Improving outcomes for resistant rectal cancer



A thesis submitted for the degree Doctor of Philosophy

by

Séan M. O'Cathail

BMedSci MB BCh BAO MSc MRCPI MRCP FRCR

St Cross College, University of Oxford

Trinity 2019

CRUK/MRC Oxford Institute of Radiation Oncology
Department of Oncology
Old Road Campus Research Building
OX3 7LE

Acknowledgements

Firstly, I would like to thank my supervisors – Prof Tim Maughan, Prof Maria Hawkins and Dr Annie Lewis. Tim and Maria gave the opportunity to pursue a research fellowship and for that I am extremely grateful. Over the last three years I have learned so much from both about research, leadership and life in clinical academia. But I am particularly grateful for allowing me the freedom to take my research into areas I was keen to pursue. I am eternally indebted to Annie for taking me into her lab, for the patience and time she took to teach me the techniques I needed and for the freedom to design and carry out my work.

Others who have contributed to my efforts over the last three years include Dr Remko Prevo and his guidance on performing colony forming assays, Mr Andrew Blake from the S:CORT group for his assistance with my queries on data and resource access and lastly to Dr Chieh His-Wu with whom I have had wonderful collaboration and from whom I have learned so very much about statistics, modelling and data rigour.

To my parents, Seamus and Kathleen, without whose tireless support of my education and unwavering faith in my ability I would never have started on a road that led to me to this point. To my brothers, Cillian, Seamus, Micheal and Colman, whom I can always count on for anything but most especially to keep me grounded and humble.

And, finally, to my wife Kara. She is my greatest supporter, my most thoughtful critic and anything I have achieved was and is only possible because of her.

For Kara, Iris, Senan and ‘baby’

Publications

O'Cathail SM, Pokrovska TD, Maughan TS, Fisher KD, Seymour LW, Hawkins MA. Combining Oncolytic Adenovirus with Radiation-A Paradigm for the Future of Radiosensitization. Front Oncol. 2017; 24(7) – 153 doi: 10.3389/fonc.2017.00153

Rayner E, Durin M, Thomas R, Moralli D, O'Cathail SM, Tomlinson I et al, CRISPR-Cas9 causes chromosomal instability and re-arrangements in cancer cell lines, detectable by cytogenetic methods. CRISPR J, 2019 (accepted manuscript, in press)

O'Cathail SM, Wu CH, Lewis A, Holmes C, Hawkins MA, Maughan T. NRF2 metagene signature is a novel prognostic biomarker in colorectal cancer. BMC Cancer, 2019 (manuscript, in review)

Presentations

O'Cathail, Wu CH, Lewis A, Maughan T, Hawkins MA. NRF2 metagene expression associated with worse response to neoadjuvant chemoradiotherapy in rectal cancer. IJROBP (2019) – **Oral presentation ASTRO, Sept 2019**

O'Cathail SM, Wu CH, Lewis A, Maughan T, Hawkins MA. Nrf2 metagene as a prognostic biomarker across all stages of colorectal cancer (CRC). Journal of Clinical Oncology (2019) 37 (4_suppl) 557 – **Poster presentation ASCO GI, Jan 2019**

O'Cathail SM, Lewis A, Hawkins MA, Maughan TS. KEAP1-NRF2 as a modulator of response to radiation in rectal cancer. IJROBP (2018); 102(3_suppl) e191-192 – **Poster discussion ASTRO, Oct 2018**

Abstract

The management of locally advanced rectal cancer involves the use of neoadjuvant chemoradiotherapy but up to 20% of patients have no discernible response. Understanding the factors that mediate resistance to treatment is therefore an important clinical research question. The work in this thesis aimed to further understand potential biological contributors to therapeutic resistance by exploring the contribution of the KEAP1-NRF2 pathway *in-vitro* and evaluating it as a potential biomarker *in-vivo*.

Using RNAi and clonogenic assays, NRF2 knockdown produced a radiosensitive phenotype in intrinsically resistant cell line models but had no effect in the most radiosensitive cell line. An attempt at CRIPSR isogenic validation of the RNAi was also made. Using a CRISPR isogenic activating mutation, radioresistance was induced in this previously sensitive cell line. Overexpression of NRF2 also induced chemoresistance to the commonly used radiosensitizers in clinical practice.

RNA sequencing was carried out on both NRF2 downregulated (radiosensitive) and NRF2 upregulated (radioresistant) cell line models to better understand the mechanism of radioresistance. NRF2 downregulation produces a marked reduction in amino acid metabolism, particularly glutamine/asparagine, relative to control conditions. Overexpression of NRF2 produced the opposite effect. The data suggest an NRF2 regulated ‘switch’ of metabolism in favour of an anabolic, proliferative phenotype which produces radioresistance.

To understand the clinical relevance of NRF2 signalling a bioinformatic approach was used. A metagene signature of 36 NRF2 regulated genes to aggregate pathway expression was derived and trained using publically available datasets. The resulting transcriptomic biomarker demonstrated that high NRF2 expression was associated with worse survival outcomes in 1,360 validation patients from public resources and the S:CORT collaboration. Interestingly NRF2 expression was significantly enriched in Consensus Molecular Subtype 4, suggesting a further refinement of the molecular transcriptomic classification of colorectal cancer.

Lastly higher NRF2 expression was shown to be associated with a worse response to neoadjuvant chemotherapy, as measured by the neoadjuvant rectal score, in a curated cohort of 127 rectal cancers who received standard of care treatment. This validates the *in-vitro* cell line findings and supports an important role for NRF2 signalling in prognosis and therapeutic resistance in colorectal cancer.

Contents

1 CHAPTER 1 1

1.1 Introduction 2

1.1.1 Rectal cancer 2

1.1.2 Radiotherapy 3

1.1.3 Radiotherapy for rectal cancer 6

1.1.4 Failures in intensification in rectal cancer 8

1.1.5 Biomarkers in oncology 10

1.1.6 Current biomarkers in colorectal cancer 12

1.1.7 Biomarkers of radioresistance in colorectal cancer 18

1.2 Overview of the KEAP1 - NRF2 pathway 22

1.2.1 NRF2 as a ‘tumour suppressor’ 24

1.2.2 NRF2 as an ‘oncogene’ 24

1.2.3 Mechanisms of KEAP1-NRF2 activation 26

1.2.4 KEAP1-NRF2 as biomarker of radioresistance 28

1.2.5 KEAP1- NRF2 in colorectal cancer 29

1.3 Aims of thesis 30

2 CHAPTER 2 32

2.1 Laboratory methods and materials 33

2.1.1 Cell culture 33

2.1.2 RNA interference reagents and conditions 34

2.1.3 Gene Expression (mRNA) analysis 36

2.1.4 Western blot analysis 38
2.1.5 Clonogenic survival 38
2.1.6 Cell viability assay 40
2.1.7 CRISPR protocols 42
2.1.8 Cell line karyotyping 49
2.2 RNA sequencing 51
2.2.1 Differential expression of genes 52
2.2.2 Gene Set Enrichment Analysis 52
2.3 Amino acid exchange experiments 53
2.3.1 Cell culture media 53
2.3.2 Technique 53
2.4 Bioinformatic materials and methods 54
2.4.1 Candidate gene selection 54
2.4.2 STRING analysis 54
2.4.3 Regulatory sequence analysis tools (RSAT) 55
2.4.4 RNA array probe matching 55
2.4.5 Unsupervised analysis for dimension reduction 55
2.4.6 Supervised analysis using multivariate Cox regression models 56
2.4.7 Construction of the NRF2 signature in the validation sets 57
2.4.8 Datasets 57
2.4.9 Aberdeen rectal cohort 58
2.4.10 Sample handling and RNA extraction for S:CORT datasets 58

2.4.11 Statistical analysis - Prognosis 59
2.4.12 Statistical analysis - Prediction 61
3 CHAPTER 3 62
3.1 Introduction 63
3.2 Results 64
3.2.1 Knockdown of NRF2 causes radiosensitization 65
3.2.2 Consistent radiosensitivity with RNAi targeting alternate NFE2L2 exons 70
3.2.3 CRISPR validation of siRNA induced changes 73
3.2.4 Resistant cell lines have increased NRF2 signalling 84
3.2.5 Induction of NRF2 signalling causes radioresistance 86
3.2.6 Overexpression of NRF2 induces chemoresistance 94
3.3 Discussion 96
4 CHAPTER 4 101
4.1 Introduction 102
4.2 Results 103
4.2.1 Transcriptomic changes of SW1463 siNFE2L2 model 106
4.2.2 Gene Set Enrichment Analysis (GSEA) 109
4.2.3 Transcriptomic changes of HCT116 E79K+ model 112
4.2.4 Gene Set Enrichment Analysis (GSEA) 115
4.2.5 KEGG pathways enrichment 118
4.2.6 Asparagine facilitates increased growth in NRF2 mutant line 127
4.3 Discussion 128

4.3.1 Glutamine metabolism....128
4.3.2 Asparagine metabolism....129
4.3.3 Ribosomal biogenesis....130
5 CHAPTER 5....133
5.1 Introduction....134
5.1.1 Gene signature Derivation....135
5.1.2 Current prognostic NRF2 signatures....138
5.1.3 Hypothesis....139
5.2 Results....139
5.2.1 Candidate gene refinement for NRF2 CRC signature....139
5.2.2 Construction of the NRF2 signature in GSE17536....143
5.2.3 NRF2 expression signature in the validation sets....146
5.2.4 High NRF2 signature is biomarker of worse survival outcomes....147
5.2.5 NRF2 signature provides additional prognostic power in MVA....152
5.2.6 NRF2 signature was not associated with response to oxaliplatin *in-vivo*....155
5.2.7 NRF2 expression enriched in Consensus Molecular Subtype 4....155
5.3 Discussion....156
5.3.1 Oxaliplatin interaction....158
5.3.2 Rapport with CMS....159
6 CHAPTER 6....162
6.1 Introduction....163
6.2 Results....166

6.2.1 Cohort characteristics 166
6.2.2 NRF2 signature pattern identified in biopsy samples 167
6.2.3 NRF2 signature is a biomarker of worse radiotherapy response 169
6.2.4 NRF2 signature provides additional explanatory power to clinical N stage 169
6.2.5 NRF2 enriched in Consensus Molecular Subtype 4 171
6.3 Discussion 172
7 CHAPTER 7 176
7.1 General summary and conclusions 177
7.2 Areas for future investigation of NRF2 182
7.3 Future strategies for treating resistant rectal cancer 185
7.4 Conclusions 186
8 REFERENCES 187
9 APPENDIX 214
9.1 Appendix 1: Supplementary RNA sequencing data 215
9.2 RNA sequencing analysis code 225
9.3 APPENDIX 2: List of probes from microarray datasets in Chapter 5 237

List of Figures

Figure 1-1. Distribution of colorectal cancers in the UK by anatomical site 2
Figure 1-2. GARD score distributions across 20 different tumour types 5
Figure 1-3. Response assessment for the standard of care arm in the CAO/ARO/AIO-04 trial 6
Figure 1-4. Proposed molecular taxonomy of colorectal cancer from the Consensus Molecular Subtypes 18
Figure 1-5. The KEAP1 – NRF2 signalling pathway 23
Figure 1-6. The context dependency of NRF2 signalling in cancer progression 25
Figure 1-7. Summary of mechanisms of NRF2 activation in cancers 27
Figure 2-1. Example of Countess™ Cell Counting System output following PLK1 transfection 35
Figure 2-2. Schematic overview of 96 well plate used for MTS cell viability assay 41
Figure 2-3. Ideogram of chromosome 2 used for FISH analysis 50
Figure 3-1. RNAi knockdown of NRF2 results in radiosensitivity of SW1463 66
Figure 3-2. RNAi knockdown of NRF2 results in radiosensitivity of HT55 68
Figure 3-3. RNAi knockdown of NRF2 results in radiosensitivity of HCT116 69
Figure 3-4. RNAi knockdown targeting second locus of NFE2L2 results in radiosensitivity of SW1463. 71
Figure 3-5. RNAi knockdown targeting third locus of NFE2L2 results in radiosensitivity of SW1463 72
Figure 3-6. Schematic overview of the CRISPR-Cas9 workflow 74
Figure 3-7. Dose response curve for SW1463 cell treated with puromycin 75
Figure 3-8. CRISPR clone screening and selection by Sanger sequencing 76
Figure 3-9. Sanger sequencing chromatograms of indel mutations in two derived clones of SW1463 77
Figure 3-10. Barplot of qPCR gene expression analysis from SW1463 clone 5 , relative to SW1463 parental line. Cell survival curves of clonogenic assay with SW1463 parental wild type and exon 4 mutant clone 5 78
Figure 3-11. Barplot of qPCR gene expression analysis from SW1463 clone 6, relative to SW1463 parental line. Cell survival curves of clonogenic assay with SW1463 parental wild type and exon 4 mutant clone 6. 79
Figure 3-12. Colony formation at 0Gy of SW1463 parental (top) and SW1463 Clone 6 (bottom) 80
Figure 3-13. DAPI karyogram of cytogenetic analysis of SW1463, parental (left) and clone 5 (right) 81
Figure 3-14. Barplots of chromosomal number from metaphase spreads for SW1463 WT, SW1463 NFE2L2 clone 5 and SW1463 NFE2L2 clone 6 82
Figure 3-15. FISH metaphase spreads from SW1463 WT line and SW1463 NFE2L2 clones 5 and 6 83
Figure 3-16. Barplot showing relative fold change in mRNA levels of NRF2 targets in control cells 85
Figure 3-17. Barplot of qPCR gene expression analysis of HCT116 following treatment with t-BHQ 50µM, relative to treatment with DMSO 50µM 86
Figure 3-18. Location and functional consequences of mutation in the *NFE2L2* gene 87
Figure 3-19. Sanger sequencing chromatrograms showing the missense point mutation in HCT116 E79K+. 88

Figure 3-20. DAPI karyogram of cytogenetic analysis of HCT116 wild type (left) and HCT116 E79K+ mutant (right) 89
Figure 3-21. Barplot of chromosomal counts in metaphase spreads for HCT116 WT and HCT116 E79K+ 90
Figure 3-22. FISH from metaphase spreads from HCT116 WT (left) and HCT116 E79K+ (right) with probes as previously indicated 90
Figure 3-23. Barplot of relative fold change (mRNA) from HCT116 WT and E79K+ mutant cells 92
Figure 3-24. Western blots (cropped) of NRF2 protein (top) with B-actin as a loading control (left) 93
Figure 3-25. Colony formation assay (top) and cell survival curves (bottom) from CFA of HCT116 E79K+ (red) and HCT116 WT (black) 94
Figure 3-26. Dose response curves for HCT116 isogenic cell viability following treatment cytotoxic drug treatment....96
Figure 4-1. Barplot of relative fold change of mRNA levels following transfection with siNFE2L2_1 ..104
Figure 4-2. Representative immunoblot of western blot for NRF2 following RNAi....104
Figure 4-3. Quality control plots for RNA sequencing of SW1463 RNAi experiment 106
Figure 4-4. Comparison of differential gene (DE) expression results between edgeR and DESeq2 packages 107
Figure 4-5. Plot smear and volcano plot for NRF2 knockdown RNA seq.....108
Figure 4-6. GSEA enrichment plot for the NFE2L2.V2 geneset in the ranked siNRF2 data 110
Figure 4-7. Stacked barchart of normalised enrichment scores (NES) for the 50 hallmark genesets following GSEA....111
Figure 4-8. Quality control plots for RNA sequencing of HCT116 WT and E79K+ cell lines 112
Figure 4-9. Comparison of differential gene (DE) expression results between edgeR and DESeq2 packages 113
Figure 4-10. Plot smear and volcano plot for HCT116 E79K+ mutant RNA seq....114
Figure 4-11. GSEA enrichment plot for the NFE2L2.V2 geneset in the ranked E79K mutant data 116
Figure 4-12. Stacked barchart of normalised enrichment scores (NES) for the 50 hallmark genesets following GSEA....117
Figure 4-13. Stacked barchart of normalised enrichment scores (NES) for the KEGG pathway GSEA in the siNRF2 model 118
Figure 4-14. Stacked barchart of normalised enrichment scores (NES) for the KEGG pathway GSEA in the E79K+ NRF2 mutant model....119
Figure 4-15. GSEA analysis enrichment plots showing significant changes in mTOR signalling 120
Figure 4-16. Venn diagram of differentially expressed genes from all four RNA seq analyses 121
Figure 4-17. STRING interaction network for the 141 shared genes with opposing directionality....122
Figure 4-18. Heatmaps of relative change in expression (z-score) of the 141 shared genes....124
Figure 4-19. Overview of amino acid metabolism, TCA cycle, carbon metabolism in a cell 125
Figure 4-20. Metabolic ‘switch’ model in response to changes in NRF2 126

Figure 4-21. Paired barplots of total cell number under different amino acid supplementation.....127
Figure 5-1. Aggregate heatmaps for Oncomine database over all over expressed and under expressed genes in 9 different colorectal datasets. 141
Figure 5-2. Known and predicted protein – protein interactions from the STRING database for the 40 genes from the refined differential expression analysis. 142
Figure 5-3. Pairwise correlation heatmap showing the degree of positive (red) and negative (blue) Pearson correlation between the 36 genes of the NRF2 pathway in the training set. 144
Figure 5-4. Regulatory sequence analysis. 145
Figure 5-5. Pairwise correlation heatmaps of the showing the degree of positive (red) and negative (blue) Pearson correlation between the 36 genes of the NRF2 pathway in the validation sets 147
Figure 5-6. Kaplan Meier curves and associated risk tables for the primary analysis of three datasets ...150
Figure 5-7. Pairwise correlation heatmap showing the degree of positive (red) and negative (blue) Pearson correlation between the 36 genes of the NRF2 pathway in GSE87211 151
Figure 5-8. Kaplan Meier plots and risk tables for survival analysis in GSE87211 dataset 152
Figure 5-9. Forest plot of the multivariate analysis of known prognostic factors in MRC FOCUS cohort 153
Figure 5-10. Distribution of NRF2 across CMS in the MRC FOCUS trial dataset 156
Figure 6-1. Pairwise Pearson correlation heatmaps between the 36 genes of the NRF2 pathway in GSE17536 (top) and the Aberdeen rectal cohort (bottom). 168
Figure 6-2. Scatter histogram of NAR scores and NRF2 signature expression 169
Figure 6-3. Ordinal logistic regression plots for NRF2 expression predicting NAR score.....170
Figure 6-4. Boxplots of NRF2 signature expression by Consensus Molecular Subtype. 171
Figure 6-5. Pathway enrichments for the 36 gene NRF2 signature 174
Figure 7-1. The tumour microenvironment.....184

List of Tables

Table 2-1. Summary table of number of cells plated 39
Table 2-2. Summary of the target gene, target sequence, primers and target location for CRISPR-Cas9 treatment 45
Table 2-3. Details of Fluorescent in situ hybridisation probes 49
Table 2-4. Details of relevant concentrations of amino acid supplements used in amino acid exchange experiment 53
Table 3-1. Molecular profiles of colorectal cancer cell lines used experimentally 64
Table 3-2. Summary of surviving fraction (SF) at 6Gy of siNRF2 and control cells 70
Table 3-3. Summary table of the FISH analysis for SW1463 parental line and clones 82
Table 3-4. Summary table of the FISH analysis for HCT116 WT and E79K+ lines 91
Table 4-1. Tapestation derived quality metrics for all replicates used in RNA sequencing analysis 105
Table 4-2. Summary table of the top 20 DE genes from SW1463 NRF2 knockdown RNA-seq 109
Table 4-3. Summary table of the top 20 DE genes from HCT116 E79K+ mutant RNA-seq 115
Table 5-1. Annotation for the 9 datasets and the anatomical sub-sites used in the Oncomine differential gene analysis. 140
Table 5-2. Cohort characteristics of the training and validation sets. 148
Table 5-3. Summary table of the Cox proportional hazard model analyses 154
Table 6-1. Aberdeen rectal cohort characteristics (N=127). 167

List of abbreviations

5-FU: 5-fluorouracil

ACPGBI: Association of Coloproctology of Great Britain and Ireland

AIC: Akaike information criteria

ALL: acute lymphoblastic leukemia

ANOVA: analysis of variance

ARE: antioxidant response element

ASNS: Asparagine synthetase

BAC: Bacterial Artificial Chromosomes

BIC: Bayesian information criteria

BLAST: Basic Local Alignment Search Tool

BRAF: murine sarcoma viral oncogene homolog B

C.I: confidence intervals

cDNA: complementary DNA

CMS: consensus molecular subtypes

CNC: cap ‘n’ collar

COSMIC: Catalogue of Somatic Mutations in Cancer

cpm: counts per million

CRC: colorectal cancer

CRISPR: clustered regularly interspaced short palindromic repeats

CRT: chemoradiotherapy

CTLA-4: cytotoxic T-lymphocyte-associated protein 4

DAPI: 4′,6-diamidino-2-phenylindole

DE: differentially expressed

DFS: disease free survival

DMEM: Dulbesco's Modified Eagles serum media

dMMR: deficient mismatch repair

DMSO: dimethyl sulfoxide

DNA: deoxy-ribonucleic acid

EC: Effective concentration

ECM: extracellular matrix

EGFR: epidermal growth factor receptor

EpRE: electrophilic response element

ER: oestrogren receptor

ERBB3: Erb-B2 Receptor Tyrosine Kinase 3

FDR: false discovery rate

FISH: Fluorescent in situ hybridisation

GARD: Genome adjusted radiation dose

GEO: Gene Expression Omnibus

GFP: green fluorescent protein

GO: gene ontology

Gy: Gray

HER2: human epidermal growth factor receptor 2

HMOX1: Heme Oxygenase 1

HR: hazard ratio

HRP: horseradish peroxidase

HRR: Homologous Recombination repair

IHC: immunhistochemistry

KEAP1: Kelch Like ECH Associated Protein 1

KEGG: Kyoto Encyclopedia of Genes and Genomes

KRAS: Kirsten rat sarcoma viral oncogene

LARC : locally advanced rectal cancer

LQ: linear quadratic

LRT: likelihood ratio test

LUAD: lung adenocarcinoma

MMP: matrix metalloproteinases

MMR: mismatch repair

mRNA: messenger Ribonucleic acid

MSI: microsatellite instability

MSigDB: molecular signature database

mTOR: mammalian target of rapamycin

NADPH: Nicotinamide adenine dinucleotide phosphate

NAMS: NFE2L2 Associated Molecular Signature

NAR: neoadjuvant rectal

NEAA: non essential amino acids

Neh: Nrf2-ECH homology

NFE2L2: Nuclear Factor, Erythroid 2 Like 2

NHEJ: Non homologous end joining

Nqo-1: NAD(P)H: quinone oxidorecductase -1

NRF2: NF-E2-related factor 2

OS: overall survival

PARP: poly ADP ribose polymerase

PBS: Phosphate Buffer Solution

PC: principal component

PC1: principal component 1

PCA: principal component analysis

pCR: pathological complete response

PCR: polymerase chain reaction

PD-1: programmed death protein 1

PDL-1: programmed death protein ligand 1

PPI: protein-protein interactions

qPCR: quantitative polymerase chain reaction

RIN: RNA Integrity Number

RIPA: Radioimmunoprecipitation assay

RNAi: Ribonucleic acid interference

rRNA: ribosomal RNA

RSAT: Regulatory sequence analysis tools

RSI: radiosenstivity index

RTOG: Radiation therapy oncology group

SBRT: stereotactic body radiotherapy

SCC: squamous cell carcinoma

SCORT: Stratification in Colorectal cancer

SER: sensitization enhancement ratio

sgRNA: single guide RNA

siRNA: small interfering ribonucleic acid

SMAD: small mothers against decapentaplegic Homolog

SNP: single nucleotide polymorphisms

t-BHQ: Tert-butyl-hydroquinone

TCGA: The Cancer Genome Atlas

TGF-β: Transforming growth factor β

TME: total mesorectal excision

TMM: trimmed mean method

TNM: tumour nodes metastasis

TP53: tumour protein 53

TRG: tumour regression grade

TXN: Thioredoxin

WT: wild type

CHAPTER 1

1.1 Introduction

1.1.1 Rectal cancer

In 2014-2016 42,042 people were diagnosed with colorectal cancer (CRC) in the UK [1], with approximately 30% of these presenting in the rectum [Figure 1-1] compared to other sub sites.

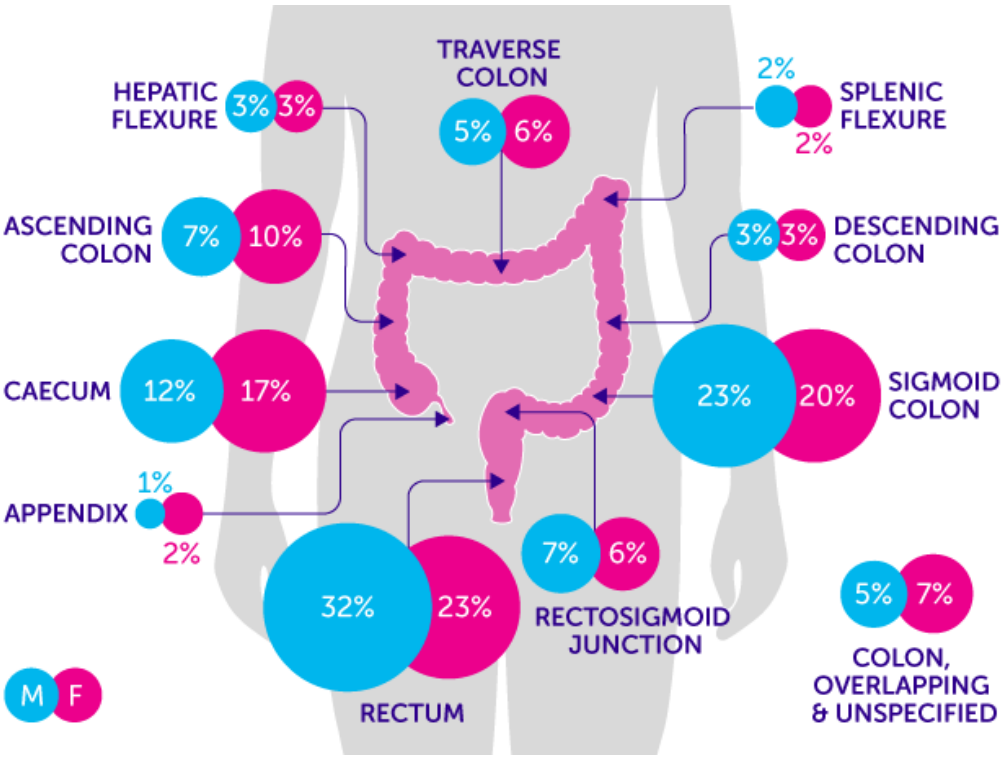


Figure 1-1. Distribution of colorectal cancers in the UK by anatomical site
[Data from Cancer Research UK, accessed 23/07/2019]

Locally advanced rectal cancer (LARC) is defined by the presence of T3/4 disease and/or nodal involvement in the absence of distal metastatic disease. Up to 30% of patients present with LARC and will require neoadjuvant or adjuvant radiation, primarily to facilitate resection with a clear margin and thus improve local control [2,3].

1.1.2 Radiotherapy

Across all tumour types radiation plays a key role and is involved in the cure of up 40% of cancers [4]. The main aim of radiation in treating tumours is to induce DNA damage which in turn mediates mitotic cell death when attempting cell division. Damage occurs in two ways; directly, where ionising radiation interacts with target atoms in the DNA, and indirectly by free radical generation. With photon radiation the latter mechanism predominates. The classical repair pathways following critical double strand break are Homologous Recombination Repair (HRR) and Non-homologous End Joining (NHEJ). Although these two pathways are key mediators of the DNA damage response pathway it has become clear recently that post translational modifications have a key role to play in coordinating the cell's repair [5]. This can include modification of the proteins themselves, phosphorylation of cyclin dependent kinases to control cell phase, ubiquitylation and sumoylation, and the regulation of checkpoints.

Modern radiotherapy techniques allow for reduced toxicity due to improved accuracy and modulation of radiation delivery. But there are limits to the improvements to be gained through technological advancements alone; mechanisms of radioresistance must be better understood to optimize the therapeutic window between efficacy and toxicity [6]. The scale of this conceptual challenge is perhaps best illustrated by Epidermal Growth Factor Receptor (EGFR) example. The scientific basis and rationale for targeting EGFR was sound [7]. The approach appeared to be validated by the positive results of the addition of anti-EGFR therapy, cetuximab, to radical head and neck radiotherapy [8] in a randomized phase III trial setting. At the time it was hailed as a success story and heralded a new era of biologically driven radiosensitization strategies. However, a subsequent negative phase III and metanalysis demonstrated that it was, in fact, not superior to cytotoxic radiosensitization with cisplatin [9,10]. The publication of two randomized phase III trials (RTOG 1016 [11] and DE-ESCALaTE HPV[12]) demonstrating inferior survival with

cetuximab, versus cisplatin, in good prognosis HPV driven oropharyngeal tumours ultimately means that no molecularly targeted therapy has demonstrable clinical utility as a radiosensitizer. Similar failures were seen in lung cancer [13] (RTOG 0617) and oesophageal cancer [14] (RTOG 0436).

With hindsight, targeting of canonical intracellular signaling pathways may not have been an attractive rationale for improving the therapeutic index and/or influencing tumour biology. Intracellular cross talk and dynamic changes in cellular biology when selective pressure is applied by such targeting make meaningful improvements difficult to maintain. It can also alter tumour biology in unforeseen ways which undoubtedly played a part in these negative studies. Another problem is the underlying heterogeneity of tumour populations. Without enriching any of the above studies for patients most likely to respond to the targeted therapy, the benefits are diluted out.

But the wider issue for the radiation community is that we do not have reliable biomarkers for the main therapy – radiation [15]. Responses to radiation, across all tumour sites, are markedly heterogeneous. Although we can broadly categorize tumour groups as sensitive (SCC head and neck, basal cell carcinomas, lymphomas) and resistant (gliomas, melanoma, sarcomas) this belies a range of patient level sensitivities. Perhaps the best illustration of this is the Genome Adjusted Radiation Dose (GARD) score [Figure 1-2] [16]. The score itself is underpinned by two factors; an mRNA based, 10 gene, individualized radiosensitivity score [17] and the linear quadratic model [18]. In a dataset of 8,271 patients across 20 sites, Scott et al demonstrated the significant variation in radiation response within different tumour types. Although each tumour has a median value in line with historical expectations of tumour sites, the wide variation within tumour types is difficult to predict.

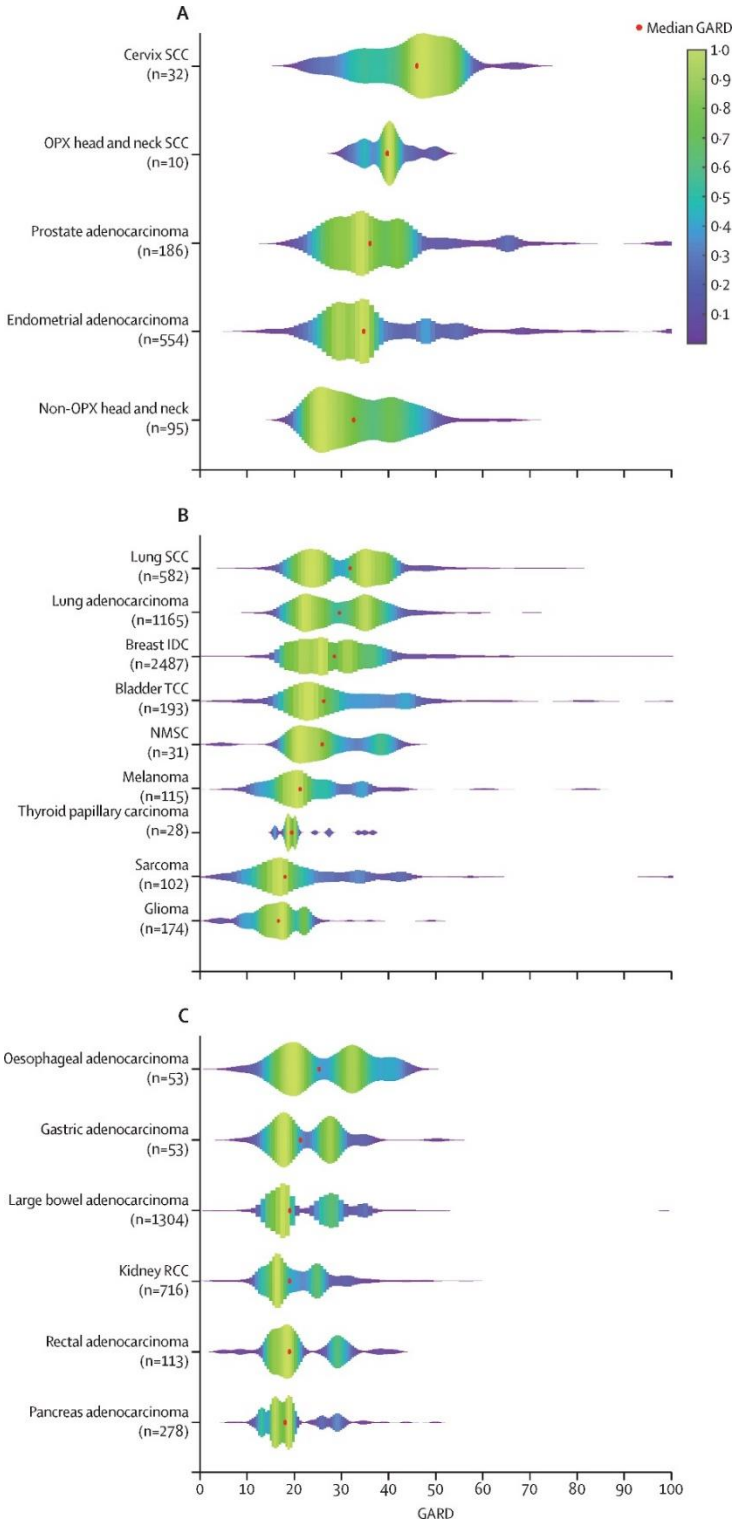


Figure 1-2. GARD score distributions across 20 different tumour types

The GARD score is shown on the x-axis, where a high score indicates higher radiosensitivity. The median for each group indicated by the red dot. The colour of the violin plot represents the sample density. The y-axis indicates the tumour type and the number of samples. Panel A= tumour types treated with >70Gy; B = 60Gy and C = 45Gy. Within each tumour group there is significant heterogeneity of radiosensitivity [Reproduced under licence 4635560581355 by Elsevier publishing from Scott et al [16]]

They also highlight that while some sites would be prescribed 45-50Gy as a standard of care, higher doses may in fact be required to overcome radioresistance. Without understanding which patients are intrinsically sensitive or resistant to radiation and why, strategies to improve radiosensitivity are likely to continue to falter.

1.1.3 Radiotherapy for rectal cancer

In the rectal cancer setting, optimal oncological outcomes are achieved with a multidisciplinary approach, with the aim of achieving local control through Total Mesorectal Excision (TME) in conjunction with successful sphincter preservation [19]. For many years, the standard of care has been a multimodal approach incorporating neoadjuvant long-course chemoradiotherapy (CRT) with concurrent fluoropyrimidines followed by total mesorectal excision surgery [3,20]. Accordingly, local recurrence rates have fallen from 30% - 45% to less than 10% in routine clinical care.

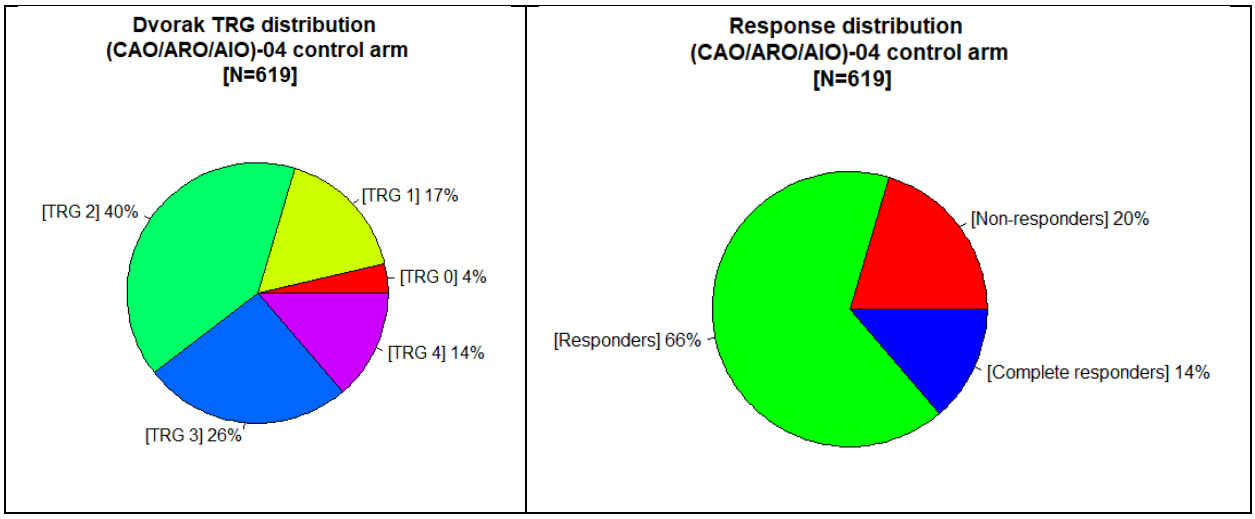


Figure 1-3. Response assessment for the standard of care arm in the CAO/ARO/AIO-04 trial

***Left**: Regression grading using the Dvorak system, where TRG 4 = complete pathological response. **Right**: Summary of the regression scores into three major response categories.*

Over this time three primary response patterns have emerged; a) those who achieve complete pathological response [~15%], b) those that have some response to radiation but have residual tumour [~65%] and c) those who have very little gross tumour response [~20%] [3,21] [Figure 1-3]. Sustained low local recurrence rates has shifted the focus to these response patterns and improving outcomes in other ways. The two high priority areas identified by the Association of Coloproctology of Great Britain and Ireland (ACPGBI) consensus, on prioritizing clinical research questions in colorectal disease, are treatment de-intensification with the option of organ preservation (surgery deferral) and treatment intensification to improve response rates [22]. While the majority of patients can continue to have standard neoadjuvant therapy and derive some benefit, alternative approaches are emerging or need to emerge for those at the extreme ends of the spectrum. Radiosensitive patients are increasingly being offered organ preservation strategies and surveillance as evidence accrues that such strategies appear to be safe and offer the patient genuine benefit of avoiding or deferring surgery [23,24]. However, this strategy has yet to be subjected to a randomised trial so level 1 evidence to support the approach is still lacking. Furthermore, the patients will still have radiotherapy and could be identified following treatment using clinical and radiological surrogates. Identifying this subgroup of patients from a pre-treatment biopsy may be considered of questionable benefit unless it materially changes their treatment course.

Predicting patients with radioresistance is potentially more beneficial. Receiving neoadjuvant chemoradiation without any meaningful response in the disease dramatically alters the risk: benefit ratio. Identifying resistant and/or poor prognostic patients up front from a pre-treatment biopsy would make a meaningful difference to their treatment, either through avoidance of radiotherapy or by enrolment in novel intensified strategies, such as radiosensitization or dose escalation, with a view to improving treatment response.

1.1.4 Failures in intensification in rectal cancer

1.1.4.1 Cytotoxic radiosensitizers

There have been many attempts at increasing response rates to neoadjuvant chemoradiotherapy, but only very few have progressed to the Phase III trial setting. The current, completed, UK based ARISTOTLE trial tested the combination of irinotecan in combination with standard of care fluorouracil chemoradiotherapy based on several early phase trials showing acceptable toxicity and pathological complete response (pCR) rates in the range of 19% to 26% [25–28]. Results are keenly awaited to see if there will be a new standard of care.

Another cytotoxic agent, oxaliplatin, has been extensively trialled at Phase III level in five separate studies [29–33]. Of these, only the German CAO/ARO/AIO-04 study resulted in a significant improvement in pCR (13% vs 17%) when oxaliplatin was added and met its primary outcome of improved 3 year DFS. However, the study mandated that oxaliplatin be given as adjuvant chemotherapy also which may have resulted in the improved survival outcomes. All studies concluded that the addition of oxaliplatin resulted in increased G3/G4 toxicity. Omission of oxaliplatin remains the standard of care.

1.1.4.2 Targeted therapy

The advent and success of biological therapies in metastatic CRC evolved into the combination of a wide variety of these agents including bevacizumab [34] (anti-VEGF), cetuximab [35] (anti-EGFR), nelfinavir [36] (Akt inhibitor) and bortezomib [37] (proteasome inhibitor) being combined with neoadjuvant chemoradiotherapy for rectal cancer and have been thoroughly reviewed elsewhere [38]. Although all were underpinned by supportive pre-clinical evidence and rational hypotheses, they all failed to meet their primary endpoints or proved to be too toxic to recommend. A recent phase Ib trial of combination veliparib (PARP inhibitor) and fluorouracil

chemoradiation (50.4Gy/28 fractions) for rectal cancer caused some excitement when it demonstrated a pCR rate of 29% and an overall response rate of 71% in 31 patients [39], without reaching a maximum tolerated dose. Unfortunately when tested in the NRG-GI002 phase II platform trial as the first experimental arm it failed to reach its primary endpoint of a decrease in median neoadjuvant rectal (NAR) score [40]. It should be noted that it was combined with more intensive chemotherapy than the phase Ib study which could have mitigated the effect of veliparib. A response rate of 71% is not entirely different from an average cohort of standard of care patients, so although the complete response rate was higher than historical comparisons, the failure at phase II level may not be unexpected.

1.1.4.3 Validity of surrogate endpoint

The failure to convert the improvements of early phase trials in rectal cancer, which appear to show some benefit, may not be entirely biological. A recent systematic review of phase II trial designs of treatment intensification in rectal cancer reported a number of methodological problems [41]. Of 92 trials that met the eligibility criteria, only 46% mandated MRI staging at baseline, 13% were randomised, 23% failed to report their sample size calculations and only 41% adequately reported their statistical design and treatment compliance. More importantly, there were 15 different primary endpoints used, although every single study reported on pCR rate. Recall that patients who achieve a complete response are the smallest group of interest (~15%) and not achieving a pathological response does not equate to radioresistance. Therefore selection of and consistency in primary endpoints are crucial.

The EXPERT-C trial [35] randomised 165 patients to 12 weeks of chemotherapy followed by chemoradiation with or without cetuximab in KRAS/BRAF wild type patients. It was a negative study for its primary outcome of complete response. The addition of cetuximab actually significantly increased the response rate to neoadjuvant therapy from 75% to 93% ($p = 0.028$)

and improved overall survival (OS) (HR 0.27, p=0.034). A response rate of 93% is far higher than that achieved or expected with standard of care treatment. If the trial had chosen a different primary endpoint it may not have been regarded as 'negative' and could have generated a new treatment paradigm for that subgroup of patients.

The distracting focus on complete response is underlined by a recent meta-analysis of 22 trials with 10,050 patients confirmed that pCR is not a surrogate of 5 year OS [42]. The individual-level surrogacy of pCR rate and 3-year DFS were poorly correlated with the 5 year OS (R=0.52; 95% C.I, 0.31 - 0.91; p=0.002; and R=0.60; 95% C.I, 0.36 - 1; p=0.002) respectively. The trial-level surrogacy analysis confirmed that the treatment effects on surrogates (ΔpCR rate and Δ3yDFS) were not strong surrogates for treatment effects on 5 year OS (R=0.2; 95% C.I, − 0.29 - 0.78; P=0.5 and R=0.64; 95% C.I, 0.29 - 1; P=0.06) respectively. They concluded that the use of pCR as a surrogate for 5 year survival is not supported.

The clear need therefore, is for biomarkers of radiation response, but the biomarkers must point towards validated surrogates of longer term outcomes. This, combined with rational high quality trial design, will prevent the ongoing failures of treatment intensification in rectal cancer.

1.1.5 Biomarkers in oncology

A biomarker is defined as "*a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes or pharmacological responses to a specified therapeutic intervention*" [43]. A prognostic biomarker is one which provides information on the risk of future event, usually disease recurrence (DFS) or death (OS) [44]. The most widespread example across all tumour sites is disease stage. For a biomarker to be

predictive it must give information on the effect of a therapy for a patient, which is usually although not always, the target of the intervention [44]. Oestrogen receptor for tamoxifen or HER2 for trastuzumab are common oncological examples. Some biomarkers are both prognostic and predictive; HER2 in breast cancer indicates a worse survival outcome and responsiveness to trastuzumab [45].

There are a number of conceptual requirements for the development of a candidate biomarker for clinical decision making; analytical validity, clinical validity and clinical utility [46]. Analytic validity refers to technical factors which could affect the results of the biomarker assay. These include fixation of tissue and length of storage which can degrade quality [47] or the manner in which the samples are processed such as batch effects from changes in laboratory reagents. Biomarkers that can demonstrate relative insensitivity to these effects, such that it can be detected in FFPE tissue and fresh tissue for example, are more likely to be clinically useful.

Clinical validity, or biological validity, refers to a biomarker's ability to divide a population into two or several distinct subgroups. This division should support either the findings from a discovery set or an *a priori* hypothesis from the underlying biology, or both. For a prognostic biomarker this would mean that respective subgroups have different survival outcomes. A predictive biomarker should separate out groups in line with the treatment response. Such findings should then be independently validated in different datasets using the original methodology [48].

The burden of proof for routine clinical use and decision making requires very high levels of evidence [49]. Ideally this should be embedded as part of a clinical trial, but this is often not

practical or cost effective. An alternative is to use clinical trial datasets, with prospectively collected tissues, to validate and determine clinical utility. Clinical utility is the most difficult hurdle in biomarker development, which is why in spite of many purported biomarkers in the literature, few are in clinical use. Recent biomarkers which have redefined the treatment paradigm are OncoType Dx in early stage, ER positive breast cancer [50,51] and KRAS mutation in metastatic CRC [52].

1.1.6 Current biomarkers in colorectal cancer

An improved understanding of the biology of CRC may provide the basis for stratification of patients into differing treatment programs, first by identifying prognostic effects. This is turn requires any novel biomarkers to be examined in the context of the current prognostic and predictive biomarkers of CRC.

1.1.6.1 Tumour location ('sidedness')

Suspected prognostic differences in observed outcomes between right and left colon were confirmed by a large meta-analysis of 66 publications and totalling 1,437,846 patients which demonstrated a pooled HR of 0.82 (95% C.I 0.79-0.84; p<0.001) for left sided colon cancer compared to right sided cancers [53]. In a separate pooled analysis from three randomised controlled trials in the first line metastatic setting, left sided tumours had superior progression free survival (PFS) (HR 0.44, 95% C.I 0.28 – 0.7; p <0.001) and OS (HR 0.52, 95% C.I 0.36 – 0.75; p <0.001) [54]. The effect was independent of mucinous histology and BRAF mutation status. This suggests that although sidedness is clearly a surrogate for biological factors, those are not necessarily fully elucidated, and not entirely the result of microsatellite instability or BRAF mutation status. Another large meta-analysis in the first line metastatic setting confirmed that there is no difference between the colon and rectum in terms of survival outcomes [55].

1.1.6.2 Microsatellite instability High (MSI-H)

MSI-H tumours have distinct clinic-pathological characteristics which distinguish them from microsatellite stable tumours [56]. Two common mechanisms result in the phenotype. Hereditary germline mutation of one mismatch repair (MMR) genes MLH1, MSH2, MSH6 or PMS2, often referred to as Lynch syndrome [57] and hypermethylation of the CpG island in the MLH1 promoter region [58]. The latter, sporadic form shows a preponderance for BRAF mutation. In early stage disease the MSI-H group accounts for about 15% of cases [59].

A meta-analysis of 32 eligible studies and a total of 7,642 cases, 1,277 of which were deficient in MMR found a combined HR estimate for OS of 0.65 (95% C.I 0.59 – 0.71, heterogeneity p = 0.16) [60]. The effect was maintained when the analysis was restricted to those patients treated in the context of a clinical trial. The effect was restricted to those patients with stage II or III disease. This confirmed that for early stage disease patients, deficient MMR is a good prognostic biomarker. Interestingly, the biomarker has the opposite effect in metastatic disease where it accounts for approximately 3-5% of cases [61]. Compared to proficient MMR cases, dMMR cases had inferior OS (HR 1.35, 95% C.I 1.13 – 1.57; p = 0.001).

The prognostic difference in MSI-H tumours has therapeutic implications, where stage II cancers are recommended to avoid adjuvant 5-FU treatment due to lack of improved survival outcomes [62,63], and decisions are based on IHC loss of mismatch repair proteins. For stage III patients, where combination chemotherapy would be recommended based on clinic-pathological factors the presence of MSI-H should not affect the treatment recommendation [62]. Therefore MSI status remains a prognostic biomarker, as opposed to predictive for adjuvant therapy.

It may have a predictive effect in the metastatic setting for immunotherapy, however. There are currently three therapies licenced – pembrolizumab, nivolumab, ipilimumab – on the basis of the

three Phase II trials [64–66]. The underlying biological responsiveness in high mutation burden tumours to PD-1/PDL-1/CTLA-4 inhibition is not fully understood as evidenced by primary resistance in some patients. Additionally, the absence of phase III randomised data recommends against its classification as a true predictive biomarker, at least for the time being.

1.1.6.3 BRAF

Hotspot mutations in codon 600 of the *BRAF* gene (*$BRAF^{V600E}$)* are present in approximately 10% of all CRC and show significant co-occurrence with microsatellite instability (MSI-H) [59]. In a retrospective analysis of the 2,559 patients with stage III disease in the adjuvant PETACC-8 trial, mutation in BRAF was associated with inferior DFS and OS (HR 1.74 95% C.I 1.14-2.69; p = .01; HR 1.84, 95% C.I 1.01-3.36; p = .046), respectively [67]. The effect was not significant in all microsatellite unstable (dMMR) tumours, due to competing risk. The poor prognosis was consistent in stage IV disease also from the meta-analysis of first line randomised controlled trials (HR 2.21, 95% C.I 2.09 - 2.34; p < .001) [55].

BRAF mutations have no predictive effect for therapy, unlike melanoma. Rapid feedback activation of EGFR, following inhibition of BRAF, causes continued proliferation of BRAF activated tumours, a mechanism lacking in melanoma cells which contributes to their sensitivity [68]. Attempts to overcome such resistance using multiple pathway inhibitors are ongoing and early results from the BEACON trial suggests some efficacy with acceptable toxicity [69].

1.1.6.4 KRAS

Mutations in *KRAS* occur mainly in codons 12 and 13 and are present in approximately 40% of cancers [59]. In stage III disease, mutations in KRAS were associated with shorter time to

recurrence (HR 3.02, 95% C.I 2.32 - 3.93; p < .001) and shorter survival after relapse (HR 1.20, 95% C.I 1.01 - 1.44; p = 0.04) in an analysis of over 4,000 patients [70], confirming results from the analysis of the PETACC – 8 trial [67]. It is also prognostic in the first line metastatic setting [55], including those patients proceeding to metastatectomy [71]. The prognostic effects are independent of BRAF status.

KRAS is also a predictive biomarker, where wild type status identifies a sub-population within which truly responsive tumours are found, for anti – EGFR therapy. The benefit was first suggested in an unselected randomised trial where KRAS mutation was a predictor of worse outcomes in the cetuximab group [72]. Although subsequent trials showed no initial OS benefit from the addition of cetuximab to standard of care cytotoxic chemotherapy in the metastatic setting, response rates were higher in the wild type cohort [73]. However, the presence of alternative activating RAS mutations which negate the benefit of EGFR suppression were not included in the primary analysis. On subsequent meta-analysis, of nine randomised controlled trials encompassing almost 6,000 patients, pan RAS wild type patients were found to have significantly improved PFS (HR 0.62, 95% C.I 0.50-0.76) and OS (HR 0.87, 95% C.I 0.77-0.99) when anti-EGFR therapy was added [74]. Interestingly, KRAS status appears to have some competing interaction with 'sidedness', where right sided cancers appear to derive less benefit in spite of having wild type status [75].

A complicating factor in the evolution of the understanding of KRAS as a biomarker in the clinical trial setting has been its dual role, where wild type isboth a prognostic factor for improved OS and predicts for improved responses to anti-EGFR therapy. The latter has been more difficult to prove without an appreciation that the biomarker was already selecting out a better performing population. Therefore it is important that potential predictive biomarkers are

assessed for both prognostic outcomes prior to or in parallel with assessment of the predictive ability.

1.1.6.5 HER2

In a pooled analysis of 3256 patients from three randomised controlled trials, amplification of *ERBB2/HER2* was present in 1.3% of stage IV patients and 2.2% of stage II-III [76]. There was no difference in PFS or OS suggesting that it has no role in prognosis, unlike breast and gastric cancer. However, HER2 signalling may provide a means of mechanistic escape or acquired resistance following selective pressure with anti-EGFR therapy [77], by de novo amplification of HER2 or higher circulating levels of ERBB3/4 ligands heregulin, respectively. The same work showed that inhibition of HER2 restored sensitivity to cetuximab in xenograft models. This hypothesis was further tested in the HERACLES trial [78], where anti-EGFR treatment refractory KRAS wild type, HER2 positive patients were enrolled in an open label phase II with dual HER2 inhibition. The combination of both trastuzumab and lapatinib had an objective response rate of 30%. Thus, although not prognostic, HER2 testing may have predictive value for non-response of RAS wild type patients to anti-EGFR therapy.

1.1.6.6 Consensus Molecular Subtypes (CMS)

All of the above biomarkers have been based on either clinical factors as biological surrogates or mutation within the DNA of the tumour itself. Stratification of tumours based on the transcriptomic profile of tumours is a generally accepted strategy in the pursuit of better biological understanding and thus biomarker discovery [79]. In this drive towards a molecular taxonomy of CRC multiple expression based classifications were published [80–83]. However, the agreement between each of these was relatively poor, with all only agreeing that a hypermutated

phenotype existed. In an effort to harmonise these efforts to synthesise a unifying transcriptomic molecular classification, the consensus molecular subtypes, were produced [84].

Just under 80% of tumours were placed into 4 separate subtypes - CMS 1 (immune), CMS 2 (canonical), CMS 3 (metabolic) and CMS 4 (mesenchymal) - with 22% unable to be classified. Each subtype has a different prognostic outlook. CMS 4 has the worse prognosis across all stages of disease. CMS 1, which has a moderate prognosis in early disease has worse outlook than CMS 4 in a restricted analysis of metastatic disease. In general, CMS 2 has the best outlook regardless of disease stage. Although the analysis was performed in an unsupervised fashion to avoid bias, in other words agnostic of the underlying biology, the authors have annotated each subtype with certain enrichments in an effort to place the subtypes in the landscape of colorectal biology. An important caveat is that these enrichments are not pathognomonic of the subtypes [Figure 1-4].

The implication is that there may be other transcriptomic mechanisms involved, captured in this global analysis, which influence the clinical outlook for these patients but which have not yet been determined. Novel expression based classifiers must be considered in the context of the CMS to help to refine the current molecular taxonomy outlined in Figure 1-4 and identify future areas of biological exploration.

CMS1 MSI immune	CMS2 Canonical	CMS3 Metabolic	CMS4 Mesenchymal
14%	37%	13%	23%
MSI, CIMP high, hypermutation	SCNA high	Mixed MSI status, SCNA low, CIMP low	SCNA high
BRAF mutations		*KRAS* mutations	
Immune infiltration and activation	WNT and MYC activation	Metabolic deregulation	Stromal infiltration, TGF-β activation, angiogenesis
Worse survival after relapse			Worse relapse-free and overall survival

Figure 1-4. Proposed molecular taxonomy of colorectal cancer from the Consensus Molecular Subtypes

Each subtypes defined by gene expression grouping has been annotated with enrichments for multiple other known factors of colorectal biology. CIMP, CpG island methylator phenotype; MSI, microsatellite instability; SCNA, somatic copy number alterations. [Reproduced under licence 4638190002060 by Elsevier publishing from Guinney et al [84]]

1.1.7 Biomarkers of radioresistance in colorectal cancer

There are currently no validated biomarkers of radioresistance in CRC. Although all of the above biomarkers are useful for prognostication, they have little or no use in stratifying patients for radiotherapy. One study suggested that the presence of KRAS codon 13 mutations predicted for a worse response to chemoradiotherapy, especially in co-occurrence with TP53 [85], a finding also suggested in a smaller recent cohort of 17 patients [86]. Interestingly, a similar phenomenon was seen in a phase II basket trial of liver metastases treated with SBRT [87]. KRAS and TP53 mutant metastases have inferior local control compared to wild type patients, in spite of identical radiation doses and high quality radiotherapy quality assurance. This data suggests that the mutation pattern may identify a more radioresistant phenotype of CRC.

One of the most widely reported radiosensitivity signatures in the literature [17,88–90], the radiosensitivity index (RSI), has been tested in a rectal cancer cohort [91]. They reported an 80% sensitivity and 82% specificity for prediction of response versus non-response. The small cohort size (N=14) and unusual definition of response, where a decrease of at least 1 'T' stage between the pre-treatment EUS and the final pathological stage, make it inherently unreliable as a biomarker however. No other groups have validated this expression based gene signature in spite of the fact that it is comprised of only 10 genes, further calling into question its clinical utility.

A recent systematic review of molecular biomarkers of radiation response in rectal cancer highlighted several important areas for future research [92]. Poynter et al identified 74 eligible studies and used Gene Ontology (GO) terms to build a hierarchical classification of each biomarker under 5 headings: 'cell cycle', 'cell death', 'response to stimuli', 'cell signalling' and 'cellular metabolism'. Intriguingly, they demonstrate that cell cycle related biomarkers such as p53 and p21 have very little association with radiation response in rectal cancers, in spite of being a fundamental component of cancer cell regulation and radiation response *in-vitro*. In contrast, the greatest number of significant biomarkers were in the 'response to stimulus' domain and included biomarkers such as MSH2 [93]. MutS Homolog 2 (MSH2) is part of the MMR system, as previously discussed, supporting the preclinical data from Yard et al [94] regarding the intrinsic radiosensitivity of dMMR cell lines. Although relatively small in number, the biomarkers that were evaluated in the 'cellular metabolism' domain demonstrated a high correlation with tumour regression (TRG), some of which are expanded upon here.

1.1.7.1 TGF-β and SMAD

Transforming growth factor β (TGF-β) is responsible for a wide variety of cellular processes including proliferation, apoptosis, immune activation/suppression and adhesion [95]. A ligand, it

binds to a heterotetrameric receptor consisting of two subunits– type 1 and II – the former of which is responsible for phosphorylation of intracellular SMAD2 (small mothers against decapentaplegic Homolog 2) and SMAD3 proteins which translocate to the nucleus and activate transcription programs [95]. In a cohort of 86 patients treated with neoadjuvant chemoradiotherapy phosphorylated SMAD3 overexpression, scored by IHC on the pre-treatment biopsy, was found to be correlated with poor tumour regression on multivariate analysis [96]. This implies that high levels of TGF-β activation at baseline promote a radioresistant phenotype. Although no specific mechanism is inferred, various known actions of the TGF-β pathway could produce this effect, by promoting an aggressive phenotype or an creating an immunosuppressive environment, increasing angiogenesis or all of the above acting in concert [97].

1.1.7.2 Matrix metalloproteinases

Matrix metalloproteinases (MMP) are part of a superfamily of proteases that are responsible for digestion and degradation of collagen in the extracellular matrix. They are increasingly recognised as important regulators of tumour cell growth, apoptosis and angiogenesis as well as interacting with stromal signalling pathways such as TGF-β [98]. In a 91 patient study, lower baseline levels of MMP-9 were seen in those patients who achieved a pCR compared to those without [99].

1.1.7.3 Nucleotide synthesis

Several papers have linked the *de novo* metabolic synthesis of nucleotides to clinical response in rectal cancer. Low mRNA levels on pre-treatment biopsies of thymidylate synthetase and thymidine phosphorylase, enzymes responsible for biosynthesis of thymidine nucleotides and its precursors, were associated with more regression in the final resection samples of 40 patients [100]. Similar findings were reported elsewhere with thymidine phosphorylase [101]. Whether the

changes in enzymatic levels induce true radiosensitivity or an increased sensitivity to co-prescribed 5-FU, as suggested by other work using single nucleotide polymorphisms (SNP) of thymidylate synthetase and associated enzymes [102], is not clear.

1.1.7.4 Immune markers

Given the radio-responsive nature of MSI-H tumours and the high immunogenicity of these tumours, it is logical to ask whether an immune infiltrated phenotype of rectal cancer exists and does it affect response to chemoradiation. An integrative comprehensive molecular characterisation of pre-treatment biopsies and post treatment resections in 17 rectal cancer patients highlights some important hypothesis generating work [86]. Although there was no difference in tumour mutation burden or neoantigen formation following chemoradiation, interferon alpha response genes were significantly enriched in pre and post samples in the responders. In non-responders there were significantly more M2 macrophages and naïve B-cells. This phenotype had previously been predicted as clinically unfavourable [103]. Furthermore, the interplay between these immune markers and known molecular phenotypes like CMS adds yet more complexity.

From this introductory overview it is clear that predicting radiation response, from the pre-treatment material available, is not straightforward. Novel avenues of enquiry must be explored and then placed in this complex landscape of epithelial, stromal and immune interactions.

1.2 Overview of the KEAP1 - NRF2 pathway

NF-E2-related factor 2 (NRF2) is a potent transcriptional activator that plays a central role in cell protection against oxidative and electrophilic stress. It is a member of the cap 'n' collar (CNC) subfamily of transcription factors with seven highly conserved domains, Neh 1-7, along its structure [104]. After translation the protein is rapidly targeted for ubiquitin mediated degradation by the proteasome in the cytoplasm [105], preventing translocation to the nucleus [Figure 1-5, A-B]. This process is facilitated by KEAP1 (Kelch Like ECH Associated Protein 1), part of the CUL3 ubiquitin ligase family[104]. KEAP1 binds to two motifs in the Neh2 domain of NRF2 – D_{29}LG and E_{79}TGE – which will be important in Chapter 3.

The exact mechanism of KEAP1-NRF2 interaction is still an area of active investigation; however two models prevail [104]. The 'hinge and latch' model, which are the ETGE and DLG motifs respectively, where KEAP1 binds, are modified by changes in thiol residues. In turn this causes misalignment with lysine residues in NRF2, resulting in dissociation or weak binding [Figure 1-5 D]. The Cul3 dissociation model suggest that the thiol residue modifications actually prevent Cul3 binding and thus prevent ubiquitination [Figure 1-5 E]. The KEAP1 – NRF2 complex remains intact and KEAP1 is cleaved at a later stage.

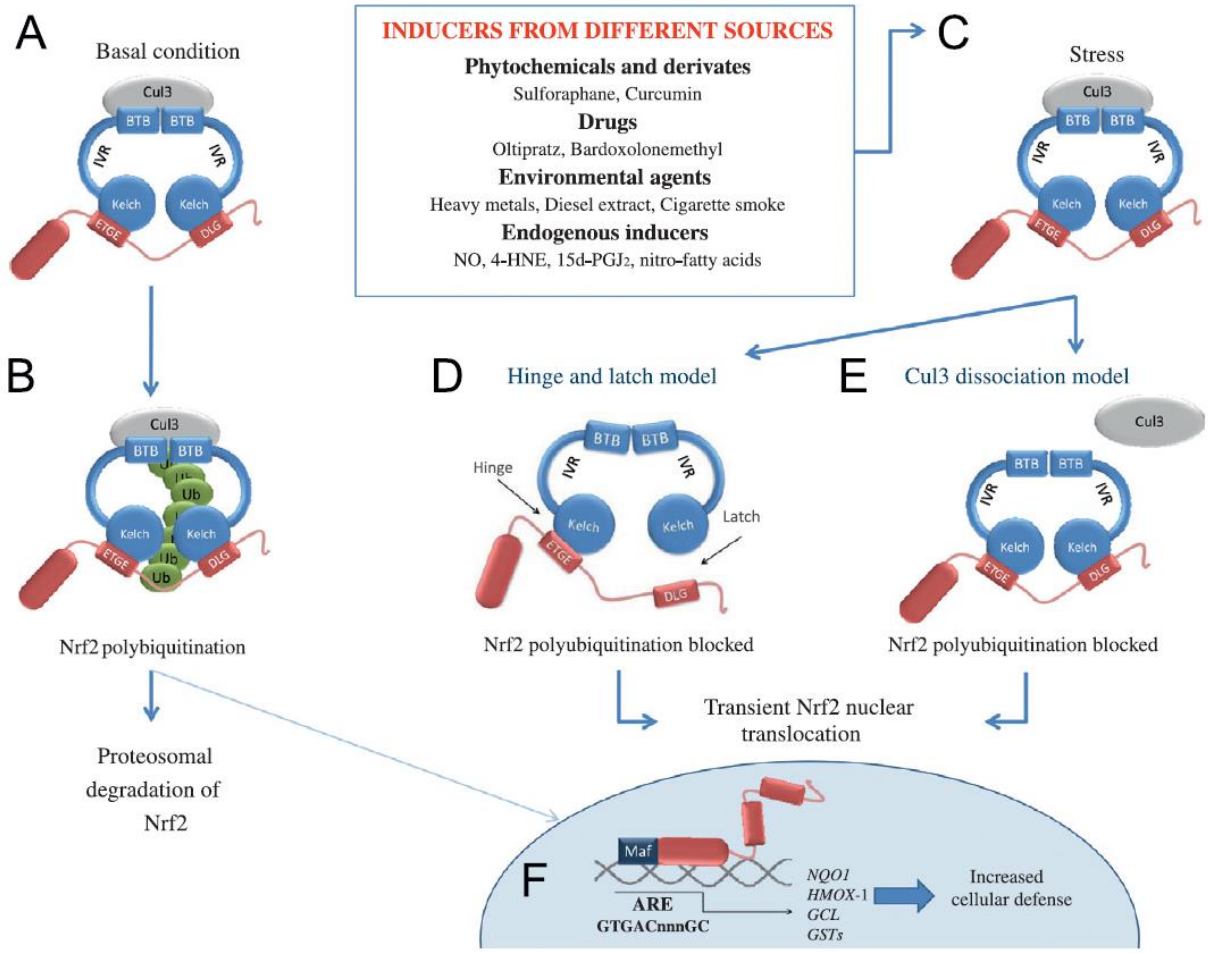


Figure 1-5. The KEAP1 – NRF2 signalling pathway

Under basal conditions, NRF2 is bound by 2 KEAP1 molecules and targeted by the Cul3 ligase for ubiquitination and subsequent proteasomal degradation. Endogenous and exogenous stresses induce conformational changes in KEAP1 or Cul3 which inhibits ubiquitination and allows for nuclear translocation where it binds to Anti-oxidant Response Element (ARE) consensus sequences in the promoter regions to transcriptionally activate cellular defence programs and maintain homeostasis. [Reproduced under licence 4635311116466 by Elsevier publishing from Kansanen et al [106]]

The net effect of either model is cytoplasmic accumulation and subsequent nuclear translocation of NRF2. Within the nucleus, the Neh7 domain of NRF2 binds to Anti-oxidant response element (ARE) [107] or electrophile response element (EpRE) [108] consensus sequences in the promoter regions of target genes. The consensus sequence of TGACnnnGC – where 'n' represents any base - is well characterised [109]. Binding is facilitated by heterodimerization with small MAF proteins [Figure 1-5 F].

During the last decade, chromatin immunoprecipitation deep sequencing performed by a number of groups has identified the target genes of the NRF2-sMAF dimer [110,111]. Independently they identified approximately 240 target genes. The transcriptional programs involving these genes include intracellular redox control, metabolic pathways, autophagy and drug transport [112]. One important group of target genes is those encoding anti-oxidant and detoxifying enzymes. NAD(P)H: quinone oxidorecductase -1(*Nqo-1*) is representative of this group and is widely used to evaluate the signalling activity of the NRF2 pathway [104]. This point is an important consideration for the work that will be presented in Chapter 3, where *Nqo-1* will be used as a surrogate of NRF2 signalling activation.

1.2.1 NRF2 as a 'tumour suppressor'

Historically the NRF2 pathway was deemed to function in a 'tumour suppressor' capacity, allowing the cell to defend against stressors such as carcinogens [113,114]. Many chemical compounds such as sulphoraphane, curcumin, resveratrol and fumaric acid have been used to induce NRF2 signalling, most commonly by reacting with cysteine residues on KEAP1 to prevent or diminish binding and release NRF2 [114]. The net effect across multiple tumour types and cancer models is a decrease in malignant potential [115–117]. *Nrf2* knockout mice show that the use of such chemoprevention approaches is entirely dependent on NRF2 production for the anti-tumorigenesis effect [118,119]. *Nrf2* deficient mice also showed an increase susceptibility to *de novo* mutations [120].

1.2.2 NRF2 as an 'oncogene'

More recent evidence shows that some tumours acquire constitutive activation of the pathway which allow it to function in a more 'oncogene' like fashion[121]. For example, *KEAP1* knockout mice have increased metastatic potential [122].

