the reference laboratories using two commonly used cfDNA extraction methods and six different mutation test methods (Table 2).

Fifteen (25.9%) false-positive or false-negative results were reported for a total of 58 analyzed genotypes (Table 2). On average, more false-negative results were reported for the *RAS* samples when compared to *EGFR*. The Capture SureSelect (Agilent) panel on the MiSeq (Illumina) sequencer was not able to detect any of the included *EGFR* or *RAS* variants in the plasma samples. In contrast, the Ion Ampliseq 50 gene hotspot panel on an Ion Proton (Life Technologies) was able to detect the single deletions in exon 19 of *EGFR* and *RAS* variants included at 5%. However, in the samples with *EGFR* p.(L858R) and p.(T790 M) (cases H, I), 6/14 tests were not able to detect at least one of the two mutations. No false-positive results were reported in the two wild-type samples or in any of the other cases as an additional variant.

The validation of these samples revealed that different ctDNA-based detection methods are able to correctly detect the genotype in 1% and 5% samples with a low

false-positive rate. Our validation procedure also revealed that for less sensitive analytical methods, the 1% samples can be challenging. As the VAFs were still relatively high, we decided to perform the pilot EQA for *KRAS/NRAS* and *EGFR* using the five samples for the *EGFR* and *KRAS/NRAS* scheme. To additionally assess the quality of the samples, DNA yield was measured by each of the five reference laboratories using the QIAamp Circulating Nucleic Acid Kit (Qiagen), and resulted in an average of 0.66 $\mu\text{g/mL}$ (min. 0.11 $\mu\text{g/mL}$, max. 4.63 $\mu\text{g/mL}$). Assessing the DNA yield is not part of the integrated workflow for cobas extraction and analysis.

In total, 32 laboratories from 16 countries participated in the pilot EQA scheme (Fig. 1). Thirty-one (97%) laboratories submitted an electronic datasheet providing details on their cfDNA extraction, analysis methods, and a list of variants tested. One of the 31 laboratories did not submit written reports, therefore their genotyping results were only scored on the entries from the electronic table. In total, 23 participants tested the samples for *KRAS/NRAS* analysis, and 31 participants for *EGFR* analysis. Three of 23 participants receiving the *KRAS/NRAS* mutation samples did not perform any *NRAS* mutation testing but did perform *KRAS* analysis in these samples.

Of all 31 participants, six different cfDNA extraction methods were used (Table 3). The majority of the participants (55%) used the QIAamp Circulating Nucleic Acid Kit (Qiagen) for cfDNA extraction. Only one laboratory used an automated cfDNA extraction method (Promega Maxwell® RSC ccfDNA Plasma Kit). For *KRAS*, *NRAS* and *EGFR* mutation analysis, the most frequently used detection methodologies were NGS (39%) and droplet digital polymerase chain reaction (ddPCR) (23%) (Table 3, Additional file 2: Table S2). A combination of platforms and panels was applied, although the largest fraction of NGS users analyzed the plasma samples with the PGM Ion Torrent (Life Technologies).

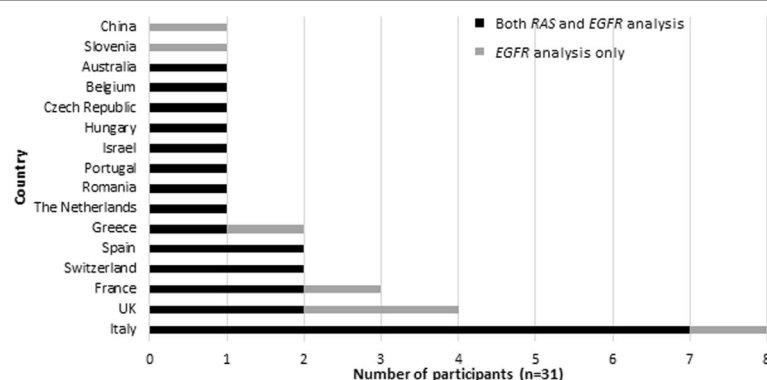


Fig. 1 Overview of the participating countries to the pilot EQA scheme. United Kingdom: One laboratory received both *RAS* (*KRAS/NRAS*) and *EGFR* samples but did not submit results for *KRAS/NRAS* as they were in the process of validation. In total, 23 participants tested the samples for *KRAS/NRAS* analysis, and 31 participants for *EGFR* analysis

Table 3 Overview of the cfDNA extraction and variant analysis methods used by the participants

	# participants to <i>KRAS</i> analysis (%) (<i>n</i> = 23)	# participants to <i>NRAS</i> analysis (%) (<i>n</i> = 20)	# participants to <i>EGFR</i> analysis (%) (<i>n</i> = 31)
cfDNA extraction method			
QIAamp Circulating Nucleic AcidKit (Qiagen)	14 (60.9)	13 (65.0)	17 (54.8)
Cobas cfDNA Sample Preparation Kit (Roche)	4 (17.4)	3 (15.0)	8 (25.8)
MagMAX Cell-Free DNA Isolation Kit (Thermo Fisher Scientific)	3 (13.0)	3 (15.0)	3 (9.7)
Maxwell® RSC ccfDNA Plasma Kit (Promega)	1 (4.3)	0 (0.0)	1 (3.2)
Nucleospin Plasma XS (Macherey-Nagel)	1 (4.3)	1 (5.0)	1 (3.2)
QIAamp DSP DNA Blood Mini Kit (Qiagen) version 2	0 (0.0)	0 (0.0)	1 (3.2)
Variant analysis method			
NGS	13 (56.5)	13 (65.0)	12 (38.7)
Commercial Kit	4 (17.4)	3 (15.0)	11 (35.5)
LDT	1 (4.3)	0 (0.0)	1 (3.2)
BEAMing	1 (4.3)	1 (5.0)	0 (0.0)
ddPCR	4 (17.4)	3 (15.0)	7 (22.6)

The LDT consisted of a 5' nuclease polymerase-chain reaction (Taqman) with peptide nucleic acid probe. For a detailed breakdown of the used methods see Additional file 2: Table S2. Abbreviations: BEAMing Beads, emulsification, amplification, and magnetics, ddPCR Droplet digital polymerase chain reaction, LDT Laboratory-developed test, NGS Next-generation sequencing

In total, 3 (1.1%) technical failures by 3 different participants, were observed from a total of 270 reported genotypes. One technical failure was classed as a partial failure as only the *NRAS* gene analysis failed to provide a reportable result. Hence the reported genotypes for this case were included for *KRAS* analysis, yielding a total of 268 analyzed samples. The reasons reported for technical failures included NGS read depth too low (Case A), problem with DNA extraction (Case G), or a defective cartridge for the partial failure (*NRAS* only, Case B). The overall scheme error rate was 54 (20.1%) on the total of 268 samples. Most errors were observed for *KRAS/NRAS* mutation testing (34/114, 29.8% samples, cases A-D), whereas error rates for *EGFR* analysis were lower (20/154 samples, 13.0%, cases F-I).

Combining the two samples containing a variant at a frequency of 5% and the two with a variant of 1%, yielded a total error rate of 15/45 (33.3%) and 19/46 (41.3%) compared to 4/61 (6.6%) and 15/62 (24.2%) for *RAS* and *EGFR* testing respectively. Sample I was withdrawn from the EQA scheme assessment but for information purposes, for *EGFR* analysis, the majority of the genotyping errors (13 of 20 errors) were observed for this sample I (Table 1) as only 18 out of 31 laboratories (58%) reported the presence of both the *EGFR* mutations at a frequency of 1%. Only one false-positive result (1/54 samples, 1.9%) was observed in the two wild-type samples (cases E and J). In the other four cases, 4/91 false-positive results were obtained for *RAS*, and 1/123 for *EGFR* analysis (Table 1).

Genotyping errors with no impact on therapeutic decisions were also observed but not included in the calculation of the error rate e.g. the detection of an incorrect

KRAS/NRAS nucleotide variant resulting in a change within the same codon, or the incorrect annotation of the *EGFR* exon 19 deletion by NGS users (Table 1). Taking into account the number of laboratories using a specific methodology, the method specific error rate over all samples was the highest for NGS (23%) compared to ddPCR (15%) and commercial kits (15%) (data not shown).