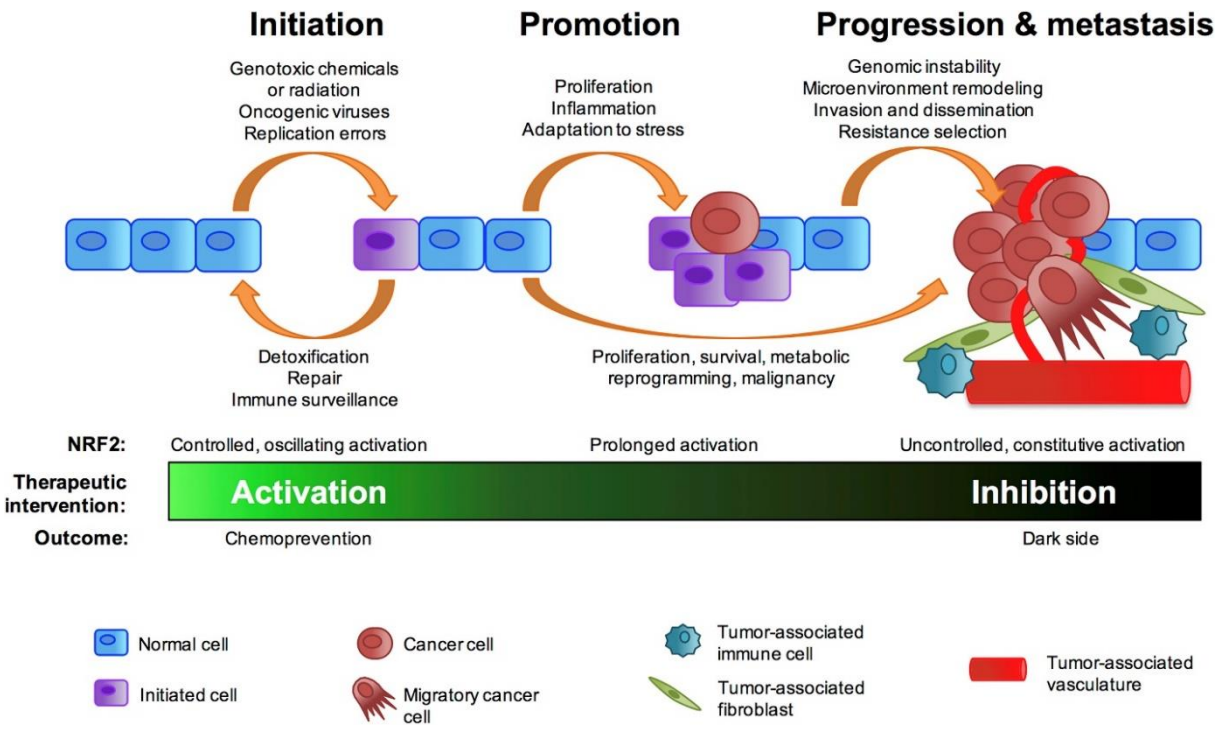


Figure 1-6. The context dependency of NRF2 signalling in cancer progression

During early malignant changes initiation of NRF2 signalling is an important mechanism of cellular defence, which is switched off by feedback mechanisms. Activation by chemopreventative strategies can alter the balance in favour of further suppressing genotoxic stresses and prevent tumorigenesis. Once 'on-off' control is lost and the signalling pathway becomes constitutively activated, the suppression of cellular stresses (radiation, chemotherapy) and regulatory control over other transcription programs creates an environment where inhibition of NRF2 could be therapeutic. [Reproduced under licence 4635310795506 by Elsevier publishing from Rojo de la Vega et al. [112]]

In premalignant states, stimulating NRF2 signalling to dampen immune activation, reduce inflammation and mutagenic stress can decrease malignant progression. But the acquisition of NRF2 control by malignant cells has the opposite effect, whereby decreasing levels of cellular stress makes them resistant to intrinsic defences and extrinsic treatment [Figure 1-6]. The situation is analogous to the role of TGFβ in some tumour types, where it appears to act in a suppressive manner in the early stages of disease before having an increasingly deleterious effect as tumours progress [97]. When this 'tipping point' is reached has not yet been resolved but it likely to influence and be influenced by a host of other cellular changes. This is supported by a recent Pan-Cancer TCGA analysis of 10 canonical signalling pathways of over 9,000 patient samples [123]. Where somatic mutations of the KEAP1-NRF2 pathway occur, loss or gain of

function in *KEAP1* and *NFE2L2* respectively were almost mutually exclusive occurrences, implying that control of the pathway by either mechanism was sufficient. Furthermore the strongest co-occurrence between the 10 canonical pathways examined was between mutations and alteration in NRF2 pathway and *PIK3CA* activation and *STK11* loss. They hypothesise that higher levels of NRF2 activity are required to tolerate the cellular stress induced by PI3K pathway activation.

1.2.3 Mechanisms of KEAP1-NRF2 activation

There are a number of distinct mechanisms by which the NRF2 pathway can become constitutively activated [Figure 1-7]. These can be grouped as genomic DNA alterations, transcriptional regulatory changes and post translational modifications. Mutations in both KEAP1 and NFE2L2 have been described in a number of solid organ malignancies. Across 226 TCGA studies from all organ systems, genetic mutations and copy number alterations occur most commonly in lung squamous cell carcinoma (31.4%), lung adenocarcinoma (24%) endometrial (20.6%), squamous oesophageal carcinoma (19.8%) and head and neck squamous cell carcinoma (17.4%) [121,123]. These tumours are often the most frequently cited NRF2 activated tumours in the literature as a result.

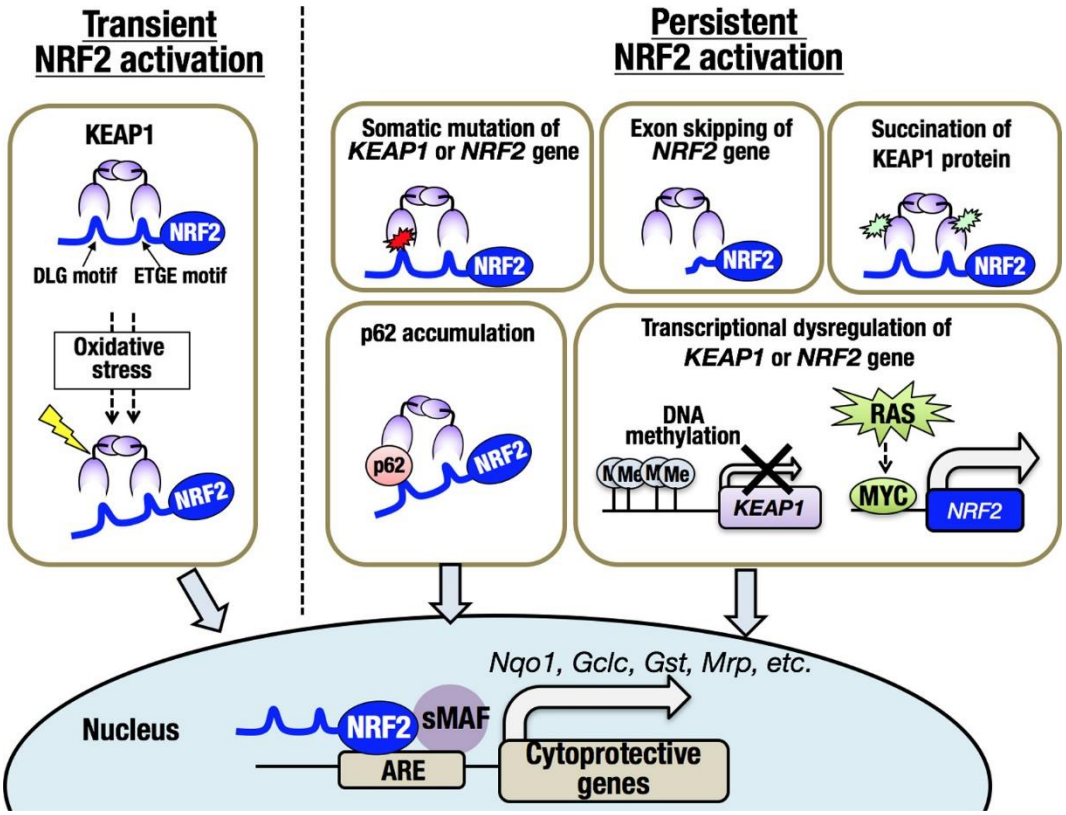


Figure 1-7. Summary of mechanisms of NRF2 activation in cancers

In the physiological, transient mode KEAP1 is bound to NRF2 (Left). During oxidative stress a NRF2 is released to translocate to the nucleus and activate cytoprotective genes. Persistent activation can occur by a number of mechanisms, by affecting the binding of NRF2 by KEAP1, the transcription of KEAP1 or by oncogene directed activation of NRF2. [Reproduced with permission from John Wiley and Sons publishing under CC BY-NC licence from Kitamura et al. [124]]

Mutations of *KEAP1* or *NFE2L2*, particularly those occurring in the crucial binding domains described in section 1.2, allow unchecked NRF2 accumulation and constitutive activation. Even heterozygous KEAP1 mutations are important because KEAP1 proteins harbouring mutations act as dominant negatives to form heterodimers with wild type protein and prevent ubiquitination of NRF2 [125].

Transcriptional changes in signalling can occur by a number of other mechanisms. Epigenetic modifications such as methylation have been demonstrated in KEAP1, silencing KEAP1's ability to regulate NRF2 [126]. More recently there is evidence of direct increase in NRF2 transcription via activated oncogenes such as $KRAS^{G12D}$, $BRAF^{V619E}$ and c-MYC [127],

highlighting that even in the presence of intact regulatory controls NRF2 can be hyper-active. These mechanisms are of particular importance in tumours where somatic mutation rates are low. A Pan-Cancer transcriptomic study of 9,000 patients in the TCGA found expected hyper-activity in NRF2 signalling in the aforementioned tumour types, but found a striking level also in tumours that lacked strong genomic explanations [128]. This included CRC.

1.2.4 KEAP1-NRF2 as biomarker of radioresistance

In Yard et al's seminal paper on radioresistance in 533 cancer lines, several interesting observations are evident [94]. Firstly, only two tumour lineages shown non-Gaussian distributions, uterine and colorectal. This binomial distribution is attributable to a discrete subset of tumours shared by both anatomical sites; hypermutated cancers with a low somatic copy number alteration [59,129]. Hypermutated tumours had increase radiosensitivity and those with high somatic copy number alteration (SCNA) significantly correlate with radioresistance. Secondly, from a list of greater than 1,600 genes *KEAP1* was ranked 9th for radioresistance. Mutations in *KEAP1* had previously been shown to confer radioresistance in prostate and squamous cell lung cancers [130,131] . Similarly, *NFE2L2* mutations are known to confer radioresistance [132]. When analysis was restricted to somatic mutations in the crucial binding domains – D_{29}LG and E_{79}TGE respectively – the highest degree of resistance was found [94]. Lastly, they demonstrated that increased levels of NRF2 transcriptional activity, as measured by an expression score of 565 genes, was strongly associated with radiation survival across all cell lineages.

These data suggest a significant and compelling role for the KEAP1-NRF2 system as a biomarker of radioresistance. In fact, the 2018 report from the American Society for Radiation Oncology/American Association of Physicists in Medicine/National Cancer Institute Precision

Medicine Conference [15] makes specific reference to NRF2 and KEAP1 as mutational biomarkers of radioresistance.

1.2.5 KEAP1- NRF2 in colorectal cancer

There is remarkably little data on the contribution of NRF2 to CRC biology [133], in spite of the clear relevance of some of the outlined mechanisms. This under investigation may be due to much of the evidence of NRF2's role in CRC suggesting a protective effect, often in the chemo-preventative setting [134–137]. Coupled with the very low somatic mutation rate in CRC which is approximately 1% for alterations in either KEAP1 or NFE2L2 [59], it may be understandable how it has been overlooked. But this ignores the context dependency of NRF2 signalling and its complex post transcriptional and post translational mechanisms of activation.

Somatic mutation rates of up to 7.8% have been reported in small cohorts [138] and are likely to be significant when they do occur. Methylation of KEAP1 occurs twice as often in CRC as it does in associated normal tissue [126]. Oncogenes KRAS and BRAF occur in 55% of non-hypermutated tumours [59], with near mutual exclusivity, while MYC is known to exert significant control over transcriptional regulation in CRC [59]. All of these oncogenes could activate NRF2 signalling. High levels of NRF2 expression within tumours, as measured by IHC, have been found to be significantly correlated with p53 expression, Duke's stage and poor clinical outcomes [139]. NRF2 promotes resistance to 5-FU in cell line models [140] while resistant cell models have higher levels of NRF2 expression [141] suggesting a role in treatment resistance in addition to aggressive biological behaviour.

The specific contribution that the KEAP1-NRF2 pathway makes to therapeutic resistance and prognosis in CRC requires active investigation. This is particularly relevant, currently, in rectal cancer where radiation is a key therapeutic modality. However, a paradigmatic shift in the use of radiation is occurring with stereotactic ablative radiotherapy [142]. Tumours which historically were not irradiated, such as right sided colon cancers, are increasingly being treated at oligometastatic sites such as the liver, pelvic nodes and lung. Therefore, understanding the relative contribution of novel pathways should apply to all anatomical colorectal sites and be viewed through the prism of current molecular phenotyping.

1.3 Aims of thesis

1. To investigate the role of the KEAP1-NRF2 pathway in mediating therapeutic resistance in CRC

It is clear from the literature that the NRF2 pathway has a role in radioresistance but that this role is under explored in CRC. To understand if NRF2 was relevant to therapeutic resistance and characterise the magnitude of its contribution, a series of *in-vitro* experiments were carried out in colorectal cell line models. The hypothesis at test was that downregulation of NRF2 expression would induce a radiosensitive phenotype in colorectal cell line models and upregulation would mediate a resistant phenotype. The results of the work in support of this aim are described in Chapter 3.

2. To understand the mechanism of radiosensitivity / resistance following changes in NRF2 pathway expression *in-vitro* in CRC models

The mechanism responsible for producing phenotypic differences in radiation response should be shared by both sensitive and resistant models. The hypothesis at test was that this mechanism

would be best detected at a transcriptomic level due to NRF2's role as a transcription factor. The results of the work in support of this aim are described in Chapter 4.

3. To derive a novel prognostic NRF2 biomarker for use in CRC to identify patients with poor outcomes

The multiple mechanisms of potential activation seen in the NRF2 pathway, coupled with the low level of somatic mutation in CRC, implies mRNA levels of *NFE2L2* are unlikely to fully reflect pathway activation. The hypothesis was that an mRNA signature of known NRF2 regulated genes could be derived to aggregate the expression of the pathway and used as a biomarker in CRC. The results of the work in support of this aim are described in Chapter 5.

4. To examine the predictive value of an NRF2 biomarker in radiation response in rectal cancer

Changes in NRF2 which mediate changes in radiation response *in-vitro* should be validated *in-vivo*. The hypothesis was, using a validated signature of NRF2 expression, higher levels of NRF2 would associate with worse tumour regression following chemoradiation in a cohort of rectal cancer patients. The results of the work in support of this aim are described in Chapter 6.

CHAPTER 2

2.1 Laboratory methods and materials

2.1.1 Cell culture

2.1.1.1 Cell lines

The cell lines used were SW146 (rectal adenocarcinoma), HT55 (rectal adenocarcinoma) and HCT116 (colon adenocarcinoma). All were retrieved from liquid nitrogen storage in the Wellcome Trust for Human Genetics, kind gift of Prof Ian Tomlinson, originally from American Type Culture Collection (ATCC) and European Collection of Authenticated Cell Cultures (ECACC). All cell lines had previously been subjected to karyotyping, comparative genome hybridisation analysis and driver mutation analysis for typing and verification by the Tomlinson lab. Isogenic cell lines $HCT116^{E79K+}$ and parental $HCT116^{WT}$ were obtained from Horizon Discovery®. $HCT116^{E79K+}$ contains a CRISPR Cas9 engineered *NFE2L2* missense point mutation c.235 G>A resulting in an amino acid substitution at position 79, glutamate for lysine.

2.1.1.2 Culture conditions

Cells were grown in Dulbesco's Modified Eagle serum media (DMEM) supplemented with 10% fetal bovine serum (Thermo Fisher, MA) and 100U/mL Penicillin, 100 mg/mL of Streptomycin. All cultures were maintained at 37 °C in a humidified 5% CO_2 atmosphere and tested regularly to ensure absence of Mycoplasma contamination MycoAlert™ PLUS Mycoplasma Detection Kit as per manufacturer's instructions. Adherent cells were grown to 70-80% confluence prior to being washed with Phosphate Buffer Solution (PBS). Cells were incubated with trypsin 0.05% for 5 minutes at 37 °C. Lifted cells were re-suspended by gentle pipetting and transferred to a Falcon tube and centrifuged for 5 minutes at 1000 rpm. The supernatant was removed and the cells re-suspended in fresh growth medium, diluted as necessary and divided into fresh tissue culture flasks.

2.1.1.3 Frozen storage of cells

Following a PBS wash, cells were detached using trypsin and centrifuged for 5 minutes at 1000 rpm. The supernatant was removed, discarded, and cells re-suspended in DMEM with 10% FBS. The cell solution was then transferred in 1mL aliquots to cryo-vials and placed in storage in the -80°C freezer for short term storage or into liquid nitrogen for long term storage.

2.1.1.4 Recovery of frozen cells

Cells were thawed and re-suspended in pre-warmed DMEM. Cell solutions were then centrifuged at 1000 rpm for 5 minutes, re-suspended in fresh DMEM and transferred to a fresh tissue culture flask and incubated at 37 °C as indicated in the cell culture methods.

2.1.1.5 Cell counting

Cells were washed with PBS, incubated with trypsin and re-suspended in fresh medium. Cell concentration was determined with Countess™ Cell Counting System (Invitrogen®) as per the manufacturer's instructions. Cell counts were done in duplicate and an average reading taken. Appropriate cell concentrations were made by diluting with DMEM.

2.1.2 RNA interference reagents and conditions

2.1.2.1 siRNA resuspension

siRNAs were supplied pre-designed as dry powder stock at 5 nmoles (Ambion®). Targeted siRNAs were obtained for NFE2L2, as well as non-targeting (NT1) for negative control and polo-like kinase 1 (PLK1) used as positive control. Visual inspection of cell death caused by PLK1 transfection was confirmed by trypan blue staining before proceeding, with cell death rate >50% sufficient [Figure 2-1]. The siRNA sequences are summarised in the table below. The

powder was suspended in 250 µL of RNA free water to create a 20 µM working stock. This working stock was stored at -20°C as per manufacturer's instructions.

siNFE2L2_1	5' GAAUGGUCCUAAAACACCAtt
siNFE2L2_2	5' CGUUUGUAGAUGACAAUGtt
siNFE2L2_3	5' CAGUCUUCAUUGCUACUAAtt
siPLK1	5' GCAAUUACAUGAGCGAGCAtt

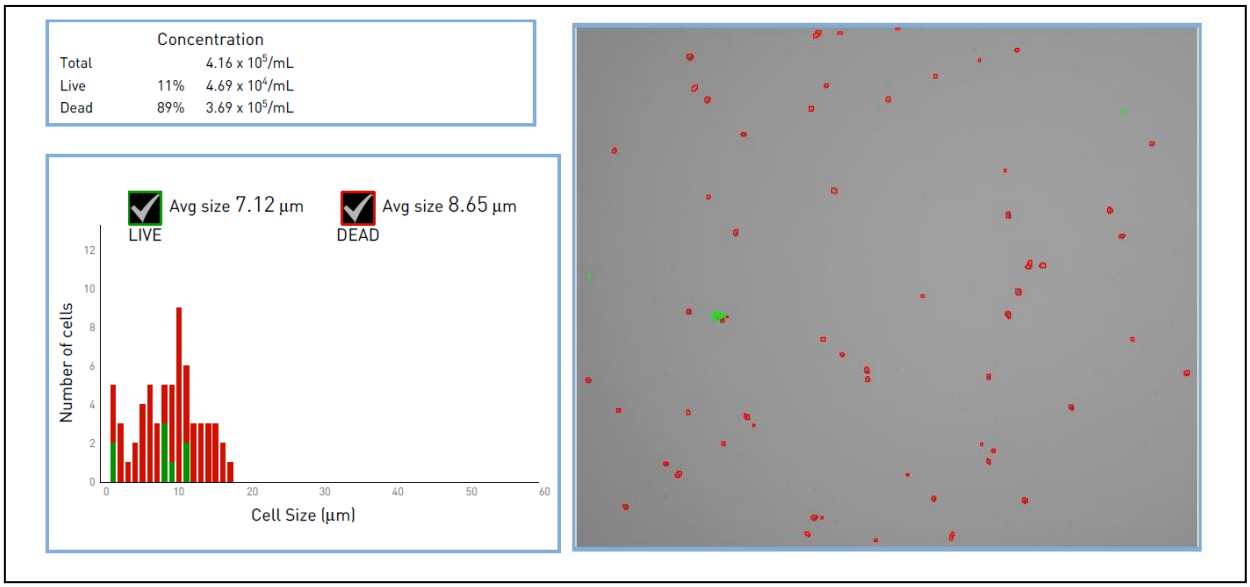


Figure 2-1. Example of Countess™ Cell Counting System output following PLK1 transfection

PLK1 was used as positive control by inducing cell death in cells. This was observed by clinical inspection by microscopy and confirmed by manually lifting and counting cells with trypan blue stain. Cell death is quantified as seen in the top left panel.

2.1.2.2 siRNA transfection conditions

Forward transfection involves plating cells and allowing them to adhere and reach a certain level of confluence, usually 50% or upwards, the day prior to proceeding. The siRNA complex and the transfection reagent is then added on top. Reverse transfection involves plating the siRNA complex and transfection reagent into a well and adding the cell suspension immediately on top. Both forward and reverse transfection procedures are used and are described for 6 well plates.

2.1.2.3 Transfection protocol

2.1.2.4 Forward

180,000 cells per well were seeded 24 hours prior to transfection. For each well, 50 µL of each siRNA from 20 µM stock was diluted in 250 µL serum free medium and mixed gently. 5µL of transfection reagent Lipofectamine 2000 (Invitrogen®) was diluted in serum free medium, mixed, and incubated for 15 minutes at room temperature. The siRNA in serum free medium, mixed gently with diluted Lipofectamine and incubated for 15mins at room temperature (final concentration 1.8 µM). The siRNA-Lipofectamine complex was then added to each well, with fresh medium and replaced in the incubator. Forward transfection was used for HCT116 and SW1463 cell lines.

2.1.2.5 Reverse

The working stock of 20 µM siRNA was diluted 10 X with RNA free water for a final concentration of 2 µM before further dilution with serum free medium. The siRNA Interferin mixture was gently pipetted up and down and incubated for 15-30 minutes at room temperature. The final siRNA concentration for transfection was 20 nM per well. The siRNA:INTERFERin-HTS complex was transferred into each well of 6-well cell culture plates with 2ml cell suspension. The plates were gently swirled and then returned to incubator for an appropriate length of time. Reverse transfection was used for the HT55 cell line, with 300,000 cells per well required.

2.1.3 Gene Expression (mRNA) analysis

2.1.3.1 Quantitative reverse transcriptase PCR

Gene expression, following siRNA treatment, was demonstrated through the measurement of mRNA. RNA was extracted using the Qiagen® RNeasy Mini Kit as per the manufacturer's

instructions. RNA concentration yields were made using the nd-1000 NanoDrop® spectrophotometer. Extracted RNA was converted to cDNA in a two-step PCR process. RNA was diluted with RNA-free water to a final volume of 10 µL and mixed with 1 µL of DNAse 1endonuclease and 1 µL of Mg^{2+}. The final solution was heated to 37 °C for thirty minutes. The reaction was terminated by adding 1 µL of EDTA solution and heated to 65 °C for ten minutes. cDNA was amplified using the High Capacity cDNA reverse transcriptase kit (Applied Biosystems™) as per the manufacturer's instructions.

Briefly, the final reaction mixture was heated to 25 °C for 10 minutes, 37 °C for two hours and 85 °C for 5 minutes before being cooled to 4 °C. Quantitative reverse transcriptase PCR was achieved using the TaqMan® system with 96 well plates as per the manufacturer's instructions. 96 well plates were loaded with triplicate wells for each experimental condition. Quantitative real time RT-PCR analyses of NFE2L2, NQO-1 and HMOX-1 were performed by using assay-on-demand primers and probe sets from Applied Biosystems. Assays were performed by using the ABI 7000 Taqman system (Applied Biosystems). Gene expression was normalised to β-Actin expression. All experimental results were confirmed with two technical replicates. Quantification of RNA gene expression was carried out as using the $\Delta\Delta$ C_t method, where:

$\Delta\Delta CT = \Delta CT$ (treated) – ΔCT (control) = [Treated ($CT_{target} - CT_{reference}$)] – [Control ($CT_{target} - CT_{reference}$)] [143].

The final quantification was expressed as $2^{-\Delta\Delta CT}$.

The resultant relative fold change was expressed graphically as mean ± s.e.m. Statistical comparison of means was performed using Welch's t-test. Significance levels are indicated as previous (*P<0.05, **P<0.01, ***P<0.001 and **** P<0.0001).

2.1.4 Western blot analysis

Whole-cell lysates were clarified by centrifugation and resuspended in 50 µL RIPA buffer with protease inhibitor. Protein quantification was carried out with Pierce™ BCA assay using the manufacturer's instructions and serial dilutions relative to a standard. Spectrophotometer readings were taken at 562 nm. Protein lystates were then diluted according to the BCA assay results to give a loading amount of 10 µg. Proteins were separated by SDS–PAGE and transferred onto 0.45 mM nitrocellulose membranes (Maine Manufacturing; Sanford, ME). After primary antibody incubation for 1–2 h at room temperature, washings, and incubation with secondary antibodies, blots were developed with a chemoluminescence system (Amersham/GE Healthcare). D1Z9C monoclonal rabbit antibody was used for NRF2 expression (Cell Signalling Technology, #12721). Anti-rabbit IgG, HRP conjugated antibody was used as a secondary (Cell Signalling Technology, #7074). β - actin was used a loading control (Abcam, ab20272). Image analysis for protein quantification was performed with open source ImageJ software.

2.1.5 Clonogenic survival

Cells were plated at appropriate dilutions [Table 2-1], irradiated at 2, 4 and 6Gy using a caesium-137 irradiator (Gamma Service GSR D1; dose rate 1.16 Gy/min). They were then incubated for 14–25 days for colony formation. Colonies were fixed in a solution of acetic acid and methanol 1:3 (v/v) and stained with 0.5% (w/v) crystal violet. A colony was defined as consisting of 50 cells or greater. Colonies were counted digitally using GELCOUNT™ (Oxford Optronix) software. Experimental results were confirmed with two technical replicates.

Table 2-1. Summary table of number of cells plated

For each plate, the number of cells seeded at a given radiation dose and its incubation time prior to staining for colony counting.

Cell line	0 Gy	2 Gy	4 Gy	6 Gy	Incubation time (days)
SW1463	200	400	800	2000	14
HT55	200	1000	2000	5000	25
HCT116	200	400	800	2000	14
HCT116 (E79K+)	200	400	800	2000	14

2.1.5.1 Linear Quadratic Modelling

For clonogenic assays involving cell irradiation, the experimental data were fitted with the linear quadratic model (LQ):

$S= \exp(-\alpha D-\beta D^2)$

where,

S is the survival probability

D the radiation dose (Gy)

α and β are the fit parameters (Gy-1 and Gy-2 respectively).

The sensitisation enhancement ratio (SER) was used to quantify radiosensitization (the SER_{10} was deduced from data by using $SER_{10} = D_{control}/ D_{treated}$, where $D_{control}$ and $D_{treated}$ doses yield 10% survival for controls and treated cells, respectively).

2.1.5.2 Statistical Analysis

Statistical analysis and plotting of clonogenic survival curves was performed using the 'CFAssay'package in R [144]. The software adopts a maximum likelihood based approach to the logarithmic survival fractions of the linear quadratic (LQ) curve, which is preferable to the least squares methods. The survival curves from two sets of treatment conditions are compared using an Analysis of Variance (ANOVA) for two model. Statistical significance was set at the 0.05 level, two way analysis. All assay data are representative of three independent experiments from triplicate wells, unless otherwise stated (*P<0.05, **P<0.01 and ***P<0.001, ****P<0.0001).

2.1.6 Cell viability assay

Cytotoxic drugs were obtained from Sigma (oxaliplatin), Flurochem (5-flurouracil) and Sellekchem (SN-38). All cytotoxic drugs were initially dissolved for a stock solution of 2mM. DMSO was used to dissolve 5-FU and SN-38 and sterile water, warmed to 25 °C, was used to dissolve oxaliplatin.

The optimum plating density for all cell lines used was 5000 cell per well in a 96 well plate. An 8 point 3.16 fold dilution of each drug was performed in a 96 well plate, with a ninth row of DMSO only. This resulted in a concentration range from 2 mM to 636 nM. Cells were plated 24 hours prior to treatment into a 96 well plate as shown in Figure 2-2. PBS was added to the outer wells to minimise evaporation. The following day 0.4 µL of drug was added to the 6 replicate wells from the stock plate. At 72 hours, the MTS assay (CellTiter® 96R AQueous Non-Radioactive Cell Proliferation Assay) was performed as per the manufacturer's instruction. In brief, 20 µL of the solution is added to each well and returned to the incubator for 2 hours. The absorbance is read at 490 nM using a photospectrometer. Absorbance shows a direct 1:1 correlation with cell viability.

PBS	PBS	PBS	PBS	PBS	PBS	PBS	PBS	PBS	PBS	PBS	PBS
PBS	2.5nM	8nM	25.28 nM	80 nM	252 nM	0.8 µM	2.52 µM	8 µM	DMSO	GM	PBS
PBS	2.5nM	8nM	25.28 nM	80 nM	252 nM	0.8 µM	2.52 µM	8 µM	DMSO	GM	PBS
PBS	2.5nM	8nM	25.28 nM	80 nM	252 nM	0.8 µM	2.52 µM	8 µM	DMSO	GM	PBS
PBS	2.5nM	8nM	25.28 nM	80 nM	252 nM	0.8 µM	2.52 µM	8 µM	DMSO	GM	PBS
PBS	2.5nM	8nM	25.28 nM	80 nM	252 nM	0.8 µM	2.52 µM	8 µM	DMSO	GM	PBS
PBS	2.5nM	8nM	25.28 nM	80 nM	252 nM	0.8 µM	2.52 µM	8 µM	DMSO	GM	PBS
PBS	PBS	PBS	PBS	PBS	PBS	PBS	PBS	PBS	PBS	PBS	PBS

Figure 2-2. Schematic overview of 96 well plate used for MTS cell viability assay

Drugs are added from serial dilution stock to the seeded 96 well plate with 100 µL of growth medium. Final concentrations of the drugs are shown. DMSO at 0.4% final concentration was used as a control.

2.1.6.1 Statistical Analysis of cell viability assay

Background absorbance from the growth media wells was averaged and subtracted from each of the test wells. Absorbance in the DMSO well was used to assess the viability of cells in the absence of drug. Absorbance of each well was calculated as a percentage of the absorbance in the DMSO well only. Percentage absorption is a direct surrogate for cell viability. Results were imported in to R for analysis with 'drc package' [145]. Dose-response curves were generated by fitting a four parameter log-logistic regression model, where the independent variable was concentration of drug whilst the dependent variable was the effect on cell viability. Plots indicate cell viability on the y-axis and the log of drug concentration of the x-axis. Effective concentration for 50% cell viability (EC_{50}) was calculated. Statistical significance was tested by analysis of variance (ANOVA) between the regression models for both sets of conditions, where $p < 0.05$ was considered statistically significant.

2.1.7 CRISPR protocols

2.1.7.1 Genomic DNA preparation

Hot sodium hydroxide and tris (HotSHOT) method was used for genomic DNA preparation for PCR prior to Sanger sequencing. The composition of reagents 1 and 2 was as follows:

Alkaline Lysis Reagent (Reagent 1)	Neutralization buffer (Reagent 2)
NaOH 25 mM	Tris-HCl 40 mM
EDTA 0.2mM	

Cultured cells were washed with PBS and typsinized as previously outlined. Equal volumes of Reagent 1 and 2 were used. Reagent 1 was added to each well of the 96 well plates. The plates were heated to 95 °C for 30 minutes, cooled to 4 °C before adding Reagent 2 to neutralise. Samples were stored at -20 °C.

2.1.7.2 Puromycin selection optimization

Cells well trypsinised and counted as previously detailed. 200,000 cells were seeded to each well of two 6 well tissue culture plates and incubated at 37 °C for 48 hours. When cells were 50-70% confluent, 100 µL stock puromycin (1mg/ µL) was diluted in 1000 µL sterile water for a final concentration working solution of 0.1 µg/µL. Escalating doses of puromycin (µg/µL) were initially added to each well as follows: 1, 2, 4, 6, 8, 10, 12, 16, 20, 30. SW1463 showed acute sensitivity to puromycin (µg/µL) so lower dose ranges as follows were used: 0, 0.1, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6. Two control wells were left without puromycin. Growth medium was replaced after 24 and 48 hours. Repeat treatments with puromycin administered when medium was changed. Cell were left over the weekend after the third treatment with puromycin. Cells were washed, manually scraped and counted to plot a puromycin kill curve. A dose response

curve was fitted to the experimental data using the 'drc' package in R, and used to estimate the effective concentration of puromycin to kill 90% of cells (EC_{90}).

2.1.7.3 Plasmid amplification and DNA isolation

Edit-R hCMV- PuroR-Cas9 expression Plasmid DNA was obtained from Dharmacon™. Stock plasmid was diluted in Buffer EB, 100x. GFP plasmid was available locally. 0.5 µL of both plasmids was diluted in 20 µL XL10 Gold Escherichia coli bacteria and incubated on ice for 30 minutes before heat shock treatment for 30 seconds at 42 °C in a water bath. 250 µL of SOC medium was added before incubation in the shaking incubator for one hour. A further 750 µL of SOC medium was then added to each plasmid tube.

Using ampicillin resistant plates, 50 µL of each plasmid was plated out and incubated upside down at 37 °C overnight and checked the next day for appropriate colony growth.

400 µL of ampicillin was diluted in 400 mL of previously prepared Lysogeny broth (LB). An appropriate plasmid colony from both the Cas9 and GFP plated was selected using a pipette tip. 3 mL of LB-ampicillin mixture was placed in a round bottomed Falcon tube with an aeration top. The pipette tip was placed in the Falcon tube and placed in the shaking incubator for ~ 6 hours. Once the mixture appeared cloudy, 1 mL of plasmid was diluted 100x with LB in round bottomed flask and replaced in the shaking incubator overnight.

Qiagen Plasmid Plus Maxi Kit (Cat No. /ID: 12963) was used as per the manufacturer's instructions for isolation and purification of plasmids thereafter. Final plasmid DNA quantification was assessed using NanoDrop spectrophotometer.

2.1.7.4 Plasmid transfection for CRISPR guides

Predesigned guide RNA were obtained from Dharmacon™ for NFE2L2 and KEAP1 as summarised in Table 2-2. Edit-R CRISPR Cas9 synthetic tracrRNA was also obtained. Cells were seeded at 200,000 cells/well into two six well plates, 24 hours prior to transfection. Stock RNAs were resuspended in RNA free water for a final concentration of 20 µM. Guide RNA, tracer RNA and Cas9 plasmid were diluted in Optimem at final concentration of 625 nM of each RNA in the following ratios for each well of the 6 well plate. DharmaFECT Duo transfection reagent was diluted in Optimem, 15 µL in 185 µL per well, combined with the guide RNA mixture and incubated at room temperature for 20 minutes before adding to the respective labelled wells. GFP plasmid was used as a positive control. Puromycin selection was carried out for 72 hours, replacing the selection medium daily, according to the optimised puromycin dose.

2.1.7.5 Primer Design

The primers were designed using the primer3 software, available online at primer3.ut.ee (release 2.4.0). The unique DNA BLAST sequence was uploaded as indicated [Table 2-2]. Primers design settings were set with optimum characteristics of size = 20, temperature = 59 and %GC = 50.

2.1.7.6 E79K+ Mutation sequencing

The primers used for sequencing the E79K+ mutation were designed using the primer3 software, available online at primer3.ut.ee (release 2.4.0). A 240 base pair sequence spanning the missense mutation at position 235 was inserted into the software which chose the optimum primers based on primer length, temperature and percentage of GC.

Forward primer	GCG ACG GAA AGA GTA TGA GC
Reverse primer	GGA GGC TGA GGT TGG AAA GT

Table 2-2. Summary of the target gene, target sequence, primers and target location for CRISPR-Cas9 treatment

Both NFE2L2 and KEAP1 were targeted with three separate sgRNA targets with a view to creating indels and functional knockout. The primers were designed as described with the size of the PCR product also indicated. The genome locus and targeted PAM sequences are highlighted in bold as these were used to identify potential clones from Sanger sequencing

Gene	Target sequence	DNA sequence	Primers	Genomic location and PAM	Exon target list
NFE2L2	ACTGGGCTCGATGTGACC	AAGTTAATAGCACCCTCCAATCCTTCCTATAAATAGGTGGTATATAAATA ATCAGAATGACTAAAGGCACTGAATATAAAAAAAATCTCCAGTATTACAT TCTATTTTAGTTACCTGTAACTCAGGAATGGATAATAGCTCCTCCCAAAC TTGCTCAATGTCCTGTTGCATACCGTCTAAATCAACAGGGGCTACCTGAG CAACAGAAGTTTCAGGTGACTGAGCCTGATTAGTAGCAATGAAG[**ACTGGG CTCTCGATGTGACCGGG**]AATATCAGGAACAAGTGACTGAAACGTAGCCGA AGAAACCTAAAATTGATAAGGCATTGATTTATGAGTCTTCCACCAACAGG GGATGAGTAAGCTCTAAGGCACCACTACAAAACAAAATGGAATCAGCTGC AGTTCTGTGTCTCTCTACTTACTAACCAGTCATGACCCCGTTTGTAAATA ACAAATCTTTTTTTCCTACCCCACTGACTCAAATTTATAATTTGCTGGTC TGTTTCAGCTT	LEFT TGTCCTGTTGCATACCGTCT	hg38 + Chr2: 177232522-177232544 GGG	NM_001145412 (Ex4)
			RIGHT CCCTGTTGGTGGAAGACTCA		NM_001145413 (Ex4)
			PCR product 193 bp		NM_006164 (Ex4)
NFE2L2	GTGTTCTGGTGATGCCACAC	TGACCAGGACTTACAGGCAATTCTTTCTCTGGTGTGTTCTCACATTGGGC ATCATGCACGTGAGTGCTCTGCCCCTGAGATGGTGACAAGGGTTGTACCA TATCCCCAGAAGAATGTACTGGTGTTTTAGGACCATTCTGTTTGACACTT CCAGGGGCACTATCTAGCTCTTCCACTTCAGAATCACTGAGGCCAAGTAG TGTGTCTCCATAGCTGGAAGATTCCACTGA[**GTGTTCTGGTGATGCCACAC TGG**]GACTTGTGTTTAGTGAAATGCCGGAGTCAGAATCATTGAATTCTGCT GTGCTTTCAGGGTGGTTTTGGTTGAAAGCTTTGCAAAGTGATAGATCAGA AACATCAATGGGCCCATTTAGAAGTTCAGAGAGTGAATGGCTTAAAGTAG CAGGTGAGGGCATGCTGTTGCTGATACTGGGCTCAGCTATGAAAGCAGAA TAAAATTCATCACCAAAATCTGTGTTGACT	LEFT AGGCCAAGTAGTGTGTCTCC	hg38 + Chr2: 177231526-177231548 TGG	NM_001145412 (Ex 5)
			RIGHT ATGCCCTCACCTGCTACTTT		NM_001145413 (Ex 5)
			PCR product 224 bp		NM_006164 (Ex 5)
NFE2L2	TATTTGACTTCAGTCAGCGA	TCTAGAACATTGGATCTGTGTTCTCATTTAGATTGTATTTGTAATCTCCC CACTTCCCACCATCAACAGTGGCATAATGTGAATTAATTTATGTGGTATC TGTCATTTAAAAAACATGAGCTCTCTCCTTCCTTTTTTTGTCTTAAACAT	LEFT TGATTGACATACTTTGGAGGCA	hg38 - Chr2: 177234188-177234210 CGG	NM_001145412 (Ex 2)

		AGGACATGGATTTGATTGACATACTTTGGAGGCAAGATATAGATCTTGGA GTAAGTCGAGAAG[**TATTTGACTTCAGTCAGCGACGG**]AAAGAGTATGAGCT GGAAAAACAGAAAAAACTTGAAAAGGAAAGACAAGAACAACTCCAAAAGG AGCAAGAGAAAGCCTTTTTCGCTCAGTTACAACTAGATGAAGAGACAGGT GAATTTCTCCCAATTCAGCCAGCCCAGCACATCCAGTCAGAAACCAGTGG ATCTGCCAACTACTCCCAGGTACAGAGTACTCAGTTCTTGGGAAA	RIGHT GCTGGCTGAATTGGGAGAAA PCR product 211 bp		NM_001145413 (Ex 2) NM_006164 (Ex 2)
KEAP1	GTGCGGTTGCCATGCTGGGA	CCTCCATGCCCTGCTCCCGCAGCCCGTTGGTGAACATGGCCTTGAAGACA GGGCTGGATGAGGCCAGCACCACCTTGTGGGCCATGAACTGGGCGGCCGG TGCATCCTGGTACTTGACCTGCAGTGTGACGTCACACAGCTGCTGGCTGA GCCGCAGCTCGTTCATGATGCCAAAGGCCTGCTTGGTATGATCCTCCAGG GTGTAGCTGAAG[**GTGCGGTTGCCATGCTGGGAGGG**]CGTCACCTCCGCCTT GCACTCAGTGGAGGCGTACATCACCGCGTCCCCTGCCCCCTCAGGGCACT GTGACTGCAGGGGCAGGAATCGGCAGCAGGCCCCAGCCCCGCTAGGCCTG GGATCTGGCTGCATGGGGTTCCAGAAGATAAGCAACACCACCACCTCTGG CACTCAGGGACCTGGAGGGGAGAGAGCACAGGGCAGAGGGCAGGG	LEFT CAGCTCGTTCATGATGCCAA RIGHT AGAGGTGGTGGTGTTGCTTA PCR product 244 bp	hg38 + Chr19: 10499882-10499904 GGG	NM_012289 (Ex 2) NM_203500 (Ex 2)
KEAP1	ATGCATGTAGATGTACTCCC	CTGGGCTGGTCTTGAACTCTTGGCCTGAAGCAATCCCCCCGCCTCAGCCC CTCAAACTGTGGAGACTACACCACCATACCCAGCCCAGAACCTCCTTTTT CTCCAGTTTCCTGCCTTGACATCTCAAGGGGAGACAGTGATGAGCACTCG TCCATCCCTGGTCCTTCTCCTGACACTGCCCCCAGCCCCACTTCCCCGCT CACCTCCCCAAA[**ATGCATGTAGATGTACTCCCGGG**]CACGCTGGTGCAACT CCACACAGCCAATCTGCTCAGCGAAGTTGGCGATGCCGATGGCATTGCTG GGGTCCAGCTGCTGCACCAGGAAGTCACTGCAGGCACGGACAACGCTGTC GATCTGGTACATGACAGCACCGTTCATGACGTGGAGGACACACTTCTCGC CCATGGAGATGGAGGCCGTGTAGGCGAATTCAATGAGGCGCTCCA	LEFT ACTACACCACCATACCCAGC RIGHT AACTTCGCTGAGCAGATTGG PCR product 214 bp	hg38 + Chr19: 10499404-10499426 GGG	NM_012289 (Ex 2) NM_203500 (Ex 2)
KEAP1	GTTGTGGTGGATGCAGCCGT	CCCCTAAGCATTTCCCAGCCCCAGGCACAGAATCAAAGGTCACTGACTAG AACTCTCCAAGGAGCTTAGCTTCATCCTGAGGCCTCCACTCCCTGAAGAC AGGAAGAGGAAACAGCCTCAGGAAGAATACCCGGATCTCAGTGTCTTGGG ACTTGCCAGGAGCAGGACCCTCCGAGCCCACCCCCAGGCCCTGCCACTCA CCTCTCCACACT[**GTTGTGGTGGATGCAGCCGTGGG**]AGCCGCCGACGGCAT AGATGTGGCCATCGATGACCCCCACCCCGATGCGGTTACGGGGCACGCTC ATGGGGGCGCAGGGCGACCACTGATTGGTCATGGGGTTGTAACAGTCCAG GGCGCTGGAGTCGGTGTTGCCGTCGGGCGAGTTGTTCCTGCCGCCCACGG CGTACAACAGCCCGCCCACCACGCAGCCGGCCAGGCCGCTCCGCG	LEFT ACCCGGATCTCAGTGTCTTG RIGHT CCTGGACTGTTACAACCCCA PCR product 223 bp	hg38 + Chr19: 10491588-10491610 GGG	NM_012289 (Ex 3 NM_203500 (Ex 3)

2.1.7.7 Polymerase Chain Reaction for Sanger sequencing

DNA previously prepared using HOTSHOT technique was amplified using PCR. A mixture of 5 µL of buffer solution, 4 µL of dNTP, 1.5 µL of Magnesium, 35 µL of Milli-Q® water and 2.5 units of BIOTAQ™ of DNA polymerase was added to 1 µL of the forward primer and 1 µL of the reverse primer for each of the primer sets detailed in Table 2-2. The final solution volume of 48 µL was added to 2 µL of HOTSHOT DNA on a per well basis. PCR was done with initial denaturation at 95°C for 5 min, 94 °C for 1 minute, 55 °C for 1 minute, 72 °C for 1 minute followed by 30 cycles at 94°C for 1 min and final elongation at 72°C for 5 min.

To increase specificity betaine was added to some reaction mixtures and run at denaturation at 95°C for 5 min, 94 °C for 1 minute, 60 °C for 1 minute, 72 °C for 1 minute followed by 30 cycles at 94°C for 1 min and final elongation at 72°C for 5 min.

2.1.7.8 Gel Electrophoresis of PCR products

The PCR products were checked for purity using 2% w/v agarose gel with ethidium bromide (0.5 µg/ml). 10µL of PCR product was mixed with 1µL of Orange G. The products were added to the agarose gel and run for 20 minutes at 120V or until bands were clearly resolved.

2.1.7.9 PCR Purification

All samples were purified using the QIAquick® PCR purification kit from QIAGEN to ensure that sequencing was as accurate as possible. In brief, each 50 µL PCR product was diluted with 250 µL of binding buffer (Buffer PB). The solution was centrifuged at 13,000 rpm for 1 minute in the collecting tube to bind the DNA to the membrane. Each sample was then washed with 750 µL of wash buffer (Buffer PE) and centrifuged again at 13,000 rom for 1 minute. In a clean Eppendorf, 50 µL of elution buffer (10mM of Tris-Cl, pH 8.5) was added to each column and the

samples centrifuged at 13,000 rpm to elute the purified DNA. The purified DNA was then diluted 1:20 with Milli-Q® water.

2.1.7.10 Sanger sequencing preparation

On a per well basis, 4 µL of diluted, purified DNA was combined with 8µL of Big-Dye terminator reaction mix (Thermofisher), 7µL of water and 1µL of the chosen primer (either forward or reverse). The final volume was pipetted into a 96 well plate and PCR carried out at 96°C for two minutes, 95°C 10 seconds, 50°C for 5 seconds, 60°C for four minutes, repeated by 30 cycles. Qiagen DyEx™ 96 spin kit was then used as per the manufacturers' instructions to remove any unwanted dye terminators from the solution. Finally, the DNA was concentrated and solvents evaporated using the SpeedVac™ Concentrator Savant ISS110.

2.1.7.11 Sanger sequencing and analysis

Sequencing of DNA was carried out by the Department of Zoology, University of Oxford. Sequence traces were returned as .ab1 files. Original primer sequences were located using the alignment tools and sequences were analysed for quality of trace with the open source 'A plasmid Editor' (ApE v2.0.55) software. Final chromatogram traces were drawn using 'sangerseqR' package [146], which also allows for identification of heterozygote calls within a sequence.

2.1.8 Cell line karyotyping

2.1.8.1 Fluorescent in situ hybridization (FISH)

This work was carried out in collaboration with Dr Daniela Moralli and the structural genomics group at the Wellcome Trust Centre for Human Genetics. The following Bacterial Artificial Chromosomes (BAC) and plasmids were used for hybridisation and identification of the NFE2L2 locus [Table 2-3].

Table 2-3. Details of Fluorescent in situ hybridisation probes

Gene locus and loci of the flanking probes used in FISH analysis assess for cytogenetic abnormality at the gene locus

BAC clone	**Colour**
pBSD4 2 alphasat: chromosome 2 specific centromeric repeat	Green
RP11-157E8 chr2:177,343,817 -177,497,286	Red
NFE2L2 chr2: 178,095,031-178,129,859	
RP11-65L3 chr2:179,258,141-179,430,959	Orange
RP11-463B12 chr2:240,911,276 -241,082,695	Purple

RP11-157E8 and RP11-65L3 flank *NFE2L2*, marking the opposite boundaries of the cytogenetic band of where the gene is located (2q31). Upon DNA condensation, the two signals should almost overlap on metaphase chromosomes. If the two flanking BACs are always present together, one can deduce that the chromosomes were not rearranged and so infer that the NFE2L2 gene was also present. Thus, each combined RP11-157E8 and RP11-65L3 signal correspond to one copy of NFE2L2.

One expects the FISH signals to be in the order illustrated [Figure 2-3]. Any chromosome showing this pattern will also carry, with high probability, a copy of NFE2L2 and will not be rearranged with respect to the normal chromosome 2.

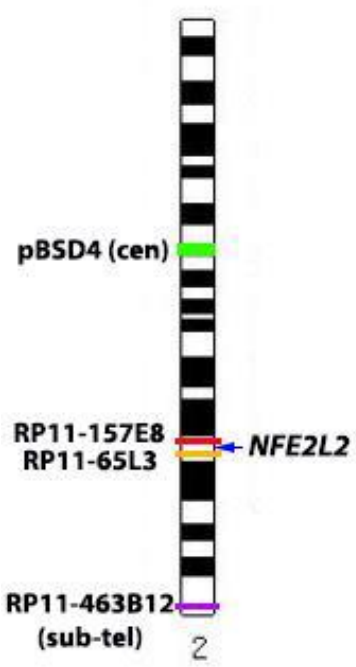


Figure 2-3. Ideogram of chromosome 2 used for FISH analysis

The FISH signals are expected to be in this order along the chromosome: the centromere, RP11-157E8 and RP11-65L3 signals on the long arm (flanking the locus of interest), and the sub-telomeric region. [Courtesy of Dr Daniela Moralli]

All of the probes were hybridised to metaphase spreads from the cell lines, and a minimum of 20 metaphases were analysed per line. As a control for the FISH experiment, the BACs were labelled and hybridised to metaphase spreads from a control cell line. All of the probes mapped to the correct genomic location in the control cells. Results were expressed as a ratio of the number of signals observed per cell.

2.2 RNA sequencing

Isogenic HCT116^{WT} and HCT116^{E79K+} mutant cell lines and siRNA transfected SW1463 were used for next generation RNA sequencing experiments. RNA was extracted using the Qiagen® RNeasy Mini Kit as per the manufacturer's instructions quantified using the nd-1000 NanoDrop® spectrophotometer. Standardised concentrations of 10 µg in 45 µL were calculated. Using the TURBO DNA-free kit (Invitrogen™ Catalog number: AM1907), 5 µL of DNase buffer was added with 1 µL of DNase enzyme and incubated at 37 °C for 20 minutes. 10 µL of inactivation reagent was then added, incubated at 25 °C for a further five minutes before being centrifuged at 10,000g for 90 seconds. The PCR products were checked for purity using 2% w/v agarose gel with ethidium bromide (0.5 µg/ml). 10µL of PCR product was mixed with 1µL of Orange G. The products were added to the agarose gel and run for 20 minutes at 120V. Using the Agilent RNA ScreenTape system, 1 µL of each sample was mixed with 5µL sample buffer against a control ladder. Samples were mixed and heated to 72 °C for three minutes before being placed on ice and centrifuged for thirty seconds. The samples were then analysed on an Agilent TapeStation 4200. The quality output score was an RNA Integrity Number (RIN) calculated by taking the ratio of the area under the 18S and 28S rRNA peaks to the total area under the graph.

All samples with a RIN of 7 or higher were processed to generate libraries for RNA sequencing following the Illumina TruSeq stranded RNA sample preparation guide. Poly-A enriched mRNA libraries were sequenced as a 12 sample multiplex on one lane of an Illumina HiSeq4000 machine, performing 75bp paired end sequencing. Approximately 20 million reads were obtained per sample. Reads were aligned to the human reference genome (GRCh37) using HISAT2 [147] and duplicate reads removed using the Picard 'MarkDuplicates' tool [148]. Reads mapping uniquely to Ensembl-annotated genes (~15 million per sample) were summarised using featureCounts [149]. The raw gene count matrix was imported into the R/BioConductor

environment for further processing and analysis [150,151]. Genome alignments and mapping were performed by the bioinformatics hub at the Wellcome Trust Centre for Human Genetics, University of Oxford. The data have been deposited to the Gene Expression Omnibus under accession number GSE136011.

2.2.1 Differential expression of genes

Differential expression analysis was performed with both 'DESeq2' [152] and 'edgeR' [153] to ensure robustness. Main analysis and plots were generated from edgeR data. Genes were initially filtered out with read counts of <1 count per million (cpm) in at least 4 samples. Normalisation factors for the library sizes were determined in using the trimmed mean method (TMM) in edgeR. Relative to the control, gene dispersion of the perturbed system was estimated before fitting to a negative binomial model. Genes with a FDR of =< 0.01 were selected to check for overlap with DESeq2 as an estimation of agreement between the two packages.

2.2.2 Gene Set Enrichment Analysis

GSEA was carried out against the gene sets from the Molecular Signatures Database v6.2 [154] (http://software.broadinstitute.org/gsea/downloads.jsp) using the 'fgsea' package [155]. The ranked gene list was compared to the *a priori* defined gene sets and the curated geneset NFE2L2.v2 [110], for a total of 1000 permutations. A FDR <0.1 was used to highlight significant pathways of enrichment [156]. Venn diagrams of overlap were created using www.interactivenn.com [157].

The R code for analysis is included in Appendix 1.

2.3 Amino acid exchange experiments

2.3.1 Cell culture media

Specialised media was obtained from ThermoFisher [Table 2-4] and supplemented with 10% FBS and 100 µg/mL of penicillin streptomycin. Media was either glutamine deplete, asparagine deplete [Nil] (catalogue number #11960) or glutamine deplete, asparagine replete [N] (catalogue number #10938), relative to the complete medium [Q].

Table 2-4. Details of relevant concentrations of amino acid supplements used in amino acid exchange experiment

Amino acid	Q	Nil	N
L-glutamine (Q) (Conc. mg/L)	584	0	0
L- asparagine (N) (Conc. mg/L)	0	0	60
D –glucose (Conc. mg/L)	4500	4500	4500

2.3.2 Technique

1 x 10^5 cells were seeded 6 well plates in complete media for 48 hours. Media was then removed, cells washed in PBS and media replaced with glutamine deplete [N], asparagine deplete [Q] or glutamine deplete, asparagine replete [Nil] media as detailed in table. Cells were grown for a further 48 hours before being lifted and viable cell number was counted using the Countess II system. Data are presented as mean +/- s.e.m.

2.4 Bioinformatic materials and methods

2.4.1 Candidate gene selection

Using the Oncomine database [158] the *a priori* identified candidate genes were analysed for differential expression. The input parameters were "Cancer type = Colorectal cancer" and "Analysis type = Cancer vs. Normal". Gene expression levels from studies where matched tumour expression and normal tissue expression data was available were used for analysis. Within each dataset, differential expression was determined by setting threshold values of fold change > 2, p-value < 0.0001 and gene rank of top 10%. The database normalised gene expression across all selected datasets to allow summative gene expression comparisons. The resulting median gene rank for the meta-analysis across all selected datasets was calculated along with its associated p-value. All p-values were corrected for multiple hypothesis testing using the false discovery rate method [159].

2.4.2 STRING analysis

The STRING database is a protein-protein interaction database (https://string-db.org/) which collates, integrates and predicts associations for a large number of organisms [160]. This includes both direct physical interactions between proteins and indirect functional interactions. As a result the database can be used to predict interactions between genes from a list of candidates through a combination of systematic co-expression analysis, detection of shared selective signals across genomes, automated text-mining of the scientific literature and computational transfer of interaction knowledge between organisms based on gene ontology. All analysis was carried out using version 11.

2.4.3 Regulatory sequence analysis tools (RSAT)

In silico analysis of putative Antioxidant Response Elements (AREs) in metabolic gene promoters were identified using RSAT [161]. The promoter sequence upstream to -5 kb lengths for named genes were retrieved from Eukaryotic Promoter Database (EPD) [162]. These sequences were analysed to identify AREs using the string-based pattern matching program 'dna pattern' [163]. The DNA patterns were entered as RTGASNNNGCR and RTGAYNNNGCR, where R = A or G, S = C or G, Y = C or T, and N = any nucleotide. These generic sequence queries were used in the query option to identify an exact match of an ARE sequence in the given 5-kb promoter sequences. The DNA patterns have been used previously [163] and are based on well-known previous publications on the consensus structure of AREs [164,165].

2.4.4 RNA array probe matching

To minimise confounding between datasets as a result of different platforms of RNA profiling, gene symbols were matched to probesets from the expression array annotation file, resulting in multiple probes per gene. Probes were matched between platforms using genomic locus references from the annotation file. Two types of matches were considered—(i) exact-match and (ii) sub/super-set-match. Exact match was where the chromosomal alignment from the annotation file of the microarray build probes matched perfectly. Sub/super-set-match occurred where the U133 probe was completely nested in a validation array probe or vice versa. Unmatched probes were filtered out.

2.4.5 Unsupervised analysis for dimension reduction

A signature approach has the advantage of preserving statistical power, which would be exhausted if all the candidate gene probes were included. It was expected that some probes

would be correlated, due to multiple probes for the same gene, and co-expression of the candidate genes which operate as part of a pathway. Correlation would duplicate information and result in confounders, such as multi-collinearity, in the downstream regression analysis. To obtain a simplified, aggregate of NRF2 representation, a principal component analysis (PCA) was applied. This produced a new set of continuous variables – principal components (PCs) - which were weighted averages of all the probe expressions. In the presence of high correlation among probes, the majority of the variation in the data can be explained by a smaller set of principal components (PCs). PCA can provide a simpler representation of the expression data by filtering out the PCs that explain little of the variation in the expression data. Only the major PC's, defined as the most variable PCs that explained up to 80% of the variation in the expression data, were retained.

2.4.6 Supervised analysis using multivariate Cox regression models

Once dimension reduction of the NRF2 signature was achieved using PCA, supervised variable selection was performed to decide on how many major principal components (PCs) would be useful for predicting prognosis in the chosen training set. A Cox proportional hazard regression model was used to model the prognosis predicted by the PCs of NRF2 expression. Cox models were fitted with increasing numbers of principal components (PC). The PCs were included in order of how much variation in NRF2 expression do they explain. For stepwise variable selection, PCs were progressively included in decreasing order of variance. Akaike information criteria (AIC) and Bayesian information criteria (BIC) were calculated to indicate the most suitable number of major PCs to keep. The selected set of major PCs were used to represent the NRF2 signature derived from the training set.

2.4.7 Construction of the NRF2 signature in the validation sets

The NRF2 signature in each validation set was constructed by performing PCA on the corresponding probe sets matched from the training set. Major PCs were identified using the same filtering method for the training set. The PCs were unsupervised summary statistics of the probe expressions. Thus, which PCs were correlated with the probe expressions of a given gene can vary purely due to sampling variation. It was not appropriate to use a fixed set of PCs to validate the biological signal discovered in the training set. For a given validation set, the correlations between the PCs and the probe expressions of genes in the represented set were calculated. Based on the PC-probe correlations, PCs were selected to represent the signature if those PCs were representative of any genes in the represented gene set. The number of PCs selected via this approach can vary across datasets.

2.4.8 Datasets

Publically available colorectal datasets were downloaded from the Gene Expression Omnibus (GEO) database. Clinical and expression data was accessed from R programming environment using the packages 'GEOquery' [166] and 'Biobase' [167] from https://bioconductor.org/biocLite.R. The datasets were divided into a training set and validation sets. GSE17536 was chosen as the training set [168]. Validation was carried out using datasets representative of non-metastatic, stage I-III disease (GSE14333 and GSE39582) [82,169], metastatic disease (MRC FOCUS trial) [170] and rectal only cancer (GSE87211) [171].

The GSE17536 dataset was chosen for training for a number of reasons. Firstly it contains RNA expression profiling from patients across all stages (I-IV) of disease. This meant that any differences in biology should be adequately reflected in the expression profiles obtained and allow for appropriate exploration and training. Secondly, it has OS as an outcome variable which

would provide more robust estimates of prognosis than disease free survival (DFS). Lastly, the data was collected from the Affymetrix U133 array which is exactly the same as the GSE14333 and GSE39582 datasets and similar to the unique Xcel array used in the S:CORT program. The MRC FOCUS trial cohort represents a balanced subset of the >2100 patients in the FOCUS study. GSE87211 was chosen in order to assess the relevance of NRF2 expression in a rectal cancer specifically, and to assess the ability to migrate between RNA platforms from Affymetrix based arrays to another type.

2.4.9 Aberdeen rectal cohort

The Aberdeen dataset comprised of rectal biopsy samples sequentially collected from patients that have been treated with radiotherapy/ chemoradiotherapy. Access to this dataset, including the RNA expression profiles, was through the MRC Stratification in Colorectal cancer (S:CORT) consortium. All samples were annotated by a single pathologist for pathological tumour (ypT) and pathological nodal stage (ypN) as well as pathological regression grade (complete, good partial, partial and minimal). Reference H&E slides were marked for areas of adenocarcinoma in the rectal biopsies and adjacent sections were cut for multi-omics profiling.

2.4.10 Sample handling and RNA extraction for S:CORT datasets

The work was carried out by multiple members of the S:CORT collaboration and provided for context.

Stored tissue blocks were transferred to Leeds University sample processing where 1 x 5 micron and 2 x 10 micron sections were cut. The 5 micron slide was H&E stained, annotated for the tumour region and all sections were shipped to Queen's University Belfast for processing. Samples were dewaxed by an automated process using the Tissue-Tek Prisma machine, using a

combination of xylene washes and ethanol. Macrodissections were performed within a designated macrodissection area treated with RNase decontamination solution. The unstained section was overlaid with the annotated H&E slide and using a clean scalpel blade the annotated area was scraped into a 1.5ml RNase-free Eppendorf tube containing the prepared Roche High Pure RNA Paraffin Kit tissue lysis reagent (catalogue 0327089001).

RNA extraction as per the Roche High Pure RNA Paraffin Kit instructions. Extracted RNA was quantified by Nanodrop and stored at -80 C. Samples which had sufficient RNA concentration were submitted for profiling by cRNA and ds-cDNA preparation before being subjected to further quality control. Finally samples were hybridised to the Xcel microarray as per the manufacturer's instructions.

RNA profiles were submitted to Oxford as raw CEL files. Quality control analysis was run on samples using the R base 'AffyQC' module (https://github.com/BiGCAT-UM/affyQCModule). The files were then processed using the R packages 'limma' (https://bioconductor.org.packages/release/bioc/html/limma.html) and 'affy' (https://bioconductor.org.packages/release/bioc/html/affy.html) to normalise the expression values using robust multiarray algorithm (RMA) and generate and expression matrix of probe intensities against the samples. All samples were batch corrected for outcome variables including date of scan.

2.4.11 Statistical analysis - Prognosis

2.4.11.1 Primary analysis

The primary analysis for each of the validation sets was to test whether the NRF2 signature has a prognostic effect on DFS and/or OS. A Cox PH regression model with no variable was constructed, representing the null model. A second Cox PH regression model was fitted with the

NRF2 signature. The likelihood ratio test (LRT) was used to provide evidence against the null hypothesis (H0), that the null model was an adequate approximation of the model with NRF2 signature. For Kaplan Meier analysis, tertiles of expression were used to categorise NRF2 signature expression into high, intermediate and low. Hazard ratios, and confidence intervals, for these curves were between the upper and lower tertiles. Significance testing was done on Cox model analysis only.

2.4.11.2 Secondary analyses

To assess whether the prognostic effect of the NRF2 signature was confounded with other known prognostic variables, adjusted analyses using multivariate Cox PH regression models were performed. Prior to the adjusted analysis, the R package MICE [172] was used to impute missing values in the adjusted variables in GSE39582 and MRC FOCUS trial data. The adjusting variables were chosen on the basis of univariate analysis in each respective dataset. Specifically, in a secondary analysis, a Cox PH regression model was fitted with the NRF2 signature alone. A further Cox PH regression model was constructed with the NRF2 signature as well as the adjusting variables. Subsequently, an LRT was performed to test the H_0 that the model with only the adjusting variables adequately approximates the model with only the adjusting variables and the NRF2 signature. If there was evidence against H_0, then it suggests that NRF2 provides additional information to explain the survival outcome. Because the adjusting variables available varied across datasets, no adjusting variables were used in the training set. Significance suggests that NRF2 expression provides additional information to explain the survival outcome.

2.4.11.3 Interaction test with Oxaliplatin

A logistic regression model was constructed to predict response to oxaliplatin, defined as radiological complete response or partial response versus stable disease and progression, at the 6

months timepoint. This was the null model (H_0). NRF2 signature expression was then added to the model. An analysis of deviance was used to test that the model with NRF2 expression provided additional information to predict for response. If there was evidence against H_0 then it would suggest that NRF2 expression was predictive of oxaliplatin response.

2.4.12 Statistical analysis - Prediction

2.4.12.1 Primary analysis

The primary analysis was to test whether the NRF2 signature explained the NAR score. An ordinal logistic regression model was constructed without any variables to represent the null model. Another ordinal logistic regression model was built with the NRF2 signature. The null hypothesis (H_0) was that NRF2 would not provide any explanatory power for NAR score. The likelihood ratio test (LRT) was employed to inform the strength of evidence against the null hypothesis.

2.4.12.2 Secondary analysis

To investigate whether the NRF2 signature was confounded with other known variables that (potentially) explained radiotherapy response, an adjusted analysis was performed using a multivariate ordinal logistic regression model. An ordinal logistic regression model was constructed with clinical nodal stage (cN). Another ordinal logistic regression models was constructed with the NRF2 metagene and clinical nodal stage (cN). A subsequent LRT was performed to test the null hypothesis that the model with the NRF2 signature did not add to the explanatory power of the model. Statistical significance from the LRT indicated that there was evidence for the NRF2 signature providing additional information to explain the NAR.

CHAPTER 3

3.1 Introduction

Chemoradiation remains the standard of care for all patients with LARC. Some patients respond poorly with little or no tumour regression [21], highlighting the need to better understand contributing factors. NRF2 is a potent transcription factor that plays a responsive role in cell protection against oxidative and electrophilic stress. As outlined in Chapter 1, in unstressed conditions NRF2 is targeted for ubiquitination by KEAP1. Constitutive activation of the pathway can occur by a number of distinct mechanisms including somatic mutation of key regulators *KEAP1*/*NFE2L2*, methylation of *KEAP1* and direct activation by oncogenes *KRAS*, *BRAF* and c-*MYC* [126,127,138]. This allows NRF2 to function in an 'oncogene' fashion, promoting survival, resisting radiation and dysregulating metabolism [114,173].

NRF2 pathway expression has been shown to contribute to radioresistance in many tumour types, in particular lung cancer [132,174], SCC head and neck [175,176], oesophageal [177] and pancreas [178]. But its role in therapeutic resistance in CRC is conspicuously under investigated, in spite of the fact that the NRF2 pathway can be overexpressed in CRC through a variety of aforementioned mechanisms [133]. This oversight is possibly due to the low rate of somatic mutation in both *KEAP1* and *NFE2L2* in CRC [59], in comparison to the previously studied tumour types, and the fact that radiation in not thought of as part of the therapeutic strategy in CRC. Nevertheless, in a PanCancer TCGA analysis the level of NRF2 pathway signalling in the COAD and READ datasets is higher than would be expected given this low somatic mutation rate suggesting that activation within tumours occurs [128] . Further investigation is therefore warranted of its role in CRC therapeutic resistance.

Previous work has focussed on the role of chemotherapeutics. NRF2 knockdown mediated by RNAi in SNUC5/5FU colorectal cells showed decreased cell viability *in vitro* and *in vivo* as well

as showing increased sensitivity to 5-FU [140]. Similar work showed that inhibition of NRF2 also rendered CRC cells sensitive to irinotecan [179]. However, no published work to date has assessed the role of NRF2 in mediating radiation resistance.

Specifically, I sought to investigate changes in radiation response using NRF2 knockdown with siRNA and CRISPR-Cas9 sgRNA approaches in CRC cell line models.

3.2 Results

An siRNA mediated knockdown of *NFE2L2* was performed in three CRC cell lines: SW1463, HT55 and HCT116. These lines have all been well characterised [180] , their main features are summarised in Table 3-1.

Table 3-1. Molecular profiles of colorectal cancer cell lines used experimentally

Colorectal lines have been well characterised in terms of mutation status, methylation profile and mutational burden. These features are important as they may influence radiosensitivity.

Cell line features	**SW1463**	**HT55**	**HCT116**
Site	Rectum	Rectum	Colon
Microsatellite status	Stable	Stable	Unstable
Methylation	CIMP –	CIMP –	CIMP +
KRAS	Mutant (G12C)	Wild type	Mutant (G13D)
BRAF	Wild type	Wild type	Wild type
TP53	Mutant (R248Q)	Mutant (R213L)	Wild type

HCT116 is a hypermutated cell line, as a result of its microsatellite unstable status, and displays a stable chromosomal number of approximately 2n [180], in essence diploid. In contrast non-hypermutated lines SW1463 and HT55 are known to be aneuploid with a modal chromosome number in excess of 46. This is important as hypermutated lines are known to be relatively more sensitive to radiation than non-hypermutated lines [94]. Conversely, cell lines with increased somatic copy number alterations are known to be more radioresistant. In addition, HCT116 is wild type for TP53 which also contributes to the increased radiosensitivity of this line. Both of the rectal lines are TP53 mutants; respectively $p53^{R248Q}$ and $p53^{R213L}$. These cell lines, through their extensive profiling, are regarded as generally representative of CRCs *in vivo*.

3.2.1 Knockdown of NRF2 causes radiosensitization

3.2.1.1 SW1463

In order to successfully transfect SW1463 with a siRNA, extensive optimisation was required. Initial attempts to reverse transfect cells with Interferin HTS, as detailed in Chapter 2, resulted in less than 50% knockdown of *NFE2L2* and no significant changes of downstream NRF2 regulated targets. Re-optimisation for forward transfection using a Lipofectamine/siRNA complex was more successful.

This resulted in approximately 75% knockdown of *NFE2L2* expression by qPCR, a statistically significant change (Welch 2-sample t-test, $p<0.0001$). In the absence of a reliable and consistent antibody to NRF2 [181], well validated transcriptional targets that contain ARE in or near the transcriptional start site [110]– NQO1 and HMOX1- were used as surrogates of pathway perturbation. This approach is widely accepted in the study of downstream NRF2 pathway effects [182]. A significant change was seen in mRNA levels of both NQO-1 (Welch 2-sample t-test, $p<0.0001$) and HMOX1 (Welch 2-sample t-test, $p<0.01$) [Figure 3-1]. The significant

changes not only confirm that NFE2L2 expression has been altered but that the translation of NFE2L2 mRNA into NRF2 protein has also been affected at a functional level.