Participants were not asked specifically to report VAFs so only a small number of laboratories provided this information. The mean VAF was calculated for the cases containing a mutation, for which the mutation was correctly detected (Fig. 2). Average VAFs closely resembled the expected frequencies for 5% and 1%, but a very broad range was observed. The average VAF for the cases with variants at 5% was 4.0% (number of genotypes = 82, minimum VAF 0.6%, maximum VAF 13.0%). For variants at 1%, the estimated VAFs were 1.4% (number of genotypes = 57, minimum VAF 0.3%, maximum VAF 10.4%). (Fig. 2). Average VAFs were closer to the expected VAF for ddPCR when compared to NGS, but not significant for either the 1% cases (Mann-Whitney-U, $p = 0.289$, $n = 11$ ddPCR and $n = 34$ NGS) or the 5% cases (Mann-Whitney-U, $p = 0.294$, $n = 17$ ddPCR and $n = 51$ NGS).

The content of the reports varied between laboratories. The most important observation was that several laboratories over-interpreted the absence of a relevant mutation without providing information on quality control (QC) metrics. It is important to state if the input DNA and LOD were appropriate to reliably interpret the results as negative. Without this information, clinical interpretation may be incorrect. For example, a negative result could be

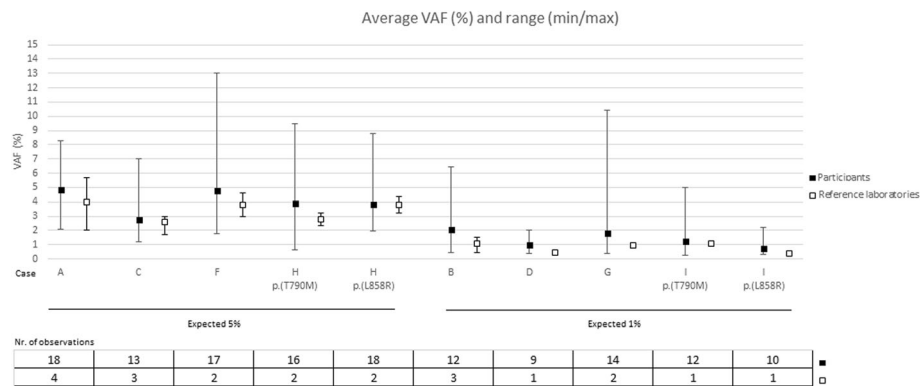


Fig. 2 Average variant allele frequencies by the pilot scheme participants and reference laboratories. Case E and J were not included since they were wild-type. Only the variant allele frequencies of correctly identified variants were taken into account. Min: minimum variant allele frequency reported, max: maximum variant allele frequency reported

interpreted as (i) the absence of a mutation indicating that the patient should receive anti-*EGFR* antibody therapy (in the case of CRC and *RAS* mutations) or that (ii) the patient would be unlikely to benefit from *EGFR*-tyrosine kinase Inhibitors (in the case of NSCLC and *EGFR* mutation). In addition, there was no standardisation in the reporting of the amount of cfDNA extracted, or the LOD. Only a small number of laboratories related the amount of input cfDNA to the assay sensitivity. Variation was also observed for several other elements, including the correct use of Human Genome Variation Society (HGVS) nomenclature [16, 17], reporting of reference sequences [18], and the specification of analysis limits of the methodology.

Discussion

Plasma cfDNA analysis is emerging as a valuable tool to complement resected solid tumour or biopsy material in targeted treatment decisions. Many of the participating laboratories have been performing ctDNA analysis for some time. As there are no current EQAs for testing clinically relevant mutations in plasma, there is an urgent need for well-designed EQA schemes to provide education and benchmarking in order to permit implementation in an accurate, highly qualitative manner [13].

The acquisition and validation of artificial material for this pilot ctDNA molecular testing EQA was harmonized between several EQA schemes. The main goal was to harmonize the minimal requirements for the implementation of a ctDNA EQA scheme, in order to score the laboratories' analytical performance and reporting, and eventually to serve as guidance for the organization of future large-scale EQA schemes. Secondly, harmonization between the four European EQA providers aimed to increase efficiency, and reduce the cost of delivery and speed of access to EQA.

This pilot EQA scheme demonstrated the feasibility of designing and delivering a technically challenging EQA. It

also demonstrated sample stability during in-house distribution, preparation and transportation, which enabled the testing laboratory to produce a reportable result.

Technical failures were reported for only 3/270 (1.1%) of samples (Table 1) and there were none reported during validation (Table 2). However, a high rate of genotype errors was observed by the participants (20.1%). Prior to distribution, in the validation process we observed that the samples with 1% VAF and cases with the two relevant *EGFR* variants were challenging. This was reflected in FN rate (Table 1). Although during validation only six different detection methods were applied (including two different NGS assays), the results indicated that the analytical sensitivity of the methods is important and could be an explanation for the poorer performance of NGS.

In the pilot scheme the participants used a wide range of detection methods, and selected arbitrary cut-offs as a LOD for their assays (when indicated). Our analysis revealed that the highest error rates (false-negative rates) occurred for less sensitive techniques for ctDNA analysis, in concordance with the validation testing and the recent German pilot scheme [19]. Interestingly, when participants were separated into those using commercially available panels ($n = 8$ for both *EGFR* and *RAS* analysis) and those using in-house primers or panels ($n = 5$ for *RAS* and $n = 4$ for *EGFR* analysis), the commercial NGS methods showed excellent scores whereas the latter demonstrated a significantly higher error rate. These findings underline the need for robust validation of in-house NGS approaches for cfDNA testing.

For the samples which yielded a reportable result, more errors were observed for *RAS* analysis when compared to *EGFR* analysis. *EGFR* mutation testing in cfDNA is already widely implemented in clinical practice, whereas *RAS* plasma testing is still an experimental procedure in many centers, a fact which may account for the error rates. In addition, more participants are using commercial, targeted

assays for *EGFR* detection compared to NGS for *RAS* analysis (Table 3). Non-NGS based methods are known to have a greater sensitivity and require less complicated bio-informatics. Despite the high error rate for case D, this sample was retained in the assessment as errors seemed to be related to a poorer performance of NGS, with more participants using this technology compared to *EGFR* analysis for case I. As recommended previously [19], we evaluated the estimated VAFs compared to the expected outcome, in order to assess scheme quality (Fig. 2). We found that average VAFs closely resembled the expected frequencies for 5 and 1%, especially for ddPCR when compared to NGS, although results were not significant as a broad range of VAFs were reported.

Many genotyping errors were observed for the two cases which included both an activating and a resistance *EGFR* variant: the majority of participants did not detect the p.(T790M) variant, especially at a VAF of 1%. Since the majority of *EGFR* mutations detected in the ctDNA of NSCLC patients are detected at < 5% allelic frequency, this would mean that a significant fraction of patients would not have received targeted treatment as result of these tests. For metastatic colorectal cancer, the likely consequence of a false negative result is that a patient inappropriately receives anti-*EGFR* treatment. The overall scheme error rate was higher than that observed in the German ctDNA EQA scheme [19]. However, we included variants at a VAF of 1% and 5% to resemble patient material as closely as possible, rather than at 5% and 10% as previously reported [19]. Furthermore, with the majority of laboratories using less sensitive techniques (Table 3), a fraction of the observed false-negative results occurred because the variant was included at a frequency below the LOD (Table 1). This observation highlights the issue of reporting mutations at low levels when the clinical significance is not known. Taking into account only the true false-negative results, the scheme error rates would be lower and therapy decision making would not always be compromised.

However, the error rates should be interpreted with some caution especially in assays used by a small number of participants, such as BEAMing and for some laboratory-developed tests (LDTs) (Table 3). To be able to draw firm conclusions on different cfDNA detection assays, an EQA with more than 500 participants is needed on a regular basis.

The high number of genotyping errors reported by this group of participants potentially indicates that the artificial material provided does not perform the same way as clinical samples. The difficulties in the implementation of this new methodology to clinical practice and the enormous variation in methods to process plasma, extract cfDNA and detect ctDNA, all compounded by a lack of guidelines, go some way to explain the observed

variations. Additionally, some laboratories reported difficulties in extracting sufficient cfDNA material or specifically reported a reduced assay sensitivity due to the limitations of the supplied material.