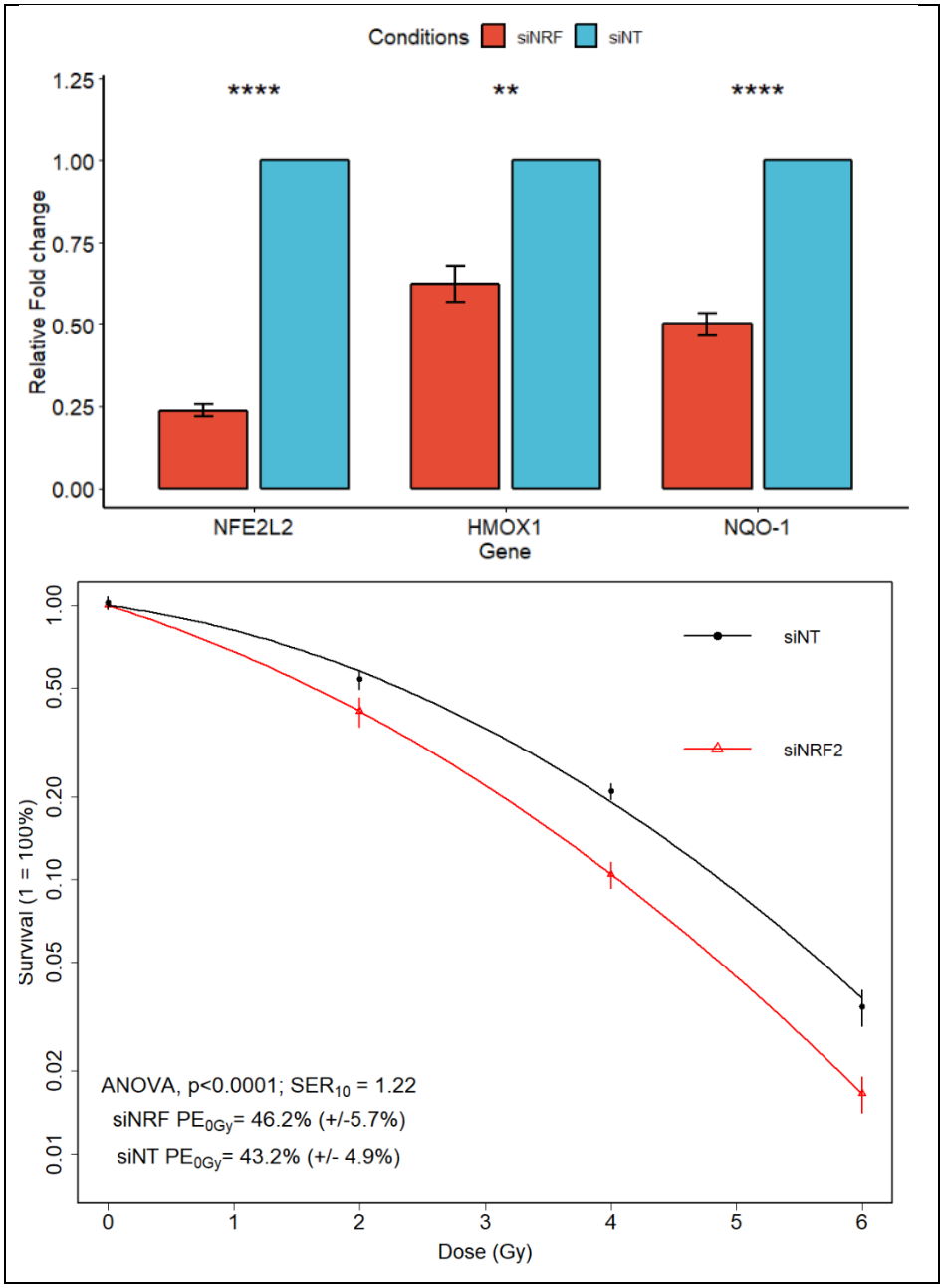


Figure 3-1. RNAi knockdown of NRF2 results in radiosensitivity of SW1463

Barplot of relative fold change in mRNA levels of NFE2L2 and two key NRF2 regulated genes following siRNA treatment (left). Data shown are mean +/- s.e.m of three repeats from triplicate wells (n=3). Cell survival curve of clonogenic assay showing increased radiosensitivity of siNRF2 treated cells (right). Statistical significance by ANOVA and plating efficiencies shown in bottom left of plot.

NRF2 knockdown SW1463 cells and controls were plated for colony forming assays. The cell survival curve demonstrated that there was a difference in radiosensitivity between cells treated with a non-targeting siRNA and that targeting NRF2 [Figure 3-1]. The observed differences in survival curves are statistically significant (ANOVA, $p<0.0001$). The sensitization enhancement ratio at 10% survival (SER_{10}) following knockdown of NRF2 was 22%, suggesting that approximately 20% less radiation dose was required to result in 10% cell survival.

3.2.1.2 HT55

Transfection of the same siRNA resulted in the decreased expression of NFE2L2 by approximately 70% [Figure 3-2], a significant decrease (Welch 2-sample t-test, $p<0.0001$). Functional pathway changes were confirmed by significant decreases in levels of both NQO-1 (Welch 2-sample t-test, $p<0.001$) and HMOX1 (Welch 2-sample t-test, $p<0.001$).

The survival curve from the resultant clonogenic assay shows that there was a significant difference in the survival of siNRF2 treated cells compared to a non-targeting control [Figure 3-2]. These differences are statistically significant (ANOVA, $p<0.01$). The SER_{10} value indicates that approximately 17% more dose is required in the presence of active NRF2 signalling. This was of a similar order of magnitude to that seen in SW1463.

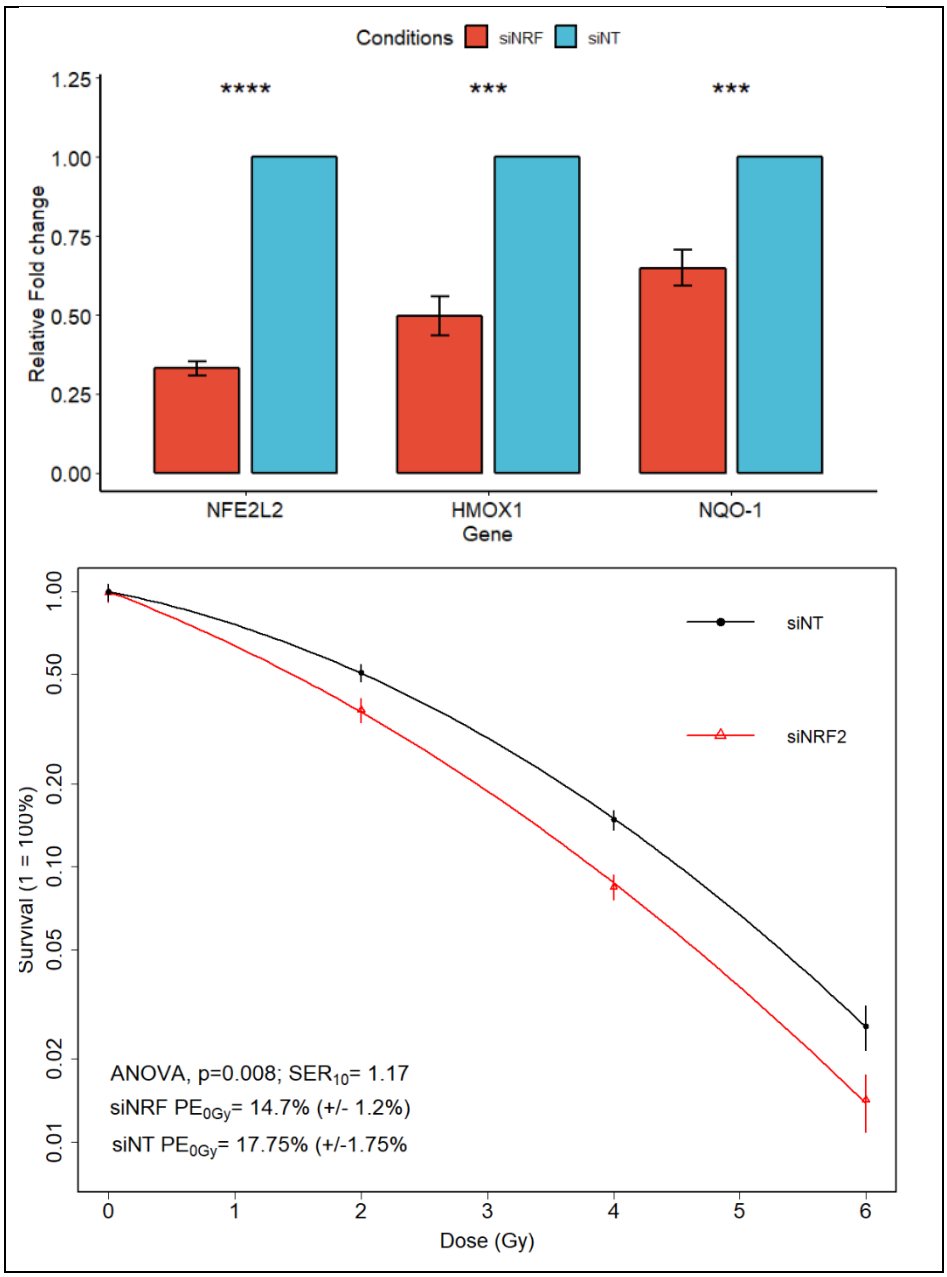


Figure 3-2. RNAi knockdown of NRF2 results in radiosensitivity of HT55

Barplot of relative fold change in mRNA levels of NFE2L2 and two key NRF2 regulated genes following siRNA treatment (left). Data shown are mean +/- s.e.m of three repeats from triplicate wells (n=3). Cell survival curve of clonogenic assay showing increased radiosensitivity of siNRF2 treated cells (right). Statistical significance by ANOVA and plating efficiencies shown in bottom left of plot.

3.2.1.3 HCT116

To further validate the findings from the previous two cell lines the colon cell line, HCT116, was targeted with the same siNRF2. Again, excellent knockdown of NFE2L2 mRNA was seen with an approximately 80% decrease [Figure 3-3]. This was highly significant (Welch 2-sample t-test,

p<0.0001). There was also a significant change in the expression of both NRF2 regulated genes indicated once again that functional knockdown had been achieved; NQO1 (Welch 2-sample t-test, p<0.001) and HMOX1 (Welch 2-sample t-test, p<0.05). Interestingly the resulting cell survival curve of the clonogenic assay showed no difference [Figure 3-3]. This was confirmed by statistical analysis of both survival curves (ANOVA, p=0.7291). There was no difference seen in the SER_{10}.

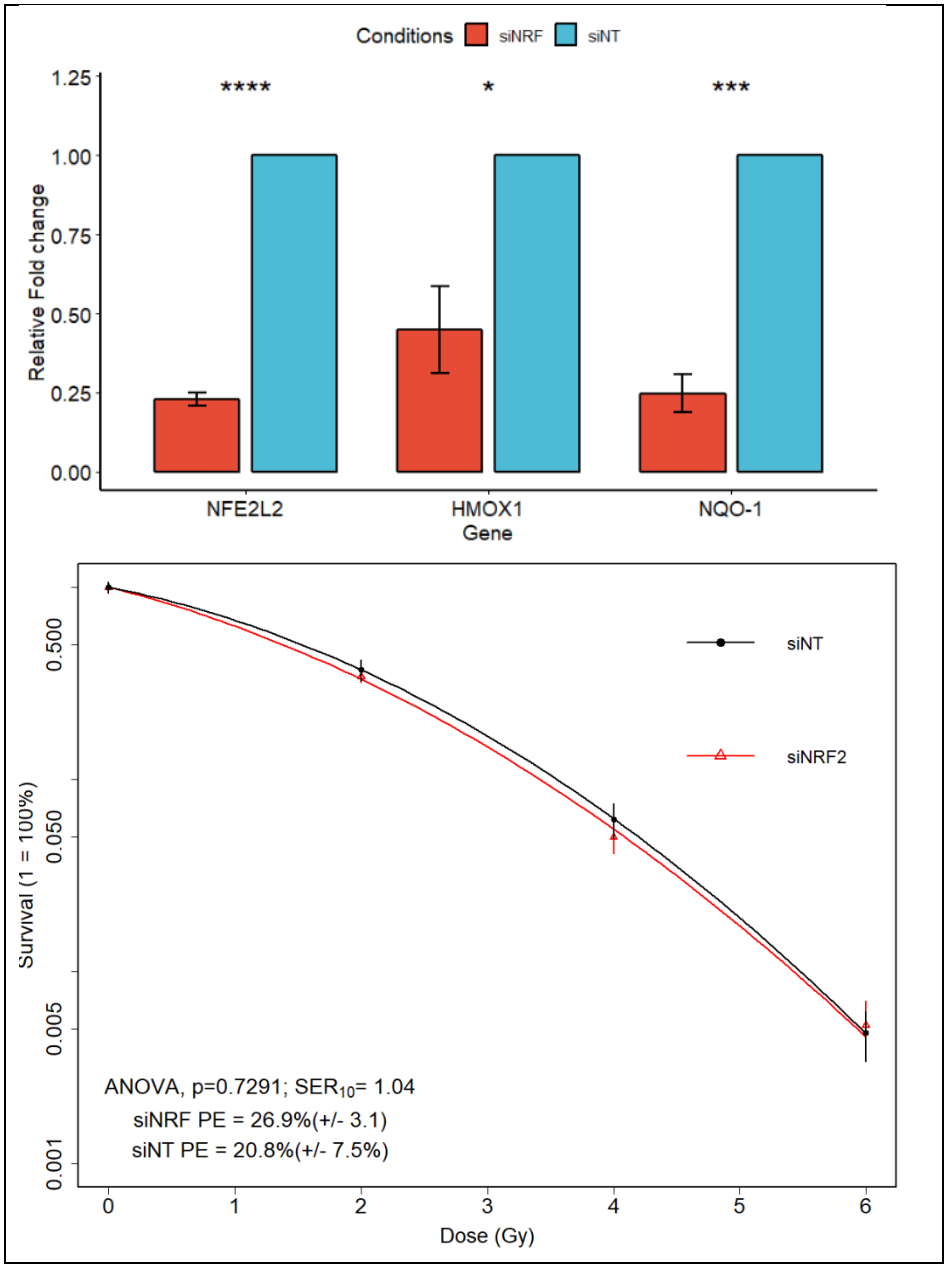


Figure 3-3. RNAi knockdown of NRF2 results in radiosensitivity of HCT116

Barplot of relative fold change in mRNA levels of NFE2L2 and two key NRF2 regulated genes following siRNA treatment (left). Data shown are mean +/- s.e.m of three repeats from triplicate wells (n=3). Cell survival curve of clonogenic assay showing no difference in radiosensitivity of siNRF2 treated cells (right). Statistical significance by ANOVA and plating efficiencies shown in bottom left of plot.

Comparing the clonogenic assay results of all showed the surviving fraction (SF) of HCT116, 0.005, was almost a full log scale lower than that SF at 6Gy of the control siRNA cells in SW1463, ~0.034 [Table 3-2]. The SF at 6Gy of HT55 was intermediate between both these points. The finding that the HCT116 was the most sensitive of the three lines is in keeping with published data [94] as discussed earlier. However, taking the three sets of results together one can see a possible trend, as illustrated by the decreasing SER_{10} values, that as cells become more intrinsically radiosensitive the role of NRF2 in mediating radioresistance potentially diminishes.

Table 3-2. Summary of surviving fraction (SF) at 6Gy of siNRF2 and control cells

The SF6Gy for HCT116 in the control treated cells is significantly lower than either of the other two lines. This is consistent with the known behaviour of these cells.

Cell line	**SF 6Gy**	
Conditions	siNT1	siNRF2
SW1463	0.034 (+/- 0.005)	0.016 (+/- 0.002)
HT55	0.024 (+/ - 0.004)	0.012 (+/- 0.001)
HCT116	0.0047 (+/- 0.001)	0.005 (+/- 0.001)

3.2.2 Consistent radiosensitivity with RNAi targeting alternate NFE2L2 exons

In order to rule out off target effects, two separate biological replicates were carried out using siRNAs which targeted NFE2L2 at different exons. The details of the siRNA are found in Chapter 2.

Transfection with either siRNA resulted in approximately 70-75% knockdown of NFE2L2 expression (Welch's t.test, $p<0.0001$) and also resulted in a significant change in expression of

NRF2 regulated genes HMOX-1 (Welch's t.test, $p<0.001$) and NQO-1 (Welch's t.test, $p<0.0001$) relative to the non-targeting control [Figure 3-3].

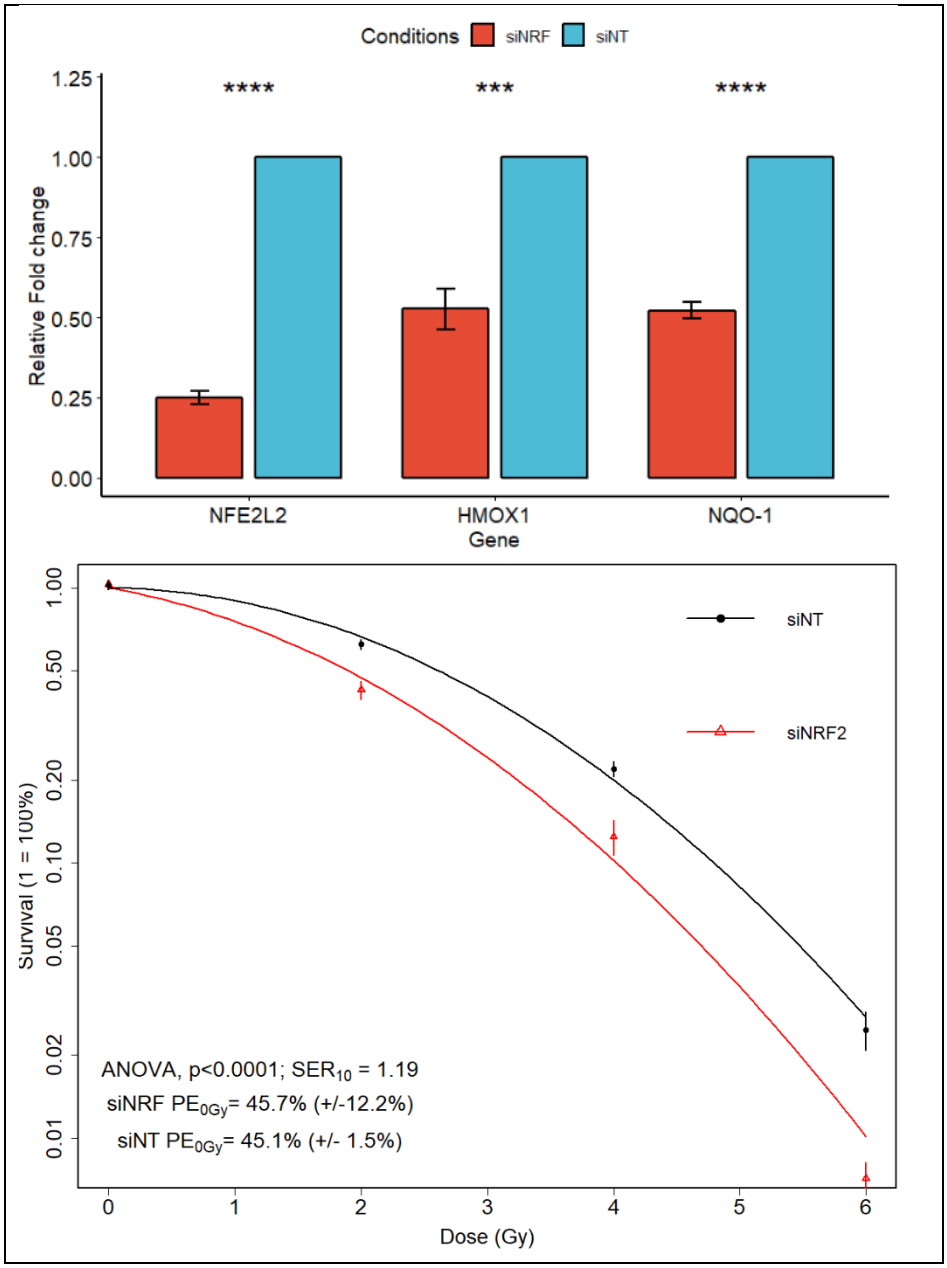


Figure 3-4. RNAi knockdown targeting second locus of NFE2L2 results in radiosensitivity of SW1463

Barplot of relative fold change in mRNA levels of NFE2L2 and two key NRF2 regulated genes following siRNA treatment (top). Cell survival curves of clonogenic assay showing a significant difference in radiosensitivity of siNFE2L2_2 treated cells. Statistical significance testing by ANOVA and plating efficiencies shown in bottom left of plot. Data shown are mean +/- s.e.m of three repeats from triplicate wells (n=3)

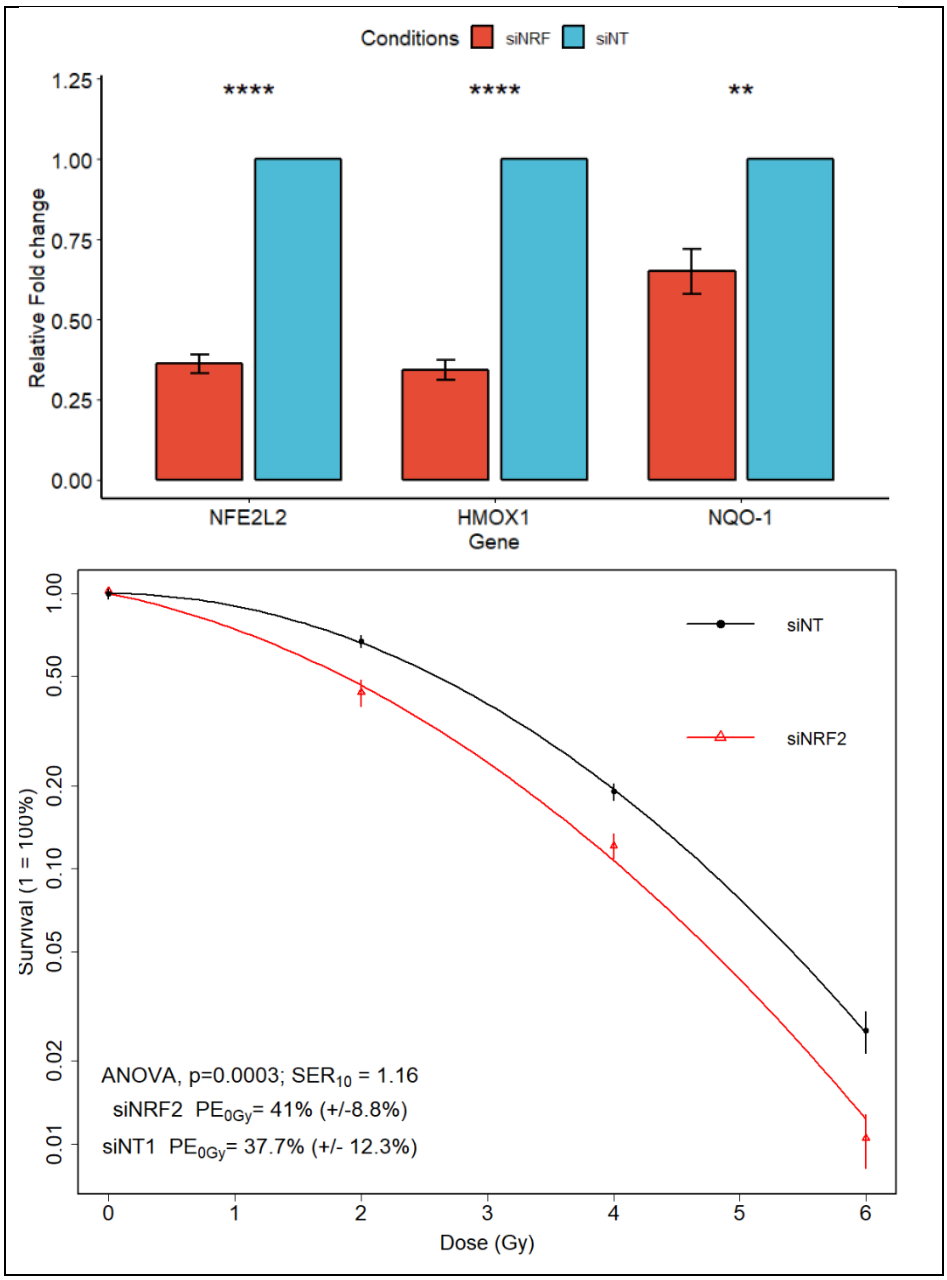


Figure 3-5. RNAi knockdown targeting third locus of NFE2L2 results in radiosensitivity of SW1463

Barplot of relative fold change in mRNA levels of NFE2L2 and two key NRF2 regulated genes (top). Cell survival curves of clonogenic assay showing a significant difference in radiosensitivity of siNFE2L2_3 treated cells. Statistical significance testing by ANOVA and plating efficiencies shown in bottom left of plot. Data shown are mean +/- s.e.m of three repeats from triplicate wells (n=3).

The cell survival curves from the resultant clonogenic assays demonstrated that there was a difference in radiosensitivity between cells treated with a non-targeting siRNA and that targeting NRF2 [Figure 3-4 & Figure 3-5]. The observed differences in surviving fraction were statistically significant (ANOVA, $p<0.0001$). The sensitization enhancement ratio at 10%

survival (SER_{10}) following knockdown of NRF2 was approximately 20%. These results were similar to the effect seen with siRNA_1.

3.2.3 CRISPR validation of siRNA induced changes

The robust and reproducible changes in gene expression and resulting radiosensitization with transient knockdown of NRF2 were encouraging, but the transient nature of the pathway interference limited more in depth study of NRF2 mediated changes. To produce a fixed aberration of NRF2 pathway signalling an attempt to engineer CRISPR-Cas9 isogenic lines with an *NFE2L2* knockout mutant and a *KEAP1* knockout mutant was made. The former was hypothesised to produce similar results to the RNAi experiments. KEAP1 knockout was hypothesised to simulate *in-vivo* methylation of *KEAP1* or a silencing mutation and increase NRF2 pathway expression by decreasing KEAP1 directed ubiquitination of NRF2.

3.2.3.1 Creation of isogenic NFE2L2 mutation in exon 4

CRISPR-Cas9 mutation was carried out on three target loci for *KEAP1* and *NFE2L2* (6 in total) of SW1463 cell line according to standard protocols [Figure 3-6]. The aim was to knockout one copy of the *NFE2L2/KEAP1* gene by causing insertions or deletions (indels) in the respective exons as detailed in Chapter 2 methods.

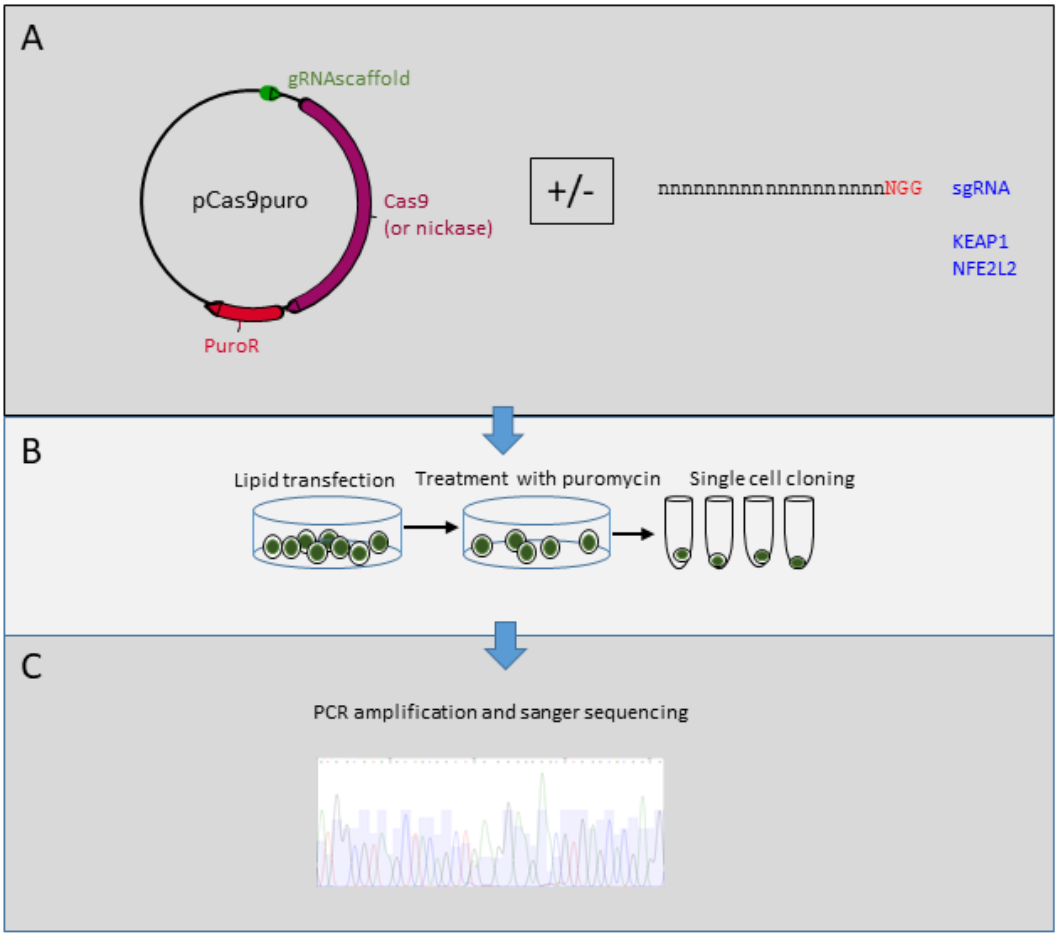


Figure 3-6. Schematic overview of the CRISPR-Cas9 workflow

A) *The guide sgRNA for targeting the desired locus and the plasmid with a puromycin resistance cassette (puroR) with the Cas9 nickase. B) The plasmid was transfected into the plated cells with a transfection reagent before selection by daily puromycin treatment based on previous optimisation titrations. Selected cells undergo serial dilution and single cell cloning. C) Surviving clones are screened by Sanger sequencing for evidence of the indel mutations at the desired locus.*

A puromycin selectable wildtype Cas9 plasmid, together with synthetic ready-to-use sgRNAs, were used in an attempt to improve efficiency. The CRISPR-Cas9 machinery was transfected into cells using lipid based DharmaFECT, followed by puromycin selection and single cell cloning. Optimisation of the puromycin dose for selection of SW1463 clones was performed prior to single cell cloning. A dose of 0.8mg of puromycin produced an estimated kill of 90% (C.I 83% - 98%) [Figure 3-7]. Therefore following transfection, 0.8mg of puromycin was added on days 2, 3 and 4 to induce cell death in non-resistant (i.e. untransfected) cells.

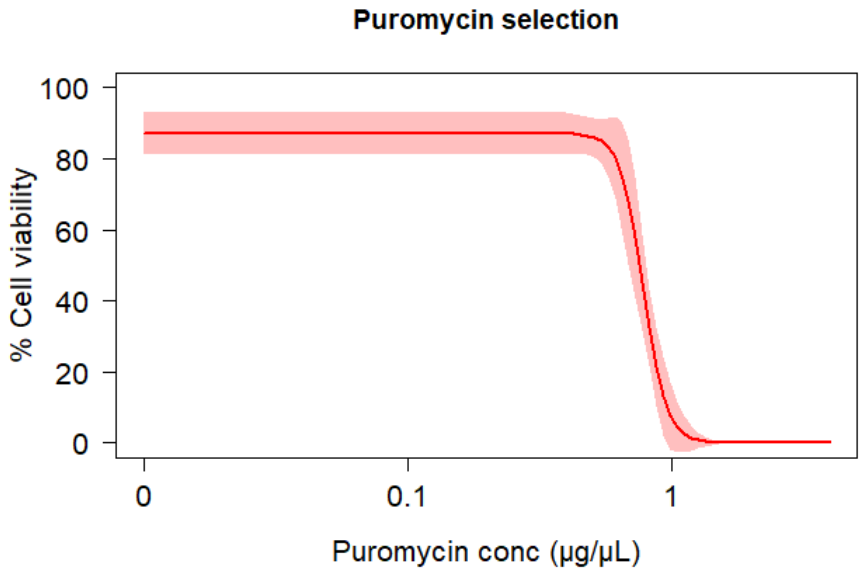


Figure 3-7. Dose response curve for SW1463 cell treated with puromycin

Serial, escalating doses of puromycin (µg/µL) were added to each well between 0.1 and 1.6 µg/µL and % cell viability estimated by trypan blue staining. A four parameter logarithmic dose response curve was fitted to the data and indicates a steep dose response between 0.6 and 0.8 µg/µL.

Single cell cloning by serial dilution in 96 well plates was then performed on surviving cells. In total, 29 *NFE2L2* and 25 *KEAP1* clones were viable and proliferating. Each of these were screened for the 3 possible indel mutations [Figure 3-8]. Clones were then split into replica plates with one kept in culture and DNA extracted from the other for mutation screening by Sanger sequencing. Of the 25 potential *KEAP1* mutant clones, no mutation was detected giving an overall efficiency of 0%. Of the 29 *NFE2L2* clones, 2 had mutations detected by Sanger sequencing [Figure 3-9] giving an overall efficiency of 2.3%.

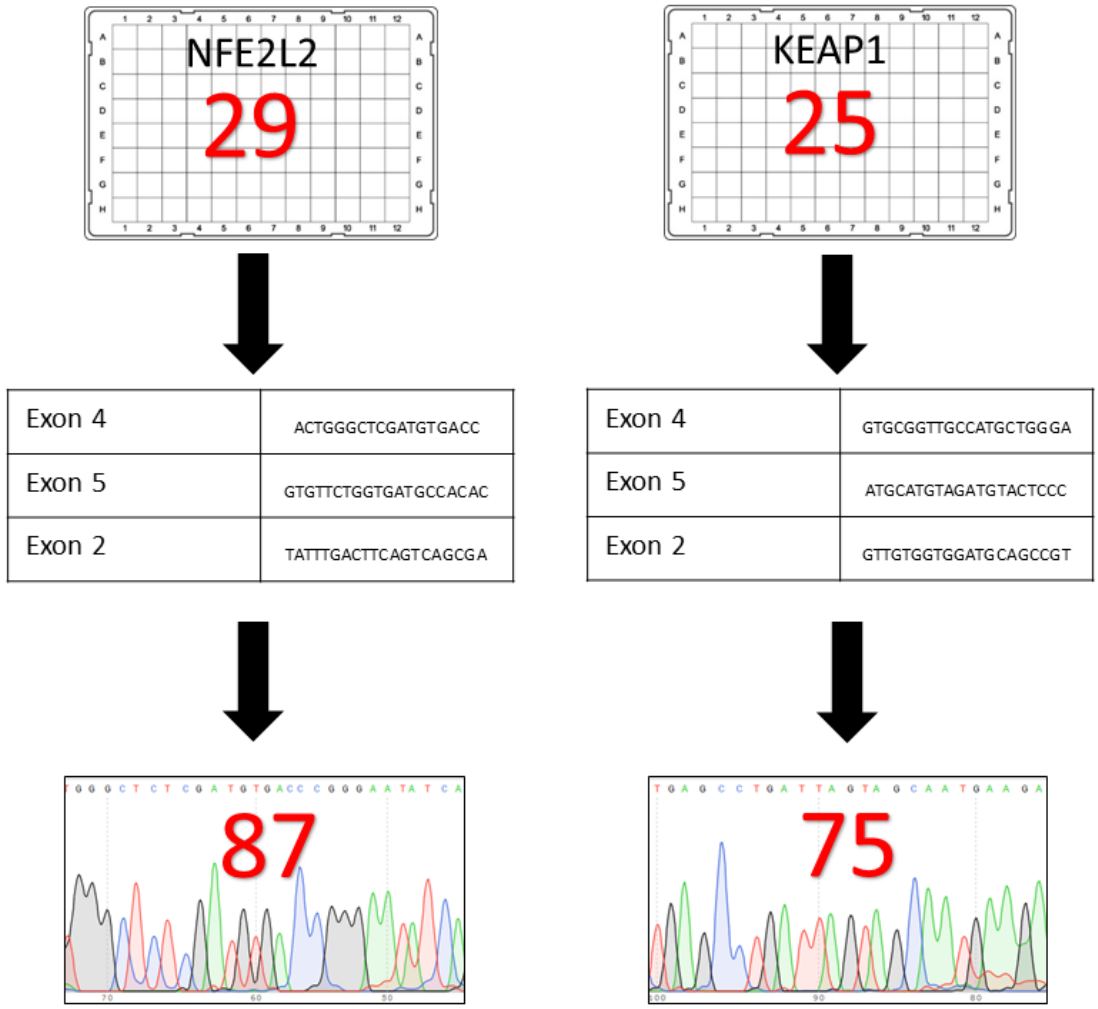


Figure 3-8. CRISPR clone screening and selection by Sanger sequencing

The surviving clones from the single cell cloning process were grown up in standard culture conditions until in sufficient number to allow DNA extraction. Each clone was screened for three potential indel mutations from the sgRNA with the total number of screening Sanger sequences shown at the bottom of the flow diagram.

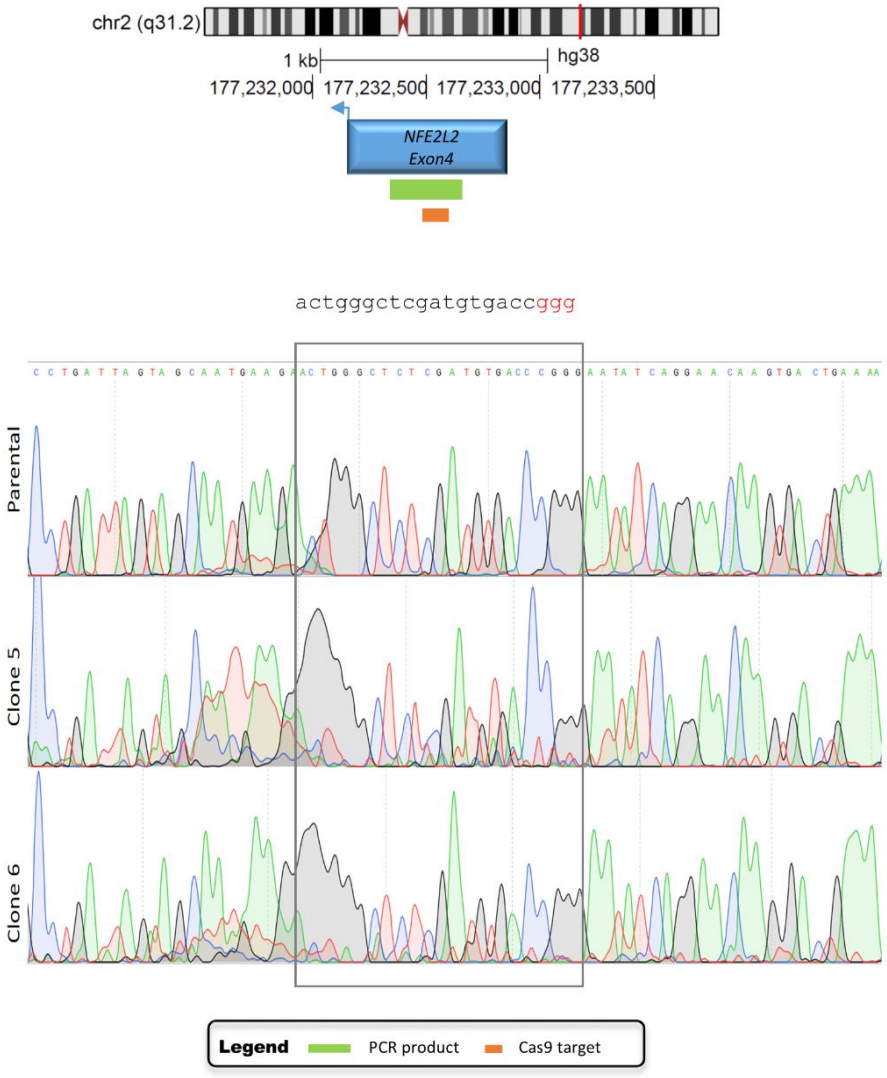


Figure 3-9. Sanger sequencing chromatograms of indel mutations in two derived clones of SW1463

Two SW1463 clones were identified. The reference sequence for exon 4 NFE2L2 on Chromosome 2 is shown on the top. The chromatograms of the mutant clones in reference to the parental line are shown underneath. The NGG protospacer adjacent motif (PAM) adjacent to the target DNA is highlighted in red, in this case 'GGG'. Approximately 4 bases upstream of the PAM the sequence trace of the clones begins to deteriorate with poor quality over the length of the target DNA sequence.

3.2.3.2 Isogenic NFE2L2 mutant line showed downregulated NRF2 target genes and increased radiosensitivity

Gene expression mRNA analysis of both derived clones and the parental line was carried out to determine if there were any changes in NRF2 regulated genes. Compared to the parental line, neither clone 5 or 6 showed significant changes in the expression of *NFE2L2* at the mRNA level

as might be expected [Figure 3-10]. However, clone 5 showed statistically significantly downregulated expression of the NRF2 regulated target gene TXN (Welch's t.test, p=0.0021), with no change in NQO-1 expression suggesting potentially lower levels of functional NRF2 protein in this clone. In contrast, clone 6 showed no statistically significant changes in expression of either NRF2 target genes.

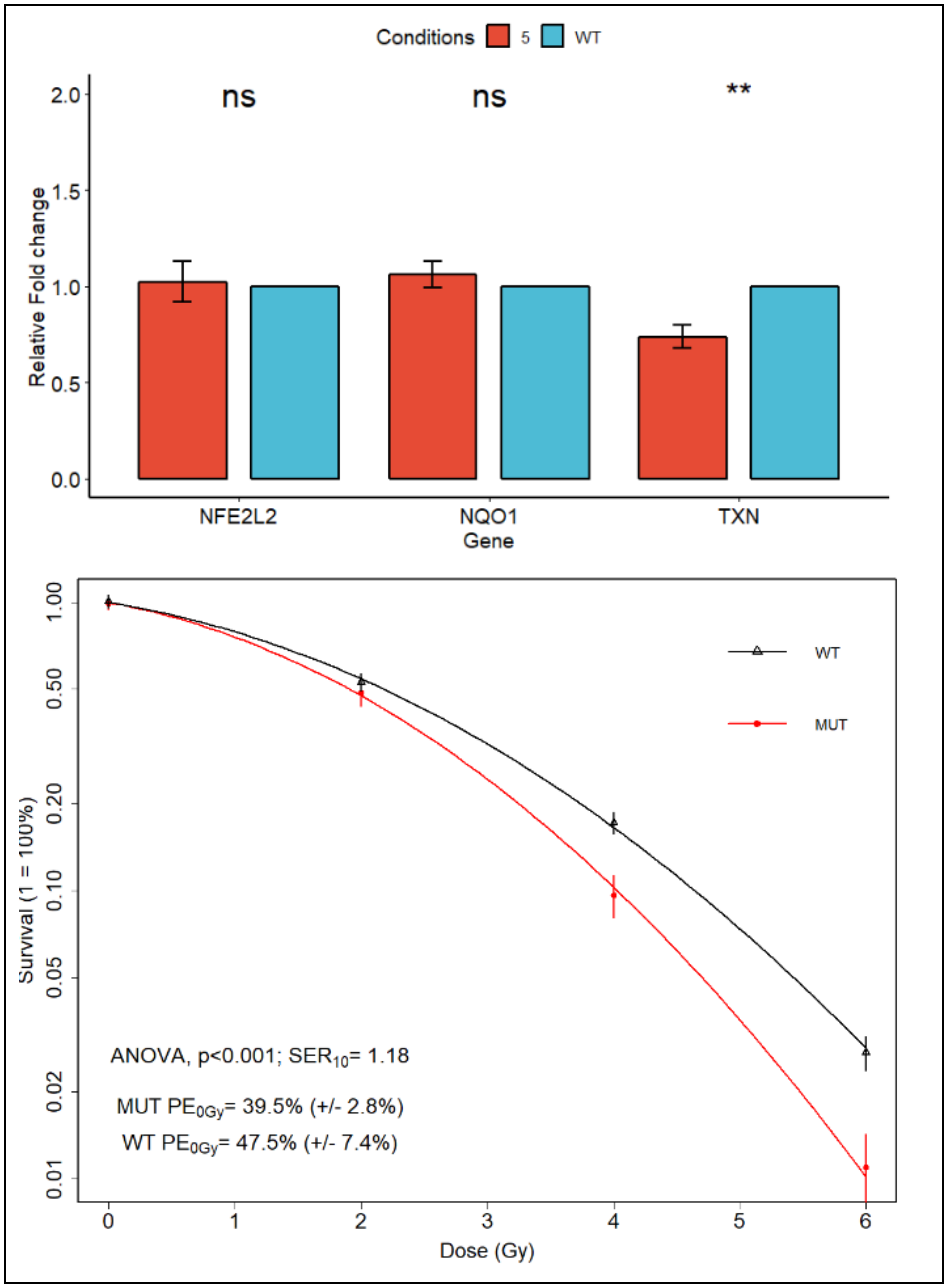


Figure 3-10. Barplot of qPCR gene expression analysis from SW1463 clone 5 , relative to SW1463 parental line. Cell survival curves of clonogenic assay with SW1463 parental wild type and exon 4 mutant clone 5

Clone 5 shows evidence of increased radiosensitization at a statistically significant level. Clone 5 shows significant downregulation of Thioredoxin (TXN) but not NQO-1.. Barplot data shown as mean +/- s.e.m (N=3, respectively).

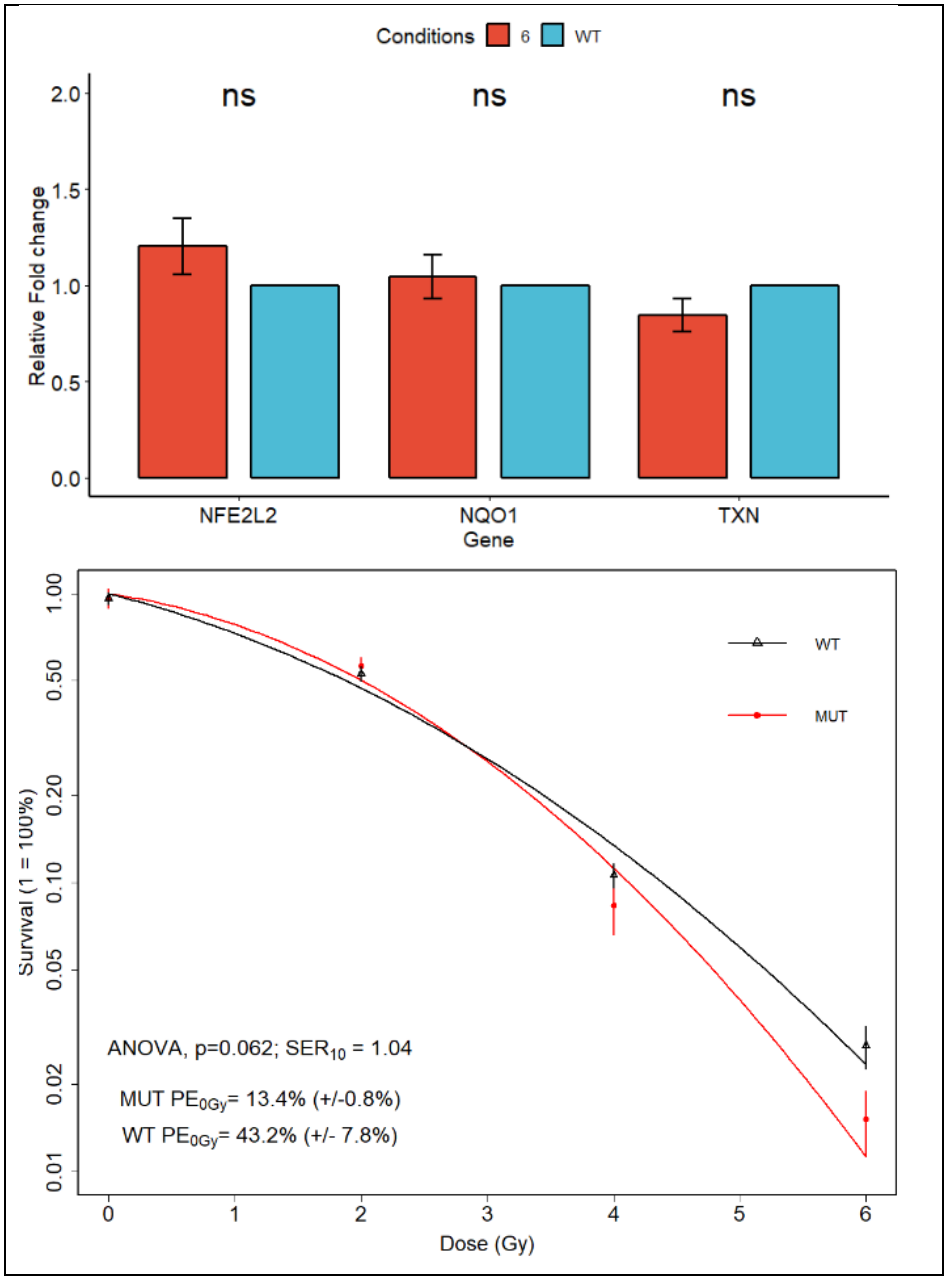


Figure 3-11. Barplot of qPCR gene expression analysis from SW1463 clone 6, relative to SW1463 parental line. Cell survival curves of clonogenic assay with SW1463 parental wild type and exon 4 mutant clone 6.

Clone 6 shows no evidence of increased radiosensitization at a statistically significant level. Clone 6 shows no significant change in expression of any NRF2 regulated targets. Barplot data shown as mean +/- s.e.m (N=3, respectively)

Having observed changes in KEAP1-NRF2 pathway expression, for clone 5 at least, changes in radiosensitivity were assessed. Clonogenic assays were carried out for both clones [Figure 3-10]. Clone 5 showed increased radiosensitivity compared to the parental SW1463 line (ANOVA, $p<0.001$) with a SER_{10} of 18%. Clone 6 shows no change in colony forming ability (ANOVA, $p=0.061$). However, the assay for clone 6 may not be reliable. The ability of clone 6 to form

colonies, even at control doses (0Gy), appeared to be significantly impaired compared to the parental line [Figure 3-12] with a plating efficiency of 13.4% compared to 43.2% for the WT parental line.

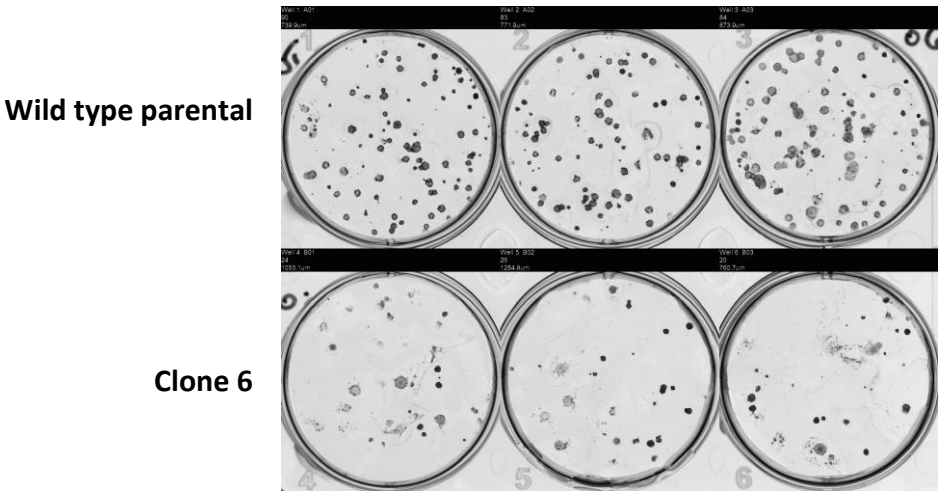


Figure 3-12. Colony formation at 0Gy of SW1463 parental (top) and SW1463 Clone 6 (bottom)

The colonies from the NFE2L2 mutant show irregular size and poor plating efficiency compared to the parental line. This poor plating efficiency probably stems from the significant chromosomal changes induced by the CRISPR treatment process as evidenced by the FISH analysis.

3.2.3.3 Karyotyping of SW1463 NFE2L2 exon 4 mutants

CRISPR-Cas9 gene editing has recently been reported to cause large scale deletions and complex genomic rearrangements [183]. These on and off target alteration are usually detected by experimental and / or computationally resource intensive techniques such as long-read sequencing, long range PCR or even whole genome sequencing. Our group has recently reported on a method for performing quality control on cancer cell lines using more straight forward techniques of chromosome karyotyping and FISH analysis (Rayner et al, in press, The CRISPR journal, 2019). Given the poor colony forming ability of one of the *NFE2L2* clones, these techniques were applied to both clones.

DAPI (4′,6-diamidino-2-phenylindole) banding showed some evidence of chromosomal instability in the mutant clones. SW1463 is an aneuploid line with a ploidy status of ~ 3n. However, the number and frequency of re-arrangements in both derived *NFE2L2* clones was higher and made analysis extremely difficult, to the point where it was not even possible for clone 6 [Figure 3-13].

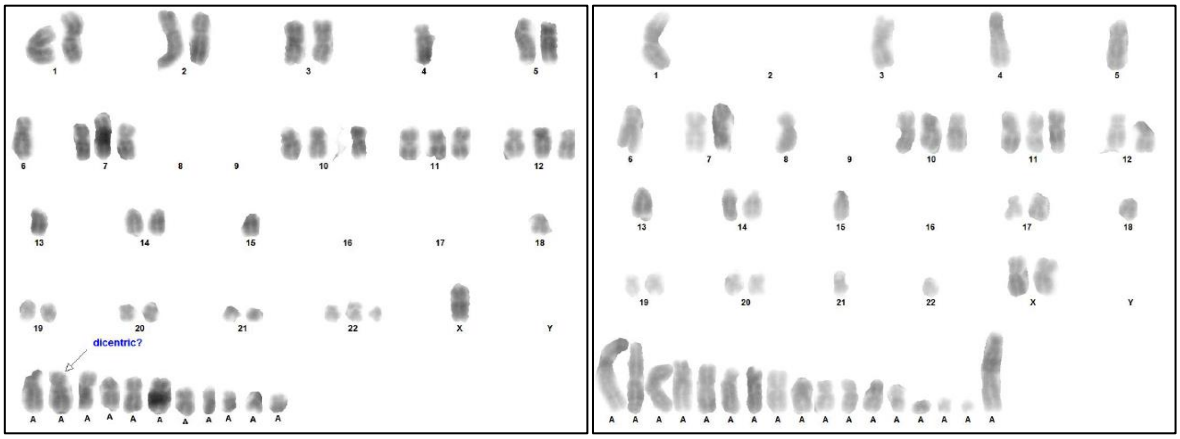


Figure 3-13. DAPI karyogram of cytogenetic analysis of SW1463, parental (left) and clone 5 (right)

The modal chromosomal numbers of SW1463 parental, clone 5 and clone 6 were also different [Figure 3-14]. The parental line has a clear modal number of 49. Although the modal number of clone 5 was also 49, the distribution of chromosomal number from across the 25 metaphase spreads was slightly different. The results from clone 6 however were significantly different, both in modal number but also in distribution. This supported the DAPI findings of chromosomal alterations, significant in the case of clone 6, following CRISPR editing.

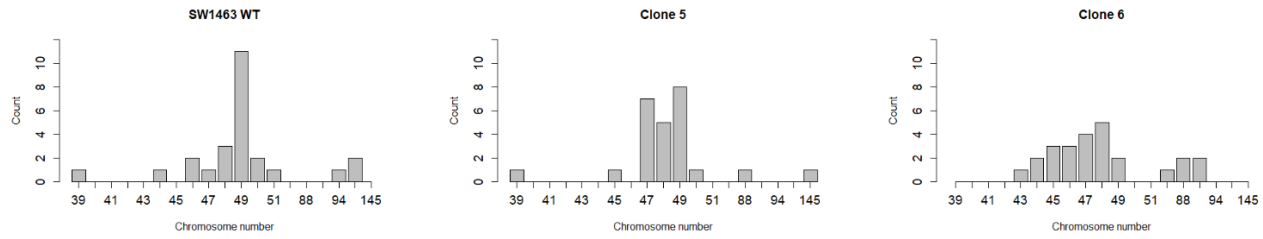


Figure 3-14. Barplots of chromosomal number from metaphase spreads for SW1463 WT, SW1463 NFE2L2 clone 5 and SW1463 NFE2L2 clone 6

The modal chromosomal number of the wild type line is clearly 49. The distribution is similar in clone 5 but significantly different in clone 6 (N=25)

3.2.3.4 FISH analysis of SW1463 CRISPR suggest significant cytogenetic abnormalities

Although the majority of cells examined contain two copies of NFE2L2, there was a suggestion that copy number of NFE2L2 was increased in the derived clones as indicated by the increased signal ratio [Table 3-3]. Rearrangements and large structural abnormalities are also more common in the derived SW1463 clones [Figure 3-15].

Table 3-3. Summary table of the FISH analysis for SW1463 parental line and clones

The signal to cell ratio observed in each cell line indicates the observed copy number of the different cell lines. The majority of cells had 2 copies of NFE2L2. All cell lines had some cytogenetic abnormalities in Chromosome 2, but this clearly increases in the CRISPR clones.

	1 signals /cell	2 signals /cell	3 signals /cell	4 signals /cell	5 signals /cell	Total nr. of cells analysed	Normal chr 2	Cells with >= 1 rearrangement
SW1463 WT	4%	88%	4%	4%	-	25	92%	8%
SW1463 clone 5	16%	68%	12%	4%	-	25	88%	12%
SW1463 clone 6	5%	80%	15%	-	-	20	64%	36%

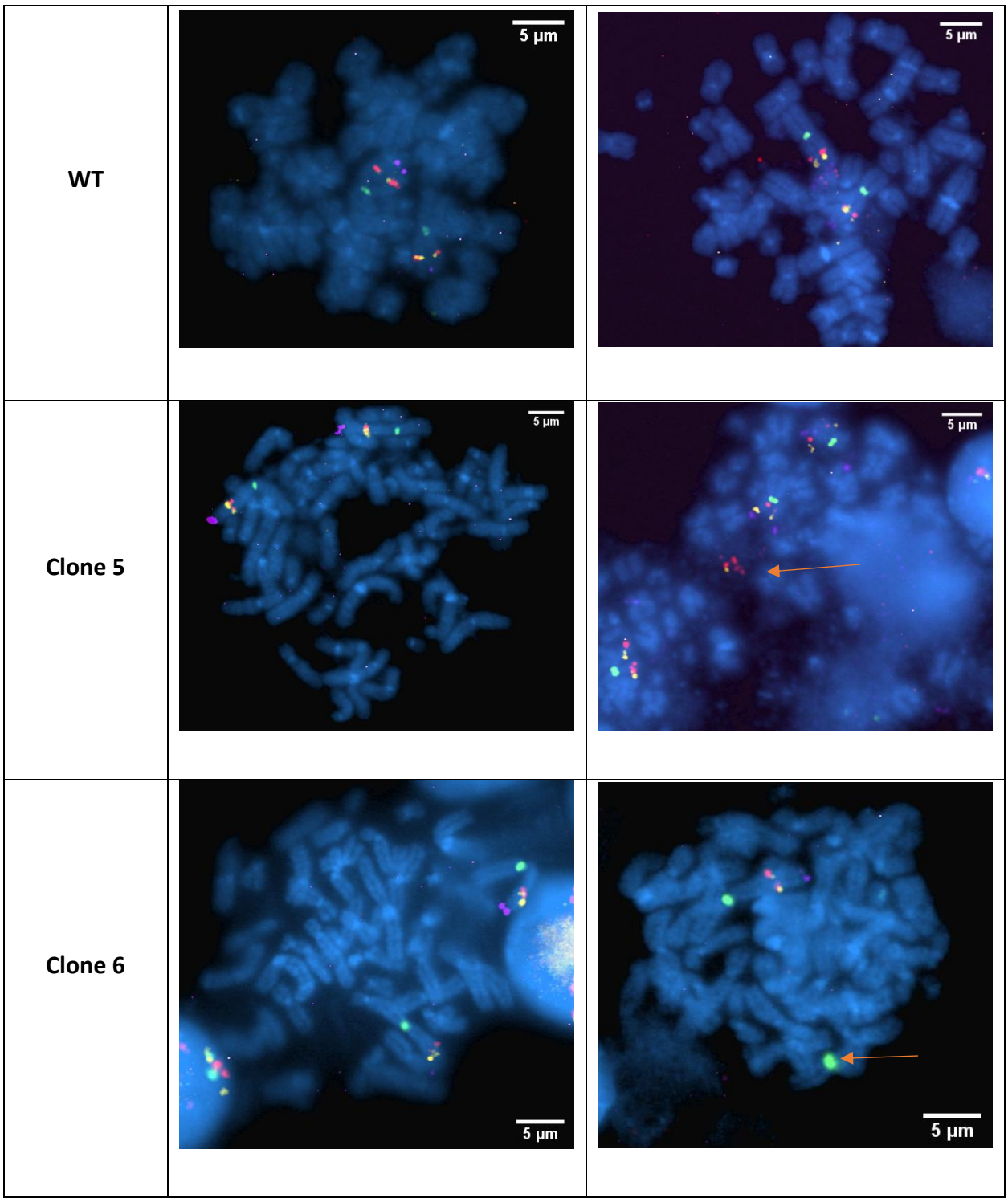


Figure 3-15. FISH metaphase spreads from SW1463 WT line and SW1463 NFE2L2 clones 5 and 6

The chromosomes are counterstained in DAPI, blue. Normal chromosome 2 are identified by white arrows. A chromosome fragment, carrying material from the NFE2L2 locus is shown by an orange arrow (middle, right). A truncated chromosome 2 in clone 6 lacking the NFE2L2 locus and telomeric region is shown by an orange arrow (bottom, left) [images courtesy of Dr Daniela Moralli]

In SW1463 WT, a cell with fragmentation of the NFE2L2 locus and one translocation was identified. Clone 5 also had a translocation and some acentric fragments. Overall, these data suggested that clone 5 was not significantly structurally different when compared to the SW1463 wild type parental line. Clone 6 however contained a higher percentage of rearrangements, including around the NFE2L2 locus, fragmentation, truncation, translocations of chromosome 2 and double stranded breaks associated with the NFE2L2 locus. As such they were regarded as significantly different from the parental line and not suitable for downstream analysis.

Within the limitations of the CRIPSR-Cas9 technique in an aneuploid cancer cell line, SW1463 clone 5 could be regarded as an isogenic mutant of its parental line. With further QC it may have been sufficient for further study of NRF2 in CRC models. However, with the express purpose of understanding the difference in radiosensitivity caused by knockdown of NRF2, this cell line was not considered ideal due to the presence of, albeit limited, chromosomal changes. Chromosomal instability is known to cause radiosensitivity [94]. The modest radiosensitization seen in the colony forming assay of clone 5, although similar to that seen with RNAi, would be difficult to ascribe solely to the effects of NFE2L2 knockdown. The generation of several further NFE2L2 mutant clones, each demonstrating similar radiosensitization, would have been required to validate this experiment.

3.2.4 Resistant cell lines have increased NRF2 signalling

To understand the potential mechanisms behind the differences between the 3 cell lines, the mRNA levels of NFE2L2 and NQO1 in the non-targeting controls were tested. The hypothesis was the resistant lines may contain higher levels of pathway expression at baseline, which when downregulated, facilitated a more sensitive phenotype.

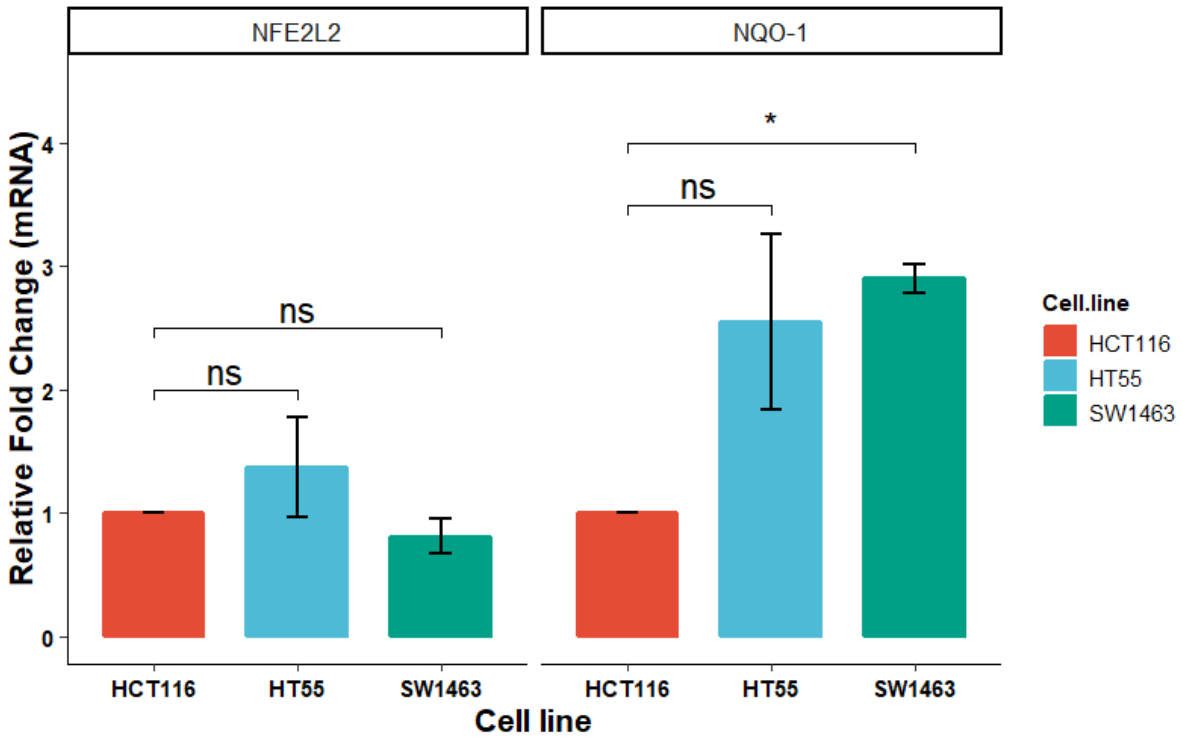


Figure 3-16. Barplot showing relative fold change in mRNA levels of NRF2 targets in control cells

To understand the levels of NRF2 signalling present at baseline, the mRNA levels of NFE2L2 and NQO-1 were measured in the control treated cells of the siRNA experiments. No significant difference was seen in NFE2L2 levels. However, there are increased levels of NQO-1 in both HT55 Welch's t.test (p = 0.27) and SW1463 (Welch t.test; p = 0.04). Data shown are mean +/- s.e.m (N=2)

Relative to the HCT116, the other two cell lines appeared to have higher basal levels of NRF2 signalling as suggested by increased NQO-1 mRNA levels. No differences in NFE2L2 expression point towards post transcriptional modifications in the pathway. This suggested that it may be possible to induce NRF2 pathway signalling which may in turn increase the radioresistance of the HCT116 cell line.

3.2.5 Induction of NRF2 signalling causes radioresistance

3.2.5.1 Tert-Butyl-Hydroquinone

Tert-butyl-hydroquinone (t-BHQ) is a known potent activator of NRF2. The mechanism is not fully understood but it is thought to do so by stabilising NRF2 in the cytoplasm [184] and increasing the phosphorylation of NRF2 to allow for improved nuclear transport [185] thus causing pathway activation. It may also be involved in regulating NRF2 at a transcriptional level[186]. Treatment with t-BHQ did induce NRF2 pathway expression as evidenced by increased expression of NRF2 regulated genes, HMOX-1 and NQO-1 [Figure 3-17]. Unfortunately the process of incubating cells with t-BHQ meant that the recovered cells were unable to successfully form colonies following plating for a clonogenic assay to assess changes in radioresistance. Therefore alternative strategies for upregulating NRF2 signalling were sought.

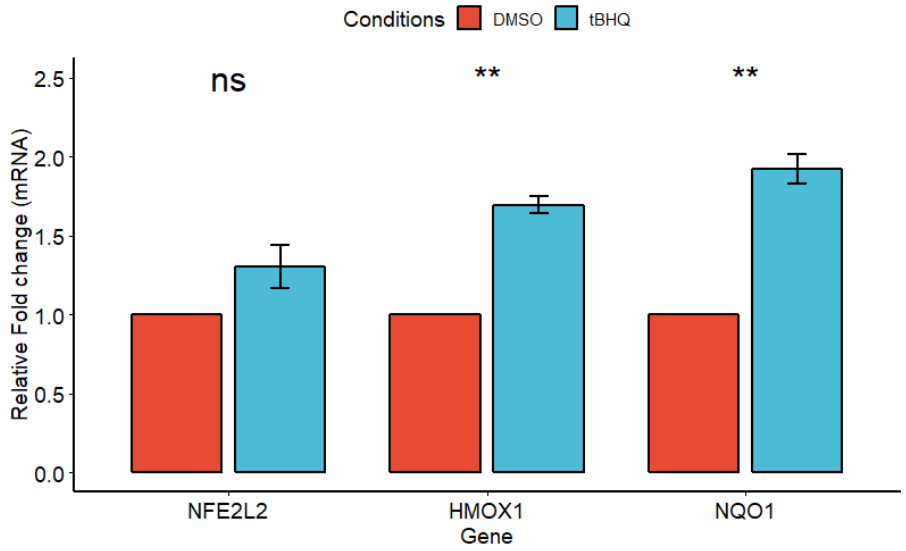


Figure 3-17. Barplot of qPCR gene expression analysis of HCT116 following treatment with t-BHQ 50µM, relative to treatment with DMSO 50µM

Cells were incubated for 72 hours and expression analysed. t-BHQ resulted in significant increases in NRF2 downstream targets. Data shown as mean +/- s.e.m (N=3)

3.2.5.2 NFE2L2 E79K+ mutation

There is a documented mutation in NFE2L2, with a Catalogue of Somatic Mutations in Cancer (COSMIC) ID COSM132851 [https://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=NFE2L2]. A missense substitution mutation occurs at base position 235 (c.235G>A) where the native guanine is substituted for adenine. The resulting coding error at an amino acid level, where lysine (K) is read instead of glutamate (E), occurs in the crucial E_{79}TGE motif. The ETGE motif is a key binding domain for KEAP1 [Figure 3-18]. The conformational change caused by this missense substitution means that KEAP1 cannot bind NRF2 for degradation, leading to accumulation of cytoplasmic NRF2 and chronic, constitutive activation of the pathway. Prior to embarking on further assays, quality control was performed on the isogenic lines to assess the presence of the missense mutation and the presence of any chromosomal aberrations following the CRISPR Cas9 process.

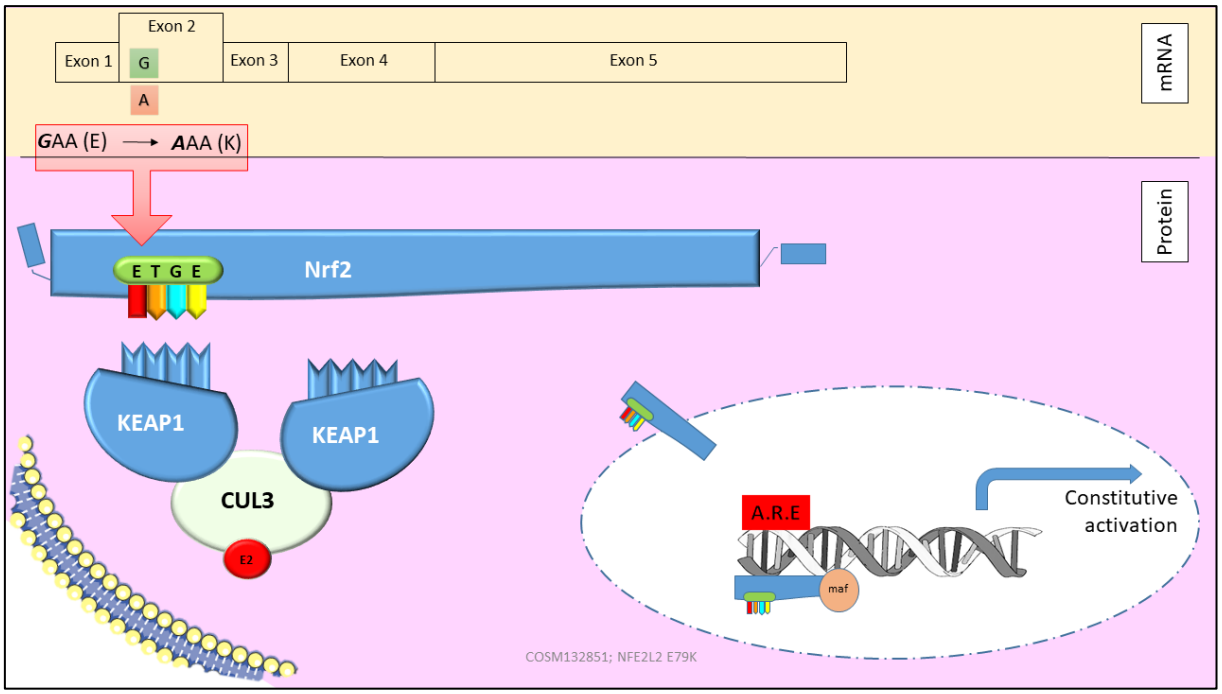


Figure 3-18. Location and functional consequences of mutation in the *NFE2L2* gene

A missense mutation in exon 2 of the gene c.235 G>A results in an amino acid coding error, pGlu79Lys, in the ETGE binding motif (top). This area is required for KEAP1 binding and in turn ubiquitination of NRF2 protein. The missense mutation prevents KEAP1 mediated degradation resulting in increased NRF2 levels in the absence of stress and constitutive activation of the pathway. [adapted from Huppke et al. [187]]

3.2.5.3 Mutation sequencing

Following DNA extraction from HCT116^{WT} and HCT116^{E79K+} and Sanger sequencing as previously outlined, the chromatograms indicated a base change consistent with the known mutation. The WT line showed a 'G' base on the chromatogram [Figure 3-19]. The E79K+ line shows a small 'G' peak but also the presence of an 'A' base call. This indicated the presence of heterozygote mutation at this base position, relative to the parental line.

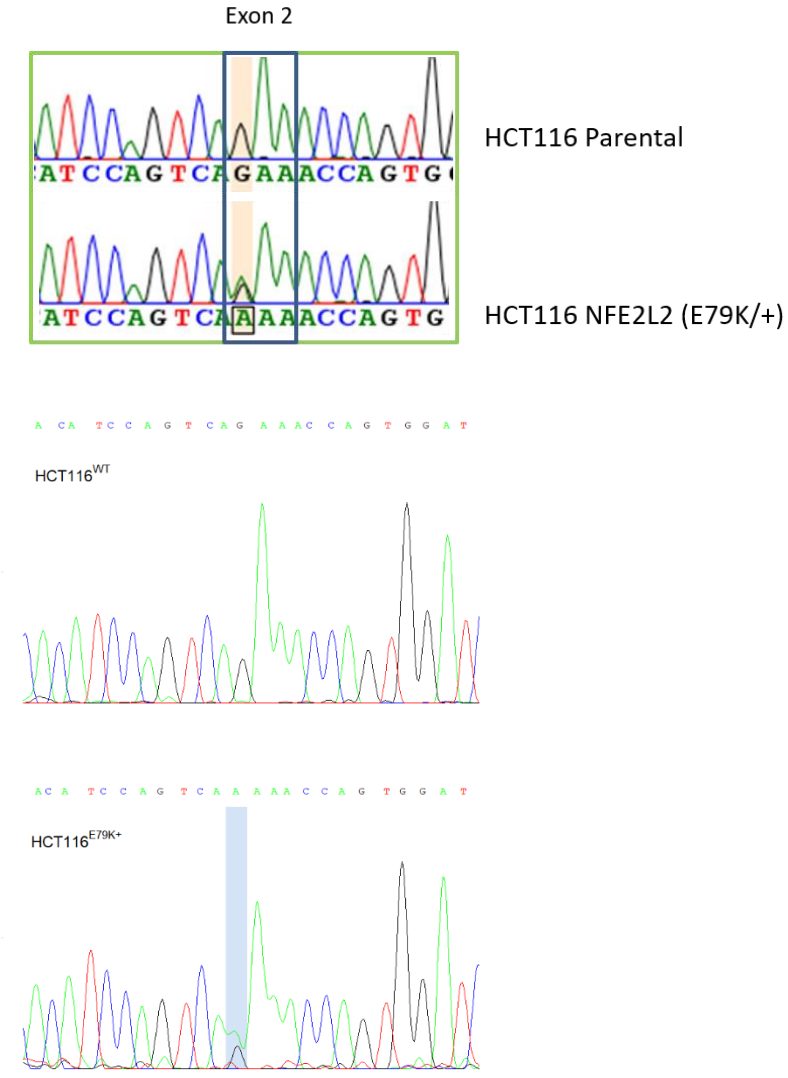


Figure 3-19. Sanger sequencing chromatrograms showing the missense point mutation in HCT116 E79K+.

The top diagram shows the reference sequence from Horizon. Below is the verification plot, highlighting the heterozygote A (green) G (black) base call in the mutant line, compared to the homozygote G (black) in the parental wild type line.

3.2.5.4 Karyotyping HCT116 CRISPR

CRISPR alteration of cell lines has the potential to introduce large deletions and complex rearrangements [183]. This was previously identified as an issue when using CRISPR Cas9 on cancer lines in SW1463, and for other cell lines in lab group. Therefore significant structural changes as a result of CRISPR editing had to be ruled out prior to functional investigation, in case any differences in gene expression or radiosensitivity could be attributed to chromosomal translocations.

In the first instance actively growing cells were harvested and chromosomes were stained in metaphase with DAPI and showed similar karyotypes [Figure 3-20]. Of 25 metaphase spreads of wild type HCT116, 48% show a chromosomal number of 45 [Figure 3-21]. 64% of HCT116^{E79K+} had a chromosomal number of 45. The distribution across the metaphase spread was also broadly similar. This was confirmed by FISH (Fluorescent in situ hybridisation) of the NFE2L2 locus at 2q31.

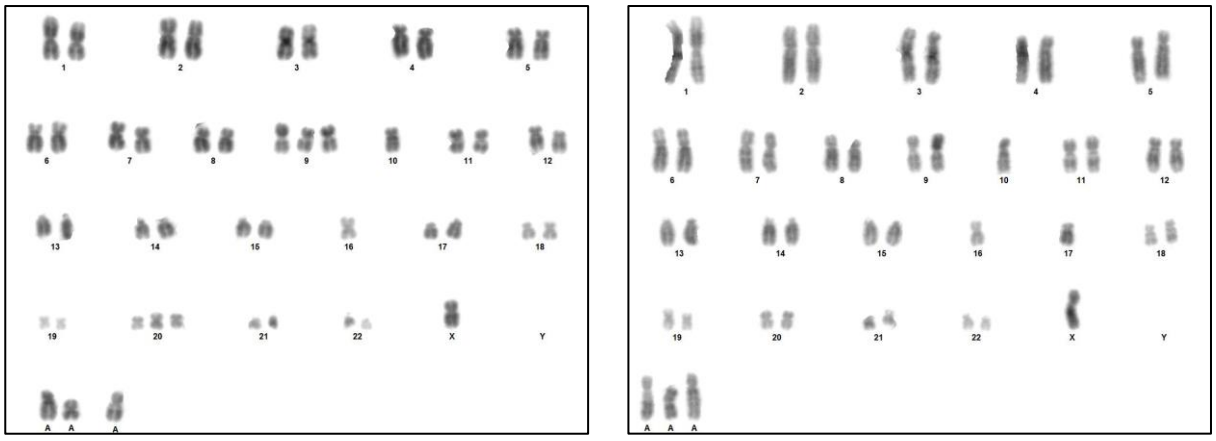


Figure 3-20. DAPI karyogram of cytogenetic analysis of HCT116 wild type (left) and HCT116 E79K+ mutant (right)

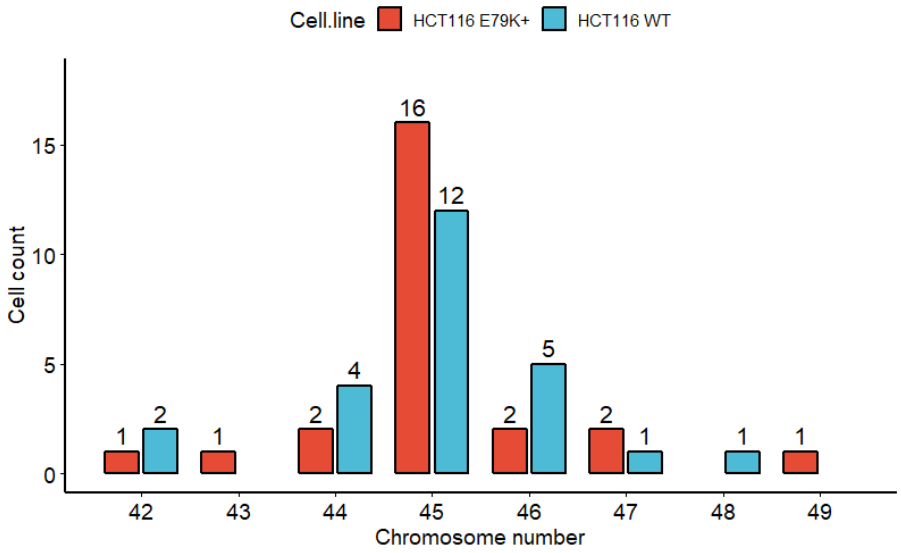


Figure 3-21. Barplot of chromosomal counts in metaphase spreads for HCT116 WT and HCT116 E79K+

The modal number for both cell lines is 45, in keeping with published modal number for HCT116 (N=25)

3.2.5.5 FISH HCT116 CRISPR

FISH analysis of the NFE2L2 locus shows that the majority of both the wild type line and the E79K+ mutant line had two copies of NFE2L2 [Figure 3-22 ; Table 3-4]. This was in keeping with a modal chromosome number of 45 or approximately diploid.

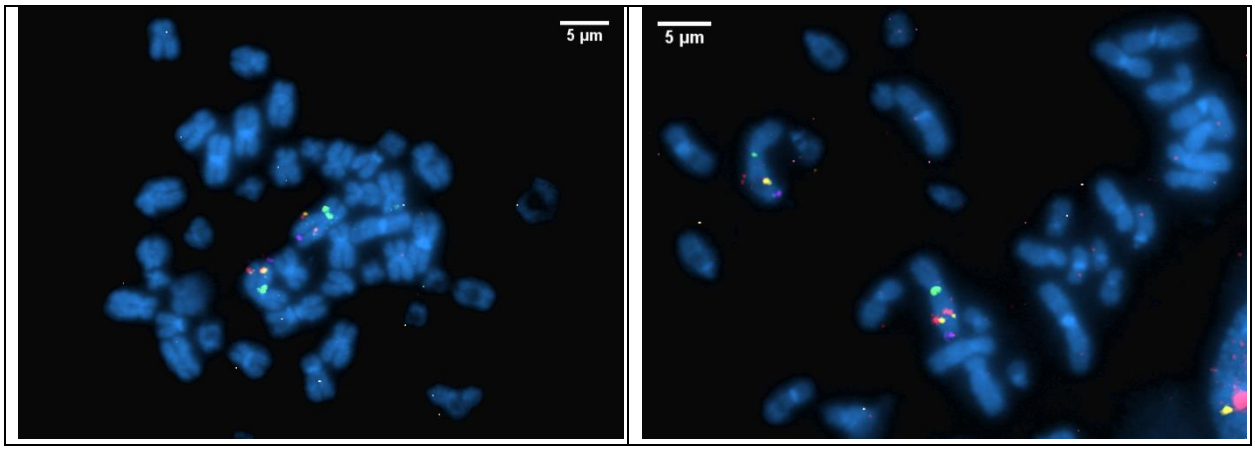


Figure 3-22. FISH from metaphase spreads from HCT116 WT (left) and HCT116 E79K+ (right) with probes as previously indicated

The chromosomes are counterstained in DAPI, blue. Normal chromosome 2 are identified by white arrows [images courtesy of Dr Daniela Moralli]

Table 3-4. Summary table of the FISH analysis for HCT116 WT and E79K+ lines

The signal to cell ratio observed in each cell line indicates the observed copy number of the different cell lines. The majority of cells had 2 copies of NFE2L2.

	1 signal /cell	2 signals /cell	3 signals /cell	4 signals /cell	5 signals /cell	Total nr. of cells analysed	Normal chr 2	Cells with >= 1 rearrangement
HCT116 WT	10%	75%	5%	5%	5%	20	100%	0
HCT116 E79K+	-	96%	-	4%	-	25	100%	0

In addition cells showed that the locus of NFE2L2 was found to be present on normal versions of chromosome 2, in so far as the technique would allow one to confirm, implying that no significant translocations or truncations had occurred in the region of NFE2L2.

3.2.5.6 HCT116 E79K+ cells have increased NRF2 activity

Having established that there were no significant structural differences between the isogenic HCT116 cell lines I sought to examine the functional differences. Gene expression analysis of $HCT116^{E79K+}$ mutant cells and $HCT116^{WT}$ parental lines was undertaken prior to investigation of changes in radiosensitivity [Figure 3-23]. Known NRF2 regulated genes TXN and HMOX1 were chosen to assess functional changes in pathway expression. KEAP1 and NFE2L2 were also assessed as the primary regulators of the pathway. E79K+ mutant cells showed evidence of statistically increased expression of HMOX1 (Welch's two sample t-test, p=0.0082) and TXN (Welch's two sample t-test, p=0.0076). Although NQO1 expression was also increased this increase was not statistically significant on testing (Welch's two sample t-test, p=0.7882).

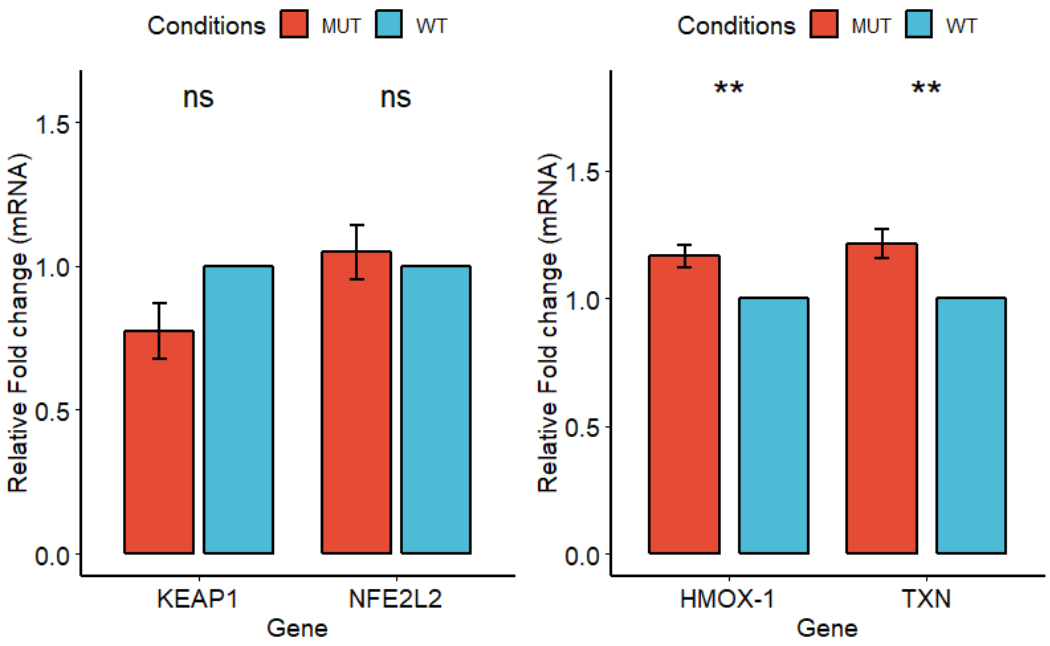


Figure 3-23. Barplot of relative fold change (mRNA) from HCT116 WT and E79K+ mutant cells

On the left, expression of pathway regulators KEAP1 and NFE2L2 are unchanged. On the right expression of downstream targets HMOX-1 and TXN are significantly increased. Data shown are mean +/- s.e.m (N=3)

There was no evidence of increased expression of KEAP1 (Welch's two sample t-test, p=0.056) or NFE2L2 (Welch's two sample t-test, p=0.616), consistent with the pathology of the mutation whereby the mutation occurs at the protein level and one would not expect to find changes in mRNA levels of *NFE2L2* itself. Increased levels of NRF2 protein were demonstrated by Western blotting from whole cell lysates [Figure 3-24].

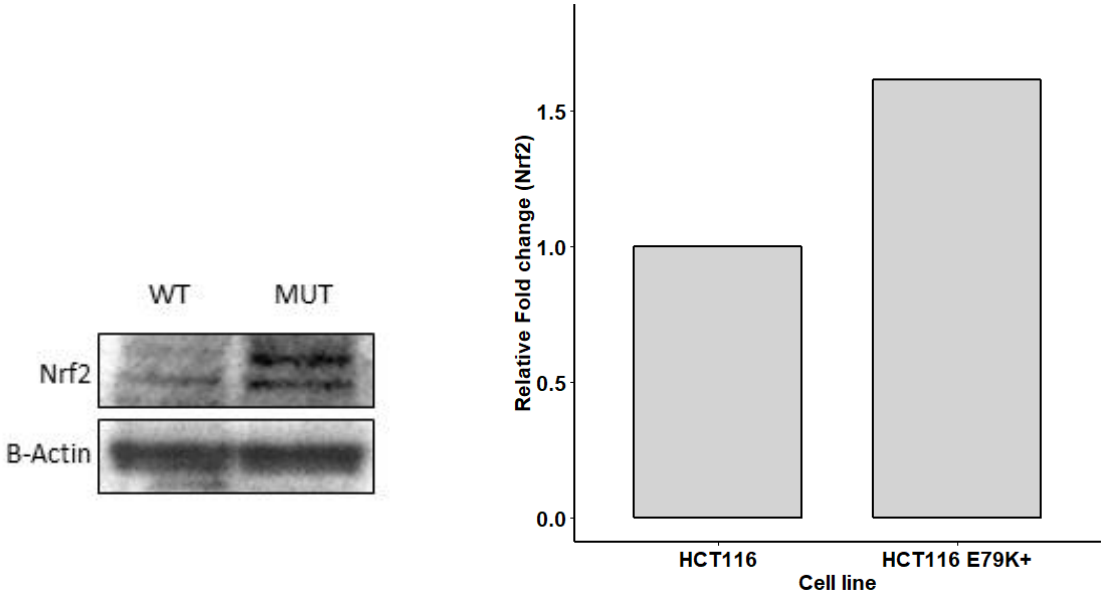


Figure 3-24. Western blots (cropped) of NRF2 protein (top) with B-actin as a loading control (left)

The double band is indicative of both the phosphorylated and unphosphorylated protein. Quantitatively more NRF2 protein is present in whole cell lysates of HCT116 E79K+ relative to HCT116 wild type cells (right) when normalised to B-actin levels

3.2.5.7 HCT116 E79K+ show radioresistance compared to WT cells

Clonogenic assays were performed to assess changes in radiosensitivity between the isogenic HCT116 cell lines [Figure 3-25]. The plotted cell survival curves show that mutant cell lines were more resistant to radiation with evidence of increased colony forming ability between doses of 2 and 6Gy. The SER_{10} value of 0.71 indicates that in mutant cells approximately 30% more dose is required to produce a surviving fraction of 10%. The survival curve of the parental wild type line was of a similar scale to the cell survival curve seen with HCT116 in the siRNA experiments, with a SF 6Gy of 0.0047 (+/- 0.001) for NT1 treated cells and 0.003 (+/- 0.0005) for $HCT116^{WT}$.

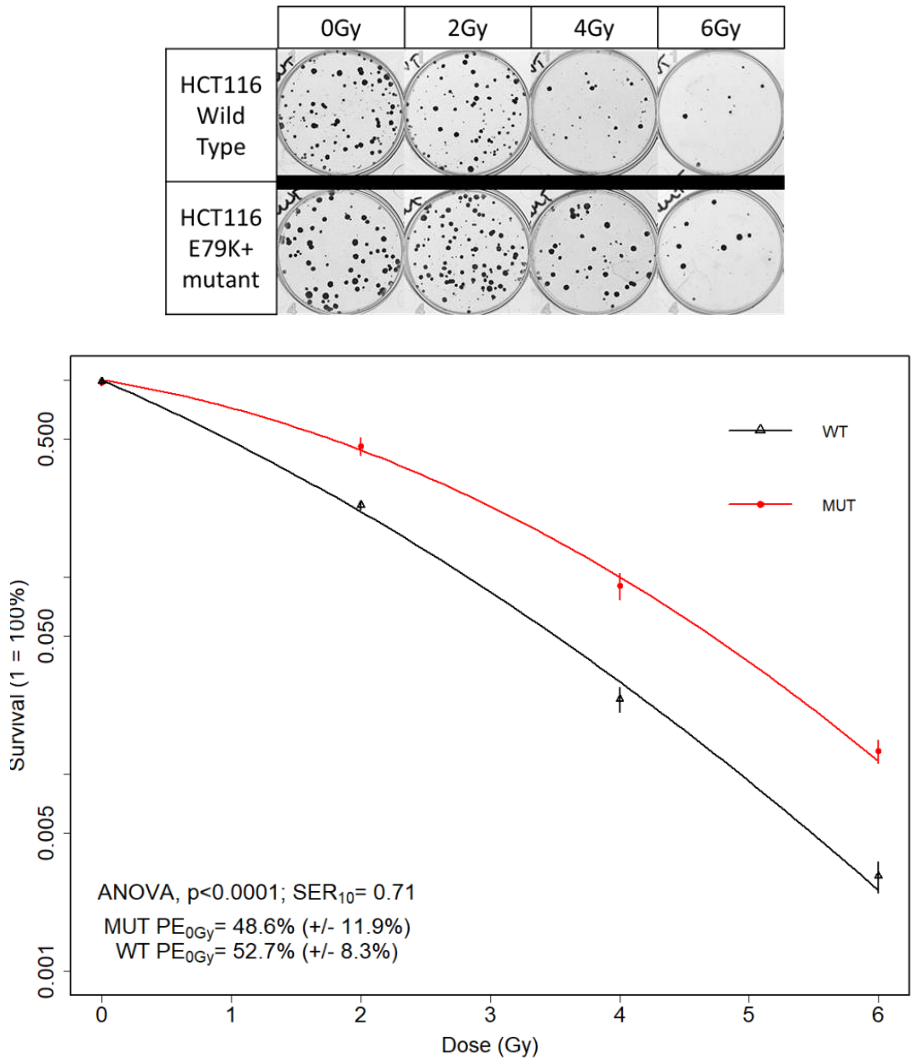


Figure 3-25. Colony formation assay (top) and cell survival curves (bottom) from CFA of HCT116 E79K+ (red) and HCT116 WT (black)

Mutant cells show evidence of increased radioresistance and increased cell survival at doses of 2, 4 and 6Gy. Data are representative of three independent experiments.

3.2.6 Overexpression of NRF2 induces chemoresistance

Having established that increased levels of NRF2 pathway expression results in radioresistance, I next sought to examine whether this mechanism could also lead to other forms of therapeutic resistance. Fluorouracil, oxaliplatin and irinotecan (the active metabolite of which is SN-38) are often used as single agents and in combination with each other in CRC and increasingly in combination with radiation in rectal cancer. The isogenic CRISPR HCT116 model was used to evaluate this further.

3.2.6.1 5-Fluorouracil

Cell viability assays were performed as indicated in the methods section. 5-FU functions by inhibiting thymidylate synthase thereby interrupting nucleotide synthesis in DNA strands. Treatment with serial doses of 5-FU were used to generate a dose response curve [Figure 3-26-A]. The estimated effective concentration to produce 50% cell death (EC_{50}) was 535 nM for wildtype and 2,572 nM for mutant cells. The logarithmic regression models used to plot the curves were compared using ANOVA and found to be significantly different ($p<0.0001$) [Figure 3-26-A].

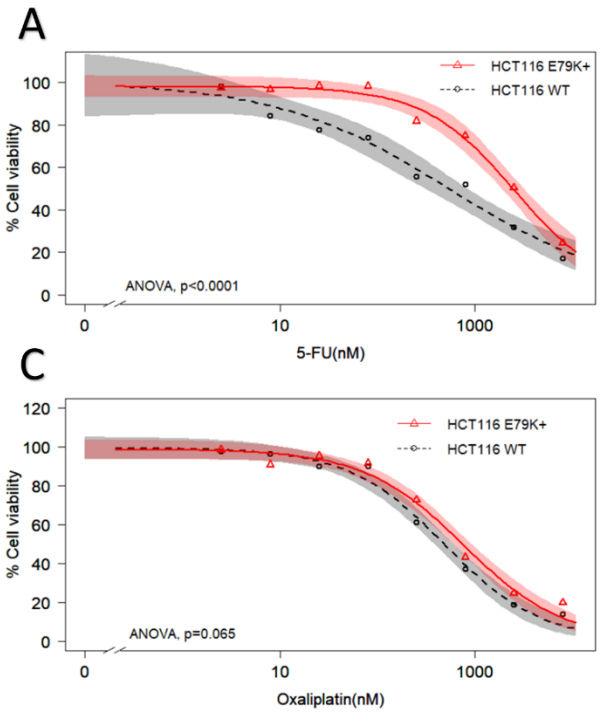
3.2.6.2 SN-38 (Irinotecan)

SN-38 is the active metabolite of Irinotecan. By mechanism, it inhibits Topoisomerase I, DNA synthesis and causes DNA single strand breaks. SN-38 was used to avoid any confounders regarding differences in metabolism between the isogenic lines. The cell kill potency of SN-38 is higher than irinotecan or camptothecin and required the original assay to be re-optimised with 10,000 cells per well. E79K+ mutant cells showed evidence of increased cell viability compared to wild type cells [Figure 3-26-B]. The EC_{50} for mutant cells was 160 nM and for wild type cells 1.6 nM. Logarithmic regression models were fitted as dose response curves to the data and were found to be significantly different ($p<0.0001$). Figure 3-26-D shows a comparison of the EC_{50} for all three chemotherapeutics and the ratios of the doses required for EC_{50}.

3.2.6.3 Oxaliplatin

By mechanism, oxaliplatin functions by causing DNA adducts which inhibit DNA synthesis. Again using the isogenic HCT116 model similar experiments were carried out using oxaliplatin. The EC_{50} for mutant cells was 378 nM and 240 nM for wild type cell [Figure 3-26 C].

Logarithmic regression models for both dose response curves show similar patterns. ANOVA shows no statistically significant difference between the two curves (p=0.065). However, the clinical significance of such a small estimated increase in cell viability in probably debatable.



D

Drug	Ratio EC50 (E79K+ / WT)	95% C.I	ANOVA p value
5-Flurouracil	4.12	1.7 – 8.9	<0.001
SN-38	4.01	2.8 – 5.2	<0.0001
Oxaliplatin	1.51	1.01 – 1.9	0.065

Figure 3-26. Dose response curves for HCT116 isogenic cell viability following treatment cytotoxic drug treatment

HCT116 E79K+ mutant lines (red curve) show increased resistance and viability compared to HCT116 WT when treated with (A) 5-fluorouracil, (B) SN-38 but not (C) oxaliplatin. (D) Summary table of ratio of drug concentration required for 50% cell viability. Cells were treated for 72 hours prior to performance of MTS assay to assess cell viability. Data shown are representative of three independent experiments.

3.3 Discussion

Herein the data demonstrated that RNAi of NRF2 meditated a radiosensitive phenotype in 2 out of 3 colorectal cell lines. This is the first demonstration of the radiosensitizing effect of NRF2 knockdown in colorectal cell line models. Although previous work has established NRF2 upregulation as mediator of radiosensitivity / radioresistance in other cancer models, its effect in CRC has been overlooked. This provides important proof of concept data that NRF2 may have a role to play in mediating therapeutic resistance *in vivo.* The absence of radiosensitivity in

HCT116 was potentially the result of lower basal levels of NRF2 signalling. The induction of radioresistance in HCT116 by an activating mutation in NFE2L2 resulting in increased NRF2 levels and highlights that gain of function in the NRF2 signalling pathway has important consequences for cellular fate. NRF2 mutant cell lines also demonstrated increased chemoresistance to 2 of 3 cytotoxic radiosensitizers in clinical use for rectal cancer.

The degree of radiosensitization seen was consistent at approximately 20%, which was mirrored in the degree of radioresistance induced by NRF2 overexpression in the sensitive cell line model. The symmetry and consistency is reassuring and supports a genuine contribution of NRF2 signalling to radiation responses in CRC models, which was the main aim of my work in this chapter.

However, an appropriate criticism of this work is that the number of colorectal lines, three, is relatively small. A larger panel from across the colorectal spectrum of phenotypes would improve the robustness of this work. The increased breadth could also assess the relevance of NRF2 to the main clinical phenotypes of colorectal cancer – MSI and MSS. The absence of radiosensitization and the apparently low levels of NRF2 signalling in HCT116 may be a feature shared by all MSI cell lines.

Further laboratory work is also needed to understand the phenotypic changes induced by interference of NRF2. Assaying of the knockdown models for generation of reactive oxygen species (ROS) would be useful given the phenotypic difference, for instance to understand if there is a difference in ROS between SW1463, HT55 and HCT116. ROS contributes to DNA damage [18] and excessive production in the absence of NRF2 could have contributed to an increase in radiosensitization in the former but not the latter.

The phenotypic implications of the mutation in the crucial KEAP1 E_{79}TGE binding motif need to be better understood also as this mutation is artificial and unusual in vivo. The model suggests weak or absent binding of NRF2. A co-immunoprecipitation experiment of the KEAP1-NRF2

complex, relative to the parental wild type line, would confirm this. NRF2 protein half-life is also theoretically increased by the mutation suggesting that further phenotypic screening using a pulse chase analysis or cycloheximide blocking to quantify this may be useful. However, the absence of a robust and reproducible antibody to NRF2 may present some challenges.

The attempts to produce CRISPR Cas9 knockout SW1463 isogenic cell lines, whilst ultimately unsuccessful, highlight several important findings. The standard CRISPR-Cas9 genome editing protocols in cancer cell lines with existing CIN, like most non MSI colorectal lines [180], are highly likely to cause unwanted chromosomal rearrangements both at the target and distant loci. The events were detected using relatively straightforward and cost-effective cytogenetic analysis. While not designed to be high throughput, the method of cytogenetic analysis and FISH were quick and relatively simple to carry out on the small numbers of clones which give the expected mutation in Sanger sequence traces. For instance small deletions outside BAC probes will be more easily detected using sequencing, whereas large deletions greater than 30kb, chromosome truncations and translocations will only be observed using FISH. Diploid cancer cells were more likely to maintain their karyotype during the CRISPR-Cas9 process whereas clones from aneuploid parental cell lines showed high levels of CIN (Rayner et al, in press, The CRISPR journal, 2019). In support of this, the commercially obtained HCT116E79K+ line was also diploid and the extensive quality control confirmed the point mutation and the absence of significant cytogenetic abnormalities.

These data have important implications for using CRIPSR Cas9 screens in radiosensitivity screens. Significant structural changes in alterations in karyotype of a cell line, outside the region of interest detected by Sanger sequencing, could lead to changes in radiosensitivity which are erroneously attribute to the indels created.

The radiosensitizing effect was only seen in the TP53 mutant, non hypermutated models: SW1463 and HT55. In HCT116 the presence of intact p53 signalling as well as MSI could both account for the high radiosensitivity and override any sensitizing effect of NRF2 knockdown. A contributing factor could be that the HCT116 model has lower levels of NRF2 pathway expression at baseline. The absence of significant NRF2 activity in HCT116 would therefore mean an RNAi knockdown was unlikely to have a significant functional change.

Situated at the regulatory hub of a suite of pathways, mutation in p53 and associated loss of function has long been recognised as a driver mutation in tumorigenesis. Only recently however has the oligomeric structure of p53 and its consequent degradation become clear. The classical model of MDM2 targeted ubiquitination of p53 by the 26S proteasome has been well validated. But this may be more relevant to the wild type protein. There is increasing evidence that mutant p53 is degraded in a ubiquitin independent manner by the 20S proteasome [188,189]. Hot spot mutations in p53, including those at position 248 as found in SW1463, show increased binding affinity to NQO-1, a direct target of the NRF2 transcription factor [190]. By binding NQO-1 mutant p53 can avoid ubiquitin independent degradation, remain in the cytoplasm and continue to promote cell survival and proliferation.

In SW1463, which is a homozygous $p53^{R248Q}$ mutant, the mRNA levels of NQO1 were significantly lowered following RNAi as measured by qPCR (approximate 40% relative fold change). On the assumption that these cells rely on high levels of mutant p53 signalling for survival, an acute decrease in NQO-1 would mean increased degradation of mutant p53 and consequent decrease in p53 pathway signalling. The decrease in p53 mediated apoptotic cell death would result in ineffective, or at least less effective, repair of cells following irradiation

and provide a basis for increased radiosensitivity. Though the exact hotspot mutation attributed to HT55 at position 213 has not been described to interact with NQO-1 it is certainly possible.

The induction of radioresistance in HCT116 could not be explained by p53 mechanics however. The single point mutation in exon 2 of HCT116^{E79K+} in the isogenic paired lines means that p53 remained intact and functional. Therefore another mechanism must be at play to support cell survival. Given the wide variety a pathways affected by NRF2 there were multiple lines of enquiry to pursue. But it was also logical that the mechanism would also be shared to an extent by the SW1463 NFE2L2 RNAi knockdown model. These two systems were taken forward for further mechanistic exploration by examining the transcriptomic changes induced by changes in NRF2 signalling.

CHAPTER 4

4.1 Introduction

In the previous chapter, changes induced in NRF2 signalling levels, both down regulation and upregulation, mediated changes in therapeutic resistance. NRF2 is known to be a highly pleiotropic transcription factor with the ability to directly and indirectly influence a wide variety of cancer hallmarks [112]. To better understand the effects in the *in-vitro* colorectal models, which in turn could explain the phenotypic changes in radiation responses, whole transcriptomic profiling RNA sequencing was performed. Specifically, 3'-polyadenylated (Poly-A) library preparation was chosen to focus on mature mRNA and ideal for comparative genomics. The analysis was performed on two models: siRNA treated SW1463 cells with demonstrable knock down of NRF2 expression which had previously demonstrated radiosensitivity and the isogenic CRISPR HCT116 model where the activating E79K+ mutation resulted in increased NRF2 expression and had demonstrated chemo-radioresistance. These models were chosen specifically because of the demonstrable phenotypic changes.

The analysis of transcriptomic profiles involved 5 fundamental steps. First, the RNA samples were fragmented and converted into small complementary DNA sequences (cDNA) and then sequenced with a high throughput next generation RNA-sequencing platform. Second, the small generated sequences were mapped to the reference transcriptome. Third, the expression levels for each gene were estimated as read counts. Fourth, the mapped data were normalized and using statistical methods, the differentially expressed (DE) genes were identified. Finally, the relevance of the produced data was evaluated from a biological context. This was the framework used for this chapter of work. Steps one through three were carried out by the Oxford Genomics Centre and the bioinformatic hub at the WTCHG and detailed in Chapter 2.

The increasing use of RNA-Seq technology for transcriptomic profiling means there are many software programs and pipelines used for differential gene expression analysis (step 4). As such there is no gold standard. In a comparison of 9 separate packages for determination of DE expressed genes from RNA seq data, Costa da Silva et al found that no single package produced optimal results in all their predetermined performance measures for the given experimental conditions [191]. They did however indicate that the DESeq2 package [152] produced one of the best individual results with 95% specificity and 84% true positive rate. A second package, edgeR [153], which is known to have high sensitivity was also used. Both packages also use differing methods of estimating and modelling the distribution of differentially expressed genes. The analysis herein was carried out with both packages to ensure robustness and increase confidence in the 'calls' for differential expression. This was followed by gene set enrichment analysis (GSEA) to understand the biological content and context of the data.

4.2 Results

SW1463 cells were transfected with siNFE2L2 as outlined in Chapter 2. To confirm that Nrf2 expression was altered prior to whole transcriptomic sequencing, the three replicates were analysed by qPCR of NFE2L2 and downstream NRF2 targets – HMOX1 and NQO-1 [Figure 4-1]. NFE2L2 was consistently downregulated by approximately 70-80% (Welch's t.test; $n = 3$; $p = 2.7 \times 10^{-9}$), confirmed by western blot [Figure 4-2].

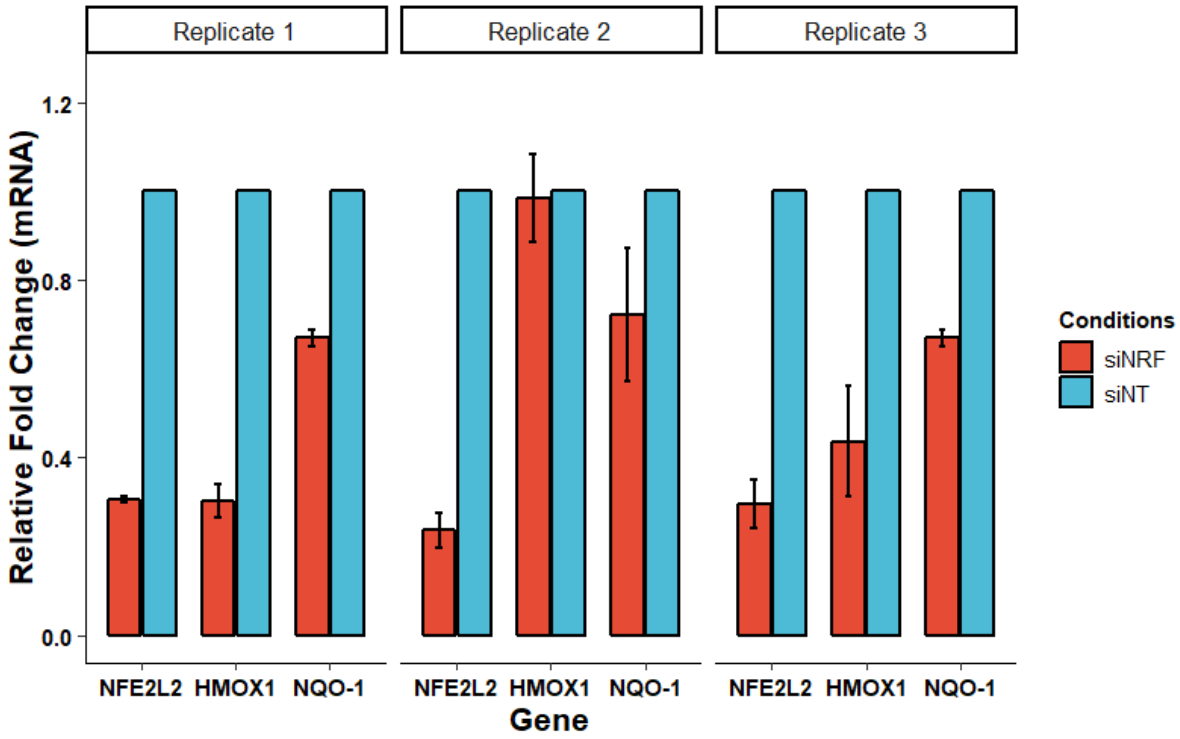


Figure 4-1. Barplot of relative fold change of mRNA levels following transfection with siNFE2L2_1

Significant changes in NFE2L2 levels were seen across all three replicates as well as decreased expression of both NRF2 regulated targets HMOX1 and NQO-1.

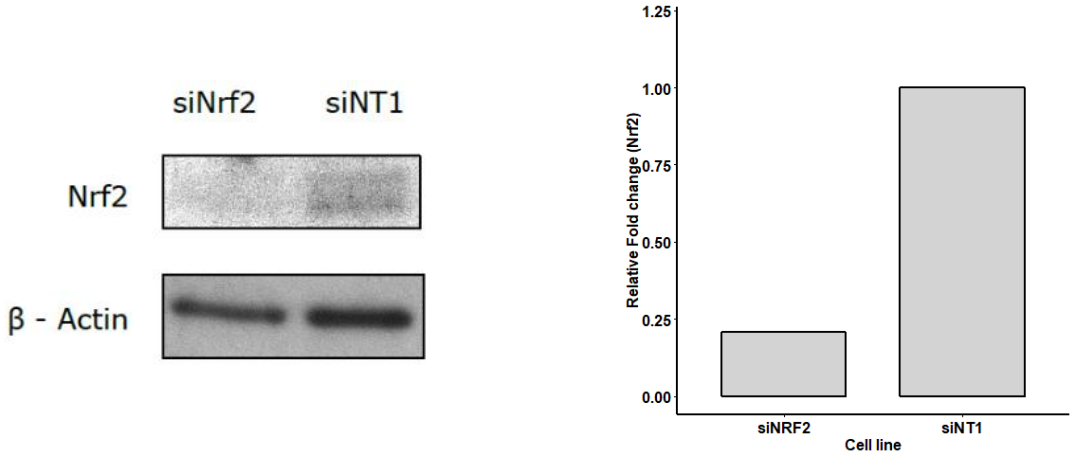


Figure 4-2. Representative immunoblot of western blot for NRF2 following RNAi.

The double band (phosphorylated and un-phosphorylated NRF2) was seen at a molecular weight of ~ 100 kDA. B-actin was used a loading control. Barplot of relative fold change in NRF2 following imageJ analysis.

HCT116^{WT} and HCT116^{E79K+} CRIPSR isogenic lines were isolated in triplicate from low passage number aliquots. These have previously been extensively quality controlled, as seen in Chapter 3, by confirming the point mutation at position c235 and confirming that NRF2 levels were increased in whole cell lysates of HCT116^{E79K+} relative to the parental wild type [**Error! Reference source not found.** Figure 3-19 & Figure 3-24].

Table 4-1. Tapestation derived quality metrics for all replicates used in RNA sequencing analysis

All but two samples had RNA Integrity Number (RIN) of 9 or above indicating extremely high quality RNA.

Well	RIN	28S/18S (Area)	Conc. [ng/µl]	Sample Description
A1	-	-	71.3	Ladder
B1	9.0	1.4	51.6	siNrf2 1.
C1	9.3	2.1	43.3	siNrf2 2.
D1	9.5	1.3	42.8	siNRF2 3.
E1	7.5	0.8	34.7	siNT1 1.
F1	9.5	1.3	39.4	siNT1 2.
G1	9.8	1.1	38.1	siNT1 3.
H1	9.5	2.2	45.1	WT1
A2	9.7	1.8	53.9	WT2.
B2	9.2	1.6	45.0	WT3.
C2	10.0	2.4	41.6	MUT1
D2	10.0	1.5	27.9	MUT2
E2	8.9	1.0	28.1	MUT3

All samples were run on the Agilent Tapestation 4200 to check the RNA quality as evaluated by RNA integrity number (RIN) prior to being sent for RNA sequencing [Table 4-1]. Poly-A enriched mRNA libraries were sequenced as a 12 sample multiplex on one lane of an Illumina HiSeq4000 machine, performing 75bp paired end sequencing. Approximately 20 million reads were obtained per sample. Reads were aligned to the human reference genome (GRCh37) using HISAT2 [147] and duplicate reads removed using the Picard 'MarkDuplicates' tool [148]. Reads mapping uniquely to Ensembl-annotated genes (~15 million per sample) were summarised using featureCounts [149].

4.2.1 Transcriptomic changes of SW1463 siNFE2L2 model

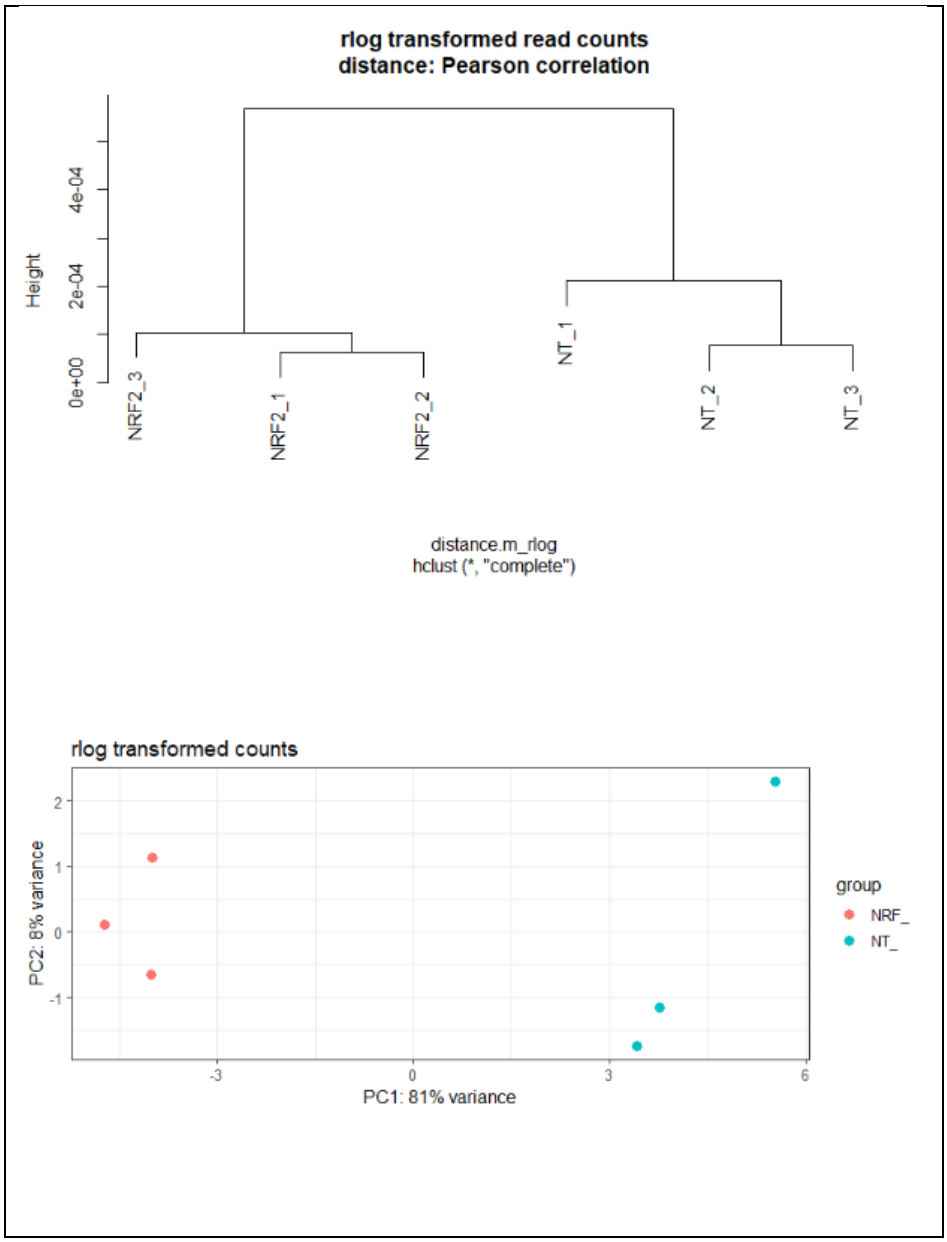


Figure 4-3. Quality control plots for RNA sequencing of SW1463 RNAi experiment

Upper panel: *Pearson correlation dendrograms of the transformed read counts for each of the replicates. The NRF2 knockdown lines and controls all cluster, respectively.* ***Lower panel:*** *Principal component analysis (PCA) indicating the greatest variance in the dataset is due to the presence of the NRF2 knockdown*

4.2.1.1 Quality control of RNA sequencing experimental conditions

Initial quality control measures were taken to assess if the experiment performed as expected using both correlation dendrograms and principal component analysis (PCA). The siNRF2 and

siNT1 replicates clustered together suggesting that those expression profiles are the most similar. PC1 which explains the greatest variation in the data separates all the siNRF2 replicates from the controls [Figure 4-3].

4.2.1.2 Comparison of differential expression R packages

As discussed in the introduction, both the DESeq2 and edgeR packages were used and the derived lists of differentially expressed (DE) genes were compared. There was a perfect correlation between assessments of logFC of genes between both packages [Figure 4-4]. In addition there was significant overlap in the genes considered to be differentially expressed. Given this very high agreement between the two packages, the remainder of the analysis and graphical output was performed with edgeR primarily.

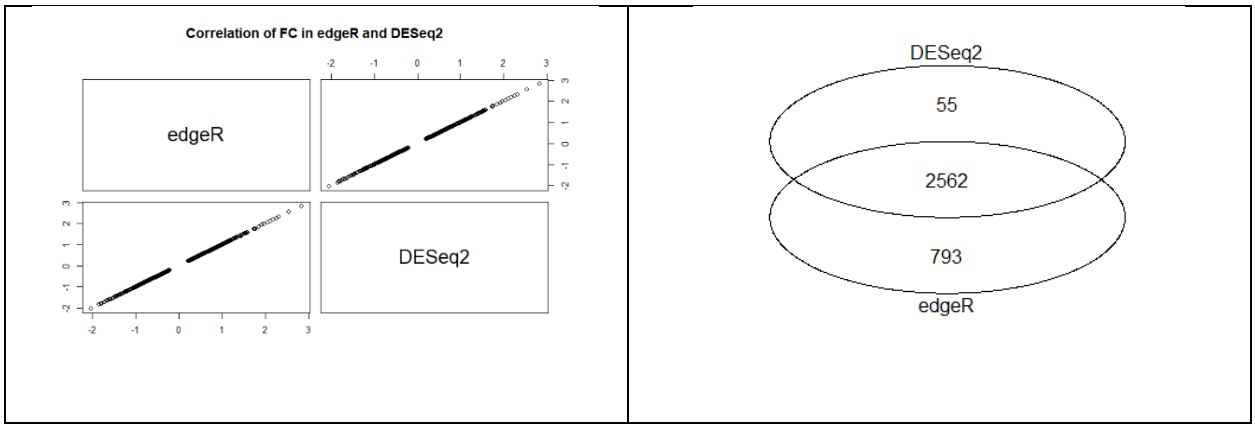


Figure 4-4. Comparison of differential gene (DE) expression results between edgeR and DESeq2 packages

***Left panel**: correlation of logFC of DE genes between both packages, indicating a perfect correlation. **Right panel**: Venn diagram of overlap between DE genes as called by DESeq2 and edgeR. Both packages broadly agree on the majority of genes with DESeq2 demonstrating more stringent results.*

4.2.1.3 Differential expression of genes following RNAi

Visual exploration of the data generated shows a significant change in both expression and fold change of large number of genes [Figure 4-5]. Notably, the top down regulated transcript was

NFE2L2 (FDR, p = 2.41 x 10^{-87}), validating the qRT-PCR results from the previous siRNA experiments. The top 20 DE genes, by false discovery rate, were tabulated [Table 4-2].

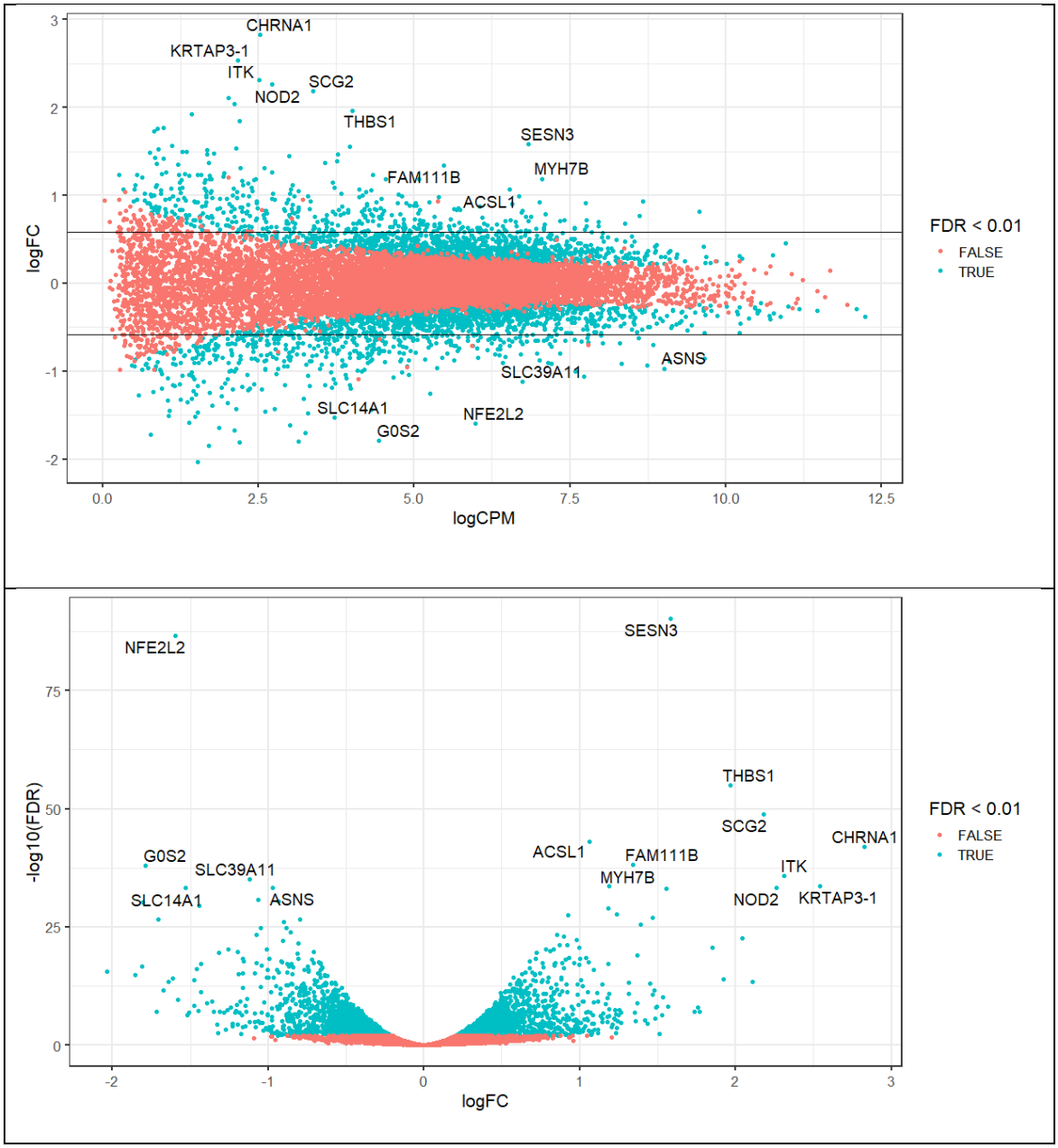


Figure 4-5. Plot smear and volcano plot for NRF2 knockdown RNA seq.

***Upper panel**: MA plot displaying the relationship between the expression (logCPM) and the logFC of DE genes. The horizontal threshold lines indicate absolute fold change of 1.5 (logFC = 0.585). The top 15 DE genes are labelled. **Lower panel**: Volcano plot of logFC of DE gene versus the negative log_{10}FDR, highlighting the most significant findings. The top 15 DE genes are labelled.*

Table 4-2. Summary table of the top 20 DE genes from SW1463 NRF2 knockdown RNA-seq

The genes are ranked according to decreasing FDR. Negative sign indicates downregulation. (logFC = log fold change; logCPM = log counts per million; FDR = False Discovery rate)

GENE	logFC	logCPM	p value	FDR
SESN3	1.582709	6.835509	4.45E-95	5.89E-91
NFE2L2	-1.59724	5.981219	3.64E-91	2.41E-87
THBS1	1.967704	4.005496	2.79E-59	1.23E-55
SCG2	2.182944	3.379758	4.27E-53	1.41E-49
ACSL1	1.064902	6.524933	2.92E-47	7.72E-44
CHRNA1	2.827382	2.527024	4.72E-46	1.04E-42
FAM111B	1.339988	5.470941	3.33E-42	6.30E-39
G0S2	-1.78682	4.437736	7.11E-42	1.18E-38
ITK	2.314698	2.507735	1.26E-39	1.85E-36
SLC39A11	-1.12193	6.743309	5.29E-39	7.01E-36
KRTAP3-1	2.540549	2.167681	1.96E-37	2.35E-34
MYH7B	1.189977	7.056783	2.48E-37	2.73E-34
ASNS	-0.96952	9.01724	5.67E-37	5.66E-34
NOD2	2.260122	2.719572	5.99E-37	5.66E-34
SLC14A1	-1.52814	3.717738	7.15E-37	6.31E-34
GFRA3	1.55377	3.960982	1.07E-36	8.86E-34
PHGDH	-1.06362	7.72055	2.69E-34	2.09E-31
SLC7A5	-0.93037	8.740206	5.32E-34	3.91E-31
IL18	-1.80316	3.148617	1.00E-33	6.98E-31
HGD	-1.44215	4.302388	4.48E-33	2.96E-30

4.2.2 Gene Set Enrichment Analysis (GSEA)

4.2.2.1 NRF2 knockdown showed enrichment for NFE2L2 targets

As a further, orthogonal, verification that NRF2 signalling had been altered by the RNAi, enrichment analysis was performed on the NFE2L2.v2 signature from the molecular signatures database [110]. This list of 481 of NRF2 target genes was generated under conditions of upregulation and downregulation of NRF2 in mouse embryonic fibroblasts and incorporates the previously described 240 target genes discussed in the introduction.

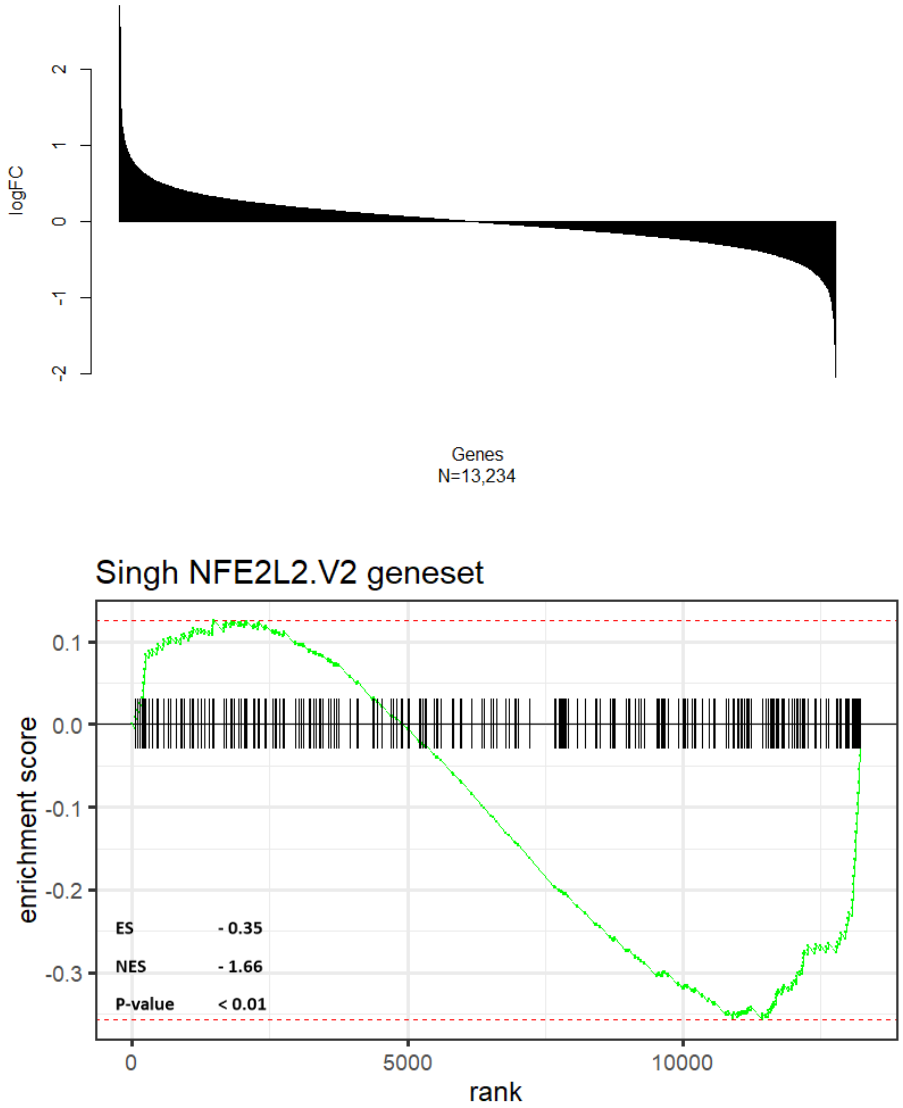


Figure 4-6. GSEA enrichment plot for the NFE2L2.V2 geneset in the ranked siNRF2 data

Upper panel: *Barplot ranked by logFC from highest to lowest of all 13,234 genes identified.*
Lower panel: *The NRF2 targets of the NFE2L2 geneset are enriched in the downregulated portion of the ranked barplot of DE genes, indicating that known NRF2 targets have been decreased in SW1463 as a result of RNAi.*

When the 13,234 genes were ranked by logFC, the downregulated genes showed significant enrichment for the *NFE2L2* targets (ES -0.35; NES-1.66; p <0.01) [Figure 4-6].

4.2.2.2 Hallmarks of Cancer enrichment in NRF2 knockdown

Gene set enrichment is particularly useful when trying to extract biological meaning from differentially expressed genes in a metabolic pathway, like NRF2, where small changes in many

genes can dramatically alter pathway flux [156]. To gain a 'high-level' view of the data and start to understand the differences mediated by downregulation of NRF2 protein, the data was analysed for enrichment using the Hallmark Gene Set Collection. A stringent FDR of 0.1 was used, in contrast to the recommended 0.25, to increase the confidence that genuine enrichments were being called.

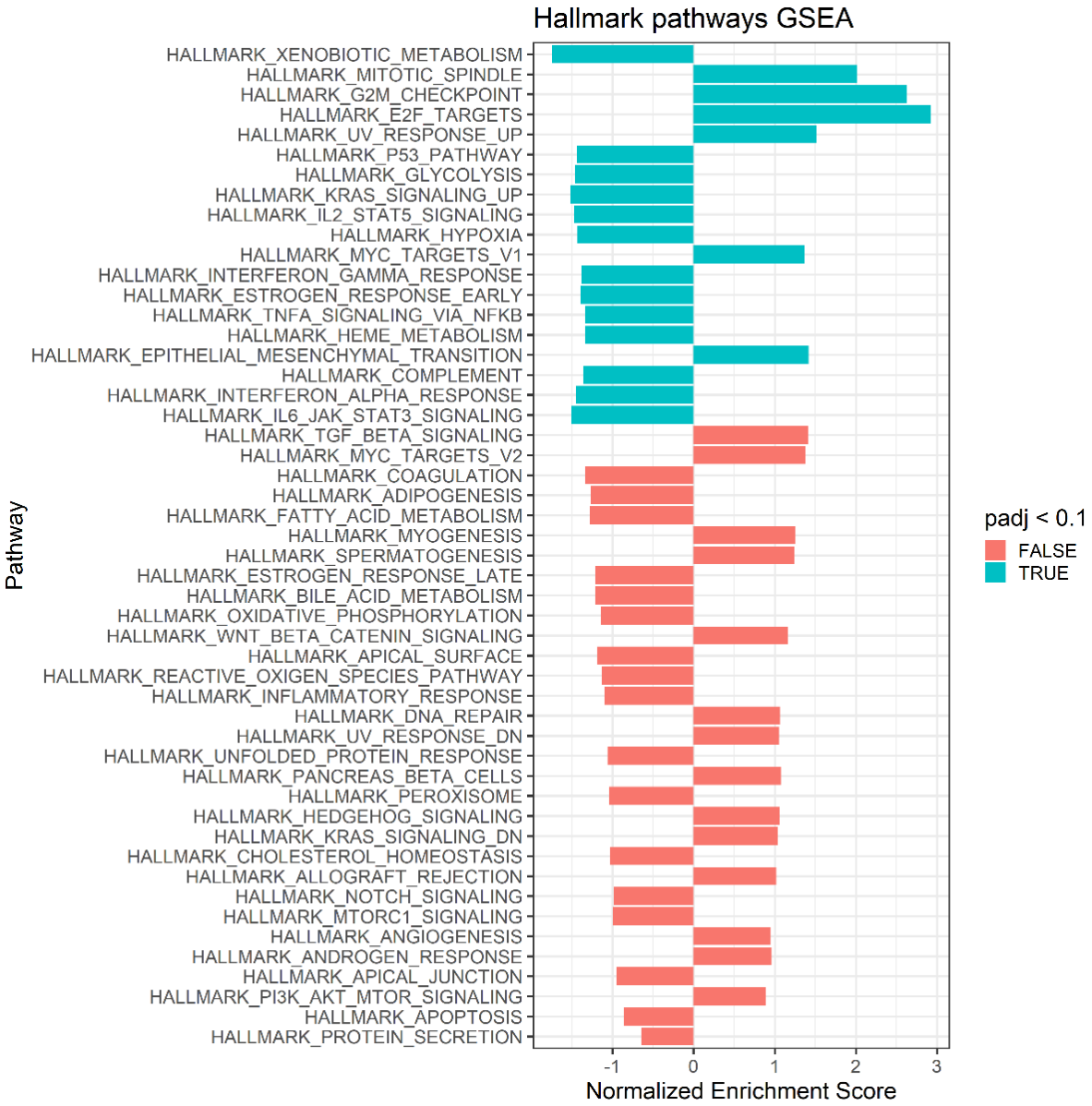


Figure 4-7. Stacked barchart of normalised enrichment scores (NES) for the 50 hallmark genesets following GSEA

A wide variety of cancer pathways are enriched by NRF2 knockdown. No clear mechanism could be elucidated.

The results show a significant number and variety of pathways enriched. Of the 50 curated gene sets, 19 show significant positive or negative enrichment at the 0.1 FDR level [Figure 4-7]. These in turn represent a majority of the classical hallmarks of cancer [192]. This underlines the pleiotropic nature of NRF2 signalling and highlights the number of different pathways affected by its perturbation.

4.2.3 Transcriptomic changes of HCT116 E79K+ model

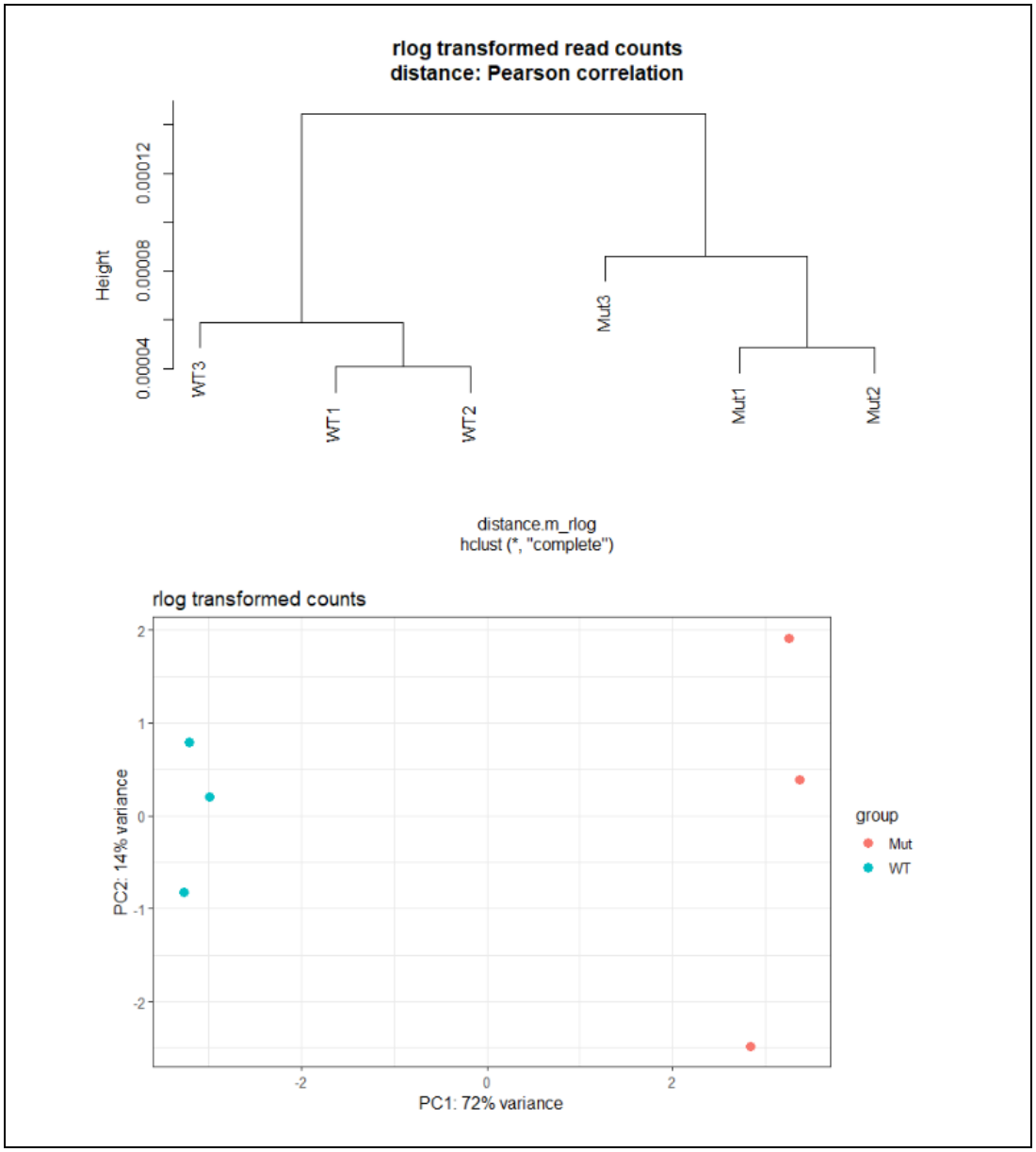


Figure 4-8. Quality control plots for RNA sequencing of HCT116 WT and E79K+ cell lines

Upper panel: *Pearson correlation dendrograms of the transformed read counts for each of the replicates. The NRF2 mutant lines and wild type all cluster together.* ***Lower panel:*** *Principal component analysis (PCA) indicating the greatest variance in the dataset is due to the presence of the NRF2 mutation*

4.2.3.1 Quality control of RNA seq experimental conditions

Initial quality control measures were taken to assess if the experiment performed as expected using both correlation dendrograms and principal component analysis (PCA) [Figure 4-8]. The $HCT116^{WT}$ and $HCT116^{E79K+}$ replicates clustered together suggesting that those expression profiles are the most similar. PC1 which explains the greatest variation in the data separates all the $HCT116^{E79K+}$ replicates from the wild type controls.