Finding sufficient plasma samples from patients with known ctDNA mutations to use in EQA is challenging, mainly due to the amount of plasma required. For this reason, EQA providers are limited to using artificial EQA samples. In this pilot EQA scheme, cell-line derived DNA was spiked into normal plasma, which has the advantage that plasma quantities can be boosted. However, it also runs the risk that different background DNA levels could be present. The fact that cell-line DNA was used instead of plasmids has the advantage of allowing stoichiometric and unbiased dilutions, including QC of the dilution steps, as well as permitting fragmentation of the DNA to resemble the structure of ctDNA observed in patients. Alternatively, artificial plasma may be used [20]. However, plasmid DNA may not be an ideal control sample as it does not represent the true genomic complexity of human tumour samples [20].

Besides the analytical assessment, EQA also assesses the post-analytical phase. The pilot EQA scheme results stress the need for standardization of several elements. Although reporting has been shown to improve across subsequent EQA schemes for formalin-fixed paraffin-embedded tissue for different EQA providers [21, 22], plasma cfDNA testing as a new technology requires the inclusion of specific content in addition to some general elements, such as the use of standardized HGVS nomenclature [16, 17] and reference sequences [18]. However, best practice guidance for cfDNA reporting is currently not available.

More specifically, this pilot EQA highlighted the need to report wild-type results, and to provide a clinical interpretation when no mutation was detected. Because, even in samples where a mutation is present, there are several reasons why a wild-type result might have been obtained.

At certain stages of cancer progression, the amount of ctDNA may be too low to detect, as there is no shedding of tumour DNA. For CRC and NSCLC, a positive association has been described between the tumour volume and the presence of ctDNA [6, 23, 24]. In addition, whether the disease is localized rather than metastatic also significantly affects the ctDNA content in gastro-intestinal stromal tumours [25]. In only 70% of NSCLC cases, the *EGFR* mutation detected in the biopsy is also detected in plasma at the base-line [26] and at progression while on therapy [3]. Therefore, in the case of negative results with sufficient cfDNA input, it is important to obtain a tissue biopsy and when this is not possible, plasma testing should be repeated on a new sample. We also recommend not to use the terms 'positive/negative' to describe the mutation status in reports, as this can be misinterpreted: rather, 'mutation detected/mutation not

detected' terminology should be employed. Secondly, a false-negative result can arise if the sensitivity and LOD of the assay is too low, and to date assay sensitivities vary between $< 0.1 - < 1\%$ [4]. Therefore, data sensitivity of mutation detection and LOD should be recorded in the report. In the pilot scheme there was a high diversity among laboratories regarding the reporting of sensitivities, which were expressed in either as copies/mL or as allelic frequency (percentage). For both options it is recommended that the amount of cfDNA extracted for a sample is included and that this should be related to the assay sensitivity because if the input of the total amount of cfDNA is too low, the test will also be negative. Thirdly, if the assay does not cover all the relevant variants and regions, a mutation might be missed. Therefore, a detailed inclusion of the list of variants, codons or exons tested should be present.

It is important to report the QC metrics of the test performance. Several laboratories reported an incorrect sequence of the deletions in *EGFR* exon 19. While this error will not compromise patients' treatment, it highlights the need for improvements of bioinformatics workflow. A false-negative result could also arise due to haemolysis during collection and processing of blood plasma, diluting the mutant DNA to non-detectable levels [3, 27]. Therefore it is clear that ctDNA testing requires additional guidelines for preanalytical processing.

The utility of circulating biomarkers in the molecular analysis of solid tumours is an exciting new mutation detection tool with many potential applications [28]. However, the highly sensitive testing technology and the handling of appropriate samples is challenging. Standardization is essential to ensure that patients receive the correct results, and so that appropriate treatment is delivered. The provision of EQA is also essential to reassure testing laboratories of the standard of their cfDNA testing service.

Conclusions

As with all EQA schemes, laboratories are encouraged to review their EQA results to ensure no errors have occurred. Errors can impact on the clinical testing service by following up on sub-optimal performance. Based on the findings of this pilot EQA scheme, the need for EQA schemes for all laboratories providing a cfDNA mutation testing service for lung and colorectal cancer has been identified. With this in mind, a second EQA round will be organized in 2018, which will be open to all laboratories from all countries.

Additional files

Additional file 1: Example of individual feedback report. (PDF 136 kb)

Additional file 2: Table S2. Description: Detailed overview of the mutation detection techniques used by the EQA participants. (XLSX 12 kb)

Abbreviations

AIOM: Associazione Italiana di Oncologia Medica; BEAMing: Beads, emulsification, amplification, and magnetics; cfDNA: Circulating cell-free DNA; CRC: Colorectal cancer; ctDNA: Circulating tumour DNA; ddPCR: Droplet digital polymerase chain reaction; *EGFR*: Epidermal growth factor receptor; EMQN: European Molecular Quality Network; EQA: External quality assessment; ESP: European Society of Pathology; HGVS: Human Genome Variation Society; IQN Path: International Quality Network for Pathology; *KRAS*: Kirsten rat sarcoma viral oncogene homolog; LDT: Laboratory-developed test; LOD: Limit of detection; NGS: Next-generation sequencing; *NRAS*: Neuroblastoma rat sarcoma viral oncogene homolog; NSCLC: Non-small-cell lung cancer; QC: Quality control; UKNEQAS: United Kingdom National External Quality Assessment Service; VAF: Variant allele frequency

Acknowledgements

This pilot EQA would not have been possible without the help of a number of organizations and individuals. The authors would like to thank IQN Path for the administrative support. We would also like to gratefully acknowledge the support given to this project by our sponsors, the IQN Path Liquid Biopsy Working Group and the following validating laboratories.

- Molecular Diagnostics Laboratory, The Royal Marsden NHS Trust and the Institute of Cancer Research, Surrey SM2 5NG, United Kingdom.
- All Wales Medical Genetics Service, The Institute of Medical Genetics, Cardiff and Vale University, LHB University Hospital of Wales, Heath Park, Cardiff CF14 4XW, United Kingdom.
- University Medical Center Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands.
- Istituto Nazionale Tumori Fondazione Pascale - CROM, via Ammiraglio Bianco 83013 Mercogliano (AV) Naples, Italy.
- AstraZeneca, Personalised Healthcare and Biomarkers, Darwin, Building 310, Cambridge Science Park, Milton Rd., Cambridge, CB4 0WG, United Kingdom.

Funding

This study was funded by our sponsors who supported the cfDNA pilot and associated workshop. These include Amgen, AstraZeneca, Boehringer Ingelheim, Biocartis, Horizon Diagnostics, Merck KGaA, Qiagen, Roche, Sysmex Inostics, Seracare and Thermo Fisher Scientific/Life technologies.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

All authors conceived and designed the pilot scheme, were involved in the acquisition, analysis and interpretation of data, and contributed to drafting the manuscript or revising it critically for important intellectual content. JAH, HW and ZCD were involved in ordering and shipment of the samples. NN, FF, RB and ES acted as a reference laboratory. CK and EMCD were responsible for collection of results in accordance to ISO17043. JAH and IQN Path provided administrative support, fundraising and sponsorship management. All authors read and approved the final manuscript, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

EMCD received research grants from Pfizer and Amgen. SJP received financial support for educational programmes from Astra Zeneca. NN received fees or research funds from Roche, Qiagen, ThermoFisher, Merck, and Astrazeneca. JAH owns stock in Vivactiv Ltd. ES performed lectures for Illumina, Novartis, Pfizer, BioCartis; is consultant in advisory boards for AstraZeneca, Pfizer, Novartis, BioCartis; and received financial support from Roche, Biocartis, BMS, Pfizer (all fees to the Institution). ZCD received financial

support for educational programmes from Astra Zeneca, Roche and Qiagen and is a member of advisory boards for Amgen, Astra Zeneca, Pfizer, Merck Serono and Roche. All other authors have nothing to declare.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Public Health and Primary Care, Biomedical Quality Assurance Research Unit, University of Leuven, Kapucijnenvoer 35d, 3000 Leuven, Belgium. ²European Society of Pathology (ESP), Anderlecht, Belgium. ³European Molecular Quality Network (EMQN), Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester M13 9WL, UK. ⁴Cell Biology and Biotherapy Unit, Instituto Nazionale Tumori "Fondazione Giovanni Pascale", IRCCS, Napoli, Italy. ⁵All Wales Genetic Laboratory, Institute of Medical Genetics, University Hospital of Wales, Heath Park, Cardiff CF14 4XW, UK. ⁶UK NEQAS for Molecular Genetics, Department of Laboratory Medicine, Royal Infirmary of Edinburgh, Little France Crescent, Edinburgh EH16 4SA, UK. ⁷International Quality Network for Pathology (IQN Path) Association Sans But Lucratif (A.S.B.L), 3A Sentier de l'Espérance, L-1474 Luxembourg City, Luxembourg. ⁸Division of Cancer, Department of Surgery and Cancer, Imperial College London, London, UK. ⁹Department of Pathology, University of Groningen, University Medical Center of Groningen, Groningen, the Netherlands.