4.2.3.2 Comparison of differential expression R packages

Again, there was a perfect correlation between assessments of logFC of genes between both packages [Figure 4-9]. In addition there was broad overlap in the genes considered to be differentially expressed. Given this very high agreement between the two packages, the remainder of the analysis and graphical output was performed with edgeR primarily.

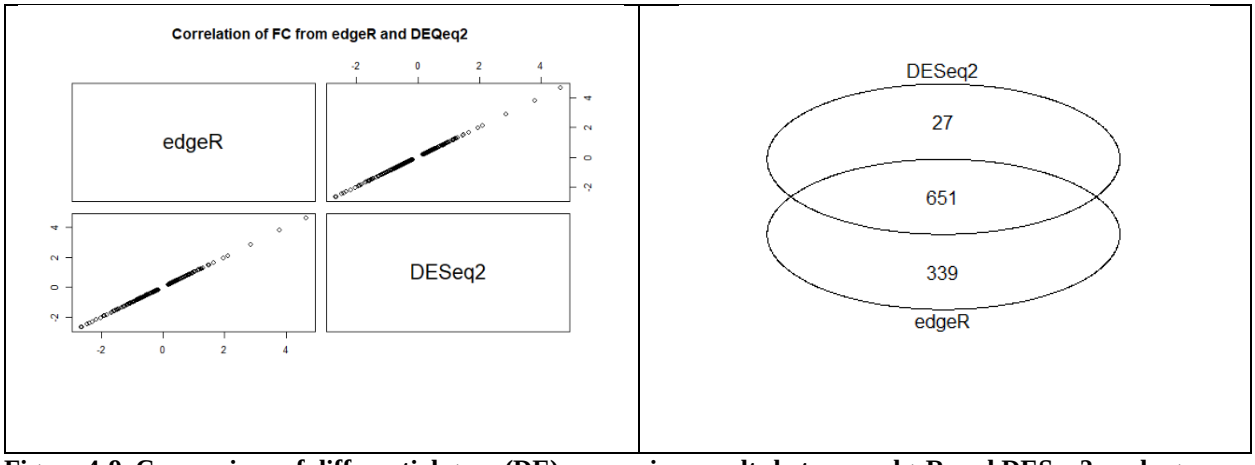


Figure 4-9. Comparison of differential gene (DE) expression results between edgeR and DESeq2 packages

Left panel: *correlation of logFC of DE genes between both packages, indicating a perfect correlation.* ***Right panel:*** *Venn diagram of overlap between DE genes as called by DESeq2 and edgeR. Both packages agree on the majority of genes with DESeq2 demonstrating more stringent results. FDR of 0.01 was used as cut-off.*

4.2.3.3 Differential expression of genes HCT116 E79K+

Visual exploration of the data generated shows a significant change in both expression and fold change of large number of genes [Figure 4-10]. There was no significant change in *NFE2L2,* validating the post transcriptional alteration of the pathway as a result of increased protein stability. The top 20 DE genes, ranked by false discovery rate, were tabulated also [Table 4-3].

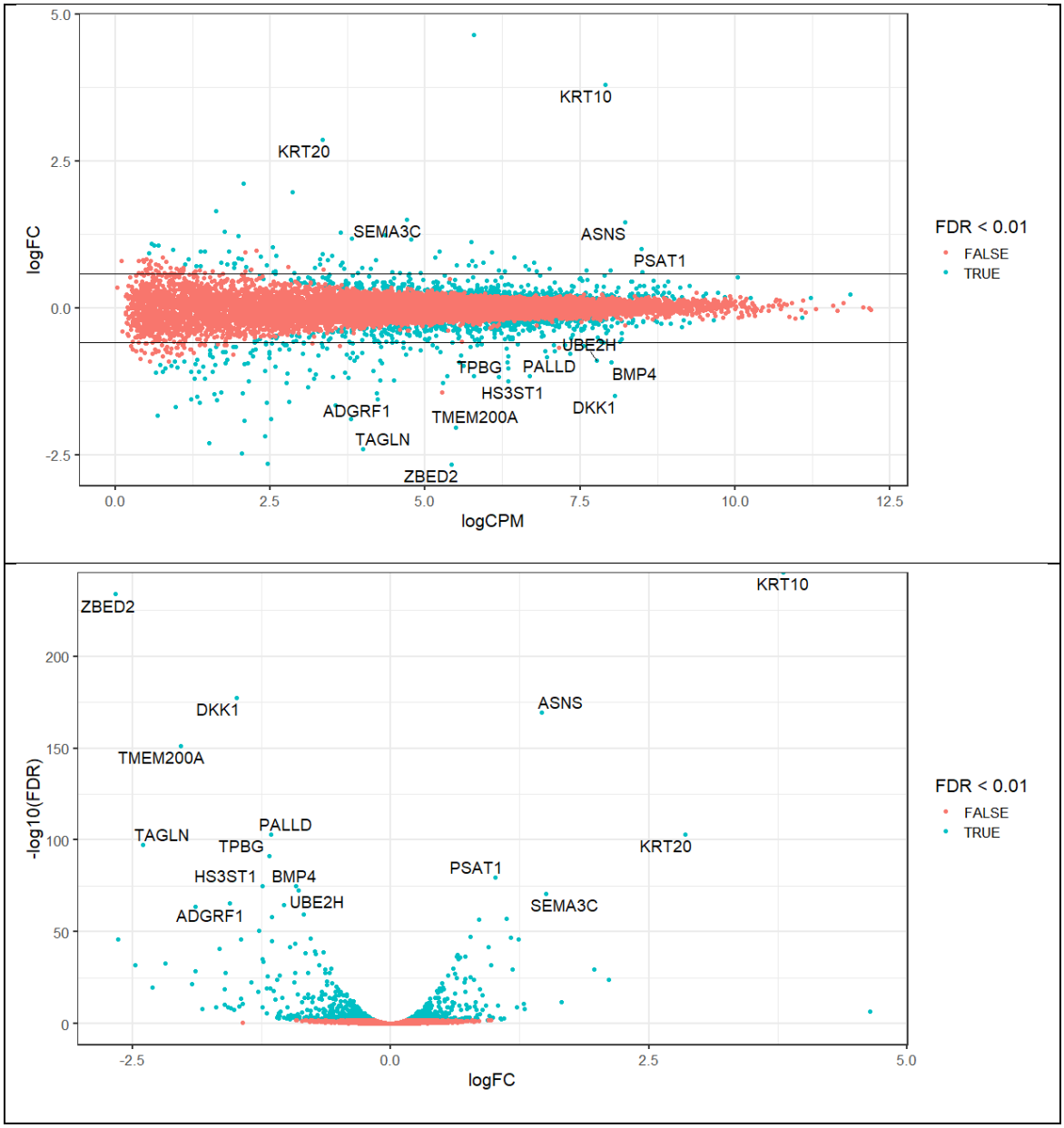


Figure 4-10. Plot smear and volcano plot for HCT116 E79K+ mutant RNA seq

Upper panel: *MA plot displaying the relationship between the expression (logCPM) and the log FC of DE genes. The horizontal threshold lines indicate absolute fold change of 1.5 (logFC = 0.585). The top 15 DE genes are labelled.* ***Lower panel:*** *Volcano plot of logFC of DE gene versus the negative log10FDR, highlighting the most significant findings. The top 15 DE genes are labelled.*

Table 4-3. Summary table of the top 20 DE genes from HCT116 E79K+ mutant RNA-seq

The genes are ranked according to decreasing FDR. Negative sign indicates downregulation. (logFC = log fold change; logCPM = log counts per million; FDR = False Discovery rate)

Gene	logFC	logCPM	PValue	FDR
KRT10	3.795434	7.909317	0	0
ZBED2	-2.66291	5.42174	1.37E-238	8.93E-235
DKK1	-1.48933	8.057236	6.99E-182	3.03E-178
ASNS	1.462906	8.227937	8.38E-174	2.72E-170
TMEM200A	-2.0313	5.500241	3.25E-155	8.45E-152
KRT20	2.854668	3.34405	3.02E-107	6.54E-104
PALLD	-1.15683	6.691634	6.18E-107	1.15E-103
TAGLN	-2.39517	3.996039	2.15E-101	3.50E-98
TPBG	-1.17559	6.193155	2.16E-95	3.12E-92
PSAT1	1.009705	8.493047	2.41E-83	3.13E-80
HS3ST1	-1.24289	6.344073	8.06E-79	9.52E-76
BMP4	-0.91945	8.011888	1.64E-78	1.78E-75
UBE2H	-0.89491	7.777434	3.59E-76	3.58E-73
SEMA3C	1.500954	4.704154	1.60E-74	1.48E-71
ADGRF1	-1.55542	4.226927	4.01E-69	3.47E-66
SUSD2	-1.02908	6.338146	2.87E-68	2.33E-65
MYO15B	-1.89277	3.807007	2.63E-67	2.01E-64
BDNF	-0.83939	6.959922	3.67E-63	2.65E-60
HTR7	-1.15254	5.780619	1.24E-61	8.47E-59
PCK2	1.124121	5.749287	1.88E-60	1.22E-57

4.2.4 Gene Set Enrichment Analysis (GSEA)

4.2.4.1 E79K+ mutant cell line shows enrichment for NFE2L2 targets

As a further, orthogonal, verification that NRF2 signalling had been altered by the introduction of activating mutation, enrichment analysis was performed on the NFE2L2.v2 signature from the molecular signatures database [110] [Figure 4-11].

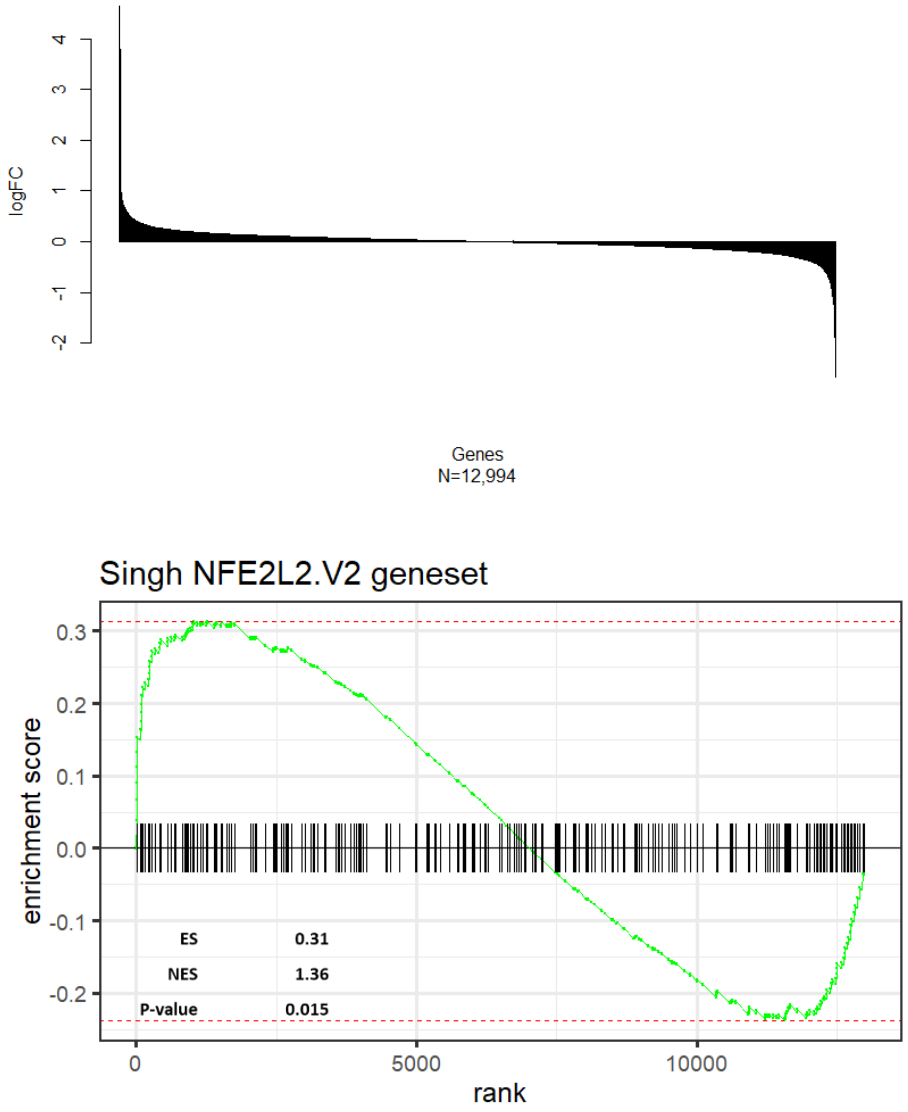


Figure 4-11. GSEA enrichment plot for the NFE2L2.V2 geneset in the ranked E79K mutant data

Upper panel: *Barplot ranked by logFC from highest to lowest of all 12,994 genes identified.* ***Lower panel:*** *The NRF2 targets of the NFE2L2 geneset are enriched in the upregulated portion of the ranked barplot of DE genes, indicating that known NRF2 targets have been decreased in HCT116 E79K mutant.*

When the 12,994 genes were ranked by logFC, the upregulated genes showed significant enrichment for the *NFE2L2* targets (ES 0.31; NES 1.36; p = 0.015).

4.2.4.2 Hallmarks of Cancer enrichment in NRF2 upregulation

Using the cut point of 0.1, 10 were significantly enriched [Figure 4-12]. Although some were of interest such a downregulation of the hypoxia transcriptional program, there were no immediate conclusions as to the mechanism of radioresistance. Furthermore, there was limited overlap with the RNAi experimental model. Only 6 pathways were mutually enriched and of these only 'Epithelial Mesenchymal Transition' had opposing directions of enrichment.

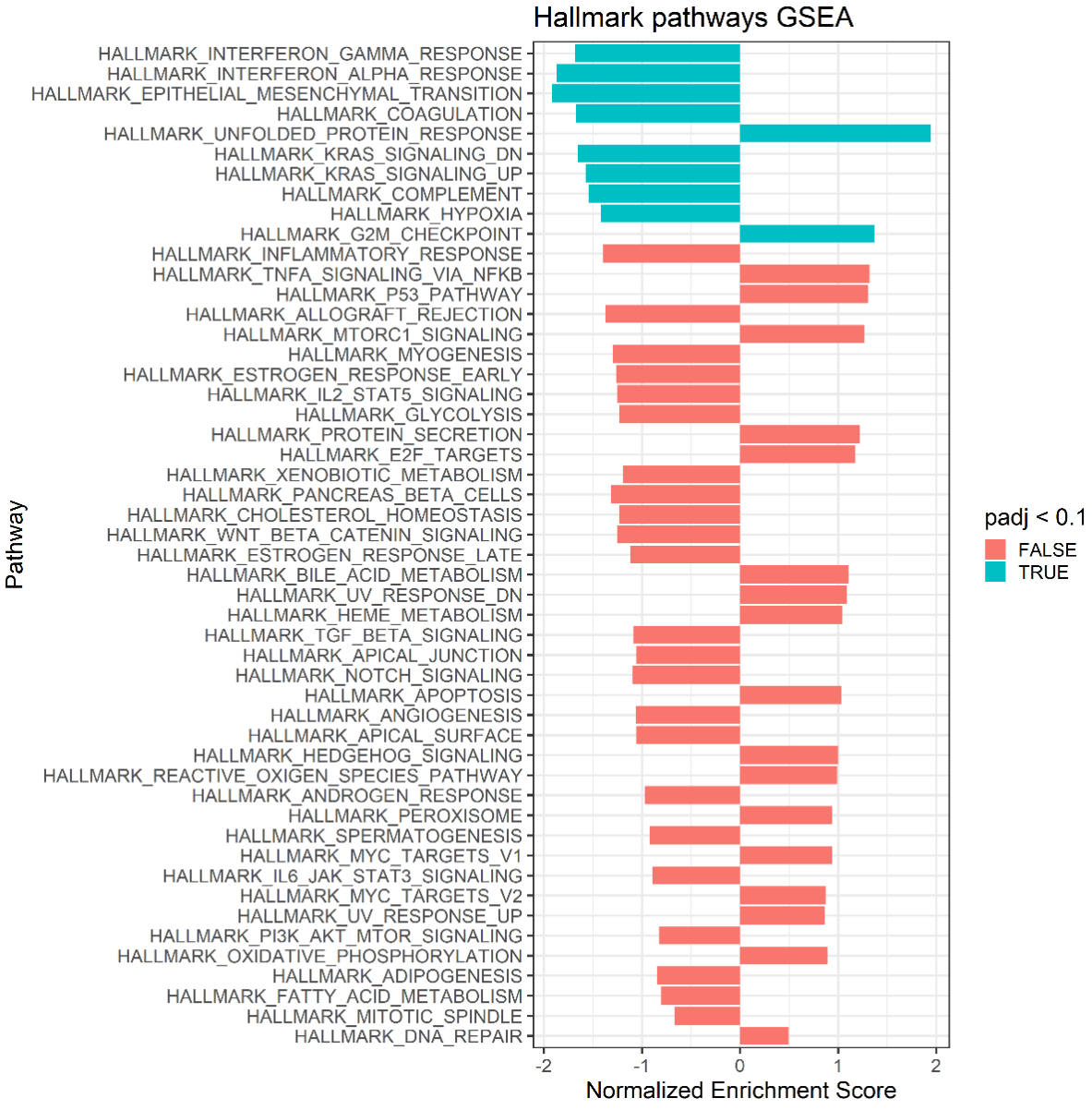


Figure 4-12. Stacked barchart of normalised enrichment scores (NES) for the 50 hallmark genesets following GSEA

A wide variety of cancer pathways are enriched by NRF2 activation. No clear mechanism can be concluded.

4.2.5 KEGG pathways enrichment

4.2.5.1 NRF2 downregulation

Further exploration was undertaken using the Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways. Interestingly all the significant downregulated pathways were involved in metabolic processes, in particular protein synthesis and pentose phosphate pathway. The top hit was 'KEGG_RIBOSOME' [Figure 4-13]. The upregulated pathways were mainly focussed on DNA repair pathways. The top hit was 'KEGG_DNA_REPLICATION'.

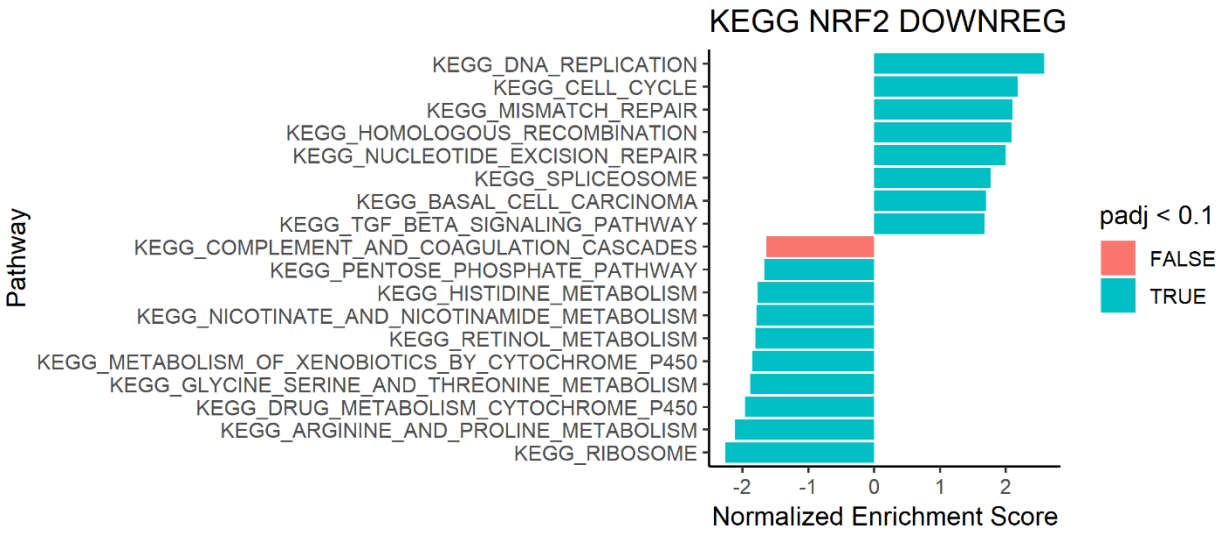


Figure 4-13. Stacked barchart of normalised enrichment scores (NES) for the KEGG pathway GSEA in the siNRF2 model

The positively enriched pathways are chiefly concerned with DNA repair and cell cycle. The negatively enriched pathways are chiefly concerned with amino acid/protein metabolism.

4.2.5.2 NRF2 upregulation

The significantly upregulated pathways were all involved in protein metabolism; KEGG AMINOACYL TRNA BIOSYNTHESIS, KEGG RIBOSOME and KEGG_GLYCINE_SERINE_THREONINE_BIOSYNTHESIS [Figure 4-14]. This was the direct opposite of the effects seen in the siNRF2 treated cells. The downregulated pathways were

all involved in immune processing, which were not significantly enriched in the siNRF2 treated cells.

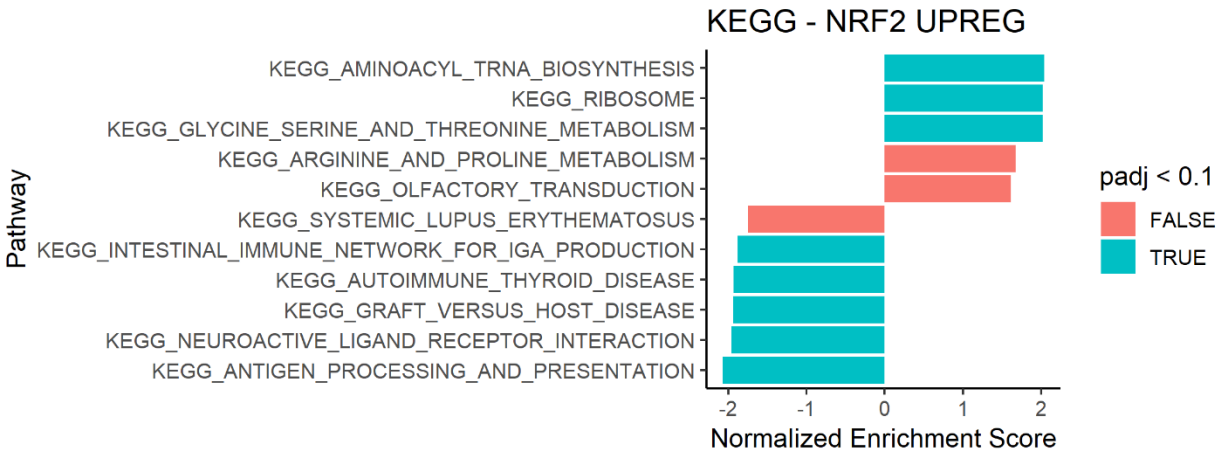


Figure 4-14. Stacked barchart of normalised enrichment scores (NES) for the KEGG pathway GSEA in the E79K+ NRF2 mutant model

The positively enriched pathways are chiefly concerned with amino acid/ protein metabolism. The negatively enriched pathways are chiefly concerned with immunomodulation.

The highly enriched, but opposing effects, on amino acid metabolism was of particular interest and prompted further investigation as a potential mechanism. One striking overlap from the volcano plots and the top 20 DE genes under both conditions was the presence of Asparagine synthetase (*ASNS*). It was the fourth down regulated target in the knockdown model (logFC -0.96, FDR 5.66 x 10^{-34}) and the second upregulated target in the NRF2 mutant model (logFC 1.46, FDR 2.72 x 10^{-170}). This gene encodes an enzyme which catalyses the conversion of aspartate and glutamine, to asparagine and glutamate respectively, in an ATP-dependent reaction. Recent data also suggests that ASNS acts in conjunction with mTOR signalling to promote protein and nucleotide synthesis and cell proliferation [193]. To test this in the current data, GSEA analysis using the MSigDB C6 Oncogenic signatures set was performed [154]. The geneset associated with increased mTOR expression was negatively enriched in the NRF2 knockdown model positively enriched in the NRF2 mutant model [Figure 4-15].

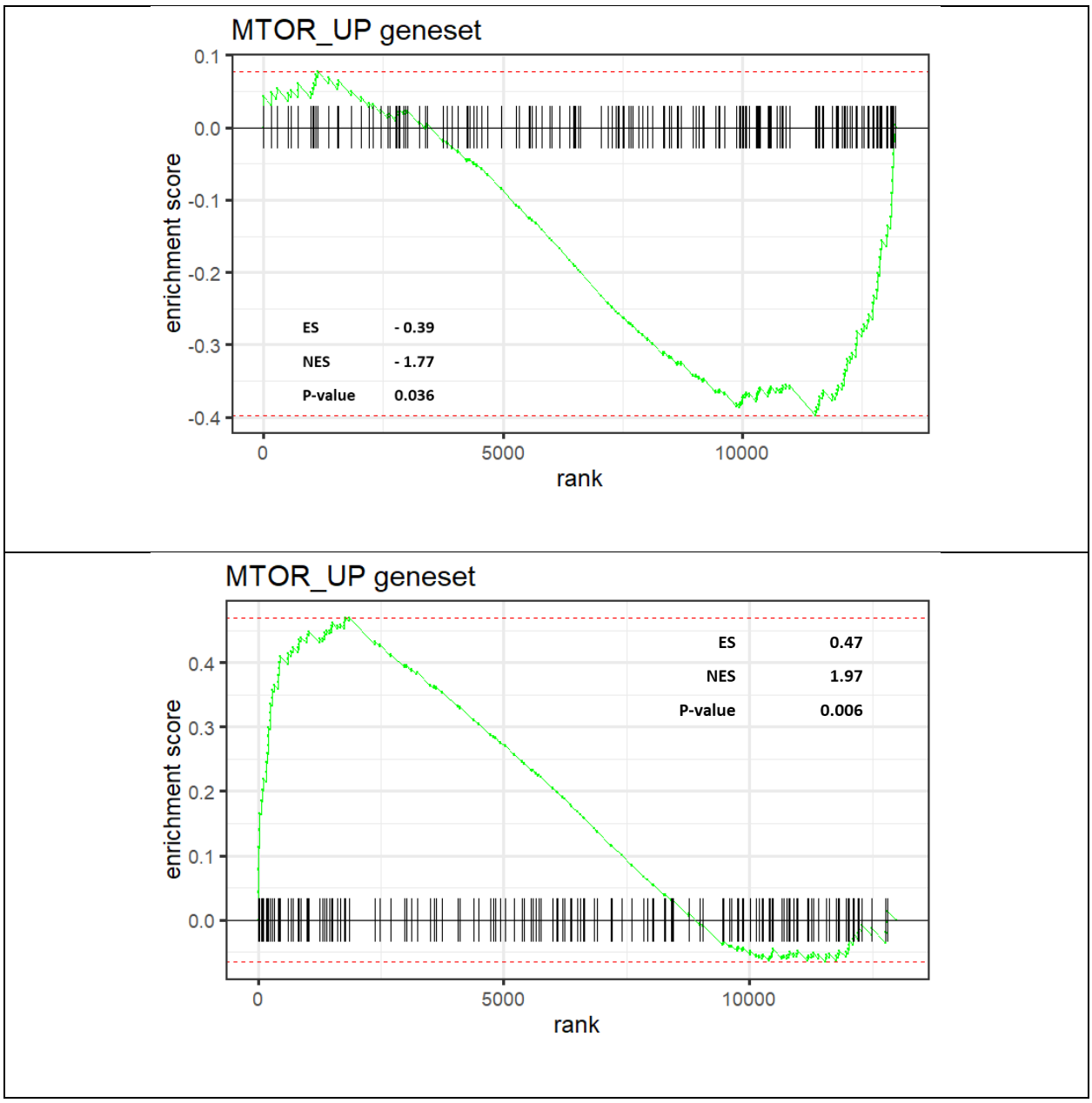


Figure 4-15. GSEA analysis enrichment plots showing significant changes in mTOR signalling

Upper: *The NRF2 knockdown, radiosensitive phenotype showed significant downregulation of mTOR signalling.* ***Lower****: In contrast the NRF2 mutant model indicated significant upregulation of the mTOR pathway.*

To examine this further and understand which genes are mediating these changes, all of the shared differentially expressed genes from the siNRF2 experiment and the E79K+ mutant analysis were tested for overlap. Recall that these genes were significantly differentially expressed DE by two separate packages using an FDR 0.01. The overlap analysis revealed that 216 genes were shared [Figure 4-16] and enriched for metabolic processes seen in the individual analyses [**Error! Reference source not found.**].

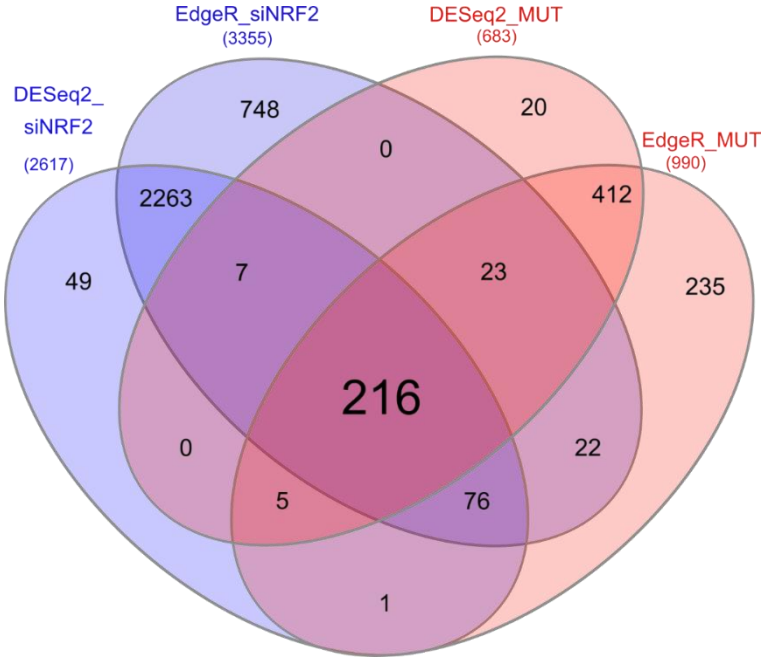


Figure 4-16. Venn diagram of differentially expressed genes from all four RNA seq analyses

Genes were selected using a FDR cut-off of 0.01. DE genes from NRF2 downregulation are highlighted in blue and NRF2 upregulation in red. In parentheses below each package and experimental condition is the number of DE genes at the FDR cut point. 216 genes were shared by all four.

This group was further refined by looking for genes which had opposing logFC changes due to the opposing directions of mTOR and ASNS. The final remaining group was 141 genes (Appendix 1, Supplementary Table 1 and Supplementary Table 2 respectively for lists of 216 and 141 genes with associated logFC and FDR). These genes were imported in the STRING database [160] to better understand their functional relationship.

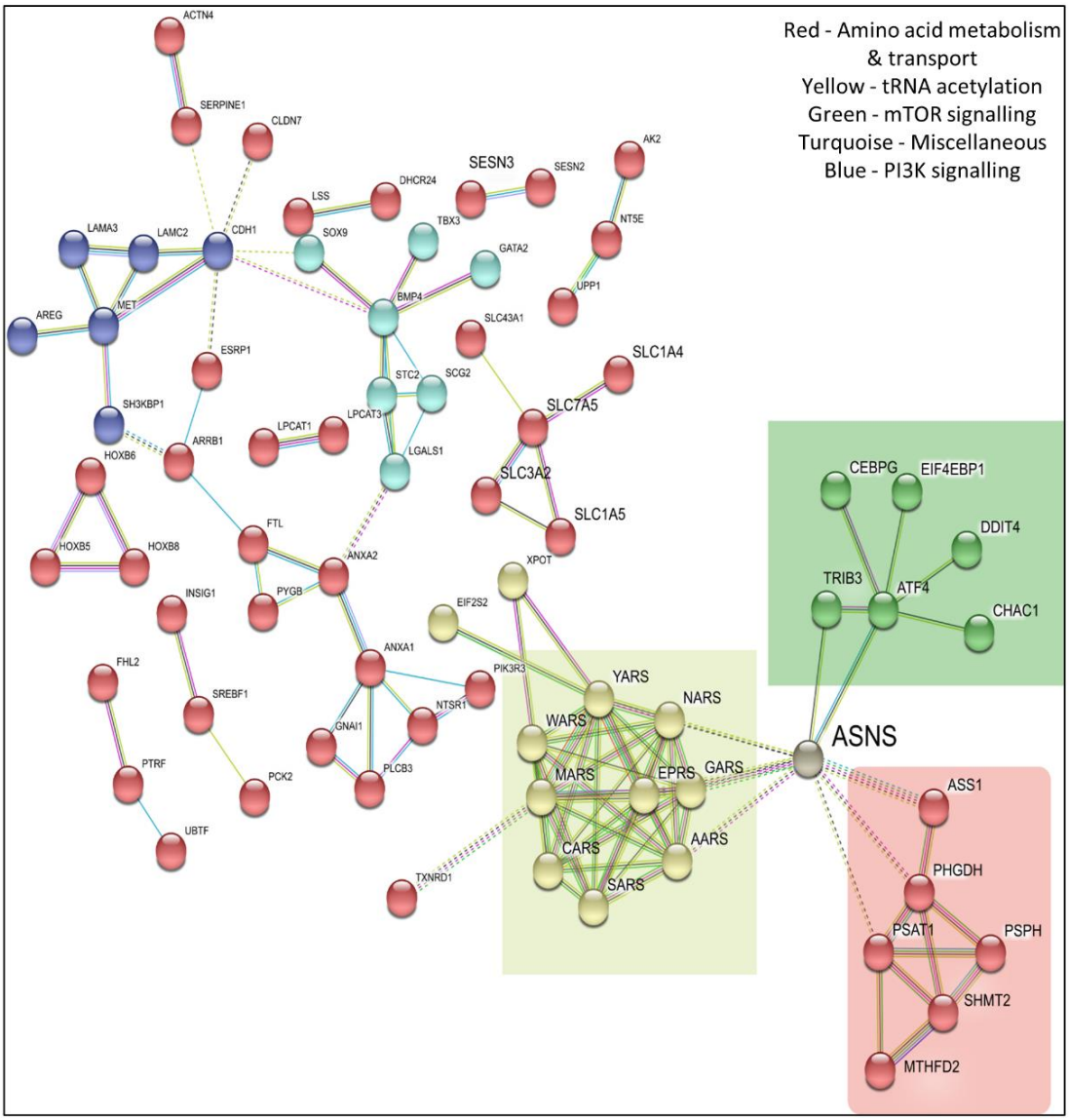


Figure 4-17. STRING interaction network for the 141 shared genes with opposing directionality

The analysis was performed using high confidence settings. Unconnected nodes are not displayed for simplicity. The genes were clustered using k-means (5) to reveal intrinsic co-expression pattern in the network, three of which are highlighted further; yellow = tRNA aminoacylation genes, green = mTOR signalling pathway; Red = serine/threonine and one carbon metabolism.

The analysis revealed a highly significant interaction network, with ASNS at the hub [Figure 4-17]. The number of interacting nodes was 117, higher that the predicted of 35 for a set of 141 genes with a protein-protein interaction enrichment p-value < 1×10^{-16}. The top biological process (GO) enrichments were GO: 0006521 'cellular amino acid metabolic process' (19/308;

FDR p-value 7.83 x 10^{-9}) and GO: 0006418 ‘tRNA aminoacylation of protein translation’ (9/48; FDR p-value 4.4 x 10^{-7}).

The interaction network highlight three key domains which correspond to the interactions seen in the KEGG GSEA analysis. The yellow genes are all primarily involved in tRNA aminoacylation, the red group in serine/threonine amino acid and one carbon metabolism while the green group are implicated in mTOR signalling and protein synthesis. The red grouping are also notable for the genes outside the core group as these are also involved in amino acid metabolism, for example the interconnected solute carrier superfamily (SLC*x*A*x*) which are transmembraneous receptors responsible for amino acid influx and efflux.

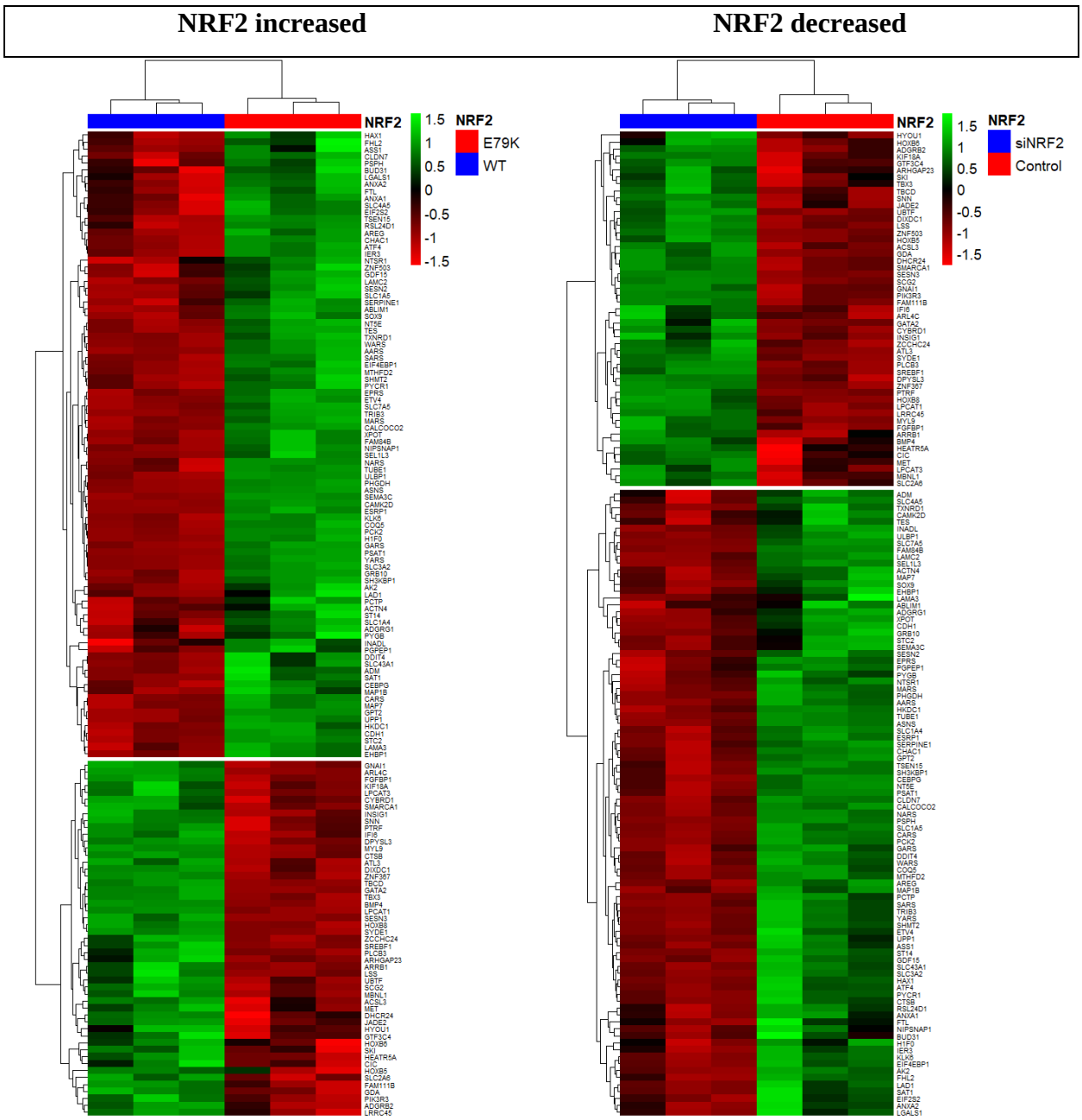


Figure 4-18. Heatmaps of relative change in expression (z-score) of the 141 shared genes

***Left**: NRF2 upregulation increased expression of metabolic genes such as ASNS, PHGDH, ASS1 while downregulating SESN3, a negative regulator of mTOR. **Right:** NRF2 downregulation produces the opposite shift in expression pattern*

Overall these data suggest a complex, interrelated anabolic processes supported by changes in glutamine, asparagine and amino acid metabolism in response to aberrant NRF2 signalling, potentially mediated by ASNS, given its position at the hub of this network. These

interconnected pathways are illustrated below, with some but not all of the 141 potential genes shown [Figure 4-19].

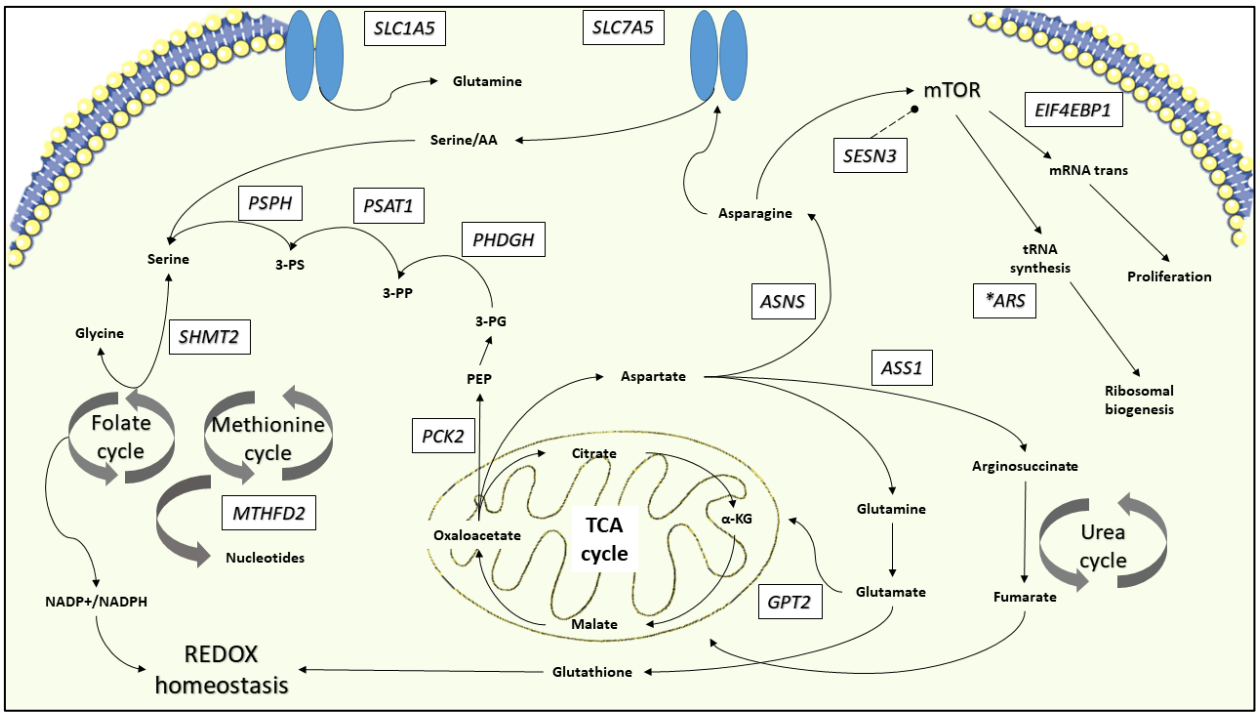


Figure 4-19. Overview of amino acid metabolism, TCA cycle, carbon metabolism in a cell

Glutamine is imported into a cell via SCL1A5 and, converted to glutamate and cycled into TCA metabolism, or converted to glutathione to maintain redox homeostasis. Oxaloacetate can be catalysed to aspartate or PEP. Aspartate is converted by ASNS to asparagine which can activate mTOR signalling or be used for exchange with extracellular amino acids (SLC7A5), as needed, to support metabolic use of serine or other amino acids. In turn, recycled carbon from PEP via PHDGH, PSAT1 and PSPH enters into folate and methionine cycles. The net effect of these interconnected processes is to maintain redox, produce nucleotides, cell proliferation and ribosomal biogenesis.

4.2.5.3 NRF2 regulated metabolic 'switch' model

That so many of these genes are significantly differentially expressed in response to aberrant NRF2 expression suggests that it controls a metabolic 'switch' in colorectal cell line models [Figure 4-20]. ASNS may be the main mechanism by which the metabolic effect of NRF2 is transduced within the cell. The majority of genes change in the same direction as NRF2 itself. *SESN3,* part of the sestrin family of proteins is responsible for negatively regulating mTORC1 [194194] and was upregulated in response to NRF2 knockdown.

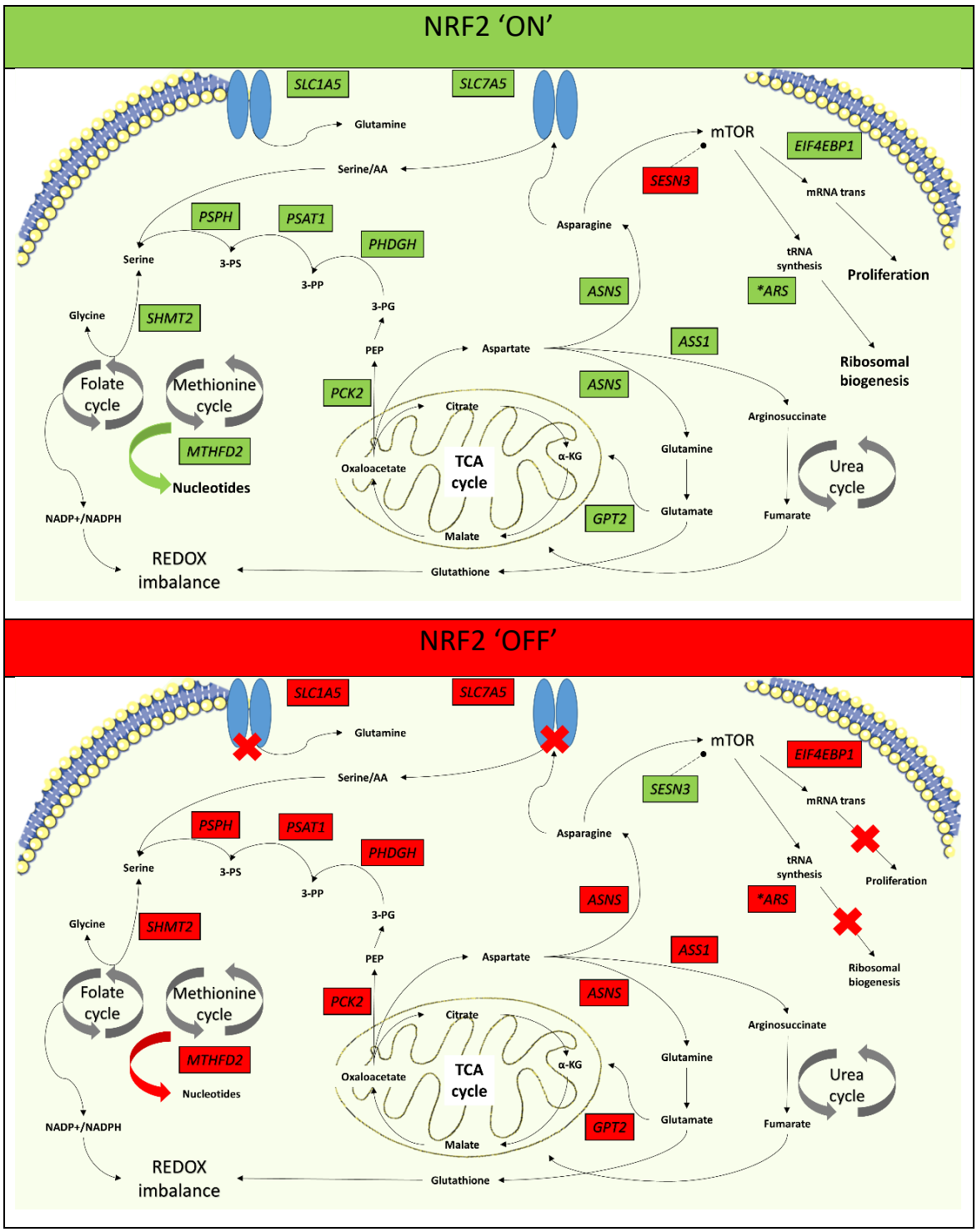


Figure 4-20. Metabolic 'switch' model in response to changes in NRF2

Upregulation of NRF2 facilitates an anabolic process of nucleotide production, mTOR signalling which increases cell proliferation and ribosomal biogenesis. It also impairs the cells ability to maintain redox homeostasis. Downregulation of NRF2 produces the reverse effect, abrogating many essential cell mechanisms required for cell survival. The cell circuit diagram summarises the changes produced by the relative change in gene expression (green = upregulated; red = downregulated)

4.2.6 Asparagine facilitates increased growth in NRF2 mutant line

The suggestion that NRF2 upregulation facilitated an anti-apoptotic phenotype and pro-survival mechanism, either directly or indirectly, by regulating ASNS and asparagine was further evaluated. Previous work suggested that knockdown of ASNS combined with low levels of asparagine in sarcoma models blunted tumour growth [195]. Other work indicated intracellular asparagine can activate mTORC1 to promote amino acid/protein synthesis [193]. On this basis, I hypothesised that the E79K+ cells, possessing increased ASNS, would make use of asparagine more effectively to support cellular growth, than the wild type line.

Cell growth was explored under 3 separate conditions: glutamine replete medium (Q), glutamine deplete medium, with asparagine supplement (N) and media without glutamine, or asparagine, but with other non-essential amino acids (Nil). These were detailed in Chapter 2.

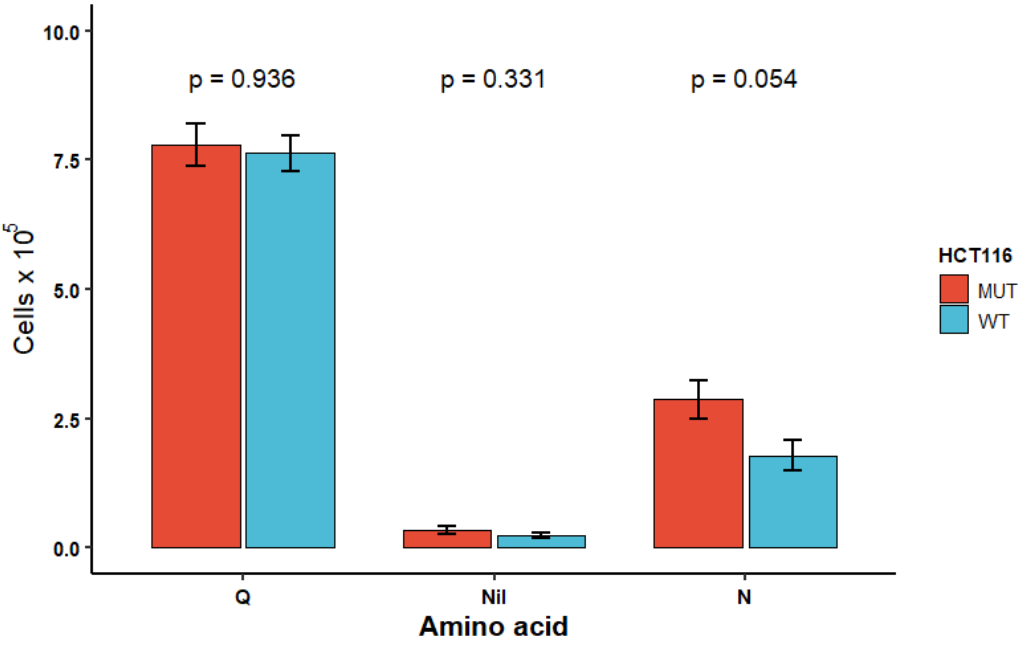


Figure 4-21. Paired barplots of total cell number under different amino acid supplementation

All cell were initially seeded at 1 x 10^5 in 6 well plates in complete medium. After 2 days, media was removed, cells washed in PBS and covered in media with 6 uM glutamine (Q), 4uM asparagine (N) or NEAA only. After 2 further days of growth cells were counted. Data representative of two technical replicates (n = 2).

There was no difference in growth between the cell lines in glutamine replete medium, as expected [Figure 4-21]. In the absence of either glutamine or asparagine, cell growth was severely curtailed. The presence of asparagine allowed for supported improved cell growth, with a trend toward statistical significance (Wilcoxon rank sum, $p = 0.054$). This suggests that NRF2, via upregulation of ASNS, supported efficient metabolism of asparagine to support intracellular, anabolic growth processes.

4.3 Discussion

4.3.1 Glutamine metabolism

Although malignant, proliferating cells are highly dependent on aerobic glycolysis ('Warburg effect'), they also rely on glutamine to support cellular metabolism. High energy requirements and anabolic demand induce dysregulation of metabolism and a switch away from glucose as the main source of energy to meet these demands [196]. The quantity of glutamine consumed is greater than that needed as a biosynthetic precursor and actually serves as a robust source of NADPH production for energy, carbon to replenish the TCA cycle intermediates (anaplerosis) and nitrogen for non-amino acid synthesis [197]. NRF2 has recently been described to play an increasingly important role in regulating this process [198], acting to redirect glucose and glutamine into anabolic pathways via metabolic reprogramming [182], and that NRF2 regulates a serine biosynthesis metabolic program to supply substrates for glutathione and nucleotide production [199]. The synthesis of both purine and pyrimidine nucleotides require the donation of the gamma position amino group from glutamine as it is converted to glutamate [200], exclusively, making it a potential rate limiting step in DNA repair. The regulation, and dysregulation, of amino acid metabolism thus provides both a source of energy as well as macromolecules to facilitate cell survival [201].

The dependence on glutamine metabolism of NRF2 driven tumours has been exploited in the lung cancer setting. Romero et al. demonstrated the metabolic requirement for glutaminolysis in KEAP1-NRF2- mutant lung adenocarcinoma and showed that reduced growth potential following treatment with two glutaminase inhibitors [174]. In the CRC setting, there are currently two phase I/II trials (NCT03263429; NCT02861300) adopting glutaminase inhibition as a therapeutic strategy. However, in the absence of a direct biomarker of NRF2 activation, the trials may prove negative and highlights the need to develop reliable methods of measuring activation in the absence of somatic mutations.

4.3.2 Asparagine metabolism

Asparagine Synthetase (ASNS) is regulated by a *cis*-regulatory consensus sequence in its promoter region (5'- TGATGAAAC- 3') known as CCAAT/enhancer – binding protein (C/EBP) [202]. The transactivating factors that bind this region are members of the basic-leucine zipper (bZIP) superfamily of transcription factors, of which NRF2 is a member. ASNS is typically known to be regulated by ATF4 (Activating Transcription Factor 4) [202], a stress response pathway that regulates the cell's unfolded protein response. But there is increasing interest in the cross talk between transcriptionally activated pathways such as NRF2 and ATF4 as downstream targets can be shared [203]. ATF4 was differentially expressed in response to NRF2 changes in this dataset (NRF2 K.D logFC -0.48, FDR p = 1.04×10^{-7}; NRF2 MUT logFC 0.25, FDR p = 2.34×10^{-6}) but at much lower levels that ASNS, suggesting that NRF2 may play a more direct role in regulating ASNS levels in colorectal models.

Recent evidence challenges the dogma that the main function of glutamine is to maintain the TCA cycle and support macromolecular biosynthesis. When glutamine deprived cells undergo cell cycle arrest and decreased protein synthesis, the process can be rescued by a single amino

acid – asparagine [204]. Krall et al suggest that asparagine, which is produced *de novo* by asparagine synthetase through the unidirectional catalysis of aspartate to asparagine and glutamine, is a key regulatory process in suppressing apoptosis and promoting cellular adaption. ASNS levels are known to correlate with higher grade and worse prognosis in some solid organ malignancies [204]. One clinical example of therapeutic exploitation is the use of L-aparaginase as a chemotherapeutic to deplete asparagine in the treatment of childhood acute lymphoblastic leukaemia (ALL) [205].

Recent *in vitro* and *in vivo* data from lung cancer models has suggested that transcriptional activation of KRAS-NRF2-ATF4 pathways controls downstream ASNS expression [206]. They suggest that NRF2 plays a previously underestimated role in global metabolic regulation, which the current data would suggest is also relevant to CRC models also.

Future work to confirm these alterations and better characterise these transcriptomic changes is warranted, in particular assessing for significant changes in ASNS at a protein level, with knockdown and over expression of NRF2. Subsequent knockdown of ASNS in the HCT116 NRF2 overexpressed model would allow greater understanding of its contribution of the radioresistance phenotype. Experimental confirmation of metabolite changes, in particular glutamine and glutamate, in the contrasting NRF2 models would support the 'switch' model described in the chapter which was solely based on changes in the transcriptome.

4.3.3 Ribosomal biogenesis

Ribosomes are responsible for the translation of the genetic information contained in mRNAs into functional proteins. Ribosome biogenesis involves multiple coordinated steps; a) synthesis and processing of ribosomal RNA (rRNA), b) synthesis of ribosomal proteins and their import into the nucleus, c) the assembly of ribosome subunits, and d) the transport of the mature 40S and 60S subunits into the cytoplasm. For some time the ribosome was viewed as a cellular

automaton, whose sole function was to synthesise cellular proteins. Over the last decade this monofunctional paradigm has been challenged [207] and ribosomal biogenesis, as a process, is increasingly known to be important in tumours even if the mechanistic detail remains more elusive. Activating mutations in cancer may have key roles. For example, virtually every stage in ribosomal biogenesis is influenced by MYC [208].

The chronic proliferation program of cancers creates a high demand for protein synthesis and increased biomass production. High levels of ribosomal biogenesis are therefore required for ongoing proliferation and cell survival. Mammalian cells normally keep ribosome production in check. For example, TP53 inhibits RNA pol I to block rRNA synthesis machinery and maintain homeostasis, a function frequently absent in tumour cells [209].

As a process, ribosomal biogenesis is the most energetic anabolic process and therefore needs close connection to energy production pathways [210]. This is facilitated by the mTOR complex which sits downstream of frequently activated mutational pathways in cancer [210,211]. Specifically in the current data, downregulation of mTOR driven genes was seen with NRF2 knockdown which contributed to the decreased ribosomal biogenesis signal. Activated AKT-mTOR complex 1 drives protein synthesis, a process that frequently becomes hyperactive in cancer cells [210]. As a result of NRF2 knockdown there was a significant downregulation in this process. Whether direct, or mediated through changes in ASNS [193], the significant downregulation of protein metabolism would starve cell cycle repair pathways of the raw materials – nucleotides – needed to continue cell cycle progression and growth. One could hypothesise that this would result in a rebound increase in the transcription of DNA repair and cell cycle genes, as seen in the upregulated pathways of the KEGG analysis.

Taken in aggregate, the data from these experiments and related recent literature suggests that NRF2 exerts regulatory control over the complex amino acid metabolic biogenesis program, with

a specific target of ASNS, possibly through co-ordination with mTOR, which results in increased therapeutic resistance.

CHAPTER 5

5.1 Introduction

As demonstrated in the previous chapters, aberrant NRF2 expression induces therapeutic resistance in colorectal cell lines, while reduced pathway expression was linked to a more sensitive phenotype. NRF2's role as a transcription factor means that it controls a large number of downstream targets who possess an Anti-Oxidant response Element (ARE) in their promoter region [107]. The RNA sequencing results from Chapter 4 from both downregulated and upregulated NRF2 models underline this finding in CRC, with a large number of genes that were differentially expressed.

Although high levels of NRF2 expression within resected tumours have been found to be significantly correlated with poor prognostic features such as p53 expression, Duke's stage [139] , tumour size, TNM stage and metastases [212] there is no data on actual survival outcomes. In the latter study, Hu et al make the association between NRF2 and survival status but their analysis did not take account of the various aspects of time to event data, including data censoring or the concept of time. In addition, it is impossible to directly measure the prognostic effect of NRF2 in presence of potential confounders in the analysis framework.

In vivo, many of the mechanisms are post translationally activated, suggesting that measuring the transcription of *KEAP1* or *NFE2L2* on their own is unlikely to provide either an accurate assessment of pathway expression or be useful as a biomarker in its own right. However, a signature – defined herein as a known set of NRF2 targets with co-ordinated mRNA expression - that accurately reflects NRF2 pathway expression could be used an appropriate surrogate of signalling in the pathway and thus as potential biomarker for prognostic and/or predictive benefit [213].

5.1.1 Gene signature Derivation

There are, generally, three accepted approaches to developing a gene signature in cancer biology [213]; a 'top-down' supervised approach, a 'bottom-up' approach and a candidate gene approach. In brief, the methods are described below highlighting the benefits and potential pitfalls.

5.1.1.1 'Top-down' approach

This is a supervised approach whereby the most differentially expressed genes are explored across a training dataset. It results in large numbers of genes, typically several thousand. Further refinement by testing correlation with the outcome of interest, for example disease relapse or OS, is then performed. The final set of genes, or signature, is compiled by progressively testing increasing numbers of differentially expressed genes that are associated with the outcome of interest until the best predictive model is selected.

A significant drawback of this type of approach is it produces a large number of false positives given the very large numbers of genes involved. Failure to appropriately correct for the high false discovery rate means that many putative signatures will fail at the validation stage as the discovered association was unique to the training dataset only. One major advantage of this type of approach is that it is completely agnostic to the underlying biology of the disease and therefore may uncover some previously unknown genetic processes associated with a particular tumour. Perhaps the best known gene signature produced by this method is the MammaPrint® breast cancer signature [214,215].

5.1.1.2 'Bottom-up' approach

This approach starts with a reasoned biological hypothesis which is known to confer survival advantage on a cancer cell, for example genomic instability or deregulated energetics or other

hallmarks of cancer [192]. Cells or cancer models that display the phenotype of interest are then interrogated for differential expression of genes, to derive the signature of interest, before applying to training set for scaling of the model and establishing the coefficients. Once the model is created it should then be validated on an independent dataset.

Although the approach has the advantage of testing a biological hypothesis, the assumption that the differentially expressed genes are linked to the phenotype of interest is not always valid [216]. The Genome Grade Index which was developed on the hypothesis that cell proliferation has predictive ability is actually based on chromosomal instability biology, rather than cell proliferation [217]. It does not take away from the clinical utility of the signature, only it highlights where possible a set of gene compiling a signature should be linked back to biology to improve understanding.

5.1.1.3 Candidate gene approach

The focus of this approach is to choose a small number of genes which, *a priori*, have a high likelihood or known association with a disease or phenotype of interest. The chosen candidates should ideally have some previous association with the outcome of interest, for example HER2 being associated with poorer prognosis in breast cancer. Once the data has been applied to a training set, or multiple sets, and the model created it can then be rapidly validated given its compact size. Having a low number of genes means that information on expression can be gained quickly from standard techniques like qPCR, validated as necessary with orthogonal platforms and analysed with less bioinformatic and statistical support than high dimensional data.

This approach has also been the most commercially successful; it was key to the development of the OncotypeDx breast cancer prognostic kit [50]. The signature compromises of only 16 genes; five to assess proliferation, two to assess HER2 expression, four to assess ER expression, two for

invasion and three individual genes. Five further genes are added as reference, normalisation genes. A similar approach in diffuse large B-cell lymphoma, using six genes only, has been validated [218]. The ingenuity behind the Oncotype test that it does not try to binarize patients into artificial groups of high and low, as intuitively one can see that those around the median cannot be arbitrarily pushed into high and low risk. To answer that very question the recently reported TAYLORx trial, which randomised over 10,000 patients with an intermediate score, to either chemotherapy or endocrine therapy showed no difference in outcomes. This provides the highest level of support of the use of the gene signature test but also further refines its' prognostic ability to optimise clinical utility.

A variation of the 'top down' model is a co-expression network model, particularly in microarray experiments. Here each probeset, in an array, or gene, is used a node and edges are mapped between nodes when their expression is significantly correlated. Gradually a network of expression is mapped out. The difficulty again is the propensity to perform multiple testing, the resulting high false positive rate and the inability to set an appropriate cut-off or stringency for the false positivity, which in turn derives from its agnostic assumptions. An overly high threshold means genuine, albeit novel or weak, associations will be overlooked but an overly lenient threshold will result in spurious findings.

Some groups have chosen to merge both approaches in an effort to minimise the downsides inherent in each model [219]. A 'neighbourhood of co-expression' around a collection of starter genes or probesets (referred to as seeds by the authors) is created and then significant overlap between the neighbourhoods is assessed. This is more robust against the inclusion of non-significant edges as seen in the network based approach. The methodology has been primarily used to assess the prognostic significance of hypoxia in large heterogeneous groups of head and neck squamous cell carcinoma patients and breast cancer patients [219,220].

5.1.1.4 Consensus Molecular Subtypes of CRC

The current transcriptomic taxonomy of CRC has been heavily influenced by the Consensus Molecular Subtypes [84], as discussed in Chapter 1. Although there are known deficiencies to the CMS classification, and while there are other classifiers available [221], it is the most widely used expression based classifier and thus provided a useful backdrop in which to frame any findings and understand NRF2's role in current transcriptomic landscape of CRC biology.

5.1.2 Current prognostic NRF2 signatures

The majority of preclinical work into the importance of the KEAP1-NRF2 axis as a biomarker has been done in lung cancer. Qian et al [222] published a 50 gene signature which they termed NAMS (NFE2L2 Associated Molecular Signature) and they validated for both progression free survival and OS in three independent, publicly available lung cancer datasets. To derive the signature they first examined the differential expression of genes from KEAP1 mutant A549 lung cancer cells and NFE2L2 knockdown A549 cells. The resulting gene list was found to be significantly enriched for cancer related terms by KEGG pathway analysis. The gene list was further refined from 240 genes to 50 genes by training cox regression models on two independent lung cancer training sets. The resulting genes and their scaled cox regression model coefficients constitute the final NAMS score.

The other validated signature was produced by Namani et al [223] who adopted a different approach to identify NRF2 regulated genes. Using microarray data from both KEAP1 over expressing A549 cell lines, versus controls, and NFE2L2 knockdown A549 cells, versus controls, they examined the commonly downregulated genes. As both KEAP1 overexpression should have similar effect to knockdown of NFE2L2 they concluded that the combinatorial analysis of the data would identify the definite NRF2 regulated genes. In total 214 were identified. They then performed function annotation of these genes using the DAVID tool

(Database for Annotation, Visualisation and Integrated Discovery) to assess the biological relevance of these genes. One of the most significant clusters of genes was a 34 gene group which aligns with 'small molecule metabolic processes'. These 34 genes were further refined by examining *in silico* for the presence of ARE in the promoter regions, which excluded 7 genes. Finally they examined the TCGA LUAD [224] dataset for differential expression of the remaining 27 genes between KEAP1 wild type tumours and KEAP1 mutant tumours. They found that 12 of the 27 genes were over-expressed in the mutant tumours. They termed the resulting 12 gene signature NRMGS and went on to validate it in eight publically available clinical datasets.

5.1.3 Hypothesis

Known NRF2 regulated genes which had previously proved of prognostic importance in lung cancer would retain prognostic significance in CRC and could be used to aggregate NRF2 pathway expression for use as a biomarker.

5.2 Results

5.2.1 Candidate gene refinement for NRF2 CRC signature

62 candidate genes were obtained from the two published lung cancer signatures. In total 9 independent datasets were available for analysis [59,225–230]. Some were subsetted into different anatomical sites for the purposes of analysis, resulting in 24 discrete sets of data and are annotated in Table 5-1. Annotation for the 9 datasets and the anatomical sub-sites used in the Oncomine differential gene analysis. Of the 62 candidate prognostic genes input into the Oncomine database, 40 were found to be differentially expressed by the stringent cut off criteria applied, 21 of which were significantly over-expressed and 20 which were significantly under-expressed, in at least one or more of the datasets. One gene, COL3A1, was shared as it appeared

as over expressed in some datasets and under expressed in others. The summary heatmaps for differential expression are shown in Figure 5-1.

Table 5-1. Annotation for the 9 datasets and the anatomical sub-sites used in the Oncomine differential gene analysis.