Received: 12 January 2018 Accepted: 25 July 2018

Published online: 09 August 2018

References

- Lo Nigro C, Ricci V, Vivenza D, Granetto C, Fabozzi T, Miraglio E, Merlano MC. Prognostic and predictive biomarkers in metastatic colorectal cancer anti-EGFR therapy. *World J Gastroenterol*. 2016;22(30):6944–54.
- Thakur MK, Gadgil SM. Predictive and prognostic biomarkers in non-small cell lung cancer. *Semin Respir Crit Care Med*. 2016;37(5):760–70.
- Normanno N, Denis MG, Thress KS, Ratcliffe M, Reck M. Guide to detecting epidermal growth factor receptor (EGFR) mutations in ctDNA of patients with advanced non-small-cell lung cancer. *Oncotarget*. 2017;8(7):12501–16.
- Diaz LA Jr, Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. *J Clin Oncol*. 2014;32(6):579–86.
- Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol*. 2013;10(8):472–84.
- Diehl F, Li M, Dressman D, He Y, Shen D, Szabo S, Diaz LA Jr, Goodman SN, David KA, Juhl H, Kinzler KW, Vogelstein B. Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A*. 2005;102(45):16368–73.
- Mouliere F, Rosenfeld N. Circulating tumor-derived DNA is shorter than somatic DNA in plasma. *Proc Natl Acad Sci U S A*. 2015;112(11):3178–9.
- Rolfo C, Castiglia M, Hong D, Alessandro R, Mertens I, Baggerman G, Zwaenepoel K, Gil-Bazo I, Passiglia F, Carreca AP, Taverna S, Vento R, Santini D, Peeters M, Russo A, Pauwels P. Liquid biopsies in lung cancer: the new ambrosia of researchers. *Biochim Biophys Acta*. 2014; 1846(2):539–46.
- Goto K, Ichinose Y, Ohe Y, Yamamoto N, Negoro S, Nishio K, Itoh Y, Jiang H, Duffield E, McCormack R, Saijo N, Mok T, Fukuoka M. Epidermal growth factor receptor mutation status in circulating free DNA in serum: from IPASS, a phase III study of gefitinib or carboplatin/paclitaxel in non-small cell lung cancer. *J Thorac Oncol*. 2012;7(1):115–21.
- Douillard JY, Ostoros G, Cobo M, Ciuleanu T, Cole R, McWalter G, Walker J, Dearden S, Webster A, Milenkova T, McCormack R. Gefitinib treatment in EGFR mutated caucasian NSCLC: circulating-free tumor DNA as a surrogate for determination of EGFR status. *J Thorac Oncol*. 2014;9(9):1345–53.
- Fenzia F, De Luca A, Pasquale R, Sacco A, Forgione L, Lambiase M, Iannaccone A, Chicchinelli N, Franco R, Rossi A, Morabito A, Rocco G, Piccirillo MC, Normanno N. EGFR mutations in lung cancer: from tissue testing to liquid biopsy. *Future Oncol*. 2015;11(11):1611–23.
- Spindler KL, Pallisgaard N, Andersen RF, Brandslund I, Jakobsen A. Circulating free DNA as biomarker and source for mutation detection in metastatic colorectal cancer. *PLoS One*. 2015;10(4):e0108247.
- Deans ZC, Williams H, Dequeker EMC, Keppens C, Normanno N, Schuurings E, Patton SJ, Cheetham M, Butler R, Hall JA. Review of the implementation of plasma ctDNA testing on behalf of IQN path ASBL: a perspective from an EQA providers' survey. *Virchows Arch*. 2017;471(6):809–13.
- IQN Path. (2017) <http://www.iqnpath.org>. Accessed 10 Jan 2018.
- International Organisation for Standardization. Conformity assessment - general requirements for proficiency testing. ISO/IEC 17043:2010. Geneva: ISO; 2010.
- Tack V, Deans ZC, Wolstenholme N, Patton S, Dequeker EM. What's in a name? A coordinated approach toward the correct use of a uniform nomenclature to improve patient reports and databases. *Hum Mutat*. 2016;37(6):570–5.
- Human Genome Variation Society (HGVS): Sequence Variant Nomenclature. (2016) <https://varnomen.hgvs.org>. Accessed 10 Jan 2018.
- Locus Reference Genomic. (2011) <http://www.lrg-sequence.org>. Accessed 10 Jan 2018.
- Haselmann V, Ahmad-Nejad P, Geilenkeuser WJ, Duda A, Gabor M, Eichner R, Patton S, Neumaier M. Results of the first external quality assessment scheme (EQA) for isolation and analysis of circulating tumour DNA (ctDNA). *Clin Chem Lab Med*. 2017; <https://doi.org/10.1515/cclm-2017-0283>.
- Whale AS, Cowen S, Foy CA, Huggett JF. Methods for applying accurate digital PCR analysis on low copy DNA samples. *PLoS One*. 2013;8(3):e58177.
- Dequeker EM, Keppens C, Egele C, Delen S, Lamy A, Lemoine A, Sabourin JC, Andrieu C, Ligtenberg M, Fétique D, Tops B, Descarpentries C, Blons H, Denoux Y, Aube C, Penault-Llorca F, Hofman P, Leroy K, Le Marechal C, Doucet L, Duranton-Tanneur V, Pedeutour F, Soubeyran I, Côté JF, Emile JF, Vignaud JM, Monhoven N, Haddad V, Laurent-Puig P, van Krieken H, Nowak F, Lonchamp E, Bellocq JP, Rouleau E. Three Rounds of External Quality Assessment in France to Evaluate the Performance of 28 Platforms for Multiparametric Molecular Testing in Metastatic Colorectal and Non-Small Cell Lung Cancer. *J Mol Diagn*. 2016;18(2):205–14.
- Tack V, Dufraing K, Deans ZC, van Krieken HJ, Dequeker EM. The ins and outs of molecular pathology reporting. *Virchows Arch*. 2017; <https://doi.org/10.1007/s00428-017-2108-0>.
- Holdhoff M, Schmidt K, Donehower R, Diaz LA Jr. Analysis of circulating tumor DNA to confirm somatic KRAS mutations. *J Natl Cancer Inst*. 2009; 101(18):1284–5.
- Schwarzenbach H, Stoehlmacher J, Pantel K, Goekkurt E. Detection and monitoring of cell-free DNA in blood of patients with colorectal cancer. *Ann N Y Acad Sci*. 2008;1137:190–6.
- Boonstra PA, ter Elst A, Tibbesma M, Bosman LJ, Mathijssen R, Atrafi F, van Coevorden F, Steeghs N, Farag D, Gelderblom H, van der Graaf WTA, Desar IME, Maier J, Overbosch J, Suurmeijer AJH, Gietema J, Schuurings E, Reyners AKL. A single digital droplet PCR assay to detect multiple KIT exon 11 mutations in tumor and plasma from patients with gastrointestinal stromal tumors. *Oncotarget*. 2018;17:13870–83.
- Thress KS, Brant R, Carr TH, Dearden S, Jenkins S, Brown H, Hammett T, Cantarini M, Barrett JC. EGFR mutation detection in ctDNA from NSCLC patient plasma: a cross-platform comparison of leading technologies to support the clinical development of AZD9291. *Lung Cancer*. 2015;90(3):509–15.
- Nikolaev S, Lemmens L, Koessler T, Blouin JL, Nospikel T. Circulating tumoral DNA: Preanalytical validation and quality control in a diagnostic laboratory. *Anal Biochem*. 2017;542:34–9.
- Siravegna G, Marsoni S, Siena S, Bardelli A. Integrating liquid biopsies into the management of cancer. *Nat Rev Clin Oncol*. 2017 Sep;14(9):531–48.