1. Rectal Adenocarcinoma vs. Normal	Gaedcke Colorectal, Genes Chromosomes Cancer, 2010
2. Colorectal Carcinoma vs. Normal	Graudens Colon, Genome Biol, 2006
3. Colorectal Carcinoma vs. Normal	Hong Colorectal, Clin Exp Metastasis, 2010
4. Cecum Adenocarcinoma vs. Normal	Kaiser Colon, Genome Biol, 2007
5. Colon Adenocarcinoma vs. Normal	Kaiser Colon, Genome Biol, 2007
6. Colon Mucinous Adenocarcinoma vs. Normal	Kaiser Colon, Genome Biol, 2007
7. Rectal Adenocarcinoma vs. Normal	Kaiser Colon, Genome Biol, 2007
8. Rectal Mucinous Adenocarcinoma vs. Normal	Kaiser Colon, Genome Biol, 2007
9. Rectosigmoid Adenocarcinoma vs. Normal	Kaiser Colon, Genome Biol, 2007
10. Colon Adenocarcinoma vs. Normal	Ki Colon, Int J Cancer, 2007
11. Colon Adenoma vs. Normal	Sabates-Bellver Colon, Mol Cancer Res, 2007
12. Rectal Adenoma vs. Normal	Sabates-Bellver Colon, Mol Cancer Res, 2007
13. Colorectal Adenocarcinoma vs. Normal	Skrzypczak Colorectal, PLoS One, 2010
14. Colorectal Carcinoma vs. Normal	Skrzypczak Colorectal, PLoS One, 2010
15. Colon Adenoma Epithelia vs. Normal	Skrzypczak Colorectal 2, PLoS One, 2010
16. Colon Adenoma vs. Normal	Skrzypczak Colorectal 2, PLoS One, 2010
17. Colon Carcinoma Epithelia vs. Normal	Skrzypczak Colorectal 2, PLoS One, 2010
18. Colon Carcinoma vs. Normal	Skrzypczak Colorectal 2, PLoS One, 2010
19. Cecum Adenocarcinoma vs. Normal	TCGA Colorectal, Nature, 2012
20. Colon Adenocarcinoma vs. Normal	TCGA Colorectal, Nature, 2012
21. Colon Mucinous Adenocarcinoma vs. Normal	TCGA Colorectal, Nature, 2012
22. Rectal Adenocarcinoma vs. Normal	TCGA Colorectal, Nature, 2012
23. Rectal Mucinous Adenocarcinoma vs. Normal	TCGA Colorectal, Nature, 2012
24. Rectosigmoid Adenocarcinoma vs. Normal	TCGA Colorectal, Nature, 2012

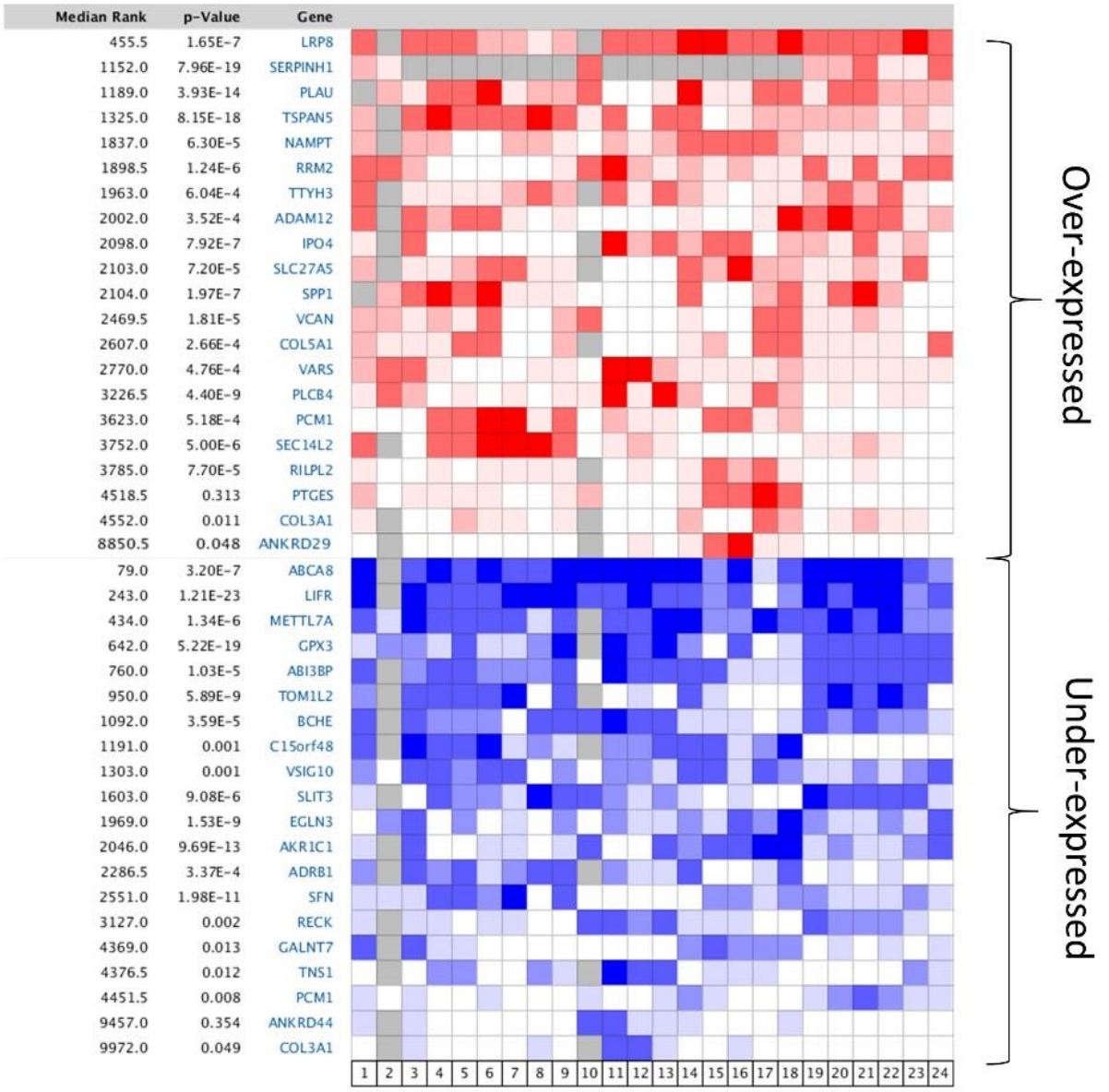


Figure 5-1. Aggregate heatmaps for Oncomine database over all over expressed and under expressed genes in 9 different colorectal datasets.

The median rank for the meta-analysis and the corrected p-value are shown on the left hand side. 40 of the 62 candidate genes are significantly overexpressed and/or underexpressed in colorectal cancer relative to normal tissue at a FDR <0.05. 36 of these were represented in the GSE17536, after filtering by probe-matching, so were taken forward for further exploration.

The 40 genes were examined in the STRING database for known and predicted protein-protein interactions (PPI) [Figure 5-2]. The analysis indicated a higher than expected number of interactions between these genes. For the given number of nodes, 40, the predicted number of edges (connections) was 51 if the proteins were chosen at random. The number of edges was 92 with an average clustering coefficient of 0.509; PPI enrichment p-value 2.17 x 10^{-7}. This

supported the hypothesis that these genes act in a highly co-ordinated fashion as would be expected for those regulated by a common factor such as NRF2. Clustering of the genes into two groups based on k-means indicated that there were a subgroup of core of genes within this network which is formed by 12 genes – *ABI3BP*, *ANGPT1*, *RRM2*, *VCAN*, *ADAM12*, *COL3A1*, *COL5A1*, *SERPINH1*, *RECK*, *PLAU*, *SPPI*, *TNS1* and *SLIT3* – which have a high degree of interaction.

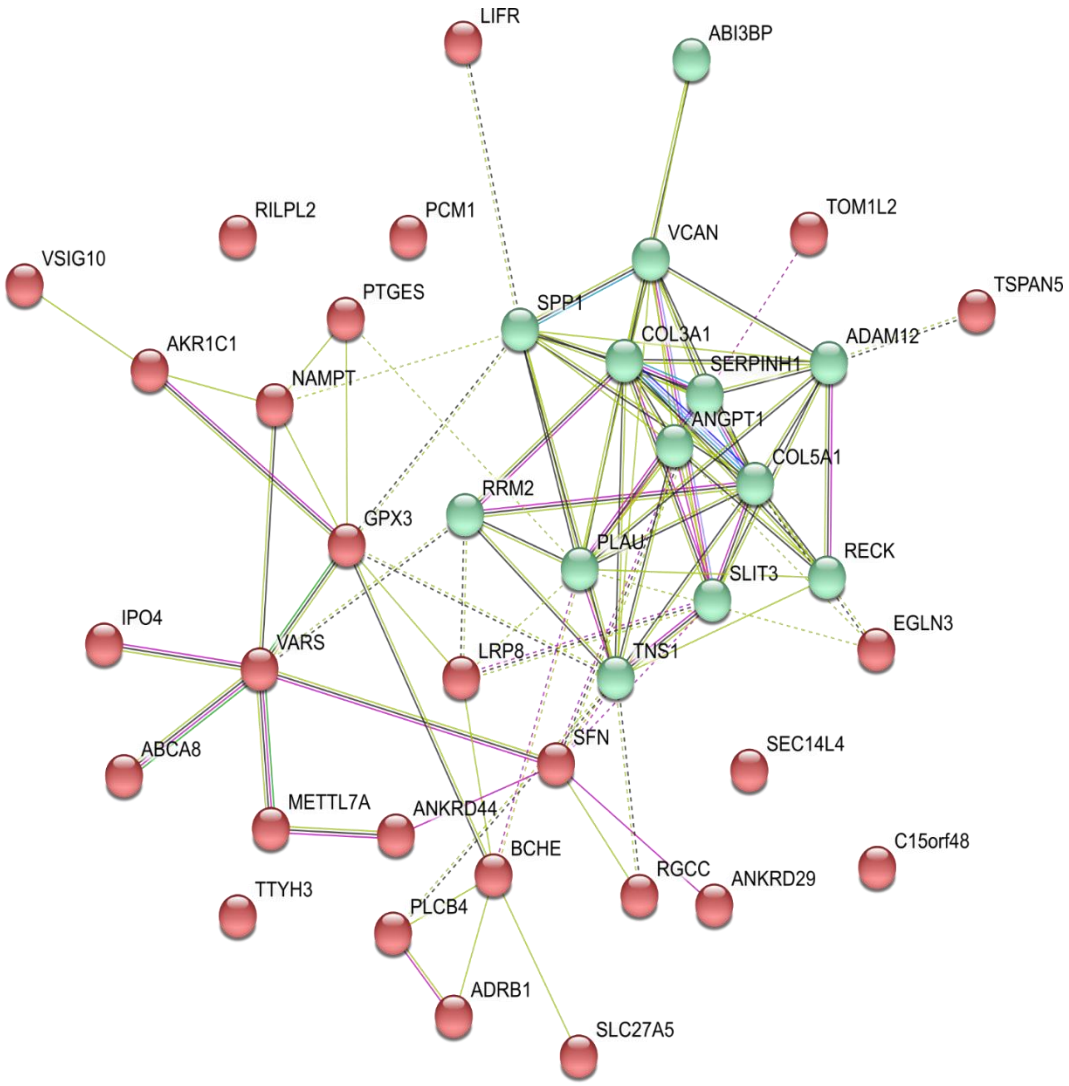


Figure 5-2. Known and predicted protein – protein interactions from the STRING database for the 40 genes from the refined differential expression analysis.

The network of genes was clustered in two groups by k-means to indicate intrinsic interactions (Green and Red). The green group at the core displayed a high degree of connectivity and were highly co-expressed in the training set. The line colours between nodes indicate the type of interaction and the thickness indicates the strength of interaction (thicker = stronger). [Cyan = known interaction from curated databases; purple = experimentally determined; green = gene neighbourhood; red = gene fusions; dark blue = gene co-occurrences; light green = textmining; black = co-expression; light blue = protein homology]

Of the 40 differentially expressed genes, four could not be matched between the training set expression platform and the validation platforms so were omitted from downstream analysis. The final group of 36 genes is: *ABCA8, ABI3BP, ADAM12, ADRB1, ANGPT1, ANKRD29, ANKRD44, BCHE, C15orf48, COL3A1, COL5A1, EGLN3, LIFR, METTL7A, PCM1, PLAU, PLCB4, RECK, RGCC, RRM2, SEC14L4, SERPINH1, SFN, SLIT3, SPP1, TNS1, TOM1L2, TSPAN5, TTYH3, VSIG10, VCAN, AKR1C1, LRP8, NAMPT, PTGES, SLC27A5*. This group of genes was designated NRF2 signature going forward.

5.2.2 Construction of the NRF2 signature in GSE17536

5.2.2.1 Principal Component Analysis (PCA)

In total the 36 genes were represented by 92 probes in Affymetrix U133 v2 array. The 92 probes in GSE17536 were used as the training dataset. 65 of the 92 probes had an absolute correlation (in expression) >0.5 with another probe. This indicated potentially strong dependencies among the probes. Thus the expression levels measured by the 92 probes could be represented by a smaller number of variables using PCA. The 11 major PCs explained up to 80% of the variation in the 92 probes. The strong dependency among the probes was confirmed by the fact that the 81 minor PCs only explained <20% of the remaining variation. The same 36 gene were represented by 140 probes in Affymetrix Xcel array and 46 probes in Agilent Human 4 x 44K v2 array.

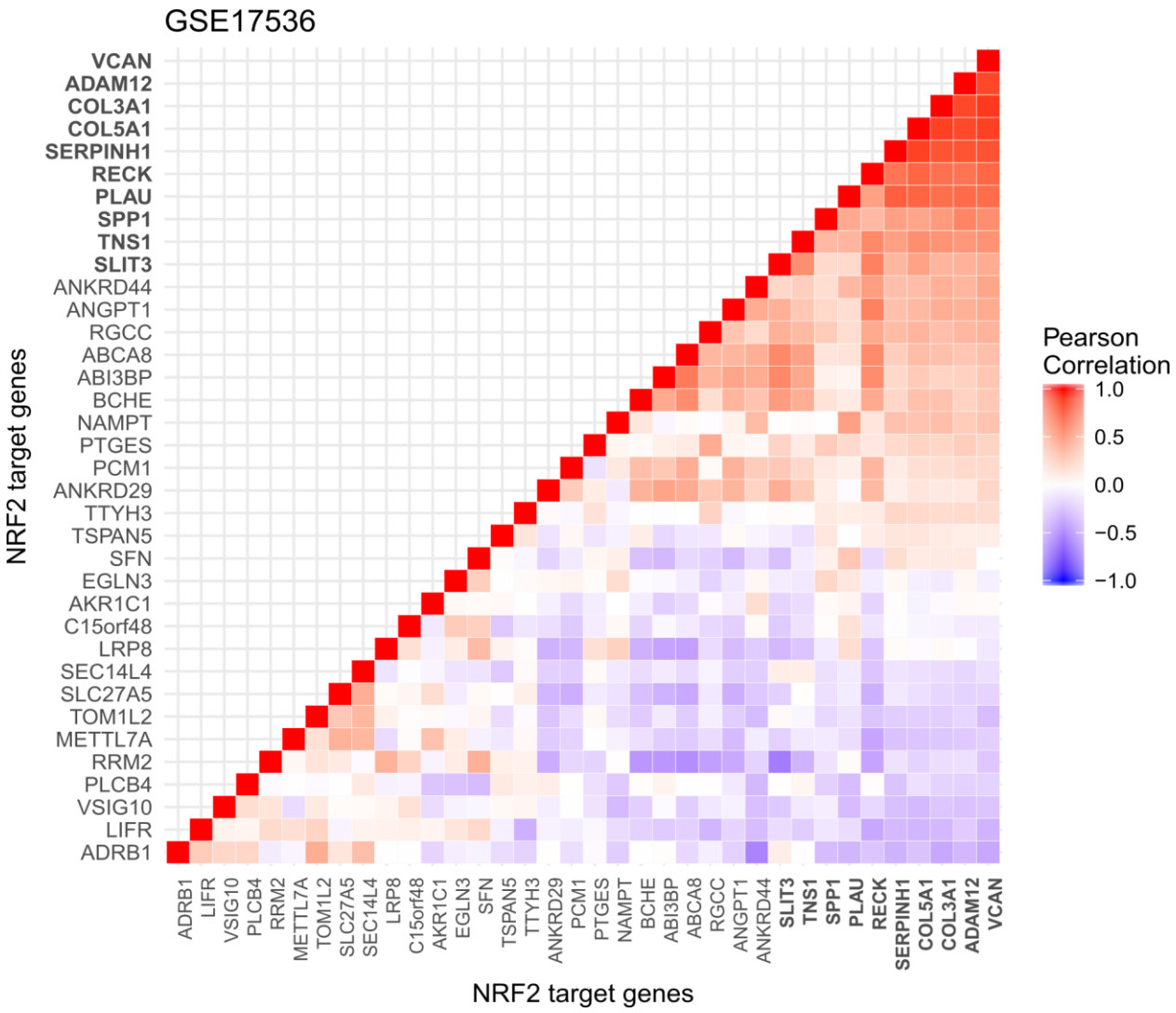


Figure 5-3. Pairwise correlation heatmap showing the degree of positive (red) and negative (blue) Pearson correlation between the 36 genes of the NRF2 pathway in the training set.

The high positive correlation between the subgroup of genes highlighted in the STRING analysis indicated a high degree of co-expression.10 of these had an absolute correlation of > 0.5 with PC1 and are highlighted in bold text.

5.2.2.2 Variable selection

AIC indicated that only the two major PCs, PC1 and PC2, were useful for predicting survival outcome, while BIC suggested that only the most variable PC (PC1) was useful. The inclusion of PC2 only slightly improved AIC, we therefore only considered PC1 in the training set. PC1 in the training set had absolute correlations >0.5 with probes that mapped to the following 10 genes: *VCAN, ADAM12, COL3A1, COL5A1, SERPINH1, RECK, PLAU, SPPI, TNS1* and *SLIT3*. These are part of predicted core gene set from the STRING analysis. Due to the high correlation

of these genes with PC1 in the training set, we hypothesised that they were of higher biological relevance for prognosis prediction than other NRF2 target genes.

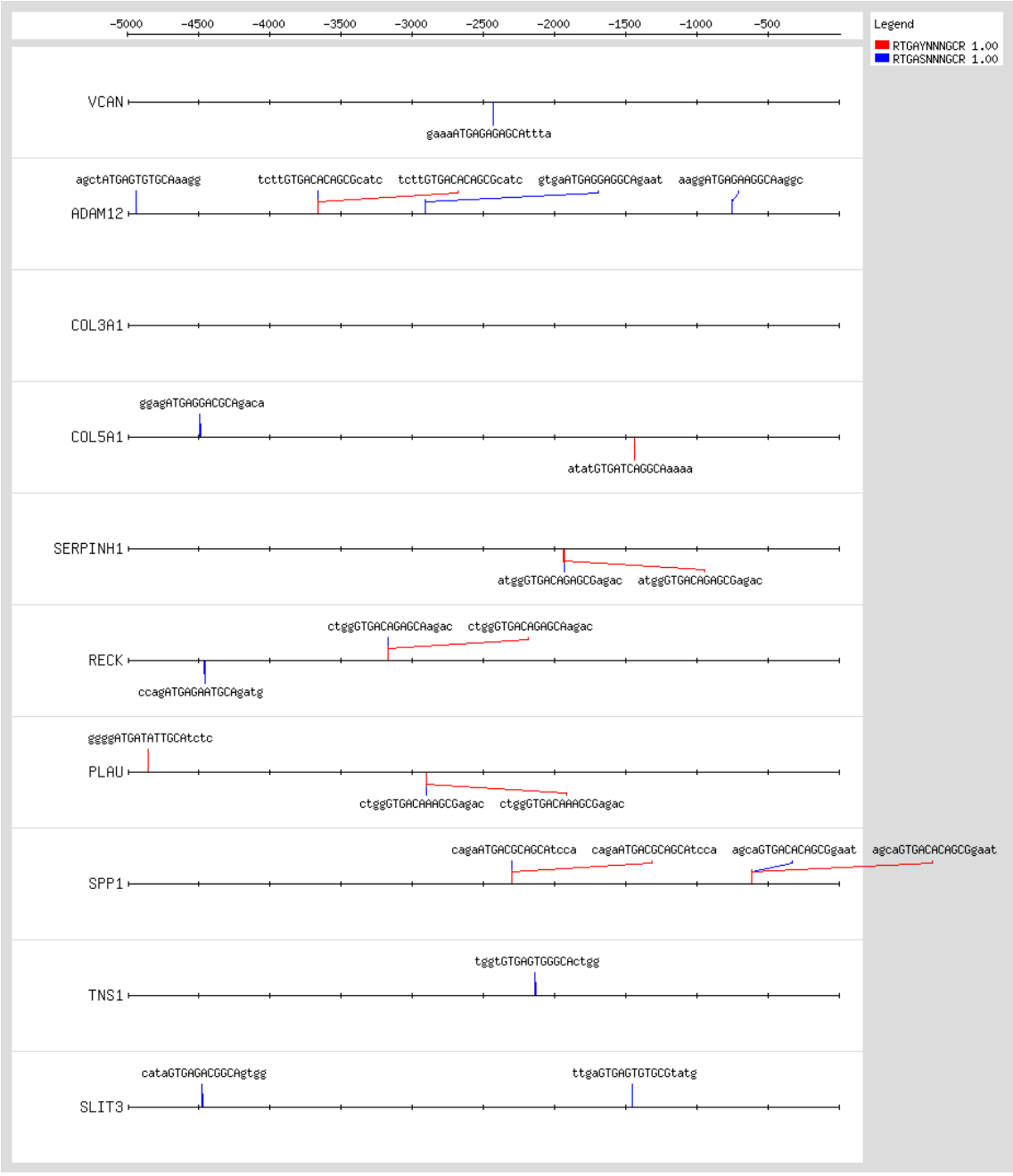


Figure 5-4. Regulatory sequence analysis.

The 5000 bp region upstream of the start site of 10 of the core gene set which have a high correlation with PC1 in the training set were analysed. 9 of these genes contain consensus ARE sites to which NRF2 can bind and directly regulate gene transcription.

These 10 genes were mooted to be the key biological representation of the pathway. Hence that was the pattern that should be sought following PCA in the validation sets, in order to ensure that the biological signal that was consistent with that demonstrated in the training set. For convenience this subset of genes are designated the core geneset. The importance of the core geneset was further underlined by the RSAT analysis [Figure 5-4]. Of the 10 genes in the core set, 9 have ARE consensus sequences in their promotor region which indicated a direct ability of NRF2 to regulate these genes.

5.2.3 NRF2 expression signature in the validation sets

Principal component analysis of all the available probes, for the 36 genes, was performed for each of the validation sets. Figure 5-5 shows the pairwise correlation of the genes with PC1 in GSE14333, GSE39582 and the MRC FOCUS datasets. GSE39582 has two representations – one for the 570 patients in the OS analysis (5B) and one for the 510 patients in the DFS analysis (5C). The overall pattern in all the datasets was similar that seen in the training set, with high correlation between the core gene set of VCAN, COL5A1, SPP1, SEPIN1H1, PLAU, TNS1, COL3A1, SLIT3, RECK and ADAM12. Given that the training set used the same array as GSE14333 and GSE39582 that was expected, but the migration to the Affymetrix array did not overly affect the signal.

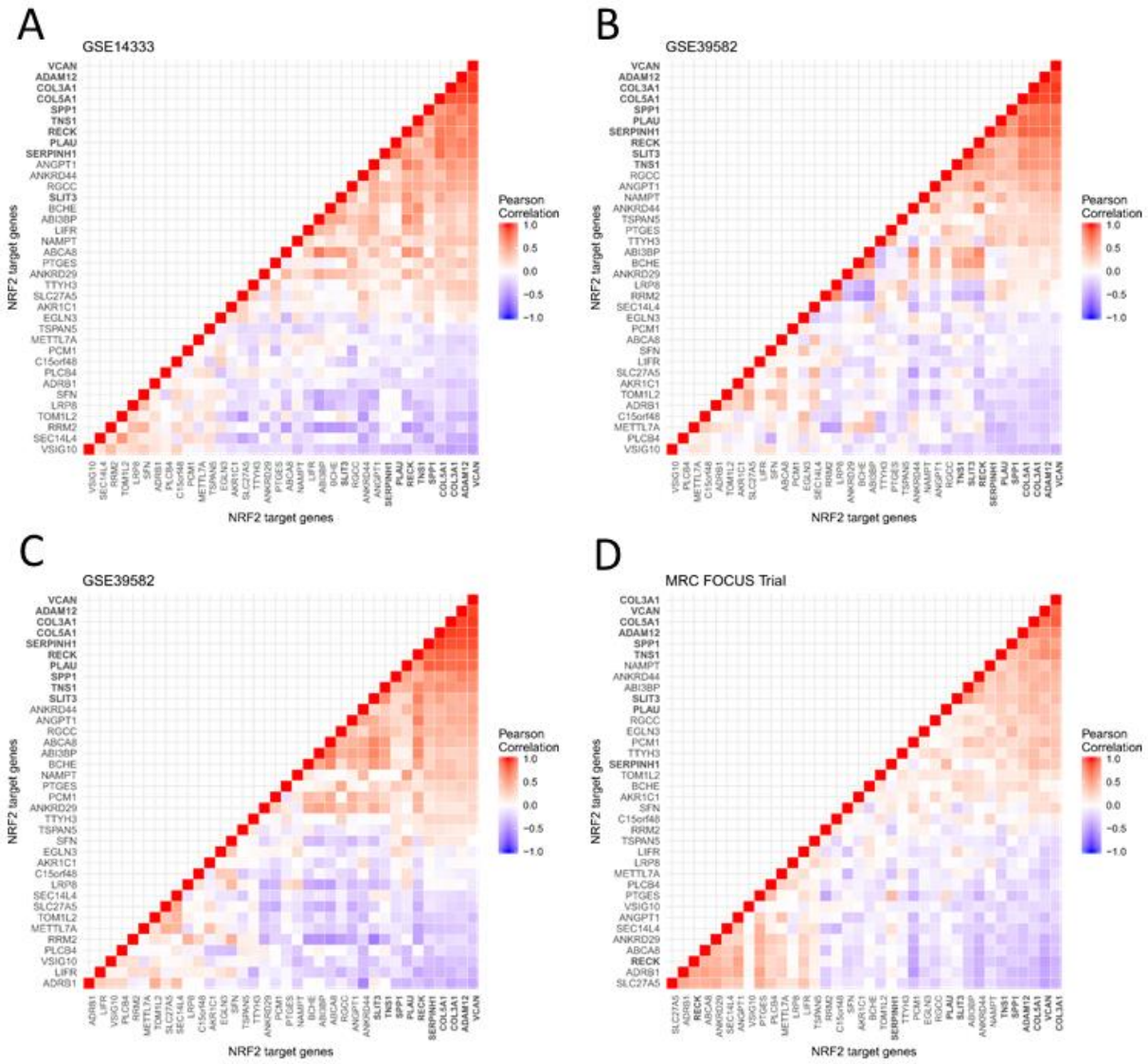


Figure 5-5. Pairwise correlation heatmaps of the showing the degree of positive (red) and negative (blue) Pearson correlation between the 36 genes of the NRF2 pathway in the validation sets

A) *GSE1433; B) GSE39582 PFS [N=510]; C) GSE39582 OS [N=570]; D) MRC FOCUS trial. The correlations show a remarkably similar pattern, more marked for the datasets with identical arrays and varying slightly in the MRC FOCUS trial. The core gene set (in bold) shows a very high correlation with the PC1 across all datasets.*

5.2.4 High NRF2 signature is biomarker of worse survival outcomes

The characteristics of the cohorts are detailed in Table 5-2. The total number of patients included in the validation analyses was 1,369 from the 4 separate cohorts. Three different platforms were used and cases consisted of early stage I-III and stage IV. The available adjusting covariates

differed between cohorts. Stage I was included in the analysis as patients from both cohorts with stage I disease (GSE14333 and GSE39582) suffered events (35 and 6 respectively).

Table 5-2. Cohort characteristics of the training and validation sets.

The numbers of cases, type of tissue, RNA expression platform, outcome variable and available covariates for adjusted analyses are indicated. (DFS = Disease Free Survival, OS = Overall survival, dMMR = deficient Mismatch repair, pMMR = proficient Mismatch Repair).

		GSE17536 (Training set)	GSE14333	GSE39582	MRC FOCUS trial	GSE87211
Patients		177	226	570	375 *(355 DNA mutant status)*	189
Tissue type		Fresh frozen	Fresh frozen	Fresh frozen	FFPE	Fresh frozen
Platform array		Affymetrix U133 v2.0	Affymetrix U133 v2.0	Affymetrix U133 v2.0	Affymetrix Xcel	Agilent Human 4 x 44k v2
Primary site		Colon	Colorectal	Colon	Colon	Rectum
Stage	I	24 (13.6%)	41 (18%)	37 (6.5%)		
	II	57 (32.2%)	95 (42%)	267 (47%)		70 (30.8%)
	III	57 (32.2%)	93 (41%)	206 (36%)		143 (63%)
	IV	39 (22%)		60 (10.5%)	375 (100%)	14 (6.2%)
Outcome variable		OS	DFS	DFS OS	OS	DFS OS
Covariates						
Chemo(radio)therapy			Yes 72	Yes 240	375 (100%)	189 (100%)
			No 154	No 326		
Site of primary				Prox. 232	Left 203	
				Dist. 351	Right 152	
BRAF V600E mutation				Mut 51	38 (10%)	
Mismatch Repair status				dMMR 77	dMMR 15	
				pMMR 459	pMMR 326	

In stage I/II/III disease, higher NRF2 signature expression corresponded to worse DFS in GSE14333 (HR=1.551, 95% C.I 1.200–2.004; LRT, p = 0. 0008) and GSE39582 (HR=1.172, 95% CI 1.008–1.362; LRT, p = 0. 0383). Including the 60 cases of stage IV disease also available in GSE39582, NRF2 expression was also associated with worse OS (HR=1.240, 95% C.I 1.086–1.416; LRT, p = 0.001). In the MRC FOCUS trial, comprised of first line stage IV metastatic patients, NRF2 expression was again associated with a worse OS (HR=1.140, 95% C.I 1.035–1.255; LRT, p = 0.008). Figure 5-6 shows that high expression corresponded with worse prognosis for DFS in GSE14333 and GSE39582 (panels A and B), and for OS in GSE39582 and MRC FOCUS trial (panels C and D).

In GSE39582, some patients did not receive adjuvant chemotherapy (i.e. untreated). However, receipt of adjuvant chemotherapy was significantly correlated with stage of disease (R^2 0.51 95% C.I 0.45 – 0.57, p <0.0001). Therefore including both adjuvant chemotherapy and stage of disease in the multivariate cox model resulted in confounding by multi-collinearity, created by the highly significant correlation between these two variables. Therefore the predictive effects over prognostic effects could not be accurately explored.

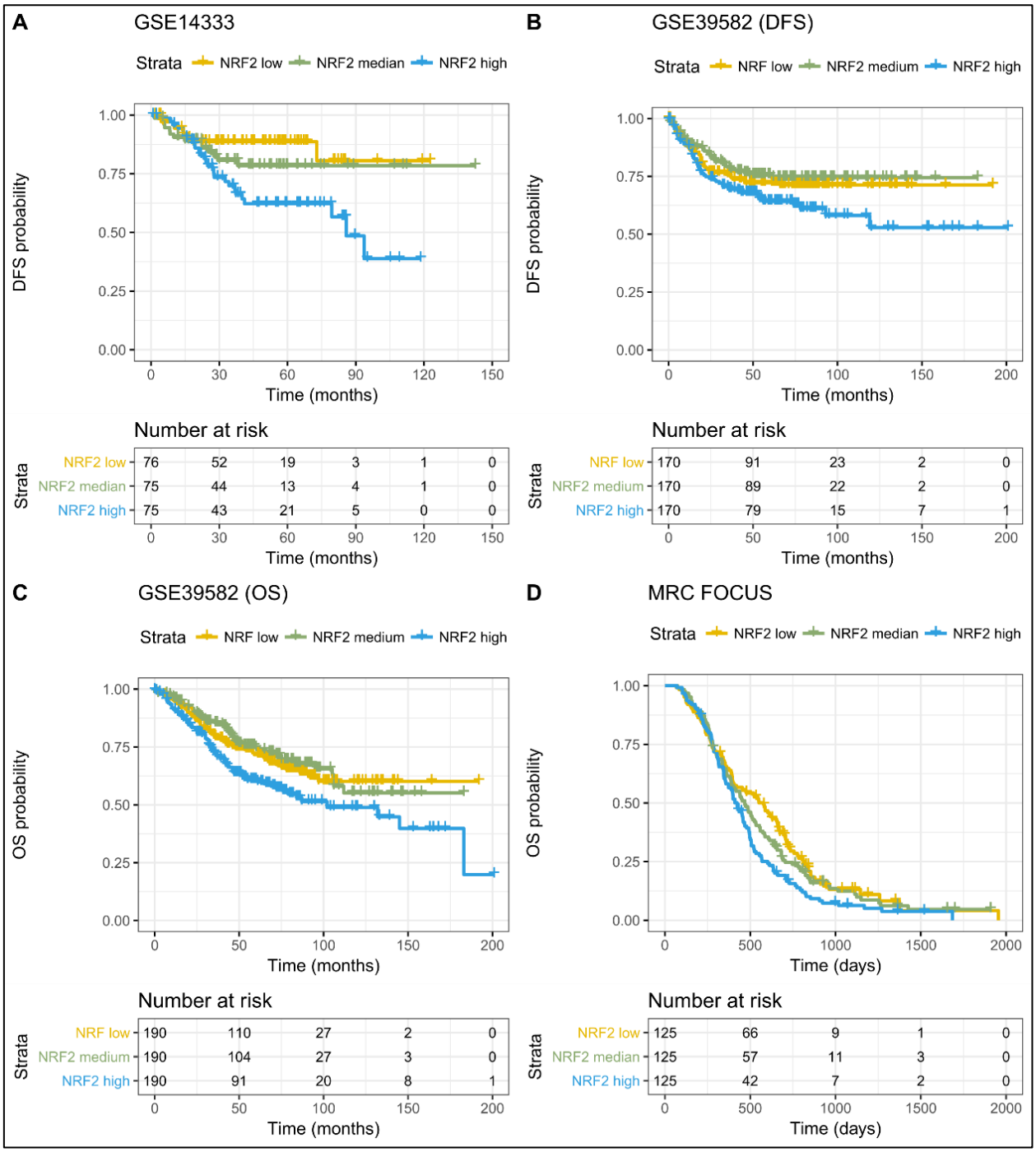


Figure 5-6. Kaplan Meier curves and associated risk tables for the primary analysis of three datasets

All show worse survival outcomes for patients with high, intermediate and low (tertiles) of NRF2 signature expression. A) GSE14333 and B) GSE39582 represents early stage I-III patients C) GSE39582 represents stage I-IV patients and C) MRC FOCUS represents stage IV first line metastatic patients. For the Kaplan Meier curves, a median cut point was used to binarise NRF2 signature expression.

In order to assess the relevance of NRF2 expression in rectal cancer specifically, and the ability to migrate between RNA expression platforms, the analysis was carried out in a rectal cancer only expression dataset, where all sampled patients received neoadjuvant chemoradiotherapy

(GSE87211). The pairwise correlation heatmap [Figure 5-7] showed that the pattern was still present, although less marked than in the other cohorts. This could be due to technical issues due to platform migration or biopsy changes or biological factors because it is a rectal only dataset. In spite of these issues, higher expression was associated with worse DFS (HR=1.431, 95% C.I 1.060–1.933; LRT, p = 0.056) but not OS (HR=1.464, 95% C.I 0.955–2.245; LRT, p = 0.197) [Figure 5-8].

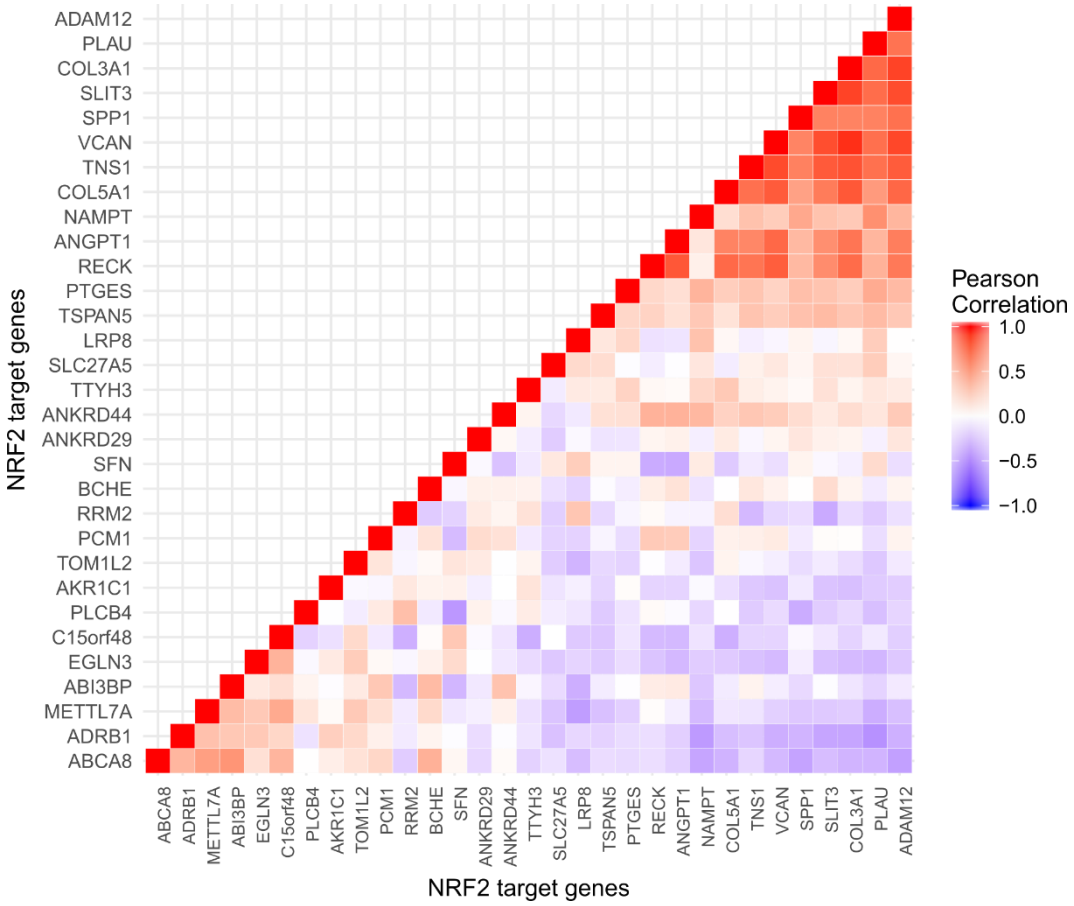


Figure 5-7. Pairwise correlation heatmap showing the degree of positive (red) and negative (blue) Pearson correlation between the 36 genes of the NRF2 pathway in GSE87211

Although the degree of co-expression between the core set of genes was present, it appears less marked than other datasets.

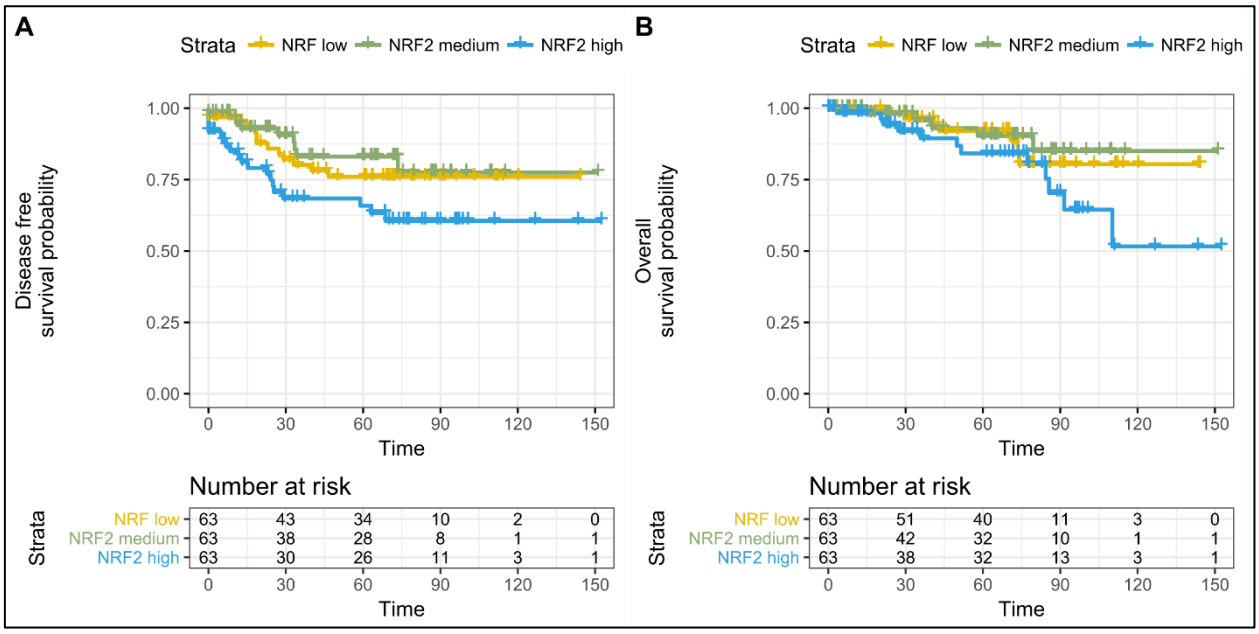


Figure 5-8. Kaplan Meier plots and risk tables for survival analysis in GSE87211 dataset

A) High expression of NRF2 was associated with worse DFS; B) not significant difference between tertiles for OS

5.2.5 NRF2 signature provides additional prognostic power in MVA

Within the publically available datasets there were additional variables that are known prognostic factors. These were specific to the individual datasets and have been summarised in Table 5-2, and include the most recognised factors such as stage, adjuvant chemotherapy, *BRAF* ^{V600E} mutation and microsatellite instability. The magnitude of their effects are summarised in the forest plot [Figure 5-9]. These known prognostic factors were used in the multivariate analysis to assess for competing effects with NRF2 signalling.

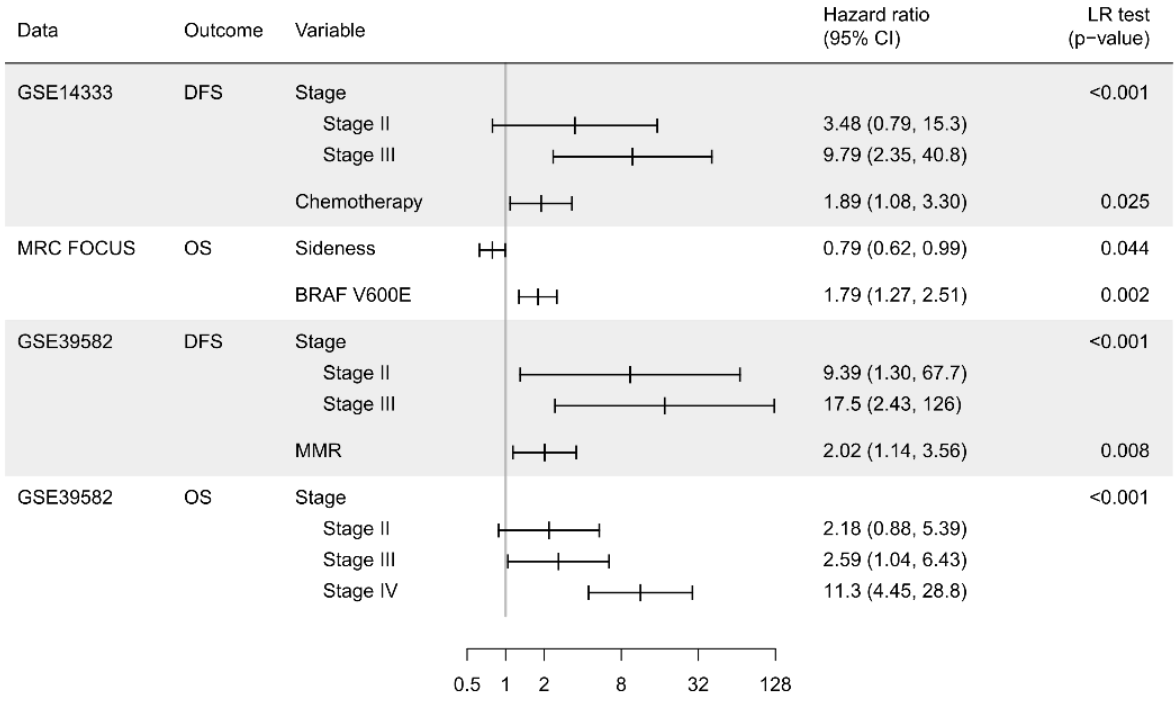


Figure 5-9. Forest plot of the multivariate analysis of known prognostic factors in MRC FOCUS cohort

It indicates the relative effects of the available prognostic variables within each dataset used in the validation analyses. These variables were used in the multivariate analysis with the NRF2 signature [Courtesy of Dr Chieh-Hsi Wu].

In GSE14333, after adjusting for the effect of stage and adjuvant chemotherapy NRF2 signature expression remained a significant predictor of worse DFS (HR=1.365, 95% C.I 1.049–1.776; LRT, p = 0.02). Similarly in GSE39582, the effect of high NRF2 expression corresponds to worse DFS (HR=1.168, 95% C.I 1.000–1.363; LRT, p = 0.049) after adjusting for the effect of stage and mismatch repair status (MMR). In the latter dataset, NRF2 expression was also significantly associated with worse OS when adjusting for stage alone (HR=1.185, 95% C.I 1.040–1.350; LRT, p = 0.01). No adjusted analysis was carried out for MMR with NRF2 expression on OS due to the known contrasting effects MMR status has on prognosis in early stage and metastatic disease, which could lead to model misspecification.

In the MRC FOCUS trial, prognostic factors within the dataset were site of the primary tumour (sidedness) and *BRAF* ^{V600E} mutation and used as adjusting variables in a multivariate analysis. Again, high NRF2 expression corresponds to worse OS (HR=1.123, 95% C.I 1.020–1.237; LRT, p = 0.0185). In summary, there was systematic evidence that NRF2 had an effect on DFS and/or OS in all available datasets [Table 5-3].

Table 5-3. Summary table of the Cox proportional hazard model analyses

It shows the numbers of patients included in each analysis, the adjusting variables where used and the p-value for the Cox model comparison (LRT = Likelihood ratio test)

	N	Analysis	Variables	HR* (C.I)	LRT (p-value)
Data set					
GSE14333	226	Primary	N/A	1.113 (1.045–1.184)	0.0008
		Secondary	Stage Adjuvant chemotherapy		0.0201
MRC FOCUS	375	Primary	N/A	1.033¥ (1.009–1.245)	0.008
		Secondary	Sidedness* BRAF ^{V600E} mutation*		0.0185
GSE39582 (OS)	570	Primary	N/A	1.054 (1.02–1.089)	0.001
		Secondary	Stage		0.01
GSE39582 (RFS)	510	Primary	N/A	1.04 (1.002–1.08)	0.0383
		Secondary	Stage Mismatch repair (MMR)*		0.049
GSE87211 (DFS)	189	Primary	N/A	1.067 (1.019–1.2445)	0.056
GSE87211 (OS)	189	Primary	N/A		0.197

* hazard ratio for event *per unit increase* in expression of NRF2 signature (continuous variable) which is different to HR between the upper and lower tertiles

¥ with imputation for missing variables

5.2.6 NRF2 signature was not associated with response to oxaliplatin *in-vivo*

Prior *in vitro* work has suggested that NRF2 expression has no effect on treatment resistance to oxaliplatin. To test this *in-vivo* data from the FOCUS trial was used. Within this trial, which randomised patients to first line fluorouracil, irinotecan plus fluorouracil or oxaliplatin plus fluorouracil, a significant difference in response rate (28% vs 57%) and progression free survival (6.3 months vs. 8.7 months) was noted between patients receiving fluorouracil and oxaliplatin plus fluorouracil. There was no difference in OS at the end of study (HR 0.93, 95% C.I 0.81 – 1.06, p = 0.26). To test for interaction, response to oxaliplatin was divided into two groups; complete response and partial response versus stable disease and progressive disease. There was no evidence of an interaction effect between oxaliplatin and NRF2 expression on the binary response to oxaliplatin (p=0.914).

5.2.7 NRF2 expression enriched in Consensus Molecular Subtype 4

In order to understand how NRF2 expression aligns with the current transcriptomic landscape of CRC, the distribution of NRF2 expression was examined across the four CMS subtypes in the MRC FOCUS trial [Figure 5-10]. While high NRF2 expression can be seen across all subtypes, there was a significant difference (Kruskal-Wallis, $p < 2.2 \times 10^{-16}$). Strikingly CMS 4 showed significantly higher NRF2 expression, followed by CMS 1. By contrast, the majority of patients in CMS 2 or 3 had low and intermediate NRF2 signature expression.

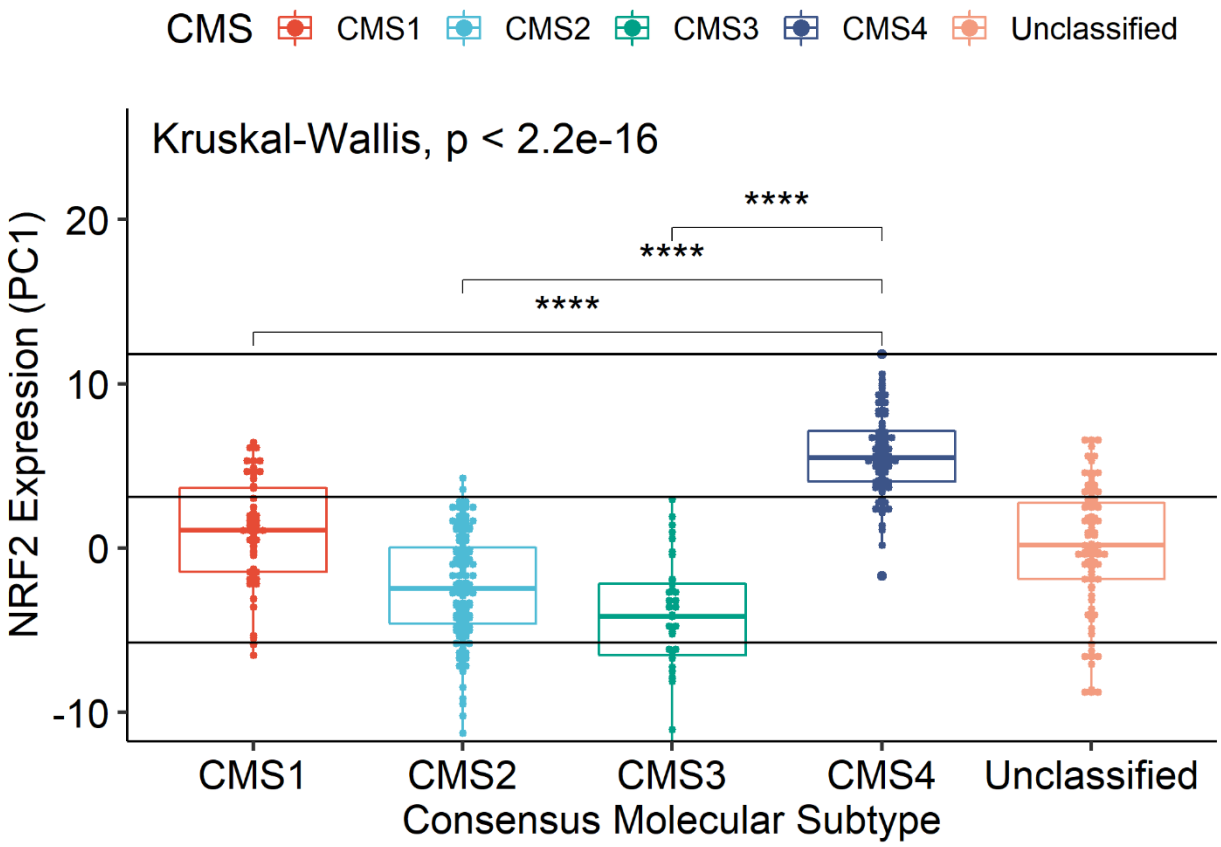


Figure 5-10. Distribution of NRF2 across CMS in the MRC FOCUS trial dataset

The boxplot and superimposed dotplots shows the distribution, median and ranges of NRF2 signature expression by each of the four CMS subtypes. The horizontal lines represent the high, medium and low tertiles of NRF2 signature expression used in the KM plots. The distribution of NRF2 was significantly different across all subtypes, with expression significantly enriched in CMS 4

5.3 Discussion

Using a candidate gene approach an NRF2 signature of pathway expression in CRC was derived and validated as an independent biomarker of poor prognosis across all stages of CRC. This is an entirely novel biological insight into CRC for several reasons.

This is the first known demonstration of NRF2 regulated genes behaving in a co-ordinated network fashion in CRC, *in-vivo*, and validates the hypothesis to use 'signature' approach to represent KEAP1/NRF2 pathway expression. The high degree of co-expression seen between genes such as *VCAN, ADAM12, SPP1, COL3A1* and *COL5A1* was consistent with the known behaviour of these genes. That the genes were selected from lung cancer signatures strengthens

the unbiased nature of the analysis. Furthermore, this pattern was seen across all four of the validation sets, including from rectal biopsies as opposed to whole tissue sections and underlines the robustness of the finding.

Secondly, the NRF2 signature has been shown as a biomarker of poor prognosis across all stages of CRC in several large, independent datasets comprising 1,360 patients making it one of the largest transcriptomic biomarkers in CRC. To place the size in context, the OS and DFS cohorts used to validate CMS were 2,129 and 1,785 cases respectively.

Thirdly, the effect was detected independently of the tissue processing technique (FFPE vs fresh frozen) or the mRNA expression platform used. This is especially important as although the signal pattern was less clear from the Agilent array, one can clearly see the trend towards worse survival in that dataset.

Lastly, and arguably most importantly, the effect was maintained when adjusting for known prognostic clinical and molecular factors including stage, adjuvant chemotherapy and microsatellite instability in the non-metastatic setting and BRAFV600E mutation and tumour sidedness in the relapsed setting. The 375 patients with stage IV, metastatic disease have been selected from a large randomised controlled phase III trial which makes the findings highly stable and robust.

The difference in prognosis between those with high and low NRF2 expression could, at least in part, be due to therapeutic resistance. Fluorouracil is the main chemotherapy drug used in both the adjuvant and metastatic setting, and was the backbone of therapy used in the FOCUS trial. Silencing of NRF2 signalling has been shown to overcome 5-FU resistance in CRC models in both an *in vitro* and *in vivo* setting [140]. We saw in Chapter 3 that NRF2 expression has a modest effect in mediating therapeutic resistance to both radiation and fluorouracil chemotherapy in CRC *in-vitro* models. Other work in CRC models also supports the idea that NRF2 signalling mediates therapeutic resistance [179].

The overall effect size of the hazard rations indicates a modest contribution of NRF2 to prognosis (1.14 – 1.55). On the one hand the relatively consistent effect size given the variety of the datasets is reassuring. Any contribution of NRF2 to prognosis in CRC is likely to have already been captured in a global analysis like CMS. The effect sizes are thus likely to be smaller than that of any individual CMS group.

However it could be argued that the clinical and biological significance of such effect sizes may be negligible in spite of the statistical significance of the analyses. If the effects are already part of a larger scale analysis then a subgroup is potentially unlikely to add further value. In the case of NRF2 driven prognosis, the absence of an intervention to target NRF2 specifically could also argue against trying to further sub-categorisation.

5.3.1 Oxaliplatin interaction

Previous work in chapter 3 has already suggested that NRF2 signalling has no role in oxaliplatin resistance, in contrast to 5-FU and SN-38. The absence of an interaction, meaning that there was no difference in radiologically assessed response at 6 months when accounting for NRF2 expression, provides further supportive *in-vivo* evidence that oxaliplatin is not sensitive to changes in NRF2. The reasons for these findings are not entirely clear. Wang et al. suggested that oxaliplatin can actually induce NRF2 signalling with *in-vitro* and *in-vivo* models and provide a protective effect [231].

Recent data also suggests that the mechanism of oxaliplatin is not due to intercalating DNA crosslinks, like the other platinum analogs, but is related to ribosome biogenesis stress [232]. Data in Chapter 4 demonstrated that upregulation of NRF2 enriched for a transcriptional program of ribosomal biogenesis. It is possible that high levels of ribosome biogenesis in NRF2 activated tumours renders them less sensitive to this mechanism of action from oxaliplatin.

5.3.2 Rapport with CMS

Although distributed across all four subtypes, NRF2 expression was significantly enriched in CMS 4 and CMS 1. Both are known to have poor prognosis in relapsed disease, consistent with the poor prognostic effect of high NRF2 expression demonstrated here. This would argue therefore that NRF2 signalling adds further biological insight into the transcriptomic taxonomy of the CMS classification. The exact role of NRF2 in interacting with the current subtype annotations, such as stromal activation and TGF-β in CMS 4, and immune activation and infiltration in CMS 1, warrants further study.

Although CMS 4 is often referred to as the mesenchymal subtype and noted for its enrichment with TGF-β activation (3.46×10^{-6}), it actually demonstrated to have higher levels of activation of 'Matrix modelling' ($0.00 \times 10^{+0}$ i.e. maximum) by GSEA [84], the latter being measured against the REACTOME extracellular matrix geneset. *SPP1, VCAN, ADAM12, COL3A1, COL5A1* and *SERPINH1* are all involved in Extracellular matrix (ECM) organisation and contain ARE consensus sequences in their promoter regions indicating a capacity for direct NRF2 regulation. Alteration of ECM constituents can activate pathways responsible for cell migration, abrogation of apoptosis and cell proliferation [233] – all hallmarks of cancer. For a cancer cell to survive its niche ECM, and embed itself at a distant metastatic site, it must achieve anchorage independence. This in turn requires cancer cells to tolerate a significant amount of stress due to caspase mediated, non-apoptotic death program known as anoikis. There is increasing evidence that cancer cells rely on mechanisms that allow it to maintain low levels of intracellular ROS to overcome anoikis [234,235]. Overexpression of the master redox pathway like NRF2 would thus create favourable, pro-survival conditions.

The diverse spectrum of processes affected by KEAP1/NRF2 pathway expression is reflected in the genes that make up the signature. *SFN* and *RGCC* are known to function within TP53

regulation of cell cycle genes. *SFN* is induced by TP53 and contributes to G2 arrest by binding to the complex of CDK1 and CCNB1 (cyclin B1) and preventing its translocation to the nucleus [236]. While promoting G2 arrest, SFN can simultaneously inhibit apoptosis by binding to BAX and preventing its translocation to mitochondria.

As a canonical pathway which functions as a transcription regulator it is a hypothetical therapeutic target, with some success already seen in directly inhibiting NRF2 [173,179]. There are however concerns over off target effects of NRF2 inhibitors. A more elegant solution may come from targeting the induced changes seen in NRF2 over expression. Romero et al. demonstrated a metabolic requirement for glutaminolysis in KEAP1-NRF2- mutant lung adenocarcinoma and showed that showed reduced growth potential following treatment with two glutaminase inhibitors [174]. The results presented in Chapter 4, where high NRF2 signalling facilitated a 'switch on' of glutamine metabolism, would support this approach. There are currently two phase I/II trials in CRC (NCT03263429; NCT02861300) adopting glutaminase inhibition as a strategy. Interestingly the former requires KRAS WT patients while the latter mandates PIK3CA mutant patients. It has been suggested that NRF2 pathway overexpression is a mechanism of acquired tolerance in the PI3K-RAS mutant cancers [123,182] so one may expect the latter study to be enriched for high NRF2 expression and thus more likely to show benefit. While these and other large scale studies may reveal on benefit from metabolic inhibition in CRC, the role of NRF2 will merely be inferred due to the absence of a biomarker.

The NRF2 signature herein is a novel, expression based prognostic biomarker which drives poor prognosis across all stages of CRC. While the KEAP1-NRF2 pathway is increasingly well understood, better characterisation of its role and relationship to other biological factors like CMS 4 in CRC is needed. Having derived and validated a method for aggregating pathway

expression in CRC datasets, I next sought to examine its utility as a biomarker of radioresistance *in-vivo* in an attempt to validate the *in-vitro* findings of Chapter 3.

Acknowledgment of contribution: This work was done in collaboration with Dr Chieh Hsi-Wu from the Dept. of Statistics. Specifically she was responsible for the PCA, imputation of missing data and advising on multivariate analysis. She is a joint author on the resulting manuscript. (https://www.biorxiv.org/content/10.1101/690974v1.full)

CHAPTER 6

6.1 Introduction

Current clinical practice in rectal cancer is that all patients who have clinical indications, such as a threatened circumferential resection margin, for neoadjuvant chemoradiation are treated with approximately the same radiation dose as standard of care of 45-50.4Gy in 1.8-2Gy per fraction with fluorouracil [19]. Whilst strategies of radiosensitization and dose escalation are continuing areas of interest, there are no validated predictive biomarkers of response that would allow stratification and/or enrichment of these studies. They therefore run the risk of giving negative results by virtue of the intrinsic heterogeneity of their study population. The ability to predict response of rectal cancers to neoadjuvant radiotherapy/chemoradiotherapy, from the pre-treatment diagnostic biopsy, is therefore an area of intense research interest.

There are a number of challenges in achieving this goal. Firstly, the prediction must be made from a pre-treatment diagnostic biopsy. This is often a small fragment of tumour, sufficient to establish a diagnosis of invasive adenocarcinoma. The sample is fixed in formalin and embedded in paraffin for the purposes of histopathological analysis which in turn affects the quality of the DNA, and particularly the RNA, making molecular diagnostics more difficult, but not impossible [237].

Secondly, there is considerable debate about what an appropriate endpoint (local control, DFS, OS) and what to use as a surrogate outcome of that endpoint. In a randomised trial setting, a surrogate is an endpoint that occurs after patient randomisation, but before the true endpoint of interest, while yielding conclusions about the effect of treatment on the true endpoint [238]. An ideal surrogate outcome requires three characteristics, first outlined by Prentice and now often referred to as the Prentice criteria [239] :

a) The surrogate must affect the true endpoint on an individual patient level
b) The effects of any treatment on the surrogate must effect with the true endpoint (referred to as trial level association).

c) The effect of the surrogate on the true endpoint does not vary by randomisation group

Consequently a true surrogate endpoint of response must correlate with OS and an improvement in response rate by dose escalation must also result in an improvement in OS.

Much of the research and literature focuses on pCR, complete clinical response (cCR) or near complete responses. It has also been the primary endpoint in multiple Phase II trials of novel therapy strategies in the neoadjuvant setting. pCR has been associated with improved outcomes [240] in a meta-analysis of over 3000 patients. Unfortunately the analysis does not demonstrate that pCR as a surrogate endpoint meet the required Prentice criteria for surrogacy. A recent meta-analysis of 22 trials with 10,050 patients confirmed that pCR is not a surrogate of 5 year OS [42]. The individual level surrogacy pCR rate and 3-year DFS were poorly correlated with the 5 year OS (R=0.52; 95% C.I, 0.31 - 0.91; p=0.002; and R=0.60; 95% C.I, 0.36 - 1; p=0.002) respectively. The trial-level surrogacy analysis confirmed that the treatment effects on surrogates (ΔpCR rate and Δ3yDFS) were not strong surrogates for treatment effects on 5 year OS (R=0.2; 95% C.I, − 0.29 - 0.78; P=0.5 and R=0.64; 95% C.I, 0.29 - 1; P=0.06) respectively. They concluded that the use of pCR as a surrogate for 5 year survival is not supported, although the use of 3 year DFS in a phase III randomised controlled trial may still be appropriate.

The other disadvantage of pCR is that it results in a crude dichotomisation of what is in reality a heterogeneous spectrum of responses. Approximately 85% of rectal cancer patients treated with neoadjuvant therapy will not achieve pCR. Considering this spectrum of responses as a single entity risks losing valuable biological data and statistical power.

Several pathological response/regression grading criteria have emerged [241–243] but none have been widely adopted for clinical trial use due to issues of inter-observer variability. Until recently none had shown individual level surrogacy for useful clinical endpoints such as DFS or OS, although the Dworak system had recently shown surrogacy for DFS in a retrospective analysis of the (CAO/ARO/AIO)-04 randomized trial [21]. In the absence of a single, universally

adopted and endorsed system it is unlikely that regression scores will be adopted into clinical trials as endpoints.

An alternative response metric is the NAR Score [244]. This relatively recent outcome measure is comprised of three clinic-pathological variables: the MRI derived clinical tumour stage (cT), the final pathological tumour stage (pT) and final pathological nodal stage (pN). Each of these is provided a weighting and combined into a pre-specified formula:

$$NAR = \frac{[5pN - 3(cT - pT) + 12]^{2}}{9.61}$$

The score was derived from the Valentini nomograms [245]. These nomograms were derived from and validated in large scale randomised clinical trials for use as surrogates of disease free and OS. However, they were never widely adopted in clinical or trial use, potentially due to their cumbersome nature. In deriving the NAR score as surrogate outcome for use in neoadjuvant rectal trials only pT and pN were retained out of the original eight variables as they are the only ones that could be affected by neoadjuvant treatment. Clinical tumour stage was incorporated to specifically capture the degree of downstaging, making it a response metric.

The score was initially validated retrospectively in the NSABP-R04 trial, a large randomised study of neoadjuvant rectal radiation, as a surrogate for OS [244]. More recently it has also been validated in the CAO/ARO/AIO-04 randomized trial as a surrogate for DFS [246]. A large retrospective study of the Dutch Cancer registry comprising 6,596 patients treated with neoadjuvant rectal radiotherapy also confirmed the prognostic power of the NAR score [247].

The calculation of the NAR variable results in a maximum of 24 discrete scores. For the purposes of Kaplan Meier analysis it has also been grouped into three categories; high > 16, intermediate < 16, > 8 and low <8. Accordingly the expected 5 year survival improved from approximately 60% to 80% and 90% respectively [244]. Across all published datasets, a high NAR

score (poor response to radiotherapy) was consistently associated with worse survival. An additional advantage is that the groupings appear to be remarkable consistent. Low NAR represents 20-25% of patients, intermediate NAR ~45% and high NAR 30-35%. Low NAR by definition encompasses those patients who achieve pCR. Because of these consistent group sizes it is possible to make use of all the available response data from any given dataset.

The NAR score has been adopted as the primary endpoint for use in the ongoing NRG platform study in the neoadjuvant treatment of rectal cancer based on its utility as a surrogate of survival outcomes (NCT02921256).

Previous *in-vitro* work in Chapter 3 demonstrated that NRF2 pathway over expression can induce radioresistance. In Chapter 5 an NRF2 metagene signature was derived and validated as a prognostic biomarker, whereby higher expression was associated with worse survival outcomes. I hypothesised that NRF2 expression could be used to predict NAR score as a surrogate outcome in a cohort of rectal cancers with RNA expression profiling. The specific aims were to see if the NRF2 signature could be identified from mRNA profiling of small rectal biopsies and to see if the detected expression pattern was predictive of a worse response to neoadjuvant radiotherapy as determined by the NAR Score.

6.2 Results

6.2.1 Cohort characteristics

A total of 127 patients were included. The patient characteristics are summarised in Table 1. 97% of patients received chemoradiotherapy (CRT) to a total dose of 45Gy in 25 daily fractions with capecitabine; only 4 didn't receive capecitabine but had 50Gy of radiation to offset the absence of a radiosensitzer. Patients who had received non-standard treatment such as neoadjuvant oxaliplatin chemotherapy were excluded from the analysis.

Table 6-1. Aberdeen rectal cohort characteristics (N=127).

		N	%
Gender	Male	77	61
	Female	50	39
CMS	1	10	8
	2	38	30
	3	23	18
	4	14	11
	Unclassified	42	33
Response	Complete	33	26
	Good partial	50	39
	Partial	38	30
	Minimal	6	5
cT stage	1	1	<1
	2	21	17
	3	100	79
	4	5	4
cN	0	63	50
	1	51	40
	2	13	10
ypT	0	33	26
	1	16	13
	2	33	26
	3	45	35
ypN	0	101	80
	1	21	16
	2	5	4
NAR category	Low	44	34
	Intermediate	56	44
	High	27	21

6.2.2 NRF2 signature pattern identified in biopsy samples

The co-expression pattern seen in chapter 5 was identified from RNA profiling of tumour sections following surgical resection. To check the pattern of expression of the NRF2 signature obtained from small biopsy samples a pairwise correlation heatmap was constructed [Figure 6-1]. This showed a similar pattern to that seen previously in training dataset (GSE17536) and the datasets used for prognostic validation. The high degree of correlation between the core network of genes of *COL3A1*, *COL5A1*, *VCAN*, *SPP1*, *ADAM12*, *SERPIN1H1*, *TNS1* and *PLAU*

with PC1 demonstrated the high degree of co-expression between these genes and their respective importance in representing the NRF2 pathway.

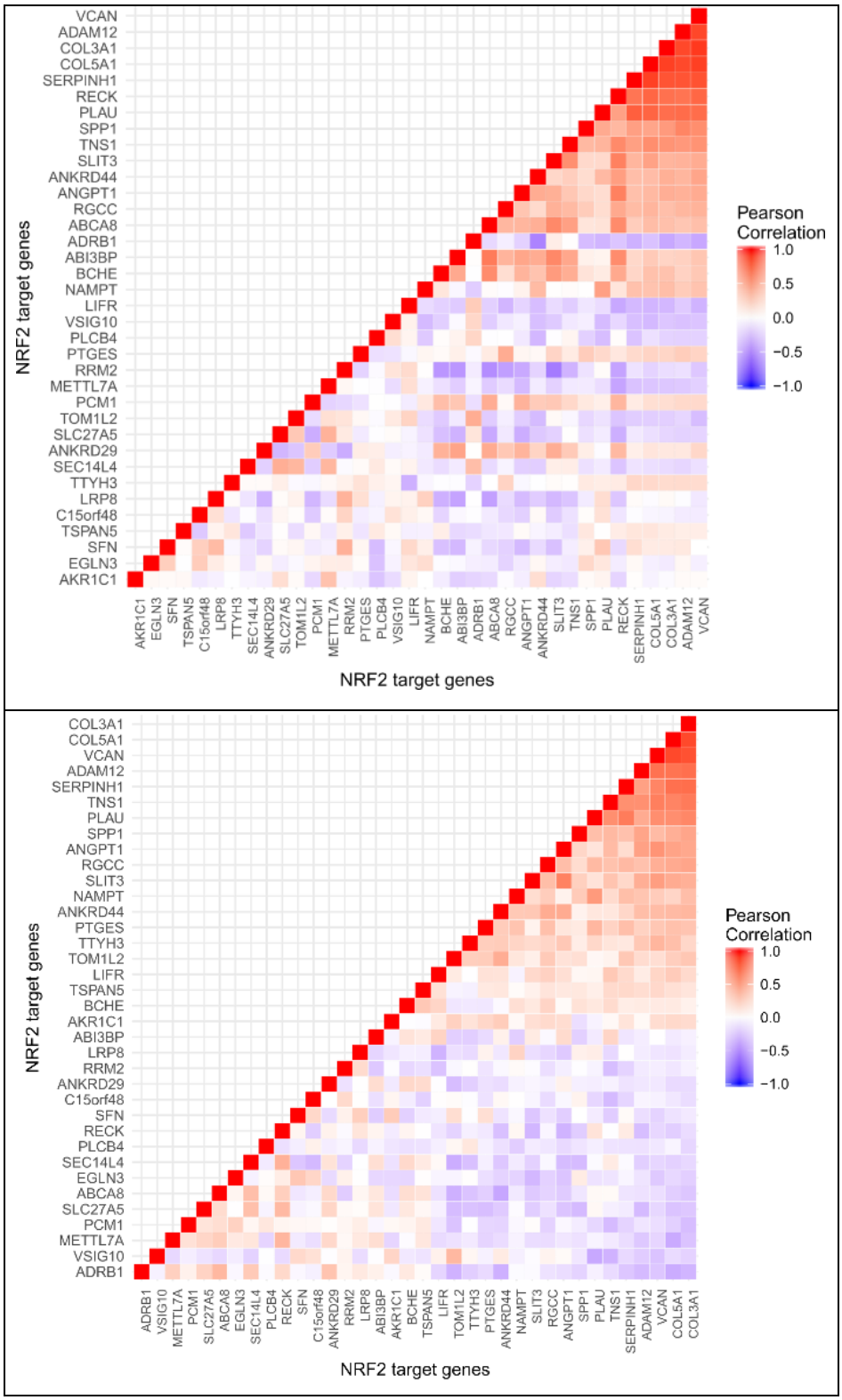


Figure 6-1. Pairwise Pearson correlation heatmaps between the 36 genes of the NRF2 pathway in GSE17536 (top) and the Aberdeen rectal cohort (bottom).

The correlation pattern shows a high degree of co-expression between the core groups of genes that was identified previously. Red = positive correlation; blue = negative correlation.

6.2.3 NRF2 signature is a biomarker of worse radiotherapy response

Across the cohort, higher NRF2 expression was associated with higher NAR [Figure 6-2]. In the ordinal regression model, this relationship was statistically significant with an odds ratio of 1.34 (95% C.I. 1.01–1.80, LRT p-value = 0.023).

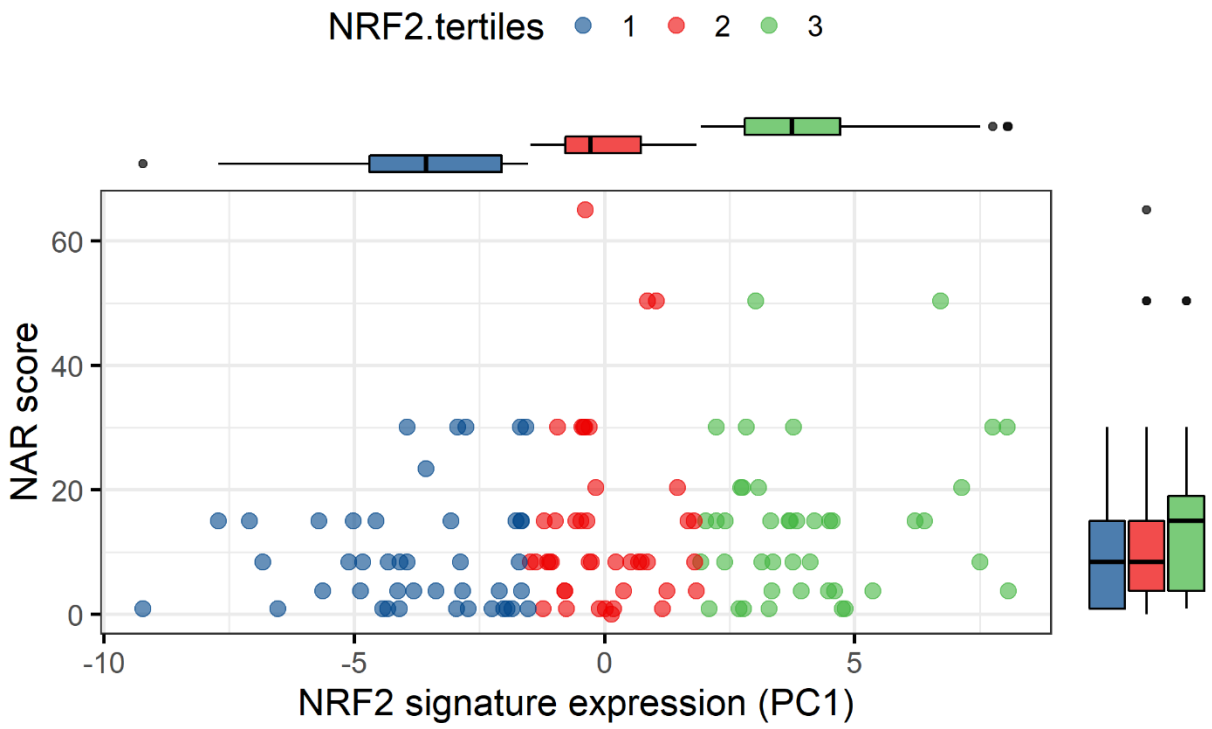


Figure 6-2. Scatter histogram of NAR scores and NRF2 signature expression

Each dot represents a patient value on the scatterplot. The outer boxplots represents tertiles with median and range. The distribution of NRF2 expression favours the higher NAR scores.

6.2.4 NRF2 signature provides additional explanatory power to clinical N stage

After adjusting for the effect of clinical N stage, there remained statistical evidence that NRF2 metagene explains the variation of NAR scores. Higher NRF2 metagene signature expression corresponded to higher NAR scores (OR = 1.35, 95% C.I. 1.00–1.81, LRT p-value = 0.0403).

Figure 6-3 shows the ordinal regression plots for the primary analysis model and the adjusted analyses. As NRF2 expression increases, the odds of a high NAR score of greater than 16 were increased. When adjusted by clinical nodal stage (cN) which is the only factor not included in the NAR score calculation, the effect becomes less marked. However, for patients with clinically involved nodes on baseline MRI and high NRF2 expression on biopsy, the odds of a high NAR score – a poor response - at the end of neoadjuvant therapy were much higher.

The lower panels of Figure 6-3 shows the opposite effect. As NRF2 expression increases the odds of an NAR score less than 8 decreases, particularly in the node negative group.

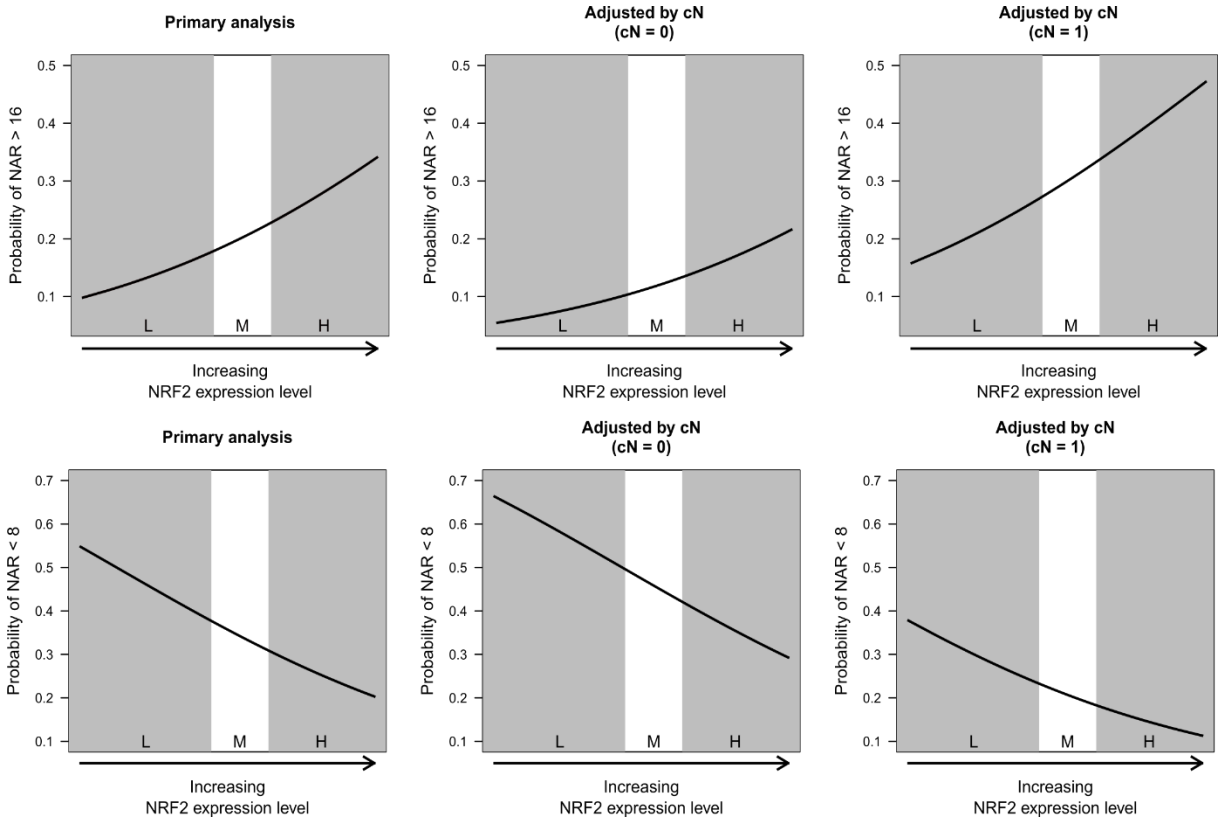


Figure 6-3. Ordinal logistic regression plots for NRF2 expression predicting NAR score

Top panel: Probability of a NAR >16. As NRF2 expression increases the odds of a high NAR score increase. Also shown are the effects of including clinical nodal stage from the MRI on the model. Clinical nodal stage is the only clinical factor that is not part of the NAR score. Node positivity at baseline and high expression of NRF2 signature has much greater odds of a high NAR score. Bottom panel: Probability of NAR < 8. As NRF2 expression increase the odds of low NAR score decrease. Adjustments for cN stage are also shown. (L, M, H are tertiles of NRF2 signature expression where L = Lowest/bottom tertile; M = median tertile; H = Highest tertile) [Courtesy of Dr Chieh-His Wu]

6.2.5 NRF2 enriched in Consensus Molecular Subtype 4

Previously, NRF2 expression was enriched in the CMS 4 subtype in the MRC FOCUS trial. The distribution of NRF2 expression was examined in the Aberdeen rectal cohort to see if the pattern was replicated given the differences between the tissue sections and anatomy. NRF2 expression was detected across all four CMS subtypes but there were significant difference in the distribution (Kruskal – Wallis test, $p = 1.9 \times 10^{-13}$). Enrichment was seen again in CMS 4 with significantly higher expression in this group compared to each of the other subtypes, followed by CMS1, 2 and 3 respectively [Figure 6-4]. The overall distribution was strikingly similar to the MRC FOCUS trial.

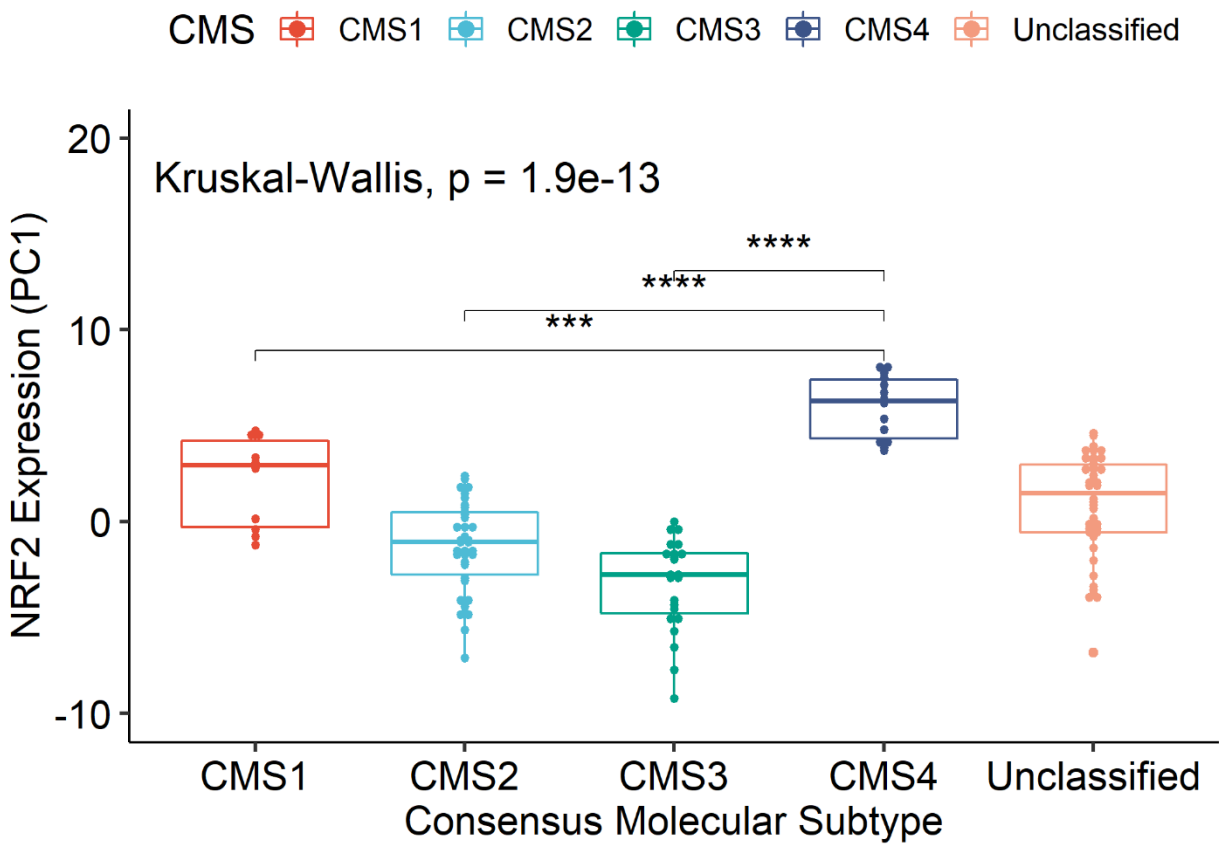


Figure 6-4. Boxplots of NRF2 signature expression by Consensus Molecular Subtype.

There was a significant difference in mean NRF2 expression between the four groups overall (Kruskal-Wallis test, $p = 1.9 \times 10^{-13}$). CMS was significantly enriched for NRF2 expression - CMS4 vs CMS 1, Wilcoxon rank sum $p < 0.001$; CMS 4 vs CMS 2, Wilcoxon rank sum $p < 0.0001$; CMS 4 vs CMS 3, Wilcoxon rank sum $p < 0.0001$)

6.3 Discussion

Using the previously validated prognostic signature of NRF2 expression, the data here supports the hypothesis that it can also be used a biomarker of worse response to neoadjuvant chemoradiotherapy in rectal cancer. Several important conclusions can be taken from this work.

Firstly, these findings validate the *in-vitro* data from Chapter 3 which suggested that high NRF2 signalling mediated a radioresistant phenotype.

Secondly, the mRNA signature was detected from small rectal biopsies that have been fixed and paraffin embedded suggesting a degree of robustness against technical and biological limitations. This is an important consideration for a potential biomarker as the clinical pathway dictates that biopsies are the only tissue available for clinical decision making.

Thirdly, the signature predicts for a worse response to chemoradiotherapy as measured by the NAR score. From a clinical perspective this can be interpreted as patients with high NRF2 expression, detected from diagnostic biopsies, will have less tumour regression and/or the development of pathologically involved nodes during the course of neoadjuvant treatment. However, in the strictest of terms, it is not a predictive biomarker with a specific cut point. Some patients with high NRF2 expression had a low NAR score meaning that they responded well to neoadjuvant treatment. The patient with the highest NAR score had an intermediate NRF2 expression, for example. As such these patients should not be treated with standard of care as routine but offered novel strategies to overcome radioresistance and/or improve outcomes.

Lastly, the response endpoint of NAR is a validated surrogate of the true clinically relevant endpoint of DFS and OS as discussed in the introduction. The prognostic outlook for patients

with high NRF2 expression would expected to be poor based on the results of Chapter 5. At least part of this poor prognosis would appear to be mediated by a poor response to radiotherapy. Previous work which demonstrated patients with lower regression, as measure by the Dworak regression grading, had worse survival outcomes [21] is consistent with this premise. The NRF2 signature demonstrated additive clinical value to clinical risk factors of involved radiological nodes making it a useful biomarker in addition to the current clinical variables. Unfortunately data on other MR based risk factors such as extramural venous invasion [248] were not available in the Aberdeen dataset to include in the ordinal regression model.

Some drawbacks should be discussed. In its present form the 36 gene signature can only be generated from bulk transcriptomic data, from a cohort of patients. It is not possible to derive an NRF2 threshold for individual patients. It is also unable to distinguish truly radioresistant patients, merely those with a higher chance of having a worse response. The OR of 1.35 is only of modest effect size in addition.

The pattern of enrichment across the CMS was significantly different between subtypes. Similar findings were noted in the MRC FOCUS trial results in Chapter 5. CMS 4 showed highly significant enrichment for NRF2 expression, followed by CMS1, with the lowest overall expression being in CMS 3. This enrichment is most likely the result of activity of ECM matrix genes in the NRF2 signature.

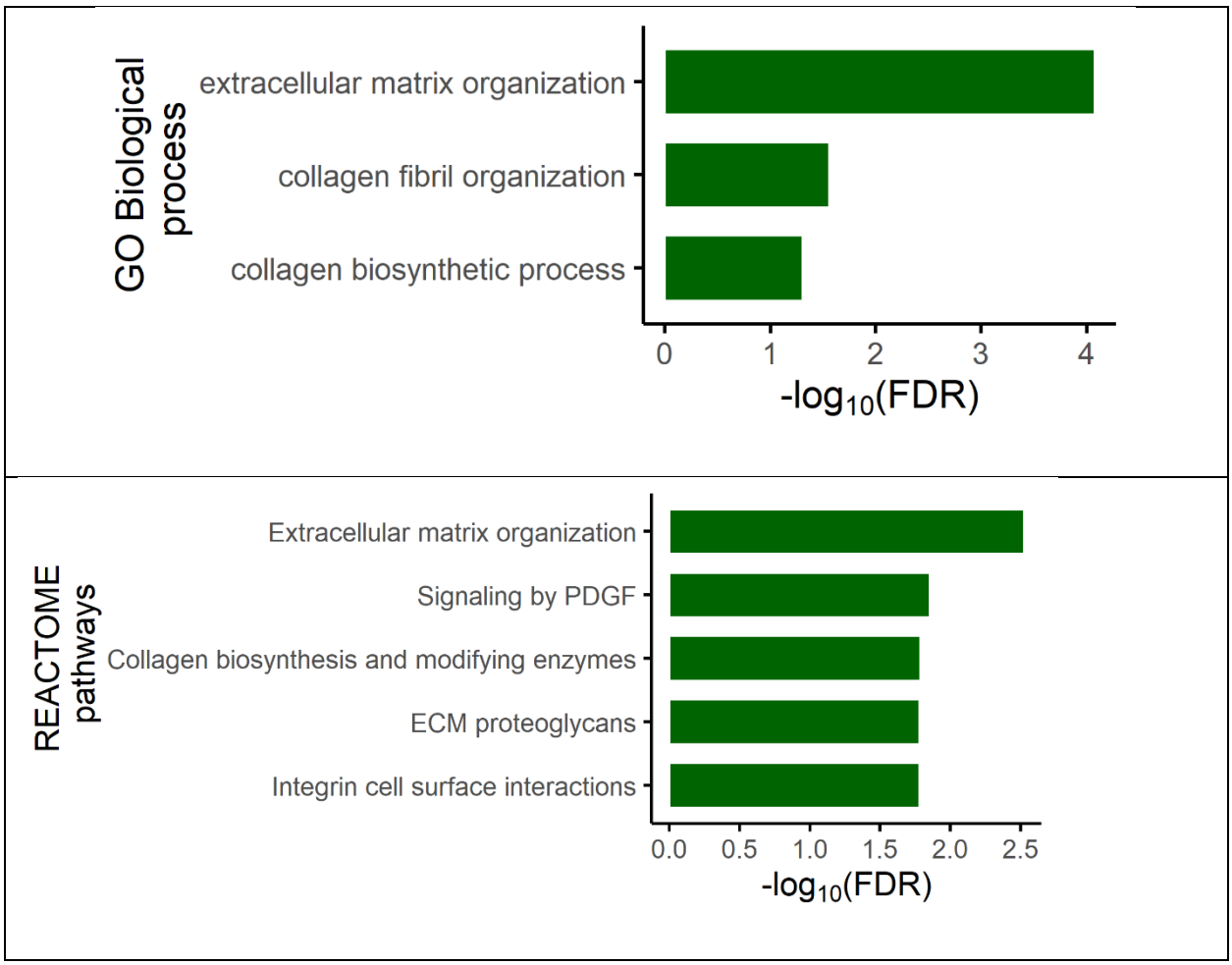


Figure 6-5. Pathway enrichments for the 36 gene NRF2 signature

Top: *Gene Ontology biological processes are significantly enriched for ECM and collagen structure.* ***Lower:*** *REACTOME pathways enriched are primarily involved in ECM processes.*

The extracellular matrix, or matrisome, is a physiologically active component of the cellular microenvironment where a complex network of macromolecules are assembled into three dimensional supramolecular structures, providing distinctive biomechanical and biochemical properties that are responsible for cell – cell communication, adhesion, motility and cell proliferation [233,249]. The primary components are collagen for architecture, cross linking proteins such as laminin, proteoglycans as functional modifiers, signalling proteins such as osteopontin and metalloproteinases [249,250]. Recent data has pointed towards a significant role for NRF2 signalling in controlling the Extracellular Matrix modelling [251]. The activation of NRF2 signalling reprogrammed fibroblasts by direct targeting of *COL3A1*, *COL5A1*, *SPP1*, *SERPINE1*

and *SERPINI1* as well as *ANGPTl2* to trigger cellular senescence which lead to increased wound healing but also skin tumorigenesis in *in-vivo* models. Three of these genes are part of the current NRF2 biomarker, while related family of genes are highly expressed in the signature (Serpin family – *SERPINH1*, A Disintegrin and Metalloproteinase Domain family – *ADAM12*, Angiopoietin family – *ANGPT1*) [Figure 6-1]. CMS 4 was also mildly enriched for 'Wound response signature' (1.79×10^{-2}) which would resonate with the above findings. The extracellular matrix has been shown to influence many of the cancer hallmarks [233]. Integrins, a component of the ECM can regulate apoptotic response to DNA damage through modulation of p53 [252] , while also inducing a malignant phenotype on the epithelium [253].

Although further work is needed to clarify the exact role of NRF2 signalling in mediating therapeutic resistance in this population, and how NRF2 signalling interacts with other known factors, the data supports the in-vitro work and confirms a role for NRF2 in being one mechanism driving radioresistance in rectal cancer. Further validation in the ARISTOTLE trial will provide further supportive evidence in the context of randomised controlled trial data. It will also provide an opportunity to explore the relationship between NRF2 expression, NAR score and longer term outcomes of DFS which was not possible in the Aberdeen cohort.

Acknowledgement of contribution: I am grateful to Dr Chieh-Hsi Wu for her assistance with the ordinal regression analysis.

CHAPTER 7

7.1 General summary and conclusions

The principal aim of this thesis was to try and improve outcomes for locally advanced resistant CRC through a better understanding of novel factors that mediate resistance and contribute to poor prognostic outcomes. I sought to achieve this aim through four strands of work.

In chapter 3 I have demonstrated that silencing of NRF2 signalling in two resistant colorectal cell line models produces increased radiosensitivity. This is the first data supporting a role for NRF2 in mediating radioresistance in CRC. The degree of enhancement on silencing NRF2 was modest but consistent across resistant cell lines. Enhancement was also only seen in cell line models which possess mutant *TP53*. The exact nature of this interaction was not characterised but previous work has suggested that TP53 hotspot mutations preferentially bind to the NRF2 target NQO-1, avoiding ubiquitination and degradation [190]. Given the prevalence of p53 mutations in CRC this mechanism requires further study.

Knockdown of NRF2 in HCT116 did not alter cell survival after radiation. Already a sensitive cell line, it appeared to have lower levels of NRF2 pathway expression compared to the other cell lines. This had been suggested in Yard et al's prior work, although not specifically in CRC setting [94]. The presence of an activating NFE2L2 mutation in a crucial KEAP1 binding domain in a CRISPR isogenic HCT116 model induced radioresistance, as well as chemoresistance. This supports the hypothesis that it is mainly in the context of overexpression that NRF2 signalling is clinically important.

An appropriate criticism of this work is that the panel of colorectal lines, three, is relatively small to accurately elucidate the relevance of NRF2 in mediating radioresistance. A larger panel from across the colorectal spectrum of phenotypes would be better. In that way one could understand the relevance of NRF2 to the main clinical phenotypes of colorectal cancer – MSI and MSS. The absence of radiosensitization and the apparently low levels of NRF2 signalling in HCT116 may be a feature shared by all MSI cell lines.

Having confirmed the relevance of NRF2 signalling to radioresistance in colorectal models, further laboratory work is needed to understand the phenotypic changes induced by interference of NRF2. Classically, NRF2 is known as the master regulator of oxidation and reduction within the cell. Assaying of the knockdown models for generation of reactive oxygen species (ROS) would be useful given the phenotypic difference, for instance to understand if there is a difference in ROS between SW1463, HT55 and HCT116. ROS contributes to DNA damage [18] and excessive production in the absence of NRF2 could have contributed to an increase in radiosensitization in the former but not the latter, although there is little change in redox signalling at a transcriptomic level.

With regard to the isogenic HCT116 CRISPR model, although the work herein has confirmed the presence of the heterozygous mutation at a DNA level, further work is warranted to better characterise the functional consequences of the mutation. The nature of the mutation in the crucial KEAP1 E_{79}TGE binding motif suggests weak or absent binding of NRF2. A co-immunoprecipitation experiment of the KEAP1-NRF2 complex, relative to the parental wild type line, would confirm this. NRF2 protein half-life is also theoretically increased by the mutation suggesting that further phenotypic screening using a pulse chase analysis or cycloheximide blocking to quantify this may be useful. However, the absence of a robust and reproducible antibody to NRF2 may present some challenges.

In Chapter 4, I explored the possible mechanisms of NRF2 mediated therapeutic resistance through whole transcriptomic profiling with RNA sequencing. Perturbations in NRF2 signalling resulted in differential expression of large numbers of genes; 2562 in the SW1463 RNAi model and 651 in the HCT116 CRISPR isogenic model. The convergence of the data suggests that NRF2 regulates a complex metabolic switch, possibly through co-regulation of ASNS, in control of amino acid metabolism, cell proliferation and nucleotide synthesis. Although some of these

aspects have been demonstrated in other cancer types such as lung [174,199], this has not been explored in detail in CRC models.

The proposed model therefore warrants further confirmatory *in-vitro* work. The significant changes in ASNS transcription should be confirmed at a protein level in the first instance, under the conditions described for the RNA sequencing experiment. The high correlation between changes in NRF2 expression and ASNS expression suggests that NRF2 may have a direct role in regulation of ASNS. But ATF4 is the accepted transcription factor responsible for ASNS regulation [202]. Further *in-vitro* work similar to that described above to establish the nature of the protein-protein interactions between ASNS, ATF4 and NRF2 in colorectal cancer would provide context and relevance of the proposed model. Although evidence in support of this exists in lung cancer [199], this understanding would allow CRC researchers to adopt treatment paradigms from more classically NRF2 associated cancers [174]. Confirmation of changes to the metabolic phenotypes, particularly in relation to asparagine, glutamine and glutamate, would be necessary before progressing to *in-vivo* murine work.

Metabolic dysregulation is garnering interest for CRC researchers but pinpointing a mechanism remains elusive due to the dynamic, interconnected nature of metabolic pathways. There are currently two phase I/II trials in CRC (NCT03263429; NCT02861300) adopting glutaminase inhibition as a strategy in CRC. The success of these two trials will depend on how enriched, intended or otherwise, the studies are for patients that are likely to respond to such interventions, and the chosen trial endpoints as discussed in chapters 1 and 6.

In Chapter 5, I derived a signature for aggregating NRF2 signalling in-vivo. For the first time in CRC, known NRF2 regulated genes were shown to be consistently expressed in a co-ordinated pattern. The NRF2 signature was reliably detected across four different validation sets, from three different mRNA platforms from both FFPE and fresh frozen tissue samples indicating

remarkable robustness. Also for the first time, using this signature, the prognostic effects of NRF2 in CRC datasets have been accurately characterised. As a prognostic biomarker it offers additional value to all other known prognostic biomarkers that were available to study in the 1,360 patients studied. Although modest in hazard ratio effect size, with hazard ratios between 1.14 – 1.55, the consistent nature of the effect supports a genuine result.

Alternatively, one could view the presented data as being of statistical significance but of low clinical or biological significance, which is entirely plausible. The low effect sizes may not contribute sufficiently to alter clinical decision making even if the signature were available as a clinical grade biomarker. It is logical that any biological effect of NRF2 is already encapsulated in a global analysis such as CMS. The data herein demonstrates enrichment of the NRF2 signature in CMS 4. A further refinement of CMS 4 in NRF2 driven or associated tumours may not offer any further clinical insights as these cancers have already been marked out as a poor prognostic group.

But HR effect size in isolation should not be a reason to further investigate the prognostic role of NRF2 in CRC. Recall that the HR effect sizes of CMS 4, compared to CMS 1, 2, 3 respectively, are 1.55 (95% C.I 1.19 – 2.01), 1.94 (95% C.I 1.58 – 2.36) and 1.72 (95% C.I 1.27 – 2.23).

In Chapter 6, the derived NRF2 signature was used to validate the *in-vitro* findings of increased radioresistance in the setting of overexpressed NRF2. High NRF2 signature expression corresponded to worse tumour regression as measured by the NAR score. The odds ratio of 1.33 was, again, modest but resonates with the modest *in-vitro* radiosensitization seen in the cell line models. Importantly it offers additional information to baseline clinical factors. The metric of response, the NAR score was chosen because it is a validated surrogate for late outcomes, as well as being the primary endpoint for NRG GI002 platform phase II trial in rectal cancer [40]. Although it did not truly dissociate responders from non-responders like a predictive biomarker

would ideally be expected to do, it does offer some potentially clinical utility. A patient with a high NRF2 signature score could be entered into a novel clinical trial with a view to improving response and lowering their NAR score.

Given the modest effect size of the OR, it is reasonable to consider whether or not NRF2 is worth pursuing as a biomarker of radioresistance. It present form it cannot accurately distinguish between a truly radioresistant patient and other groups. Furthermore its enrichment in CMS 4, again, could argue against further interrogation given that this group of tumours are already felt to have worse response to neoadjuvant chemoradiotherapy. The stromal rich, immune excluded CMS 4 classifier creates an unfavourable microenvironment for treatment response [103]. In order to simply enrich trials for patients who are likely to respond poorly to standard of care treatments why not simply assess for the presence of CMS 4. This is especially true in the absence of a therapeutic intervention in tumours that are felt to be NRF2 driven.

It must be considered that a biomarker of true 'non-response' is not possible in radiotherapy. Even in the context of poor macroscopic response there may still be a microscopic response to therapy and thus longer term benefit. Rectal cancer is an ideal illustration of this conundrum, where 20% of people appear not to respond by tumour regression grading but <5% have a local recurrence following pre-operative or post-operative radiotherapy. If the 20% of patients whose cancers do not shrink with treatment were true 'non-responders' then one would expect the local recurrence rate to mirror that of surgery alone which was approximately 25% [254]. As a discipline, lower GI radiation oncologists may have to accept that biomarkers of 'worse regression' can be clinically useful if deployed correctly.

7.2 Areas for future investigation of NRF2

The role of NRF2 in radiosensitizing normal tissues was not explored *in-vitro*. Such experiments would be useful in determining the potential therapeutic index of targeting NRF2 as a treatment strategy. Although constitutive activation of NRF2 is not present in normal tissues, inhibition of NRF2 may preferentially affect certain tissue types that are more dependent on NRF2 as a defence mechanism. Downregulation of NRF2 in the wild type HCT116 model by RNAi did not increase radiosensitization, presumably due to low levels of pathway activation, so repeated experiments in normal tissue cell lines such as MRC5 and HFL-1 would be instructive.

Although robust and reproducible, the current NRF2 signature was derived from whole transcriptome mRNA data profiling. From a prognostic and radioresistant point of view it appears to be the highest expressing tertile from each cohort that have negative outcomes. To be useful on an individual basis a cut-off threshold would need to be determined in order to decide which patients have activated levels of NRF2 signalling.

The small number of genes (36) that comprise the signature make it potentially useful as a diagnostic tool, especially as the signal was detected in both biopsy samples and whole tissue sections. If the signal could be generated from qPCR, with appropriate normalisation to housekeeping genes in a similar fashion to the OncoType Dx signature in breast cancer [50], then it would have potential as a risk adaption tool.

There may also be a smaller number of genes than the 36 herein that provide as much information. The principal component analysis indicated that in each cohort the most informative genes, those with the greatest correlation with PC1, are about 10 in number. It may therefore be

possible to derive an immunohistochemistry based score of NRF2 expression using these few genes. This would shorten the number of steps required to bring it to clinical use as such scoring could be done from FFPE sample taken as part of routine clinical care. Such a project, though exciting in prospect, would need significantly sized training and validation sets. It may be possible in the context of the S:CORT data collaboration where the Aberdeen cohort described in Chapter 6 would be complimented by the ARISTOTLE trial (ISRCTN09351447) control arm in due course.

The enrichment of NRF2 signalling in CMS 4 is an entirely new biological insight into the transcriptomic profile of CRC. This most likely reflects the enrichment of CMS 4 for the effect of extracellular matrix (ECM) signalling, noted in the original CMS work as 'REACTOME matrix modelling' [84].

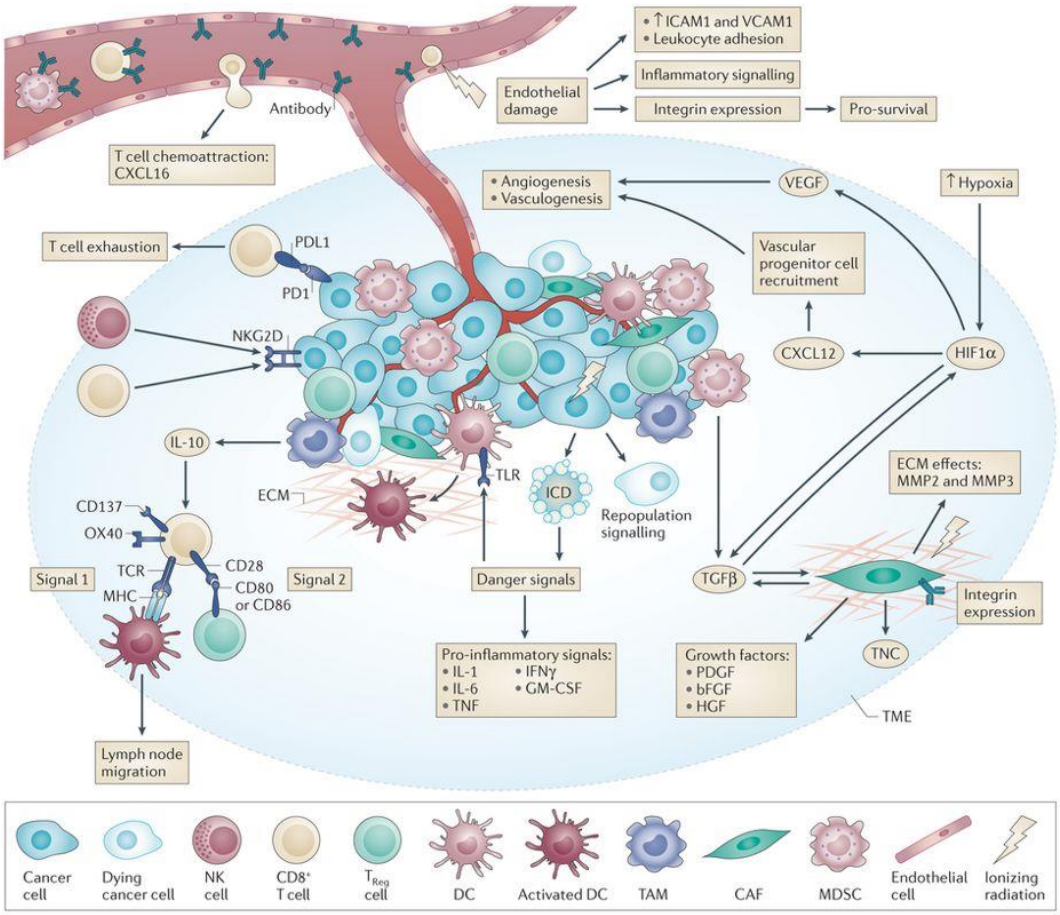


Figure 7-1. The tumour microenvironment

The interface between the cancer cell, fibroblasts, vasculature, the extracellular matrix and the immune system is intensely complex and dynamic. Most of those facets have demonstrated a role in influencing cancer cell biology and responses to radiation. [Reproduced with permission from Springer Nature under licence 4644371095137 from Barker et al [255]]

Classical radiobiology and cancer biology, for many years, focussed on the intrinsic effects within the tumour cell. The focus has widened in the last number of years to appreciate the complicated and dynamic role of the tumour microenvironment in determining a cancer cell's fate and indeed its response to radiation [Figure 7-1] [255]. NRF2's part in this conversation has yet to be fully understood but its ability to influence the ECM seems a logical starting point. Open questions include whether the cancer cell exerts an outward influence on the ECM by increasing NRF2 signalling and creating a pro-tumorigenic environment, or is it an inward effect, with the ECM altering the biology of the cancer cell. Recent proteogenomic work from squamous cell lung cancer offers interesting insights [256]. Somatic mutations of *NFE2L2/KEAP1*

are enriched for transcriptional programs in ECM and glutathione metabolism – similar to the data presented here - and represent a distinct subgroup from 'inflamed' immune rich tumours. Stewart et al suggest that high levels of ECM deposition could prevent T cell migration, in line with previous data from lung cancer [257], and thus altering the microenvironment. The poor prognosis and poor radiation response associated with these biological effects highlights the need to fully understand the complex interaction.

The inherent novelty of the work herein implies more elaboration is required before being translated into a clinical setting. Difficulties targeting NRF2 with direct, specific inhibitors [113] will need to be overcome or novel strategies developed by a detailed understanding of the regulation of NRF2. Very recent data has suggested that in order to be active and promote transcription, NRF2 must be de-glycated by Fructosamine-3-kinase (FN3K) – a protein kinase [258]. Sanghvi et al suggest this may present an opportunity for therapeutic intervention. Other attractive future strategies could include using an NRF2 signature as a biomarker for enriching studies of metabolic interventions such as glutaminase inhibitors[174] or targeting the ECM [259]. Without such biomarkers, any attempts would have a higher likelihood of faltering.

7.3 Future strategies for treating resistant rectal cancer

Two recent high profile consensus statements on novel drug radiotherapy combinations [260,261] highlighted five key aims: 1) the potential of combinations to improve outcomes; 2) the importance of communication between industry, academia, regulatory agencies and patient advocates; 3) an intelligent trial design; 4) the use of validated endpoints and 5) novel approaches including immune-oncology combined with radiotherapy should be prioritised. Several of these have already been elaborated on throughout the thesis. Previous attempts to

improve the outcomes for resistant rectal cancer were summarised in Chapter 1 and many failed because they did not address these points in detail.

The UK wide ARTEMIS (Augmenting RadioTherapy in REctal Cancer to Minimise Invasive Surgery) is a multi-arm multistage design phase II trial proposal asking a scientific question about intensified pre-operative therapy to increase the rate of organ preservation in rectal cancer. It aims to address the shortcomings of previous trials and create a platform for the rapid evaluation of novel therapies in the neoadjuvant setting.

In the US the NRG GI-002 phase II trial provides a platform the evaluation novel drug radiotherapy combinations. These trials highlight a global shift in perspective in the management of neoadjuvant rectal cancer management to avoid the repeated failures of the past and accommodate the key recommendations from these two position papers. Coupled with strong translational underpinnings this work will hopefully reveal the true nature of radiation responses in rectal cancer.

7.4 Conclusions

Locally advanced, resistant rectal cancer is a fascinating and challenging disease. The diverse tumour microenvironment responsible for the phenomenon must be reflected in an equally diverse multi-disciplinary community of clinicians, biologist, immunologists, and statisticians and of course patients, willing to tackle this problem proactively.

REFERENCES

1. Bowel cancer incidence statistics. *Cancer Research UK* https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/incidence (2015).
2. Krook, J. E. *et al.* Effective Surgical Adjuvant Therapy for High-Risk Rectal Carcinoma. *N. Engl. J. Med.* **324**, 709–715 (1991).
3. Sauer, R. *et al.* Preoperative versus Postoperative Chemoradiotherapy for Rectal Cancer. *N. Engl. J. Med.* **351**, 1731–1740 (2004).
4. Ringborg, U. *et al.* The Swedish Council on Technology Assessment in Health Care (SBU) systematic overview of radiotherapy for cancer including a prospective survey of radiotherapy practice in Sweden 2001--summary and conclusions. *Acta Oncol. Stockh. Swed.* **42**, 357–365 (2003).
5. Branzei, D. & Foiani, M. Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Biol* **9**, 297–308 (2008).
6. Higgins, G. S., O'Cathail, S. M., Muschel, R. J. & McKenna, W. G. Drug radiotherapy combinations: review of previous failures and reasons for future optimism. *Cancer Treat. Rev.* **41**, 105–113 (2015).
7. Bussink, J., van der Kogel, A. J. & Kaanders, J. H. Activation of the PI3-K/AKT pathway and implications for radioresistance mechanisms in head and neck cancer. *Lancet Oncol.* **9**, 288–296 (2008).
8. Bonner, J. A. *et al.* Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. *N Engl J Med* **354**, 567–78 (2006).
9. Huang, J., Zhang, J., Shi, C., Liu, L. & Wei, Y. Survival, recurrence and toxicity of HNSCC in comparison of a radiotherapy combination with cisplatin versus cetuximab: a meta-analysis. *BMC Cancer* **16**, (2016).

10. Ang, K. K. *et al.* Randomized Phase III Trial of Concurrent Accelerated Radiation Plus Cisplatin With or Without Cetuximab for Stage III to IV Head and Neck Carcinoma: RTOG 0522. *J. Clin. Oncol.* **32**, 2940–2950 (2014).
11. Gillison, M. L. *et al.* Radiotherapy plus cetuximab or cisplatin in human papillomavirus-positive oropharyngeal cancer (NRG Oncology RTOG 1016): a randomised, multicentre, non-inferiority trial. *The Lancet* **393**, 40–50 (2019).
12. Mehanna, H. *et al.* Radiotherapy plus cisplatin or cetuximab in low-risk human papillomavirus-positive oropharyngeal cancer (De-ESCALaTE HPV): an open-label randomised controlled phase 3 trial. *The Lancet* **393**, 51–60 (2019).
13. Bradley, J. D. *et al.* Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage IIIA or IIIB non-small-cell lung cancer (RTOG 0617): a randomised, two-by-two factorial phase 3 study. *Lancet Oncol.* **16**, 187–199 (2015).
14. Suntharalingam, M. *et al.* Effect of the Addition of Cetuximab to Paclitaxel, Cisplatin, and Radiation Therapy for Patients With Esophageal Cancer: The NRG Oncology RTOG 0436 Phase 3 Randomized Clinical Trial. *JAMA Oncol.* **3**, 1520–1528 (2017).
15. Hall, W. A. *et al.* Precision Oncology and Genomically Guided Radiation Therapy: A Report From the American Society for Radiation Oncology/American Association of Physicists in Medicine/National Cancer Institute Precision Medicine Conference. *Int. J. Radiat. Oncol.* **101**, 274–284 (2018).
16. Scott, J. G. *et al.* A genome-based model for adjusting radiotherapy dose (GARD): a retrospective, cohort-based study. *Lancet Oncol.* **0**, (2016).
17. Eschrich, S. *et al.* Systems Biology Modeling of the Radiation Sensitivity Network: A Biomarker Discovery Platform. *Int. J. Radiat. Oncol. • Biol. • Phys.* **75**, 497–505 (2009).

18. Hall, E.J. Radiobiology for the Radiologist, 7E (2012) [PDF] [UnitedVRG]. *Scribd* https://www.scribd.com/doc/252571940/Radiobiology-for-the-Radiologist-7E-2012-PDF-UnitedVRG.
19. Rectal Cancer: ESMO Clinical Practice Guidelines | ESMO. http://www.esmo.org/Guidelines/Gastrointestinal-Cancers/Rectal-Cancer.
20. Bosset, J.-F. *et al.* Chemotherapy with Preoperative Radiotherapy in Rectal Cancer. *N. Engl. J. Med.* **355**, 1114–1123 (2006).
21. Fokas, E. *et al.* Tumor Regression Grading After Preoperative Chemoradiotherapy as a Prognostic Factor and Individual-Level Surrogate for Disease-Free Survival in Rectal Cancer. *JNCI J. Natl. Cancer Inst.* **109**, (2017).
22. Tiernan, J. *et al.* Use of a modified Delphi approach to develop research priorities for the association of coloproctology of Great Britain and Ireland. *Colorectal Dis. Off. J. Assoc. Coloproctology G. B. Irel.* **16**, 965–970 (2014).
23. Creavin, B. *et al.* Organ preservation with local excision or active surveillance following chemoradiotherapy for rectal cancer. *Br. J. Cancer* **116**, 169–174 (2017).
24. Habr-Gama, A. *et al.* Watch and wait approach following extended neoadjuvant chemoradiation for distal rectal cancer: are we getting closer to anal cancer management? *Dis. Colon Rectum* **56**, 1109–1117 (2013).
25. Iles, S. M. *et al.* Irinotecan+5-fluorouracil with concomitant pre-operative radiotherapy in locally advanced non-resectable rectal cancer: a phase I/II study. *Br. J. Cancer* **98**, 1210–1216 (2008).
26. Mehta, V. K. *et al.* Phase II trial of preoperative 3D conformal radiotherapy, protracted venous infusion 5-fluorouracil, and weekly CPT-11, followed by surgery for ultrasound-staged T3 rectal cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **55**, 132–137 (2003).

27. Mohiuddin, M. *et al.* Randomized phase II study of neoadjuvant combined-modality chemoradiation for distal rectal cancer: Radiation Therapy Oncology Group Trial 0012. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **24**, 650–655 (2006).
28. Sato, T. *et al.* A Phase II trial of neoadjuvant preoperative chemoradiotherapy with S-1 plus irinotecan and radiation in patients with locally advanced rectal cancer: clinical feasibility and response rate. *Int. J. Radiat. Oncol. Biol. Phys.* **79**, 677–683 (2011).
29. Aschele, C. *et al.* Primary tumor response to preoperative chemoradiation with or without oxaliplatin in locally advanced rectal cancer: pathologic results of the STAR-01 randomized phase III trial. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **29**, 2773–2780 (2011).
30. Rödel, C. *et al.* Oxaliplatin added to fluorouracil-based preoperative chemoradiotherapy and postoperative chemotherapy of locally advanced rectal cancer (the German CAO/ARO/AIO-04 study): final results of the multicentre, open-label, randomised, phase 3 trial. *Lancet Oncol.* **16**, 979–989 (2015).
31. O'Connell, M. J. *et al.* Capecitabine and Oxaliplatin in the Preoperative Multimodality Treatment of Rectal Cancer: Surgical End Points From National Surgical Adjuvant Breast and Bowel Project Trial R-04. *J. Clin. Oncol.* **32**, 1927–1934 (2014).
32. Gérard, J.-P. *et al.* Comparison of two neoadjuvant chemoradiotherapy regimens for locally advanced rectal cancer: results of the phase III trial ACCORD 12/0405-Prodige 2. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **28**, 1638–1644 (2010).
33. Schmoll, H. J. *et al.* 505PDPREOPERATIVE CHEMORADIOTHERAPY AND POSTOPERATIVE CHEMOTHERAPY WITH CAPECITABINE +/- OXALIPLATIN IN LOCALLY ADVANCED RECTAL CANCER: INTERIM ANALYSIS FOR DISEASE-FREE SURVIVAL OF PETACC 6. *Ann. Oncol.* **25**, iv170–iv171 (2014).

34. Salazar, R. *et al.* A randomized phase II study of capecitabine-based chemoradiation with or without bevacizumab in resectable locally advanced rectal cancer: clinical and biological features. *BMC Cancer* **15**, 60 (2015).

35. Dewdney, A. *et al.* Multicenter randomized phase II clinical trial comparing neoadjuvant oxaliplatin, capecitabine, and preoperative radiotherapy with or without cetuximab followed by total mesorectal excision in patients with high-risk rectal cancer (EXPERT-C). *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **30**, 1620–1627 (2012).

36. Buijsen, J. *et al.* Phase I trial of the combination of the Akt inhibitor nelfinavir and chemoradiation for locally advanced rectal cancer. *Radiother. Oncol.* **107**, 184–188 (2013).

37. O'Neil, B. H. *et al.* A Phase I Study of Bortezomib in Combination With Standard 5-Fluorouracil and External-Beam Radiation Therapy for the Treatment of Locally Advanced or Metastatic Rectal Cancer. *Clin. Colorectal Cancer* **9**, 119–125 (2010).

38. Greenhalgh, T. A., Dearman, C. & Sharma, R. A. Combination of Novel Agents with Radiotherapy to Treat Rectal Cancer. *Clin. Oncol.* **28**, 116–139 (2016).

39. Czito, B. G. *et al.* Safety and tolerability of veliparib combined with capecitabine plus radiotherapy in patients with locally advanced rectal cancer: a phase 1b study. *Lancet Gastroenterol. Hepatol.* **2**, 418–426 (2017).

40. George, T. J. *et al.* NRG-GI002: A phase II clinical trial platform using total neoadjuvant therapy (TNT) in locally advanced rectal cancer (LARC)—First experimental arm (EA) initial results. *J. Clin. Oncol.* **37**, 3505–3505 (2019).

41. Teo, M. T. W., McParland, L., Appelt, A. L. & Sebag-Montefiore, D. Phase 2 Neoadjuvant Treatment Intensification Trials in Rectal Cancer: A Systematic Review. *Int. J. Radiat. Oncol.* **100**, 146–158 (2018).

42. Petrelli, F. *et al.* Pathologic complete response and disease-free survival are not surrogate endpoints for 5-year survival in rectal cancer: an analysis of 22 randomized trials. *J. Gastrointest. Oncol.* **8**, 39–48 (2017).

43. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* **69**, 89–95 (2001).

44. Oldenhuis, C. N. A. M., Oosting, S. F., Gietema, J. A. & de Vries, E. G. E. Prognostic versus predictive value of biomarkers in oncology. *Eur. J. Cancer* **44**, 946–953 (2008).

45. Piccart-Gebhart, M. J. *et al.* Trastuzumab after Adjuvant Chemotherapy in HER2-Positive Breast Cancer. *http://dx.doi.org/10.1056/NEJMoa052306* https://www.nejm.org/doi/10.1056/NEJMoa052306?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub%3Dwww.ncbi.nlm.nih.gov (2009) doi:10.1056/NEJMoa052306.

46. Teutsch, S. M. *et al.* The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: methods of the EGAPP Working Group. *Genet. Med.* **11**, 3–14 (2009).

47. Do, H. & Dobrovic, A. Sequence Artifacts in DNA from Formalin-Fixed Tissues: Causes and Strategies for Minimization. *Clin. Chem.* **61**, 64–71 (2015).

48. Henry, N. L. & Hayes, D. F. Cancer biomarkers. *Mol. Oncol.* **6**, 140–146 (2012).

49. Simon, R. M., Paik, S. & Hayes, D. F. Use of Archived Specimens in Evaluation of Prognostic and Predictive Biomarkers. *JNCI J. Natl. Cancer Inst.* **101**, 1446–1452 (2009).

50. Paik, S. *et al.* A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. *N. Engl. J. Med.* **351**, 2817–2826 (2004).

51. Sparano, J. A. *et al.* Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. *N. Engl. J. Med.* (2018) doi:10.1056/NEJMoa1804710.

52. Allegra, C. J. *et al.* American Society of Clinical Oncology provisional clinical opinion: testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **27**, 2091–2096 (2009).

53. Petrelli, F. *et al.* Prognostic Survival Associated With Left-Sided vs Right-Sided Colon Cancer: A Systematic Review and Meta-analysis. *JAMA Oncol.* **3**, 211–219 (2017).

54. Loupakis, F. *et al.* Primary Tumor Location as a Prognostic Factor in Metastatic Colorectal Cancer. *JNCI J. Natl. Cancer Inst.* **107**, (2015).

55. Sjoquist, K. M. *et al.* Personalizing Survival Predictions in Advanced Colorectal Cancer: The ARCAD Nomogram Project. *JNCI J. Natl. Cancer Inst.* **110**, 638–648 (2018).

56. Vilar, E. & Gruber, S. B. Microsatellite instability in colorectal cancer—the stable evidence. *Nat. Rev. Clin. Oncol.* **7**, 153–162 (2010).

57. Peltomäki, P. & Vasen, H. Mutations Associated with HNPCC Predisposition — Update of ICG-HNPCC/INSiGHT Mutation Database. *Dis. Markers* **20**, 269–276 (2004).

58. Weisenberger, D. J. *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat. Genet.* **38**, 787–793 (2006).

59. Network, T. C. G. A. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).

60. Popat, S., Hubner, R. & Houlston, R. s. Systematic Review of Microsatellite Instability and Colorectal Cancer Prognosis. *J. Clin. Oncol.* **23**, 609–618 (2005).

61. Venderbosch, S. *et al.* Mismatch Repair Status and BRAF Mutation Status in Metastatic Colorectal Cancer Patients: A Pooled Analysis of the CAIRO, CAIRO2, COIN, and FOCUS Studies. *Clin. Cancer Res.* **20**, 5322–5330 (2014).

62. Early Colon Cancer: ESMO Clinical Practice Guidelines | ESMO. https://www.esmo.org/Guidelines/Gastrointestinal-Cancers/Early-Colon-Cancer.

63. Ribic, C. M. *et al.* Tumor Microsatellite-Instability Status as a Predictor of Benefit from Fluorouracil-Based Adjuvant Chemotherapy for Colon Cancer. *N. Engl. J. Med.* **349**, 247–257 (2003).

64. Le, D. T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).

65. Diaz, L. A. *et al.* Pembrolizumab therapy for microsatellite instability high (MSI-H) colorectal cancer (CRC) and non-CRC. *J. Clin. Oncol.* **35**, 3071–3071 (2017).

66. Overman, M. J. *et al.* Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *Lancet Oncol.* **18**, 1182–1191 (2017).

67. Taieb, J. *et al.* Prognostic Effect of BRAF and KRAS Mutations in Patients With Stage III Colon Cancer Treated With Leucovorin, Fluorouracil, and Oxaliplatin With or Without Cetuximab: A Post Hoc Analysis of the PETACC-8 Trial. *JAMA Oncol.* **2**, 643–653 (2016).

68. Prahallad, A. *et al.* Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature* **483**, 100–103 (2012).

69. Van Cutsem, E. *et al.* Binimetinib, Encorafenib, and Cetuximab Triplet Therapy for Patients With BRAF V600E-Mutant Metastatic Colorectal Cancer: Safety Lead-In Results From the Phase III BEACON Colorectal Cancer Study. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **37**, 1460–1469 (2019).

70. Taieb, J. *et al.* Prognostic Value of BRAF and KRAS Mutations in MSI and MSS Stage III Colon Cancer. *JNCI J. Natl. Cancer Inst.* **109**, (2016).

71. Tosi, F. *et al.* Effect of KRAS and BRAF Mutations on Survival of Metastatic Colorectal Cancer After Liver Resection: A Systematic Review and Meta-Analysis. *Clin. Colorectal Cancer* **16**, e153–e163 (2017).

72. Tol, J. *et al.* Chemotherapy, Bevacizumab, and Cetuximab in Metastatic Colorectal Cancer. *http://dx.doi.org/10.1056/NEJMoa0808268* https://www.nejm.org/doi/10.1056/NEJMoa0808268?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub%3Dwww.ncbi.nlm.nih.gov (2009) doi:10.1056/NEJMoa0808268.

73. Maughan, T. S. *et al.* Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *The Lancet* **377**, 2103–2114 (2011).

74. 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).

75. Arnold, D. *et al.* Prognostic and predictive value of primary tumour side in patients with RAS wild-type metastatic colorectal cancer treated with chemotherapy and EGFR directed antibodies in six randomized trials†. *Ann. Oncol.* **28**, 1713–1729 (2017).

76. Richman, S. D. *et al.* HER2 overexpression and amplification as a potential therapeutic target in colorectal cancer: analysis of 3256 patients enrolled in the QUASAR, FOCUS and PICCOLO colorectal cancer trials. *J. Pathol.* **238**, 562–570 (2016).

77. Yonesaka, K. *et al.* Activation of ERBB2 signaling causes resistance to the EGFR-directed therapeutic antibody cetuximab. *Sci. Transl. Med.* **3**, 99ra86 (2011).

78. Sartore-Bianchi, A. *et al.* Dual-targeted therapy with trastuzumab and lapatinib in treatment-refractory, KRAS codon 12/13 wild-type, HER2-positive metastatic colorectal cancer (HERACLES): a proof-of-concept, multicentre, open-label, phase 2 trial. *Lancet Oncol.* **17**, 738–746 (2016).

79. Hoadley, K. A. *et al.* Multi-platform analysis of 12 cancer types reveals molecular classification within and across tissues-of-origin. *Cell* **158**, 929–944 (2014).

80. Budinska, E. *et al.* Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. *J. Pathol.* **231**, 63–76 (2013).

81. Schlicker, A. *et al.* Subtypes of primary colorectal tumors correlate with response to targeted treatment in colorectal cell lines. *BMC Med. Genomics* **5**, 66 (2012).

82. Marisa, L. *et al.* Gene Expression Classification of Colon Cancer into Molecular Subtypes: Characterization, Validation, and Prognostic Value. *PLOS Med.* **10**, e1001453 (2013).

83. Sadanandam, A. *et al.* A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat. Med.* **19**, 619–625 (2013).

84. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).

85. Duldulao, M. P. *et al.* Mutations in specific codons of the KRAS oncogene are associated with variable resistance to neoadjuvant chemoradiation therapy in patients with rectal adenocarcinoma. *Ann. Surg. Oncol.* **20**, 2166–2171 (2013).

86. Kamran, S. C. *et al.* Integrative molecular characterization of resistance to neoadjuvant chemoradiation in rectal cancer. *Clin. Cancer Res.* clincanres.0908.2019 (2019) doi:10.1158/1078-0432.CCR-19-0908.

87. Hong, T. S. *et al.* Phase II Study of Proton-Based Stereotactic Body Radiation Therapy for Liver Metastases: Importance of Tumor Genotype. *JNCI J. Natl. Cancer Inst.* **109**, (2017).

88. Strom, T. *et al.* Radiosensitivity index predicts for survival with adjuvant radiation in resectable pancreatic cancer. *Radiother. Oncol.* **117**, 159–164 (2015).

89. Ahmed, K. A. *et al.* The radiosensitivity index predicts for overall survival in glioblastoma. *Oncotarget* **6**, 34414–34422 (2015).

90. Ahmed, K. A. *et al.* Radiosensitivity Differences Between Liver Metastases Based on Primary Histology Suggest Implications for Clinical Outcomes After Stereotactic Body Radiation Therapy. *Int. J. Radiat. Oncol.* **95**, 1399–1404 (2016).

91. Eschrich, S. A. *et al.* A Gene Expression Model of Intrinsic Tumor Radiosensitivity: Prediction of Response and Prognosis After Chemoradiation. *Int. J. Radiat. Oncol. • Biol. • Phys.* **75**, 489–496 (2009).

92. Poynter, L. *et al.* Network Mapping of Molecular Biomarkers Influencing Radiation Response in Rectal Cancer. *Clin. Colorectal Cancer* **18**, e210–e222 (2019).

93. Bertolini, F. *et al.* Prognostic and predictive value of baseline and posttreatment molecular marker expression in locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **68**, 1455–1461 (2007).

94. Yard, B. D. *et al.* A genetic basis for the variation in the vulnerability of cancer to DNA damage. *Nat. Commun.* **7**, 11428 (2016).

95. Massagué, J., Blain, S. W. & Lo, R. S. TGFβ Signaling in Growth Control, Cancer, and Heritable Disorders. *Cell* **103**, 295–309 (2000).

96. Huang, M.-Y. *et al.* Relationships Between SMAD3 Expression and Preoperative Fluoropyrimidine-Based Chemoradiotherapy Response in Locally Advanced Rectal Cancer Patients. *World J. Surg.* **39**, 1257–1267 (2015).

97. Siegel, P. M. & Massagué, J. Cytostatic and apoptotic actions of TGF-β in homeostasis and cancer. *Nat. Rev. Cancer* **3**, 807 (2003).

98. Jabłońska-Trypuć, A., Matejczyk, M. & Rosochacki, S. Matrix metalloproteinases (MMPs), the main extracellular matrix (ECM) enzymes in collagen degradation, as a target for anticancer drugs. *J. Enzyme Inhib. Med. Chem.* **31**, 177–183 (2016).

99. Kurt, A. *et al.* Low Mmp 9 and VEGF levels predict good oncologic outcome in mid and low rectal cancer patients with neoadjuvant chemoradiation. *BMC Clin. Pathol.* **12**, 27 (2012).

100. Jakob, C. *et al.* Thymidylate synthase, thymidine phosphorylase, dihydropyrimidine dehydrogenase expression, and histological tumour regression after 5-FU-based neo-adjuvant chemoradiotherapy in rectal cancer. *J. Pathol.* **204**, 562–568 (2004).

101. Min, B. S. *et al.* Cyclooxygenase-2 Expression in Pretreatment Biopsy as a Predictor of Tumor Responses After Preoperative Chemoradiation in Rectal Cancer. *Arch. Surg.* **143**, 1091–1097 (2008).

102. Terrazzino, S. *et al.* A haplotype of the methylenetetrahydrofolate reductase gene predicts poor tumor response in rectal cancer patients receiving preoperative chemoradiation. *Pharmacogenet. Genomics* **16**, 817–824 (2006).

103. Roelands, J. *et al.* Immunogenomic Classification of Colorectal Cancer and Therapeutic Implications. *Int. J. Mol. Sci.* **18**, (2017).

104. Taguchi, K., Motohashi, H. & Yamamoto, M. Molecular mechanisms of the Keap1–Nrf2 pathway in stress response and cancer evolution. *Genes Cells* **16**, 123–140 (2011).

105. Kobayashi, A. *et al.* Oxidative Stress Sensor Keap1 Functions as an Adaptor for Cul3-Based E3 Ligase To Regulate Proteasomal Degradation of Nrf2. *Mol. Cell. Biol.* **24**, 7130–7139 (2004).

106. Kansanen, E., Kuosmanen, S. M., Leinonen, H. & Levonen, A.-L. The Keap1-Nrf2 pathway: Mechanisms of activation and dysregulation in cancer. *Redox Biol.* **1**, 45–49 (2013).

107. Rushmore, T. H. & Pickett, C. B. Transcriptional regulation of the rat glutathione S-transferase Ya subunit gene. Characterization of a xenobiotic-responsive element controlling inducible expression by phenolic antioxidants. *J. Biol. Chem.* **265**, 14648–14653 (1990).

108. Friling, R. S., Bensimon, A., Tichauer, Y. & Daniel, V. Xenobiotic-inducible expression of murine glutathione S-transferase Ya subunit gene is controlled by an electrophile-responsive element. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 6258–6262 (1990).

109. Hirotsu, Y. *et al.* Nrf2–MafG heterodimers contribute globally to antioxidant and metabolic networks. *Nucleic Acids Res.* **40**, 10228–10239 (2012).

110. Malhotra, D. *et al.* Global mapping of binding sites for Nrf2 identifies novel targets in cell survival response through ChIP-Seq profiling and network analysis. *Nucleic Acids Res.* **38**, 5718–5734 (2010).

111. Chorley, B. N. *et al.* Identification of novel NRF2-regulated genes by ChIP-Seq: influence on retinoid X receptor alpha. *Nucleic Acids Res.* **40**, 7416–7429 (2012).

112. Rojo de la Vega, M., Chapman, E. & Zhang, D. D. NRF2 and the Hallmarks of Cancer. *Cancer Cell* **34**, 21–43 (2018).

113. Menegon, S., Columbano, A. & Giordano, S. The Dual Roles of NRF2 in Cancer. *Trends Mol. Med.* **22**, 578–593 (2016).

114. Sporn, M. B. & Liby, K. T. NRF2 and cancer: the good, the bad and the importance of context. *Nat. Rev. Cancer* **12**, 564–571 (2012).

115. Shen, G. *et al.* Chemoprevention of Familial Adenomatous Polyposis by Natural Dietary Compounds Sulforaphane and Dibenzoylmethane Alone and in Combination in ApcMin/+ Mouse. *Cancer Res.* **67**, 9937–9944 (2007).

116. Dinkova-Kostova, A. T. *et al.* Protection against UV-light-induced skin carcinogenesis in SKH-1 high-risk mice by sulforaphane-containing broccoli sprout extracts. *Cancer Lett.* **240**, 243–252 (2006).

117. Conaway, C. C. *et al.* Phenethyl Isothiocyanate and Sulforaphane and their N-Acetylcysteine Conjugates Inhibit Malignant Progression of Lung Adenomas Induced by Tobacco Carcinogens in A/J Mice. *Cancer Res.* **65**, 8548–8557 (2005).

118. Hayes, J. D., McMahon, M., Chowdhry, S. & Dinkova-Kostova, A. T. Cancer Chemoprevention Mechanisms Mediated Through the Keap1–Nrf2 Pathway. *Antioxid. Redox Signal.* **13**, 1713–1748 (2010).

119. Hu, R., Saw, C. L.-L., Yu, R. & Kong, A.-N. T. Regulation of NF-E2-Related Factor 2 Signaling for Cancer Chemoprevention: Antioxidant Coupled with Antiinflammatory. *Antioxid. Redox Signal.* **13**, 1679–1698 (2010).

120. Khor, T. O. *et al.* Increased Susceptibility of Nrf2 Knockout Mice to Colitis-Associated Colorectal Cancer. *Cancer Prev. Res. Phila. Pa* **1**, 187–191 (2008).

121. Cloer, E. W., Goldfarb, D., Schrank, T. P., Weissman, B. E. & Major, M. B. NRF2 Activation in Cancer: From DNA to Protein. *Cancer Res.* (2019) doi:10.1158/0008-5472.CAN-18-2723.

122. Satoh, H. *et al.* Nrf2-deficiency creates a responsive microenvironment for metastasis to the lung. *Carcinogenesis* **31**, 1833–1843 (2010).

123. Sanchez-Vega, F. *et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173**, 321-337.e10 (2018).

124. Kitamura Hiroshi & Motohashi Hozumi. NRF2 addiction in cancer cells. *Cancer Sci.* **0**, (2018).

125. Ohta, T. *et al.* Loss of Keap1 Function Activates Nrf2 and Provides Advantages for Lung Cancer Cell Growth. *Cancer Res.* **68**, 1303–1309 (2008).

126. Hanada, N. *et al.* Methylation of the KEAP1 gene promoter region in human colorectal cancer. *BMC Cancer* **12**, 66 (2012).

127. DeNicola, G. M. *et al.* Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. *Nature* **475**, 106–109 (2011).

128. Levings, D. C., Wang, X., Kohlhase, D., Bell, D. A. & Slattery, M. A distinct class of antioxidant response elements is consistently activated in tumors with NRF2 mutations. *Redox Biol.* **19**, 235–249 (2018).

129. Integrated Genomic Characterization of Endometrial Carcinoma. *Nature* **497**, 67–73 (2013).

130. Zhang, P. *et al.* Loss of Kelch-Like ECH-Associated Protein 1 Function in Prostate Cancer Cells Causes Chemoresistance and Radioresistance and Promotes Tumor Growth. *Am. Assoc. Cancer Res.* **9**, 336–346 (2010).

131. Abazeed, M. E. *et al.* Integrative radiogenomic profiling of squamous cell lung cancer. *Cancer Res.* **73**, (2013).

132. Singh, A., Bodas, M., Wakabayashi, N., Bunz, F. & Biswal, S. Gain of Nrf2 Function in Non-Small-Cell Lung Cancer Cells Confers Radioresistance. *Antioxid. Redox Signal.* **13**, 1627–1637 (2010).

133. Gonzalez-Donquiles, C. *et al.* The NRF2 transcription factor plays a dual role in colorectal cancer: A systematic review. *PLOS ONE* **12**, e0177549 (2017).

134. Yang, Y. *et al.* Chemoprevention of dietary digitoflavone on colitis-associated colon tumorigenesis through inducing Nrf2 signaling pathway and inhibition of inflammation. *Mol. Cancer* **13**, 48 (2014).

135. Pandurangan, A. K., Sadagopan, S. K. A., Dharmalingam, P. & Ganapasam, S. Luteolin, a bioflavonoid inhibits Azoxymethane-induced colorectal cancer through activation of Nrf2 signaling. *Toxicol. Mech. Methods* **24**, 13–20 (2014).

136. G.T., W. *et al.* The Cinnamon-Derived Dietary Factor Cinnamic Aldehyde Activates the Nrf2-Dependent Antioxidant Response in Human Epithelial Colon Cells. *Molecules* **15**, 3338–3355 (2010).

137. Lu, M.-C. *et al.* An inhibitor of the Keap1-Nrf2 protein-protein interaction protects NCM460 colonic cells and alleviates experimental colitis. *Sci. Rep.* **6**, (2016).

138. Yoo, N. J., Kim, H. R., Kim, Y. R., An, C. H. & Lee, S. H. Somatic mutations of the KEAP1 gene in common solid cancers. *Histopathology* **60**, 943–952 (2012).

139. Ji, L., Wei, Y., Jiang, T. & Wang, S. Correlation of Nrf2, NQO1, MRP1, cmyc and p53 in colorectal cancer and their relationships to clinicopathologic features and survival. *Int. J. Clin. Exp. Pathol.* **7**, 1124–1131 (2014).

140. Kang, K. A. *et al.* Epigenetic modification of Nrf2 in 5-fluorouracil-resistant colon cancer cells: involvement of TET-dependent DNA demethylation. *Cell Death Dis.* **5**, e1183 (2014).

141. Kang, K. A. *et al.* Interaction of DNA demethylase and histone methyltransferase upregulates Nrf2 in 5-fluorouracil-resistant colon cancer cells. *Oncotarget* **7**, 40594–40620 (2016).

142. Palma, D. A. *et al.* Stereotactic ablative radiotherapy versus standard of care palliative treatment in patients with oligometastatic cancers (SABR-COMET): a randomised, phase 2, open-label trial. *The Lancet* **393**, 2051–2058 (2019).

143. Livak, K. J. & Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* **25**, 402–408 (2001).

144. Braselmann, H., Michna, A., Heß, J. & Unger, K. CFAssay: statistical analysis of the colony formation assay. *Radiat. Oncol.* **10**, 223 (2015).

145. Ritz, C., Baty, F., Streibig, J. C. & Gerhard, D. Dose-Response Analysis Using R. *PLOS ONE* **10**, e0146021 (2015).

146. Hill, J. T. *et al.* Poly Peak Parser: Method and software for identification of unknown indels using Sanger Sequencing of PCR products. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **243**, 1632–1636 (2014).

147. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).

148. http://broadinstitute.github.io/picard/. Picard Tools - By Broad Institute. http://broadinstitute.github.io/picard/.

149. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

150. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2016).

151. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).

152. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, (2014).

153. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

154. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

155. Sergushichev, A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv* 060012 (2016) doi:10.1101/060012.

156. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

157. Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P. & Minghim, R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, (2015).

158. Rhodes, D. R. *et al.* ONCOMINE: A Cancer Microarray Database and Integrated Data-Mining Platform. *Neoplasia N. Y. N* **6**, 1–6 (2004).

159. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci.* **100**, 9440–9445 (2003).

160. Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368 (2017).

161. Medina-Rivera, A. *et al.* RSAT 2015: Regulatory Sequence Analysis Tools. *Nucleic Acids Res.* **43**, W50–W56 (2015).

162. Dreos, R., Ambrosini, G., Périer, R. C. & Bucher, P. The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools. *Nucleic Acids Res.* **43**, D92–D96 (2015).

163. Abdullah, A. *et al.* Analysis of the role of Nrf2 in the expression of liver proteins in mice using two-dimensional gel-based proteomics. *Pharmacol. Rep. PR* **64**, 680–697 (2012).

164. Motohashi, H., O'Connor, T., Katsuoka, F., Engel, J. D. & Yamamoto, M. Integration and diversity of the regulatory network composed of Maf and CNC families of transcription factors. *Gene* **294**, 1–12 (2002).

165. Rushmore, T. H., Morton, M. R. & Pickett, C. B. The antioxidant responsive element. Activation by oxidative stress and identification of the DNA consensus sequence required for functional activity. *J. Biol. Chem.* **266**, 11632–11639 (1991).

166. Davis, S. & Meltzer, P. S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847 (2007).

167. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).

168. Smith, J. J. *et al.* Experimentally Derived Metastasis Gene Expression Profile Predicts Recurrence and Death in Patients With Colon Cancer. *Gastroenterology* **138**, 958 (2010).

169. Jorissen, R. N. *et al.* Metastasis-associated gene expression changes predict poor outcomes in patients with Dukes' stage B and C colorectal cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **15**, 7642–7651 (2009).

170. Seymour, M. T. *et al.* Different strategies of sequential and combination chemotherapy for patients with poor prognosis advanced colorectal cancer (MRC FOCUS): a randomised controlled trial. *The Lancet* **370**, 143–152 (2007).

171. Hu, Y. *et al.* Colorectal cancer susceptibility loci as predictive markers of rectal cancer prognosis after surgery. *Genes. Chromosomes Cancer* **57**, 140–149 (2018).

172. Buuren, S. van & Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations in R. *J. Stat. Softw.* **45**, 1–67 (2011).

173. Jaramillo, M. C. & Zhang, D. D. The emerging role of the Nrf2–Keap1 signaling pathway in cancer. *Genes Dev.* **27**, 2179–2191 (2013).

174. Romero, R. *et al. Keap1* loss promotes *Kras*-driven lung cancer and results in dependence on glutaminolysis. *Nat. Med.* **23**, 1362–1368 (2017).

175. Wang, T. *et al.* Role of Nrf2 signaling pathway in the radiation tolerance of patients with head and neck squamous cell carcinoma: an in vivo and in vitro study. *OncoTargets Ther.* **10**, 1809–1819 (2017).

176. Namani, A., Rahaman, M. M., Chen, M. & Tang, X. Gene-expression signature regulated by the KEAP1-NRF2-CUL3 axis is associated with a poor prognosis in head and neck squamous cell cancer. *BMC Cancer* **18**, 46 (2018).

177. The Cancer Genome Atlas Research Network. Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169–175 (2017).

178. Chio, I. I. C. *et al.* NRF2 Promotes Tumor Maintenance by Modulating mRNA Translation in Pancreatic Cancer. *Cell* **166**, 963–976 (2016).

179. Evans, J. P. *et al.* The Nrf2 inhibitor brusatol is a potent antitumour agent in an orthotopic mouse model of colorectal cancer. *Oncotarget* **9**, 27104–27116 (2018).

180. Mouradov, D. *et al.* Colorectal Cancer Cell Lines Are Representative Models of the Main Molecular Subtypes of Primary Cancer. *Cancer Res.* **74**, 3238–3247 (2014).

181. Kemmerer, Z. A., Ader, N. R., Mulroy, S. & Eggler, A. L. Comparison of human Nrf2 antibodies: A tale of two proteins. *Toxicol. Lett.* **238**, 83–89 (2015).

182. Mitsuishi, Y. *et al.* Nrf2 Redirects Glucose and Glutamine into Anabolic Pathways in Metabolic Reprogramming. *Cancer Cell* **22**, 66–79 (2012).

183. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* (2018) doi:10.1038/nbt.4192.

184. Chen, N., Zhang, R., Konishi, T. & Wang, J. Upregulation of NRF2 through autophagy/ERK 1/2 ameliorates ionizing radiation induced cell death of human osteosarcoma U-2 OS. *Mutat. Res. Toxicol. Environ. Mutagen.* **813**, 10–17 (2017).

185. Apopa, P. L., He, X. & Ma, Q. Phosphorylation of Nrf2 in the transcription activation domain by casein kinase 2 (CK2) is critical for the nuclear translocation and transcription activation function of Nrf2 in IMR-32 neuroblastoma cells. *J. Biochem. Mol. Toxicol.* **22**, 63–76 (2008).

186. Imhoff, B. R. & Hansen, J. M. Tert-butylhydroquinone induces mitochondrial oxidative stress causing Nrf2 activation. *Cell Biol. Toxicol.* **26**, 541–551 (2010).

187. Huppke, P. *et al.* Activating de novo mutations in NFE2L2 encoding NRF2 cause a multisystem disorder. *Nat. Commun.* **8**, 818 (2017).

188. Tsvetkov, P., Reuven, N. & Shaul, Y. Ubiquitin-independent p53 proteasomal degradation. *Cell Death Differ.* **17**, 103–108 (2010).

189. Katz, C. *et al.* Wild-type and cancer-related p53 proteins are preferentially degraded by MDM2 as dimers rather than tetramers. *Genes Dev.* **32**, 430–447 (2018).

190. Asher, G. *et al.* p53 hot-spot mutants are resistant to ubiquitin-independent degradation by increased binding to NAD(P)H:quinone oxidoreductase 1. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 15065–15070 (2003).

191. Costa-Silva, J., Domingues, D. & Lopes, F. M. RNA-Seq differential expression analysis: An extended review and a software tool. *PLOS ONE* **12**, e0190152 (2017).

192. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).

193. Krall, A. S., Xu, S., Graeber, T. G., Braas, D. & Christofk, H. R. Asparagine promotes cancer cell proliferation through use as an amino acid exchange factor. *Nat. Commun.* **7**, 11457 (2016).

194. Chantranupong, L. *et al.* The Sestrins interact with GATOR2 to negatively regulate the amino acid sensing pathway upstream of mTORC1. *Cell Rep.* **9**, 1–8 (2014).

195. Hettmer, S. *et al.* Functional genomic screening reveals asparagine dependence as a metabolic vulnerability in sarcoma. *eLife* **4**,.

196. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science* **324**, 1029–1033 (2009).

197. DeBerardinis, R. J. *et al.* Beyond aerobic glycolysis: Transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc. Natl. Acad. Sci.* **104**, 19345–19350 (2007).

198. Lee, S. B., Sellers, B. N. & DeNicola, G. M. The Regulation of NRF2 by Nutrient-Responsive Signaling and Its Role in Anabolic Cancer Metabolism. *Antioxid. Redox Signal.* **29**, 1774–1791 (2017).

199. DeNicola, G. M. *et al.* NRF2 regulates serine biosynthesis in non-small cell lung cancer. *Nat. Genet.* **47**, 1475–1481 (2015).

200. Wise, D. R. & Thompson, C. B. Glutamine Addiction: A New Therapeutic Target in Cancer. *Trends Biochem. Sci.* **35**, 427–433 (2010).

201. Possemato, R. *et al.* Functional genomics reveals serine synthesis is essential in PHGDH-amplified breast cancer. *Nature* **476**, 346–350 (2011).

202. Balasubramanian, M. N., Butterworth, E. A. & Kilberg, M. S. Asparagine synthetase: regulation by cell stress and involvement in tumor biology. *Am. J. Physiol.-Endocrinol. Metab.* **304**, E789–E799 (2013).

203. Zgheib, E. *et al.* Investigation of Nrf2, AhR and ATF4 Activation in Toxicogenomic Databases. *Front. Genet.* **9**, (2018).

204. Zhang, J. *et al.* Asparagine Plays a Critical Role in Regulating Cellular Adaptation to Glutamine Depletion. *Mol. Cell* **56**, 205–218 (2014).

205. Avramis, V. I. Asparaginases: Biochemical Pharmacology and Modes of Drug Resistance. *Anticancer Res.* **32**, 2423–2437 (2012).

206. Gwinn, D. M. *et al.* Oncogenic KRAS regulates amino acid homeostasis and asparagine biosynthesis via ATF4 and alters sensitivity to L-asparaginase. *Cancer Cell* **33**, 91-107.e6 (2018).

207. Mauro, V. P. & Edelman, G. M. The ribosome filter hypothesis. *Proc. Natl. Acad. Sci.* **99**, 12031–12036 (2002).

208. van Riggelen, J., Yetil, A. & Felsher, D. W. MYC as a regulator of ribosome biogenesis and protein synthesis. *Nat. Rev. Cancer* **10**, 301–309 (2010).

209. Whittaker, S., Martin, M. & Marais, R. All Roads Lead to the Ribosome. *Cancer Cell* **18**, 5–6 (2010).

210. Pelletier, J., Thomas, G. & Volarević, S. Ribosome biogenesis in cancer: new players and therapeutic avenues. *Nat. Rev. Cancer* **18**, 51–63 (2018).

211. Pourdehnad, M. *et al.* Myc and mTOR converge on a common node in protein synthesis control that confers synthetic lethality in Myc-driven cancers. *Proc. Natl. Acad. Sci.* **110**, 11988–11993 (2013).

212. Hu, T. *et al.* Clinicopathologic significance of CXCR4 and Nrf2 in colorectal cancer. *J. Biomed. Res.* **27**, 283–290 (2013).

213. Chibon, F. Cancer gene expression signatures – The rise and fall? *Eur. J. Cancer* **49**, 2000–2009 (2013).

214. Veer, L. J. van 't *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).

215. van de Vijver, M. J. *et al.* A Gene-Expression Signature as a Predictor of Survival in Breast Cancer. *http://dx.doi.org/10.1056/NEJMoa021967* https://www.nejm.org/doi/10.1056/NEJMoa021967?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub%3Dwww.ncbi.nlm.nih.gov (2009) doi:10.1056/NEJMoa021967.

216. Sotiriou, C. *et al.* Gene Expression Profiling in Breast Cancer: Understanding the Molecular Basis of Histologic Grade To Improve Prognosis. *JNCI J. Natl. Cancer Inst.* **98**, 262–272 (2006).

217. Carter, S. L., Eklund, A. C., Kohane, I. S., Harris, L. N. & Szallasi, Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat. Genet.* **38**, 1043–1048 (2006).

218. Lossos, I. S. *et al.* Prediction of Survival in Diffuse Large-B-Cell Lymphoma Based on the Expression of Six Genes. *http://dx.doi.org/10.1056/NEJMoa032520* https://www.nejm.org/doi/10.1056/NEJMoa032520?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub%3Dwww.ncbi.nlm.nih.gov (2009) doi:10.1056/NEJMoa032520.

219. Buffa, F. M., Harris, A. L., West, C. M. & Miller, C. J. Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. *Br. J. Cancer* **102**, 428–435 (2010).

220. Winter, S. C. *et al.* Relation of a Hypoxia Metagene Derived from Head and Neck Cancer to Prognosis of Multiple Cancers. *Cancer Res.* **67**, 3441–3449 (2007).

221. Dunne, P. D. *et al.* Cancer-cell intrinsic gene expression signatures overcome intratumoural heterogeneity bias in colorectal cancer patient classification. *Nat. Commun.* **8**, 15657 (2017).

222. Qian, Z. *et al.* Nuclear factor, erythroid 2-like 2-associated molecular signature predicts lung cancer survival. *Sci. Rep.* **5**, (2015).

223. Namani, A. *et al.* NRF2-regulated metabolic gene signature as a prognostic biomarker in non-small cell lung cancer. *Oncotarget* **5**, (2017).

224. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).

225. Gaedcke, J. *et al.* Mutated KRAS results in overexpression of DUSP4, a MAP-kinase phosphatase, and SMYD3, a histone methyltransferase, in rectal carcinomas. *Genes. Chromosomes Cancer* **49**, 1024–1034.

226. Graudens, E. *et al.* Deciphering cellular states of innate tumor drug responses. *Genome Biol.* **7**, R19 (2006).

227. Hong, Y., Downey, T., Eu, K. W., Koh, P. K. & Cheah, P. Y. A 'metastasis-prone' signature for early-stage mismatch-repair proficient sporadic colorectal cancer patients and its implications for possible therapeutics. *Clin. Exp. Metastasis* **27**, 83–90 (2010).

228. Kaiser, S. *et al.* Transcriptional recapitulation and subversion of embryonic colon development by mouse colon tumor models and human colon cancer. *Genome Biol.* **8**, R131 (2007).

229. Sabates-Bellver, J. *et al.* Transcriptome Profile of Human Colorectal Adenomas. *Mol. Cancer Res.* **5**, 1263–1275 (2007).

230. Skrzypczak, M. *et al.* Modeling Oncogenic Signaling in Colon Tumors by Multidirectional Analyses of Microarray Data Directed for Maximization of Analytical Reliability. *PLOS ONE* **5**, e13091 (2010).

231. Wang, X. J. *et al.* Oxaliplatin activates the Keap1/Nrf2 antioxidant system conferring protection against the cytotoxicity of anticancer drugs. *Free Radic. Biol. Med.* **70**, 68–77 (2014).

232. Bruno, P. M. *et al.* A subset of platinum-containing chemotherapeutic agents kills cells by inducing ribosome biogenesis stress. *Nat. Med.* **23**, 461–471 (2017).

233. Pickup, M. W., Mouw, J. K. & Weaver, V. M. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep.* **15**, 1243–1253 (2014).

234. Schafer, Z. T. *et al.* Antioxidant and oncogene rescue of metabolic defects caused by loss of matrix attachment. *Nature* **461**, 109–113 (2009).

235. Buchheit, C. L., Rayavarapu, R. R. & Schafer, Z. T. The regulation of cancer cell death and metabolism by extracellular matrix attachment. *Semin. Cell Dev. Biol.* **23**, 402–411 (2012).

236. Hermeking, H. *et al.* 14-3-3σIs a p53-Regulated Inhibitor of G2/M Progression. *Mol. Cell* **1**, 3–11 (1997).

237. Ludyga, N. *et al.* Nucleic acids from long-term preserved FFPE tissues are suitable for downstream analyses. *Virchows Arch. Int. J. Pathol.* **460**, 131–140 (2012).

238. Baker, S. G. Five criteria for using a surrogate endpoint to predict treatment effect based on data from multiple previous trials. *Stat. Med.* **37**, 507–518 (2018).

239. Prentice, R. L. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat. Med.* **8**, 431–440 (1989).

240. Martin, S. T., Heneghan, H. M. & Winter, D. C. Systematic review and meta-analysis of outcomes following pathological complete response to neoadjuvant chemoradiotherapy for rectal cancer. *BJS* **99**, 918–928 (2012).

241. Dworak, O., Keilholz, L. & Hoffmann, A. Pathological features of rectal cancer after preoperative radiochemotherapy. *Int. J. Colorectal Dis.* **12**, 19–23 (1997).

242. Mandard, A. M. *et al.* Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer* **73**, 2680–2686 (1994).

243. Rödel, C. *et al.* Prognostic significance of tumor regression after preoperative chemoradiotherapy for rectal cancer. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **23**, 8688–8696 (2005).

244. George, T. J., Allegra, C. J. & Yothers, G. Neoadjuvant Rectal (NAR) Score: a New Surrogate Endpoint in Rectal Cancer Clinical Trials. *Curr. Colorectal Cancer Rep.* **11**, 275–280 (2015).

245. Valentini, V. *et al.* Nomograms for Predicting Local Recurrence, Distant Metastases, and Overall Survival for Patients With Locally Advanced Rectal Cancer on the Basis of European Randomized Clinical Trials. *J. Clin. Oncol.* **29**, 3163–3172 (2011).

246. Fokas, E. *et al.* Neoadjuvant rectal score as individual-level surrogate for disease-free survival in rectal cancer in the CAO/ARO/AIO-04 randomized phase III trial. *Ann. Oncol.* **29**, 1521–1527 (2018).

247. van der Valk, M. J. M. *et al.* Disqualification of Neoadjuvant Rectal Score Based on Data of 6596 Patients From the Netherlands Cancer Registry. *Clin. Colorectal Cancer* **18**, e231–e236 (2019).

248. Sun, Y. *et al.* Predictive value of MRI-detected extramural vascular invasion in stage T3 rectal cancer patients before neoadjuvant chemoradiation. *Diagn. Interv. Radiol.* **24**, 128–134 (2018).

249. Frantz, C., Stewart, K. M. & Weaver, V. M. The extracellular matrix at a glance. *J. Cell Sci.* **123**, 4195–4200 (2010).

250. Walker, C., Mojares, E. & del Río Hernández, A. Role of Extracellular Matrix in Development and Cancer Progression. *Int. J. Mol. Sci.* **19**, (2018).

251. Hiebert, P. *et al.* Nrf2-Mediated Fibroblast Reprogramming Drives Cellular Senescence by Targeting the Matrisome. *Dev. Cell* **46**, 145-161.e10 (2018).

252. Lewis, J. M., Truong, T. N. & Schwartz, M. A. Integrins regulate the apoptotic response to DNA damage through modulation of p53. *Proc. Natl. Acad. Sci.* **99**, 3627–3632 (2002).

253. Weaver, V. M. *et al.* β4 integrin-dependent formation of polarized three-dimensional architecture confers resistance to apoptosis in normal and malignant mammary epithelium. *Cancer Cell* **2**, 205–216 (2002).

254. Adjuvant radiotherapy for rectal cancer: a systematic overview of 8507 patients from 22 randomised trials. *The Lancet* **358**, 1291–1304 (2001).

255. Barker, H. E., Paget, J. T. E., Khan, A. A. & Harrington, K. J. The Tumour Microenvironment after Radiotherapy: Mechanisms of Resistance and Recurrence. *Nat. Rev. Cancer* **15**, 409–425 (2015).

256. Stewart, P. A. *et al.* Proteogenomic landscape of squamous cell lung cancer. *Nat. Commun.* **10**, 1–17 (2019).

257. Salmon, H. *et al.* Matrix architecture defines the preferential localization and migration of T cells into the stroma of human lung tumors. *J. Clin. Invest.* **122**, 899–910 (2012).

258. Sanghvi, V. R. *et al.* The Oncogenic Action of NRF2 Depends on De-glycation by Fructosamine-3-Kinase. *Cell* **178**, 807-819.e21 (2019).

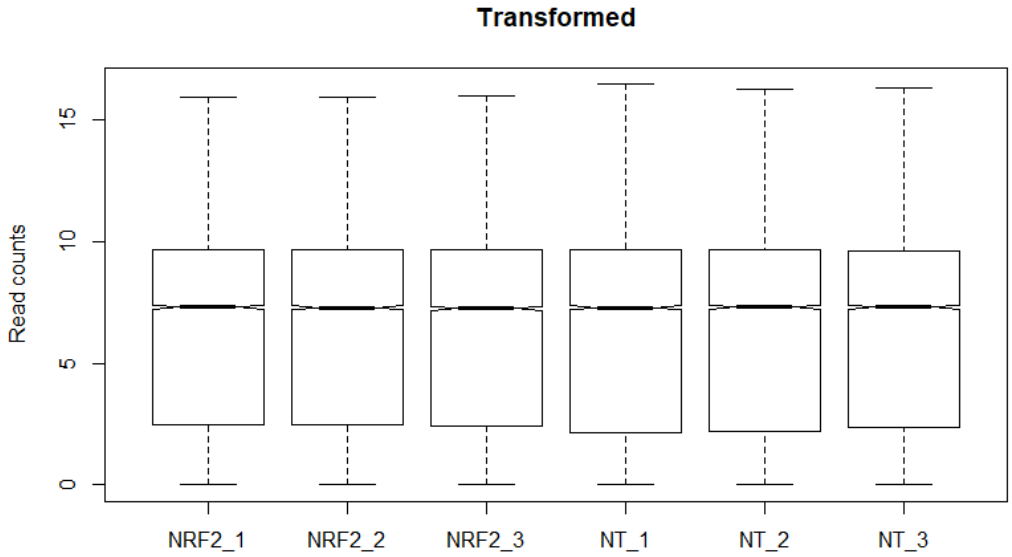
259. Reardon, D. A. *et al.* Phase II Trial of Murine 131I-Labeled Antitenascin Monoclonal Antibody 81C6 Administered Into Surgically Created Resection Cavities of Patients With Newly Diagnosed Malignant Gliomas. *J. Clin. Oncol.* **20**, 1389–1397 (2002).

260. Sharma, R. A. *et al.* Clinical development of new drug-radiotherapy combinations. *Nat. Rev. Clin. Oncol.* **13**, 627–642 (2016).

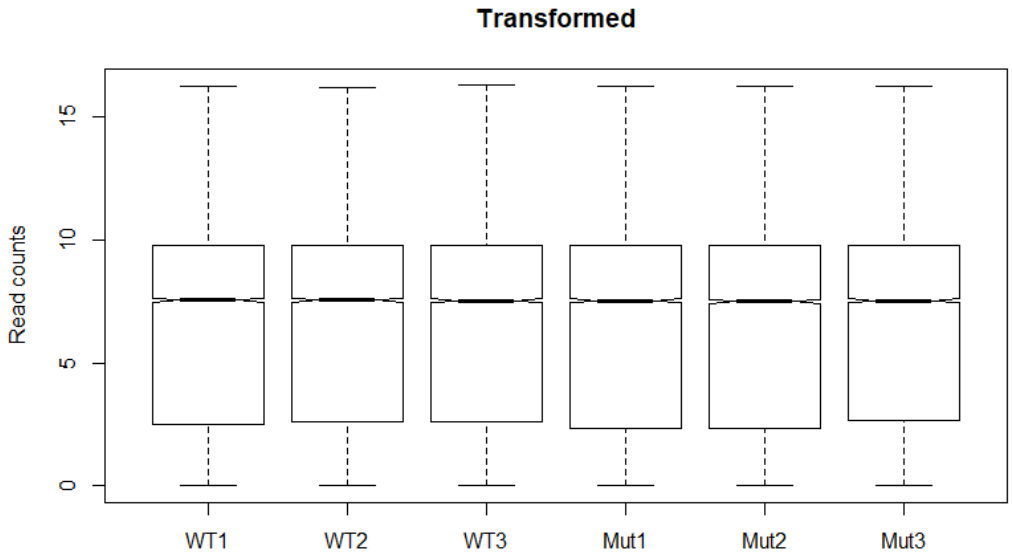
261. Ahmad, S. S. *et al.* Clinical Development of Novel Drug-Radiotherapy Combinations. *Clin. Cancer Res.* clincanres.2466.2018 (2018) doi:10.1158/1078-0432.CCR-18-2466.

APPENDIX

9.1 Appendix 1: Supplementary RNA sequencing data



Supplementary Figure 1. Boxplots for log transformed and normalised read counts from siRNA SW1463 RNA sequencing experiment



Supplementary Figure 2. Boxplots for log transformed and normalised read counts from HCT116 isogenics RNA sequencing experiment

Supplementary Table 1. List of the 216 differentially expressed genes that are shared between edgeR and DESeq2 analysis of both experimental conditions.

The fold changes and false discovery rate for both conditions from the edgeR analysis are included for illustration of the relative effects on each gene under each condition

Gene	*edgeR.logFC.NRF*	*edgeR.FDR.NRF*	*edgeR.logFC.MUT*	*edgeR.FDR.MUT*
AARS	-0.374152298	1.75E-06	0.456832032	1.01E-17
ABCC2	-0.36369221	0.001818765	-0.346982721	4.13E-05
ABLIM1	-0.355954336	0.001383153	0.332074843	6.13E-10
ACSL3	0.366529016	4.92E-06	-0.189792776	0.003709941
ACTN4	-0.265434617	0.001075692	0.206061251	0.002527408
ADGRB2	0.722294823	3.84E-10	-0.425421652	3.02E-05
ADGRG1	-0.247979863	0.003118282	0.272992544	0.001948235
ADM	-0.392049663	0.001290694	0.34477213	2.35E-06
AK2	-0.310896637	6.59E-05	0.175971894	0.004775446
ALDOC	-0.503018865	4.43E-06	-0.975433485	1.52E-42
ANGPTL4	-1.027118643	7.30E-10	-1.11638782	1.03E-13
ANXA1	-0.283711368	0.001503068	0.28238283	1.36E-05
ANXA2	-0.270240655	0.003064255	0.178503959	0.004028867
AREG	-0.670765014	5.48E-08	0.297614425	2.17E-07
ARHGAP23	0.279034147	0.004098514	-0.27708504	0.001698061
ARHGAP29	0.435777035	0.000169884	0.641132447	2.35E-37
ARL4C	0.489183622	0.000129908	-0.431039215	4.18E-09
ARRB1	0.367218144	0.001060427	-0.309364597	0.000186429
ASAH1	-0.302759992	0.000475849	-0.260247269	0.000772089
ASNS	-0.969520371	5.66E-34	1.462906147	2.72E-170
ASS1	-0.999569612	5.37E-15	0.271946702	0.002011535
ATF4	-0.480392933	1.04E-07	0.254687202	2.34E-06
ATL3	0.38892008	0.000284013	-0.239150837	0.000522185
BCL3	-0.51523043	0.000202633	-0.473139726	1.39E-06
BHLHE40	-0.34713831	6.05E-05	-0.366633576	8.90E-08
BMP4	0.340658699	0.000307822	-0.919446246	1.78E-75
BTN3A2	-0.39354628	0.000263865	-0.464379401	7.20E-07
BUD31	-0.314462047	0.006113368	0.305424782	0.001738968
CALCOCO2	-0.441962125	2.66E-08	0.285363357	1.73E-07
CAMK2D	-0.277087786	0.002887886	0.853459289	2.14E-57
CAPG	-0.388082397	0.000588274	-0.34592774	3.18E-08
CARS	-0.500683273	2.29E-11	0.4338088	5.78E-15
CDH1	-0.607665024	2.68E-10	0.463516769	1.72E-12
CDK1	0.369451906	7.95E-06	0.219396802	0.000350267
CEBPG	-0.405490772	2.79E-07	0.283978607	4.84E-06
CHAC1	-0.452343238	2.12E-06	1.178614729	3.44E-30
CIC	0.627556005	9.29E-06	-0.258434554	0.000659567
CLDN7	-1.433844224	6.73E-18	0.324374912	3.68E-08
COQ5	-0.537618235	5.13E-08	0.368557042	7.86E-06
CRABP2	0.361270562	0.000202633	0.182259666	0.005060873
CTSB	-0.320184538	0.000132979	-0.390093771	1.58E-13
CYBRD1	0.70970845	1.53E-08	-0.355513318	0.000157885
CYR61	1.315833481	6.87E-14	0.378322207	1.12E-05

DDIT4	-0.565348623	4.62E-11	0.22350306	0.00049027
DEPDC1	0.525545458	1.44E-06	0.255538458	1.02E-05
DHCR24	0.578875878	1.17E-12	-0.270300187	0.00016198
DHRS2	1.04644465	2.48E-06	0.886093143	5.82E-16
DIXDC1	0.51056776	1.37E-06	-0.338462087	0.000374697
DNMT3B	0.435487368	0.00105077	0.439158156	1.34E-06
DPYSL3	0.71381883	1.67E-15	-0.922226961	3.53E-28
EHBP1	-0.370719268	6.52E-05	0.358699868	4.35E-08
EHF	-0.447205321	2.04E-07	-0.532289208	4.35E-18
EIF2S2	-0.379484654	0.000129908	0.268383268	1.79E-05
EIF4EBP1	-0.912511418	2.94E-18	0.472021887	2.12E-12
ELF3	-0.424371958	1.68E-06	-0.327279537	1.37E-09
EMP3	-0.460079278	1.29E-05	-0.362543888	1.59E-07
EPB41L1	-0.345826373	0.000579976	-0.368649517	7.97E-10
EPRS	-0.303412992	0.000226243	0.364954952	7.77E-12
ERMP1	0.42351312	2.24E-06	0.278162821	0.00014562
ESRP1	-0.453520176	8.97E-09	0.496914014	9.57E-23
ETV4	-0.355587846	4.51E-05	0.355453226	8.62E-09
FAM111B	1.339988355	6.30E-39	-0.279544756	0.000349534
FAM83D	0.675031916	8.20E-12	0.24826894	0.000104147
FAM84B	-0.739709535	1.17E-09	0.305355948	0.000113508
FDXR	-0.353605086	0.002114438	-0.234353739	0.000546654
FGFBP1	1.229700176	3.59E-07	-1.193445043	2.15E-26
FHL2	-0.39923591	1.12E-05	0.327564079	0.000223294
FLOT1	-0.341760706	0.00051512	-0.476322536	1.13E-16
FTL	-0.299727578	0.005222283	0.183542355	0.003663785
FUS	0.42308218	1.41E-07	0.378866222	1.36E-10
GARS	-0.331079654	5.82E-05	0.607096434	1.27E-30
GATA2	0.829344997	2.63E-05	-1.150432994	8.17E-46
GDA	0.609034402	5.96E-05	-0.41093225	1.21E-07
GDF15	-0.529496715	7.15E-10	0.352072144	6.33E-07
GLO1	0.61788512	2.29E-13	0.17219098	0.008837071
GNAI1	0.634962622	9.61E-12	-0.213750738	0.000165681
GPT2	-0.360495564	3.75E-06	0.623442132	5.34E-28
GRB10	-0.343249002	0.00018054	0.304457556	1.23E-05
GSN	-0.371138178	6.00E-05	-0.218384234	0.001191945
GTF3C4	0.549197499	6.96E-07	-0.307338241	0.001294738
H1F0	-0.24134638	0.009430299	0.383145956	2.36E-14
HAS3	0.743843509	6.59E-16	0.593486693	2.38E-12
HAX1	-0.600114836	5.89E-11	0.269723667	0.000294097
HEATR5A	0.36873236	0.002649149	-0.357912702	2.03E-05
HKDC1	-0.546566031	2.88E-11	1.650047481	1.46E-12
HLA-E	-0.641500463	2.21E-09	-0.227231185	0.000431572
HNF4A	-0.240199882	0.005337979	-2.178571057	1.68E-33
HOXB5	0.813781858	4.64E-11	-0.556265421	0.001262074
HOXB6	0.395371443	0.000892029	-0.418273002	0.000743778
HOXB8	0.484759452	1.06E-07	-0.545683241	2.68E-15
HSPA8	0.815470034	6.86E-13	0.176733898	0.005388623

HYOU1	0.282821758	0.004344749	-0.226279713	0.003492772
ID1	-0.350352163	0.002618455	-0.313049561	1.33E-05
IER3	-0.401333781	3.14E-05	0.244563445	9.33E-06
IFI6	0.545786158	4.30E-05	-0.853864386	1.47E-12
IGFBP3	-0.880830207	8.00E-07	-1.920628033	4.89E-22
INADL	-0.468004283	2.91E-07	0.296849988	0.001602945
INSIG1	0.320729561	0.000446964	-0.414438087	5.55E-12
ITGA3	-0.25200831	0.003784897	-0.231281322	0.001220541
ITGB4	-0.307513944	0.000222407	-0.306346827	1.32E-06
JADE2	0.426542076	6.75E-05	-0.34917348	4.26E-07
KIF18A	0.322364875	0.00045797	-0.229610515	0.00039545
KLF5	-0.264078493	0.003159094	-0.689923566	2.41E-32
KLK6	-0.641824484	1.26E-13	0.722631932	4.76E-25
L3MBTL3	0.482254227	6.83E-05	-0.526787237	7.93E-21
LAD1	-0.282615545	0.00062496	0.258443952	0.001782622
LAMA3	-0.455153279	0.000905059	0.736238509	9.11E-09
LAMC2	-0.631273098	5.33E-14	0.598905029	2.93E-10
LGALS1	-0.428847966	6.13E-05	0.213781942	0.001426844
LPCAT1	0.355759912	1.18E-05	-0.362597363	5.02E-13
LPCAT3	0.26073734	0.002632106	-0.29121597	9.48E-06
LRP11	-0.273826207	0.002562871	-0.394822444	2.59E-05
LRRC45	0.654588283	5.28E-12	-0.307087733	2.22E-05
LSS	0.52174199	3.55E-10	-0.231218192	0.000397589
MAGI1	0.295390938	0.004055836	0.320082375	0.000127444
MAP1B	-0.249756662	0.001894687	0.587500837	2.22E-08
MAP7	-0.485123809	1.40E-07	0.347206186	6.14E-07
MARS	-0.435528111	6.76E-08	0.382826965	2.40E-12
MBNL1	0.391472078	1.39E-05	-0.160198514	0.009747253
MCC	1.390714074	2.77E-26	0.481412634	8.92E-09
MET	0.368567258	0.00017802	-0.205693153	0.003127748
MTHFD2	-0.457709938	2.74E-09	0.447393037	5.90E-15
MYL9	0.833252164	5.57E-12	-0.827358265	3.78E-39
NARS	-0.288407042	0.000133958	0.243910904	1.48E-05
NDRG1	-0.489039813	1.90E-07	-0.221495278	0.001797427
NIPSNAP1	-0.318726033	0.00085633	0.256395148	9.63E-06
NRAS	0.281574118	0.002171798	0.194341178	0.002810802
NT5E	-1.852075954	1.75E-15	0.863880457	1.23E-19
NTSR1	-0.407172764	4.50E-07	0.199232482	0.003524665
NUP155	0.604698477	4.95E-08	0.240669177	0.00384765
NUP160	0.374054497	0.000127209	0.255427534	7.25E-05
OBFC1	-0.600898525	1.40E-10	-0.292829147	0.002012851
PALLD	-0.364563605	0.000122664	-1.156832247	1.15E-103
PCK2	-0.902188914	7.44E-27	1.124120531	1.22E-57
PCTP	-0.578425568	5.80E-07	0.361725239	0.000512588
PGPEP1	-0.38463948	2.97E-05	0.351235816	0.000631352
PHGDH	-1.063617462	2.09E-31	0.646490595	5.74E-38
PHTF2	0.45668427	9.33E-07	0.281338419	0.000154337
PIK3R3	0.964117188	4.51E-18	-1.570725374	1.69E-09

PLCB3	0.432444677	3.69E-07	-0.237882662	0.000173906
PRKCDBP	-0.449339203	0.000454205	-0.44983636	0.000427989
PRRC2B	0.334316206	0.001154011	0.231314806	0.000297433
PRSS33	-1.32810301	5.45E-05	-2.304717568	2.08E-20
PSAT1	-0.501498287	3.73E-10	1.009704939	3.13E-80
PSME1	-0.457562622	4.47E-06	-0.236518013	0.000278268
PSPH	-0.795078024	2.94E-27	0.351435451	9.86E-05
PTPN1	0.42321084	3.29E-08	0.303885688	1.28E-07
PTPN14	0.885191813	4.66E-12	0.523969934	5.81E-14
PTRF	0.44622913	9.64E-09	-0.282545488	1.44E-06
PYCR1	-0.452022637	7.45E-08	0.24684034	2.22E-05
PYGB	-0.315449218	0.000275642	0.216925093	0.001841211
RAD18	0.5104188	3.15E-06	0.266462059	0.000877821
RASSF8	0.607730885	0.000270271	-0.36322458	5.43E-05
RBCK1	-0.285226526	0.001810251	0.284280389	1.94E-05
RNF145	0.287695338	0.000534192	-0.389673341	2.80E-11
RSL24D1	-0.278536984	0.001143614	0.376582245	6.83E-09
RUNX1	0.320106572	0.000143614	0.42985589	4.38E-06
SAMD9	1.016977515	3.11E-13	-0.933464589	4.88E-23
SARS	-0.52105551	3.53E-09	0.48579318	1.48E-18
SAT1	-0.465715202	1.66E-06	0.622709678	4.52E-17
SCG2	2.182943798	1.41E-49	-0.574436573	2.26E-07
SDCBP	-0.278800546	0.000818552	-0.314545541	1.81E-08
SEL1L3	-0.504200538	5.89E-11	0.24187554	5.09E-05
SEMA3C	-1.229761164	3.53E-05	1.500954073	1.48E-71
SEMA4B	-0.569249188	1.62E-12	-0.371141013	4.26E-06
SERPINB6	-0.332723452	0.001228612	-0.280963446	3.87E-05
SERPINE1	-0.467893349	3.42E-08	0.765173627	1.26E-09
SERPINH1	-0.376276665	3.20E-05	-0.334348645	4.45E-08
SESN2	-0.406602626	3.65E-06	0.552316412	2.05E-09
SESN3	1.58270947	5.89E-91	-0.601443606	1.82E-05
SGOL2	0.311906887	0.001779677	0.26835426	0.000123399
SH3BP4	0.247710113	0.002904101	0.222622443	0.000324846
SH3KBP1	-0.336901462	1.98E-05	0.250255824	1.25E-05
SHMT2	-0.562683161	5.26E-11	0.209376227	0.000456574
SKI	0.382422631	0.001706156	-0.21329745	0.002691039
SLC16A5	-0.324338728	0.001186602	-0.366244761	1.15E-06
SLC1A4	-0.797316986	7.18E-13	0.450695891	2.46E-05
SLC1A5	-0.353685858	4.03E-06	0.338917869	1.15E-07
SLC2A6	0.712941164	2.70E-05	-0.453140073	2.07E-06
SLC3A2	-0.465973115	1.36E-08	0.47036512	5.58E-20
SLC43A1	-0.495341233	8.53E-09	0.867870947	9.15E-06
SLC4A5	-0.33927104	0.002184354	0.387101642	1.92E-05
SLC7A5	-0.930367143	3.91E-31	0.516016975	2.13E-19
SMARCA1	0.488081575	1.22E-09	-0.289612546	5.19E-06
SNN	0.774785329	3.46E-08	-0.401666004	1.14E-07
SNTB1	-0.254825585	0.007672856	-0.592933758	6.77E-10
SOX9	-0.310138655	0.000157302	0.32752157	2.82E-08

SREBF1	0.26508967	0.000575596	-0.25630629	1.42E-05
SRSF6	0.290742583	0.000100687	0.16735707	0.005239783
ST14	-0.410952593	1.17E-06	0.187911382	0.002788706
STC2	-0.288941408	0.001678988	0.638471386	1.46E-25
SYDE1	0.569431318	1.28E-07	-0.565549379	7.34E-10
TAPBP	-0.392689106	6.35E-07	-0.485060979	7.06E-18
TBCD	0.417166465	3.65E-06	-0.773014137	2.88E-47
TBX3	0.446697725	0.000612255	-0.49166746	7.84E-06
TES	-0.236124247	0.008492483	0.252653781	4.87E-06
TGFBR3	-0.388414765	0.000309755	-0.57608236	2.18E-09
TMEM2	0.308550726	0.001469577	0.444281816	1.01E-06
TRIB3	-0.70246489	2.08E-14	0.946878335	1.64E-42
TRIP13	0.419775488	1.11E-07	0.217201252	0.001113544
TSEN15	-0.461863239	2.11E-08	0.462360539	6.04E-10
TSPAN5	0.415876263	4.96E-06	0.252519578	0.000199395
TUBE1	-0.708029455	3.71E-18	0.549164021	1.55E-09
TXNRD1	-0.336369334	4.70E-05	0.352636452	4.74E-10
UBE2H	-0.324179537	6.75E-05	-0.894907018	3.58E-73
UBTF	0.297745594	0.000219466	-0.187156995	0.003536891
ULBP1	-0.649636686	6.29E-10	0.591658143	1.46E-09
UNC5B	-0.545476964	2.96E-06	-0.358930994	2.14E-07
UPP1	-0.648800411	5.87E-10	0.416462952	2.07E-05
WARS	-0.247596231	0.001844683	0.504677777	1.39E-20
XPOT	-0.371006933	6.83E-06	0.359004419	3.55E-11
YARS	-0.331377143	3.82E-05	0.447267658	1.15E-18
YOD1	0.981705309	5.22E-23	0.253113284	0.001229064
ZCCHC24	0.398208514	0.000202501	-0.513880086	2.98E-07
ZNF367	1.002164938	5.08E-19	-0.316637876	1.78E-06
ZNF503	0.381664141	4.28E-06	0.304072131	0.000165644

Supplementary Table 2. List of 141 differentially expressed genes with opposing directions of effect.

The fold changes and false discovery rate for both conditions from the edgeR analysis are included for illustration of the relative effects on each gene under each condition

Gene	*edgeR.logFC.NRF*	*edgeR.FDR.NRF*	*edgeR.logFC.MUT*	*edgeR.FDR.MUT*
AARS	-0.374152298	1.75E-06	0.456832032	1.01E-17
ABLIM1	-0.355954336	0.001383153	0.332074843	6.13E-10
ACSL3	0.366529016	4.92E-06	-0.189792776	0.003709941
ACTN4	-0.265434617	0.001075692	0.206061251	0.002527408
ADGRB2	0.722294823	3.84E-10	-0.425421652	3.02E-05
ADGRG1	-0.247979863	0.003118282	0.272992544	0.001948235
ADM	-0.392049663	0.001290694	0.34477213	2.35E-06
AK2	-0.310896637	6.59E-05	0.175971894	0.004775446
ANXA1	-0.283711368	0.001503068	0.28238283	1.36E-05
ANXA2	-0.270240655	0.003064255	0.178503959	0.004028867
AREG	-0.670765014	5.48E-08	0.297614425	2.17E-07
ARHGAP23	0.279034147	0.004098514	-0.27708504	0.001698061
ARL4C	0.489183622	0.000129908	-0.431039215	4.18E-09
ARRB1	0.367218144	0.001060427	-0.309364597	0.000186429
ASNS	-0.969520371	5.66E-34	1.462906147	2.72E-170
ASS1	-0.999569612	5.37E-15	0.271946702	0.002011535
ATF4	-0.480392933	1.04E-07	0.254687202	2.34E-06
ATL3	0.38892008	0.000284013	-0.239150837	0.000522185
BMP4	0.340658699	0.000307822	-0.919446246	1.78E-75
BUD31	-0.314462047	0.006113368	0.305424782	0.001738968
CALCOCO2	-0.441962125	2.66E-08	0.285363357	1.73E-07
CAMK2D	-0.277087786	0.002887886	0.853459289	2.14E-57
CARS	-0.500683273	2.29E-11	0.4338088	5.78E-15
CDH1	-0.607665024	2.68E-10	0.463516769	1.72E-12
CEBPG	-0.405490772	2.79E-07	0.283978607	4.84E-06
CHAC1	-0.452343238	2.12E-06	1.178614729	3.44E-30
CIC	0.627556005	9.29E-06	-0.258434554	0.000659567
CLDN7	-1.433844224	6.73E-18	0.324374912	3.68E-08
COQ5	-0.537618235	5.13E-08	0.368557042	7.86E-06
CTSB	-0.320184538	0.000132979	-0.390093771	1.58E-13
CYBRD1	0.70970845	1.53E-08	-0.355513318	0.000157885
DDIT4	-0.565348623	4.62E-11	0.22350306	0.00049027
DHCR24	0.578875878	1.17E-12	-0.270300187	0.00016198
DIXDC1	0.51056776	1.37E-06	-0.338462087	0.000374697
DPYSL3	0.71381883	1.67E-15	-0.922226961	3.53E-28
EHBP1	-0.370719268	6.52E-05	0.358699868	4.35E-08
EIF2S2	-0.379484654	0.000129908	0.268383268	1.79E-05
EIF4EBP1	-0.912511418	2.94E-18	0.472021887	2.12E-12
EPRS	-0.303412992	0.000226243	0.364954952	7.77E-12
ESRP1	-0.453520176	8.97E-09	0.496914014	9.57E-23
ETV4	-0.355587846	4.51E-05	0.355453226	8.62E-09
FAM111B	1.339988355	6.30E-39	-0.279544756	0.000349534
FAM84B	-0.739709535	1.17E-09	0.305355948	0.000113508

FGFBP1	1.229700176	3.59E-07	-1.193445043	2.15E-26
FHL2	-0.39923591	1.12E-05	0.327564079	0.000223294
FTL	-0.299727578	0.005222283	0.183542355	0.003663785
GARS	-0.331079654	5.82E-05	0.607096434	1.27E-30
GATA2	0.829344997	2.63E-05	-1.150432994	8.17E-46
GDA	0.609034402	5.96E-05	-0.41093225	1.21E-07
GDF15	-0.529496715	7.15E-10	0.352072144	6.33E-07
GNAI1	0.634962622	9.61E-12	-0.213750738	0.000165681
GPT2	-0.360495564	3.75E-06	0.623442132	5.34E-28
GRB10	-0.343249002	0.00018054	0.304457556	1.23E-05
GTF3C4	0.549197499	6.96E-07	-0.307338241	0.001294738
H1F0	-0.24134638	0.009430299	0.383145956	2.36E-14
HAX1	-0.600114836	5.89E-11	0.269723667	0.000294097
HEATR5A	0.36873236	0.002649149	-0.357912702	2.03E-05
HKDC1	-0.546566031	2.88E-11	1.650047481	1.46E-12
HOXB5	0.813781858	4.64E-11	-0.556265421	0.001262074
HOXB6	0.395371443	0.000892029	-0.418273002	0.000743778
HOXB8	0.484759452	1.06E-07	-0.545683241	2.68E-15
HYOU1	0.282821758	0.004344749	-0.226279713	0.003492772
IER3	-0.401333781	3.14E-05	0.244563445	9.33E-06
IFI6	0.545786158	4.30E-05	-0.853864386	1.47E-12
INADL	-0.468004283	2.91E-07	0.296849988	0.001602945
INSIG1	0.320729561	0.000446964	-0.414438087	5.55E-12
JADE2	0.426542076	6.75E-05	-0.34917348	4.26E-07
KIF18A	0.322364875	0.00045797	-0.229610515	0.00039545
KLK6	-0.641824484	1.26E-13	0.722631932	4.76E-25
LAD1	-0.282615545	0.00062496	0.258443952	0.001782622
LAMA3	-0.455153279	0.000905059	0.736238509	9.11E-09
LAMC2	-0.631273098	5.33E-14	0.598905029	2.93E-10
LGALS1	-0.428847966	6.13E-05	0.213781942	0.001426844
LPCAT1	0.355759912	1.18E-05	-0.362597363	5.02E-13
LPCAT3	0.26073734	0.002632106	-0.29121597	9.48E-06
LRRC45	0.654588283	5.28E-12	-0.307087733	2.22E-05
LSS	0.52174199	3.55E-10	-0.231218192	0.000397589
MAP1B	-0.249756662	0.001894687	0.587500837	2.22E-08
MAP7	-0.485123809	1.40E-07	0.347206186	6.14E-07
MARS	-0.435528111	6.76E-08	0.382826965	2.40E-12
MBNL1	0.391472078	1.39E-05	-0.160198514	0.009747253
MET	0.368567258	0.00017802	-0.205693153	0.003127748
MTHFD2	-0.457709938	2.74E-09	0.447393037	5.90E-15
MYL9	0.833252164	5.57E-12	-0.827358265	3.78E-39
NARS	-0.288407042	0.000133958	0.243910904	1.48E-05
NIPSNAP1	-0.318726033	0.00085633	0.256395148	9.63E-06
NT5E	-1.852075954	1.75E-15	0.863880457	1.23E-19
NTSR1	-0.407172764	4.50E-07	0.199232482	0.003524665
PCK2	-0.902188914	7.44E-27	1.124120531	1.22E-57
PCTP	-0.578425568	5.80E-07	0.361725239	0.000512588
PGPEP1	-0.38463948	2.97E-05	0.351235816	0.000631352

PHGDH	-1.063617462	2.09E-31	0.646490595	5.74E-38
PIK3R3	0.964117188	4.51E-18	-1.570725374	1.69E-09
PLCB3	0.432444677	3.69E-07	-0.237882662	0.000173906
PSAT1	-0.501498287	3.73E-10	1.009704939	3.13E-80
PSPH	-0.795078024	2.94E-27	0.351435451	9.86E-05
PTRF	0.44622913	9.64E-09	-0.282545488	1.44E-06
PYCR1	-0.452022637	7.45E-08	0.24684034	2.22E-05
PYGB	-0.315449218	0.000275642	0.216925093	0.001841211
RSL24D1	-0.278536984	0.001143614	0.376582245	6.83E-09
SARS	-0.52105551	3.53E-09	0.48579318	1.48E-18
SAT1	-0.465715202	1.66E-06	0.622709678	4.52E-17
SCG2	2.182943798	1.41E-49	-0.574436573	2.26E-07
SEL1L3	-0.504200538	5.89E-11	0.24187554	5.09E-05
SEMA3C	-1.229761164	3.53E-05	1.500954073	1.48E-71
SERPINE1	-0.467893349	3.42E-08	0.765173627	1.26E-09
SESN2	-0.406602626	3.65E-06	0.552316412	2.05E-09
SESN3	1.58270947	5.89E-91	-0.601443606	1.82E-05
SH3KBP1	-0.336901462	1.98E-05	0.250255824	1.25E-05
SHMT2	-0.562683161	5.26E-11	0.209376227	0.000456574
SKI	0.382422631	0.001706156	-0.21329745	0.002691039
SLC1A4	-0.797316986	7.18E-13	0.450695891	2.46E-05
SLC1A5	-0.353685858	4.03E-06	0.338917869	1.15E-07
SLC2A6	0.712941164	2.70E-05	-0.453140073	2.07E-06
SLC3A2	-0.465973115	1.36E-08	0.47036512	5.58E-20
SLC43A1	-0.495341233	8.53E-09	0.867870947	9.15E-06
SLC4A5	-0.33927104	0.002184354	0.387101642	1.92E-05
SLC7A5	-0.930367143	3.91E-31	0.516016975	2.13E-19
SMARCA1	0.488081575	1.22E-09	-0.289612546	5.19E-06
SNN	0.774785329	3.46E-08	-0.401666004	1.14E-07
SOX9	-0.310138655	0.000157302	0.32752157	2.82E-08
SREBF1	0.26508967	0.000575596	-0.25630629	1.42E-05
ST14	-0.410952593	1.17E-06	0.187911382	0.002788706
STC2	-0.288941408	0.001678988	0.638471386	1.46E-25
SYDE1	0.569431318	1.28E-07	-0.565549379	7.34E-10
TBCD	0.417166465	3.65E-06	-0.773014137	2.88E-47
TBX3	0.446697725	0.000612255	-0.49166746	7.84E-06
TES	-0.236124247	0.008492483	0.252653781	4.87E-06
TRIB3	-0.70246489	2.08E-14	0.946878335	1.64E-42
TSEN15	-0.461863239	2.11E-08	0.462360539	6.04E-10
TUBE1	-0.708029455	3.71E-18	0.549164021	1.55E-09
TXNRD1	-0.336369334	4.70E-05	0.352636452	4.74E-10
UBTF	0.297745594	0.000219466	-0.187156995	0.003536891
ULBP1	-0.649636686	6.29E-10	0.591658143	1.46E-09
UPP1	-0.648800411	5.87E-10	0.416462952	2.07E-05
WARS	-0.247596231	0.001844683	0.504677777	1.39E-20
XPOT	-0.371006933	6.83E-06	0.359004419	3.55E-11
YARS	-0.331377143	3.82E-05	0.447267658	1.15E-18
ZCCHC24	0.398208514	0.000202501	-0.513880086	2.98E-07

ZNF367	1.002164938	5.08E-19	-0.316637876	1.78E-06
ZNF503	0.381664141	4.28E-06	0.304072131	0.000165644

9.2 RNA sequencing analysis code

Load packages and set working directory for files

Load text file from WTCHG with count data

```
eset <- read.table("P180572_correct_featurecounts_dedup.txt", sep = "\t", header = T,
    row.names = 1)
# Subset read count data
read.counts <- subset(eset, select = c("NRF2_1", "NRF2_2", "NRF2_3", "NT_1",
    "NT_2", "NT_3"))

orig_names <- names(read.counts)
str(read.counts)

## 'data.frame':    26577 obs. of  6 variables:
##  $ NRF2_1: int  0 211 0 0 0 0 0 0 0 0 ...
##  $ NRF2_2: int  0 206 0 0 0 0 0 0 0 0 ...
##  $ NRF2_3: int  0 136 0 0 0 0 0 0 0 0 ...
##  $ NT_1  : int  0 147 0 0 0 0 0 0 0 0 ...
##  $ NT_2  : int  0 146 0 0 0 0 0 0 0 0 ...
##  $ NT_3  : int  0 132 0 0 0 0 0 0 0 0 ...

head(read.counts)

##           NRF2_1 NRF2_2 NRF2_3 NT_1 NT_2 NT_3
## DDX11L1        0      0      0    0    0    0
## WASH7P       211    206    136  147  146  132
## MIR6859-3      0      0      0    0    0    0
## MIR6859-2      0      0      0    0    0    0
## MIR6859-1      0      0      0    0    0    0
## MIR6859-4      0      0      0    0    0    0
```

Make a data frame with meta-data where row.names should match the individual sample names

```
sample_info <- data.frame(condition = gsub("[0-9]", "", names(read.counts)),
    row.names = names(read.counts))
sample_info

##        condition
## NRF2_1      NRF_
## NRF2_2      NRF_
## NRF2_3      NRF_
## NT_1         NT_
## NT_2         NT_
## NT_3         NT_
```

Differential expression using DESeq packages

```
DESeq.ds <- DESeqDataSetFromMatrix(countData = read.counts, colData = sample_info,
    design = ~condition)
# check the result using the accessors
colData(DESeq.ds) %>% head

## DataFrame with 6 rows and 1 column
##        condition
##         <factor>
```

```
## NRF2_1      NRF_
## NRF2_2      NRF_
## NRF2_3      NRF_
## NT_1         NT_
## NT_2         NT_
## NT_3         NT_

assay(DESeq.ds) %>% head

##           NRF2_1 NRF2_2 NRF2_3 NT_1 NT_2 NT_3
## DDX11L1        0      0      0    0    0    0
## WASH7P       211    206    136  147  146  132
## MIR6859-3      0      0      0    0    0    0
## MIR6859-2      0      0      0    0    0    0
## MIR6859-1      0      0      0    0    0    0
## MIR6859-4      0      0      0    0    0    0

rowRanges(DESeq.ds) %>% head

## GRangesList object of length 6:
## $DDX11L1
## GRanges object with 0 ranges and 0 metadata columns:
##    seqnames    ranges strand
##       <Rle> <IRanges>  <Rle>
##
## $WASH7P
## GRanges object with 0 ranges and 0 metadata columns:
##      seqnames ranges strand
##
## $MIR6859-3
## GRanges object with 0 ranges and 0 metadata columns:
##      seqnames ranges strand
##
## ...
## <3 more elements>
## -------
## seqinfo: no sequences

# test what counts returns
counts(DESeq.ds) %>% str

##  int [1:26577, 1:6] 0 211 0 0 0 0 0 0 0 0 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : chr [1:26577] "DDX11L1" "WASH7P" "MIR6859-3" "MIR6859-2" ...
##   ..$ : chr [1:6] "NRF2_1" "NRF2_2" "NRF2_3" "NT_1" ...

# remove genes without any counts
DESeq.ds <- DESeq.ds[rowSums(counts(DESeq.ds)) > 0, ]

# investigate diff library sizes
colSums(counts(DESeq.ds))

##   NRF2_1   NRF2_2   NRF2_3     NT_1     NT_2     NT_3
## 16292669 16154605 14346616 13374895 17033710 14681979

colSums(read.counts)

##   NRF2_1   NRF2_2   NRF2_3     NT_1     NT_2     NT_3
## 16292669 16154605 14346616 13374895 17033710 14681979

# These should be the same!
```

Calculate the size factor and add it to the dataset - this is the default method of DESeq to normalise read counts to account for differences in sequencing depths

```
DESeq.ds <- estimateSizeFactors(DESeq.ds)
sizeFactors(DESeq.ds)

##    NRF2_1    NRF2_2    NRF2_3      NT_1      NT_2      NT_3
## 1.0922694 1.0829438 0.9308973 0.8522439 1.1125282 0.9706204

# counts () allows you to immediately retrieve the normalised read counts
counts.sf_norm <- counts(DESeq.ds, normalized = TRUE)
```

Transformation of sequencing depth normalised read counts

The transformation to log2 should be done in addition to depth normalisation

Transform size factor normalized read counts to log2 scale using a psuedocount of 1

```
log.norm.counts <- log2(counts.sf_norm + 1)
# View non transformed and transformed counts
boxplot(counts.sf_norm, notch = TRUE, main = "Untransformed", ylab = "Read counts")

boxplot(log.norm.counts, notch = TRUE, main = "Transformed", ylab = "Read counts")

# Explore read counts between replicates
plot(log.norm.counts[, 1:2], cex = 0.1, main = "Normalised log2(read counts)")

plot(log.norm.counts[, 2:3], cex = 0.1, main = "Normalised log2(read counts)")

plot(log.norm.counts[, 1:3], cex = 0.1, main = "Normalised log2(read counts)")
```

Use this package to check for homoskedasticty ie that all variables have similar variance

```
library(vsn)
library(ggplot2)
msd_plot <- meanSdPlot(log.norm.counts, ranks = FALSE, plot = FALSE)
msd_plot$gg + ggtitle("Sequencing depth normalized log2(read counts)") + ylab("standa
rd deviation") +
    theme_bw()
```

Clear hump seen on the left implies that for read counts of less than 10 the variance is higher than for those with greated counts

Use DESeq2 rlog function to reduce variance and increase stability

Obtain regularised log-transformed values

```
DESeq.rlog <- rlog(DESeq.ds, blind = TRUE)
rlog.norm.counts <- assay(DESeq.rlog)

# Mean-sd plot for rlog transformed data
```

```
msd_plot2 <- meanSdPlot(rlog.norm.counts, ranks = FALSE, plot = FALSE)
msd_plot2$gg + ggtitle("rlog-transformed(read counts)") + ylab("standard deviation")
+
    theme_bw()
```

Exploring read count patterns

```
# A dendrogram
distance.m_rlog <- as.dist(1 - cor(rlog.norm.counts, method = "pearson"))
# Plot () can directly interpret the output of the hclust
plot(hclust(distance.m_rlog), labels = colnames(rlog.norm.counts), main = "rlog trans
formed read counts\ndistance: Pearson correlation")

# PCA
pc <- prcomp(t(rlog.norm.counts))
plot(pc$x[, 1], pc$x[, 2], col = colData(DESeq.ds)[, 1], main = "PCA of seq depth nor
malized\n and rlog-transformed read counts")

# DESeq2 also offers a convenience function based on ggplot2 to do PCA
# directly on a DESeqDataSet
P <- plotPCA(DESeq.rlog)
```

Differential Gene expression analysis

DESeq2 workflow

DESeq uses the levels of the condition to determine the order of the comparison

```
str(colData(DESeq.ds)$condition)

##  Factor w/ 2 levels "NRF_","NT_": 1 1 1 2 2 2

# set NT_ (non-targeted) as the first level factor ie control
colData(DESeq.ds)$condition <- relevel(colData(DESeq.ds)$condition, "NT_")
# Run the DGE analysis
DESeq.ds <- DESeq(DESeq.ds)

## using pre-existing size factors

## estimating dispersions

## gene-wise dispersion estimates

## mean-dispersion relationship

## final dispersion estimates

## fitting model and testing

DGE.results <- results(DESeq.ds, independentFiltering = TRUE, alpha = 0.01)
summary(DGE.results)

##
## out of 20121 with nonzero total read count
## adjusted p-value < 0.01
## LFC > 0 (up)       : 1264, 6.3%
## LFC < 0 (down)     : 1353, 6.7%
## outliers [1]       : 1, 0.005%
## low counts [2]     : 6242, 31%
```

```
## (mean count < 11)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results

# The DESeqResult object can be handled like a dataframe
head(DGE.results)

## log2 fold change (MLE): condition NRF  vs NT
## Wald test p-value: condition NRF  vs NT
## DataFrame with 6 rows and 6 columns
##                      baseMean      log2FoldChange             lfcSE
##                     <numeric>           <numeric>         <numeric>
## WASH7P        161.53459664297   0.277293631389568 0.171952414072121
## LOC729737    6.33189673667095   -1.00585108559335 0.788255928022701
## MIR6723      20.1575376790078 -0.0142227097385121 0.423778872835132
## LOC100133331 85.6835902983698   0.489464019419812 0.778492947129832
## LOC100288069 7.93879928539805   0.362533515820237 0.694522609578418
## FAM87B       3.12330644168653   0.404071928970592  1.03094561320859
##                           stat            pvalue              padj
##                      <numeric>         <numeric>         <numeric>
## WASH7P          1.61261842635872 0.106827432718892 0.225160323694309
## LOC729737      -1.27604633195271 0.201939141357643                NA
## MIR6723      -0.0335616300155797 0.973226719830834 0.985537380352725
## LOC100133331   0.628732760167424 0.529524026940787 0.680245561608795
## LOC100288069   0.521989508794102 0.601677635202744                NA
## FAM87B         0.391943012117787 0.695100321327339                NA

table(DGE.results$padj < 0.01)

##
## FALSE  TRUE
## 11261  2617

head(rownames(subset(DGE.results, padj < 0.01)))

## [1] "KLHL17"   "TNFRSF18" "ACAP3"    "ATAD3B"   "ATAD3A"   "CDK11A"
```

Exploratory plots following DGE analysis

```
# Histogram of the p values
hist(DGE.results$pvalue, col = "grey", border = "white", xlab = "", ylab = "",
    main = "Frequencies of p-values")

# MA plot provides a global view of the relationship b/n expression changes
# and the conditions, average strength of the expression and the ability of
# the algorithim to detect DE ie the gens that pass the significance
# threshold (adj p val < 0.01) are colored in blue
plotMA(DGE.results, alpha = 0.01, main = "NRF2 vs NT1", ylim = c(-5, 5), colSig = "bl
ue")
```

Export list for use as a txt file

write.table(out.df, file = “DESeq2results_WT-vs-Mut.txt”, sep = “”, quote = FALSE, row.names

= FALSE)

Export list of genes with padj <0.05

EdgeR analysis

```
library(edgeR)
head(read.counts)

##           NRF2_1 NRF2_2 NRF2_3 NT_1 NT_2 NT_3
## DDX11L1        0      0      0    0    0    0
## WASH7P       211    206    136  147  146  132
## MIR6859-3      0      0      0    0    0    0
## MIR6859-2      0      0      0    0    0    0
## MIR6859-1      0      0      0    0    0    0
## MIR6859-4      0      0      0    0    0    0

# Specify sample types and make WT the reference
sample_info.edger <- factor(c(rep("NRF", 3), rep("NT", 3)))
sample_info.edger <- relevel(sample_info.edger, ref = "NT")
# Convert count matrix into an R object
edgeR.DGElist <- DGEList(counts = read.counts, group = sample_info.edger)
# Check
head(edgeR.DGElist$counts)

##           NRF2_1 NRF2_2 NRF2_3 NT_1 NT_2 NT_3
## DDX11L1        0      0      0    0    0    0
## WASH7P       211    206    136  147  146  132
## MIR6859-3      0      0      0    0    0    0
## MIR6859-2      0      0      0    0    0    0
## MIR6859-1      0      0      0    0    0    0
## MIR6859-4      0      0      0    0    0    0

edgeR.DGElist$samples

##        group lib.size norm.factors
## NRF2_1   NRF 16292669            1
## NRF2_2   NRF 16154605            1
## NRF2_3   NRF 14346616            1
## NT_1      NT 13374895            1
## NT_2      NT 17033710            1
## NT_3      NT 14681979            1
```

Remove genes with almost no coverage. To determine sensible cutoff we plot a histogram of the counts per million with cpm function

```
hist(log2(rowSums(cpm(edgeR.DGElist))))

summary(log2(rowSums(cpm(edgeR.DGElist))))

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    -Inf  -4.026   2.719    -Inf   7.723  14.820

# Focus on genes that have at least 1 cpm in at least 4 samples
keep <- rowSums(cpm(edgeR.DGElist) >= 1) >= 4
edgeR.DGElist <- edgeR.DGElist[keep, ]
# recompute the library sizes after filtering
edgeR.DGElist$samples$lib.size <- colSums(edgeR.DGElist$counts)
head(edgeR.DGElist$samples)

##        group lib.size norm.factors
## NRF2_1   NRF 16263983            1
## NRF2_2   NRF 16127280            1
```

```
## NRF2_3   NRF 14324912            1
## NT_1      NT 13356152            1
## NT_2      NT 17009852            1
## NT_3      NT 14659838            1

# Calculate the normalisation factors for the library sizes using the
# Trimmed mean method of M-values
edgeR.DGElist <- calcNormFactors(edgeR.DGElist, method = "TMM")
edgeR.DGElist$samples

##        group lib.size norm.factors
## NRF2_1   NRF 16263983    1.0181007
## NRF2_2   NRF 16127280    1.0163549
## NRF2_3   NRF 14324912    0.9915949
## NT_1      NT 13356152    0.9729863
## NT_2      NT 17009852    0.9985044
## NT_3      NT 14659838    1.0031662
```

To determine DE in a gene wise manner edgeR first estimates the disperison and subsequently tests whether the gene counts fit the negative binomial model. Following commands are only suitable for experiment with a single factor design

Specify the design setup

```
design <- model.matrix(~sample_info.edger)
edgeR.DGElist <- estimateDisp(edgeR.DGElist, design)
# fit the negative binomial model
edger_fit <- glmFit(edgeR.DGElist, design)
# Perform testing for every gene using the neg binomial model
edger_lrt <- glmLRT(edger_fit)
# Extract the results from edger_lrt
DGE.results_edgeR <- topTags(edger_lrt, n = Inf, sort.by = "PValue", adjust.method =
"BH")
head(DGE.results_edgeR$table)

##            logFC   logCPM       LR       PValue          FDR
## SESN3   1.582709 6.835509 427.9884 4.453544e-95 5.893820e-91
## NFE2L2 -1.597236 5.981219 410.0118 3.643916e-91 2.411179e-87
## THBS1   1.967704 4.005496 263.6183 2.791441e-59 1.231398e-55
## SCG2    2.182944 3.379758 235.2515 4.267028e-53 1.411746e-49
## ACSL1   1.064902 6.524933 208.5010 2.916884e-47 7.720408e-44
## CHRNA1  2.827382 2.527024 202.9614 4.716311e-46 1.040261e-42

# Export all genes with FDR rate <0.01 to txt file
```

Visualisation of results

Plot smear is similar to MA plot in DEQeq2

Summarise and pull out the DE genes at the corrected p value of 0.01

use ggplot function instead of intrinsic plotSmear as it uses existing objects in environment

Annotate top 15

```
## Convert rownames of DGE.results_edgeR$table into a column for use in
## labels
DGE.results_edgeR$table$gene <- rownames(DGE.results_edgeR$table)

library(ggrepel)
ggplot(DGE.results_edgeR$table, aes(x = logCPM, y = logFC, col = FDR < 0.01)) +
    geom_point() + theme_bw(base_size = 16) + geom_hline(yintercept = c(0.585,
    -0.585)) + geom_text_repel(data = head(DGE.results_edgeR$table, 15), aes(label =
gene),
    size = 5, col = "black")  #Add lines at FC of 1.5

# Volcano plots are used to display the measure of significance on the
# y-axis and fold change on the x-axis
ggplot(DGE.results_edgeR$table, aes(x = logFC, y = -log10(FDR), col = FDR <
    0.01)) + geom_point() + theme_bw(base_size = 16)
```

Heatmap for edgeR results

Transform the normalised read counts in log counts per million

```
y <- cpm(edgeR.DGElist, log = TRUE, prior.count = 1)
head(y)

##                  NRF2_1    NRF2_2     NRF2_3      NT_1        NT_2     NT_3
## WASH7P       3.6789877 3.6591449 3.26899634 3.5080566  3.11460814 3.176499
## MIR6723      0.2783258 0.9226538 0.07167127 0.2863958  0.56402815 0.578499
## LOC100133331 2.6520845 2.1563883 3.20435251 0.2863958  2.18269922 3.086973
## LINC00115    1.0414394 0.3623405 0.62642475 0.3746670 -0.07628992 0.578499
## LINC01128    3.8213797 3.9823671 3.77852021 4.1817956  4.10915692 4.165822
## NOC2L        7.4578505 7.5111282 7.61618272 7.5485069  7.46802220 7.364197

# Select the 1% DE genes and produce a matrix for a heatmap of FC > 2
# absolute (ie logFC > 1)
selY <- y[rownames(DGE.results_edgeR$table)[DGE.results_edgeR$table$FDR < 0.01 &
    abs(DGE.results_edgeR$table$logFC) > 0.0585], ]
head(selY)

##          NRF2_1   NRF2_2   NRF2_3      NT_1      NT_2      NT_3
## SESN3  7.436381 7.400351 7.413663 5.7091464 5.9196388 5.8613563
## NFE2L2 5.032878 4.987729 4.899398 6.5760246 6.6013320 6.5352284
## THBS1  4.612580 4.716945 4.643105 2.4669721 2.7097002 2.8760053
## SCG2   4.140739 4.089652 3.952610 1.6871971 1.8936483 2.0749252
## ACSL1  6.961795 6.986193 6.925715 5.8075713 5.9377667 5.9289337
## CHRNA1 3.180116 3.194664 3.501994 0.7495712 0.4442814 0.3665879

# Create a dataframe for column annotations
sample_col <- data.frame(NRF2 = rep(c("siNRF2", "Control"), c(3, 3)))
row.names(sample_col) <- colnames(read.counts)
head(sample_col)

##           NRF2
## NRF2_1  siNRF2
## NRF2_2  siNRF2
## NRF2_3  siNRF2
## NT_1   Control
## NT_2   Control
## NT_3   Control
```

```
# Plot the heatmap
require(pheatmap)

## Loading required package: pheatmap

# Change the color scheme
pheatmap(selY, legend = TRUE, scale = "row", fontsize_row = 6, cellwidth = 30,
    show_rownames = FALSE, cutree_cols = TRUE, annotation_col = sample_col,
    show_colnames = F, cluster_cols = T, color = colorRampPalette(c("navy",
        "white", "firebrick3"))(50), fontsize = 14, cluster_rows = T)
```

select a subset of genes which regulate glut metab by name for a heatmap

```
All.opp <- read.csv("Heatmap.csv", header = T)
selN <- y[rownames(y) %in% All.opp$Gene, ]
# List with colors for each annotation.
my_colour = list(NRF2 = c(siNRF2 = "Blue", Control = "Red"))
# 141 gene heatmap
pheatmap(selN, legend = TRUE, scale = "row", fontsize_row = 7, cellwidth = 45,
    show_rownames = T, cutree_cols = T, show_colnames = F, cluster_cols = T,
    color = colorRampPalette(c("Red", "black", "Green"))(100), fontsize = 14,
    angle_col = 45, annotation_col = sample_col, cluster_rows = T, annotation_colors
= my_colour,
    border_color = NA, cutree_rows = 2)
```

Judging and visualising the results

Make a venn diagram of results

```
library(gplots)

##
## Attaching package: 'gplots'

## The following object is masked from 'package:IRanges':
##
##     space

## The following object is masked from 'package:S4Vectors':
##
##     space

## The following object is masked from 'package:stats':
##
##     lowess

DE_list <- list(edgeR = rownames(subset(DGE.results_edgeR$table, FDR <= 0.01)),
    DESeq2 = rownames(subset(DGE.results, padj <= 0.01)))
gplots::venn(DE_list)
# Examine intersections b/n groups
x <- venn(DE_list, intersections = TRUE)

intsect <- attr(x, "intersection")
str(intsect)

## List of 3
##  $ edgeR:DESeq2: chr [1:2562] "SESN3" "NFE2L2" "THBS1" "SCG2" ...
##  $ edgeR       : chr [1:793] "KCNQ1OT1" "SYT8" "AVPR2" "LINC01589" ...
##  $ DESeq2      : chr [1:55] "PTP4A2" "RGS4" "VIM" "DKK1" ...
```

```
head(intsect$`edgeR:DESeq2`)
```

```
## [1] "SESN3"  "NFE2L2" "THBS1"  "SCG2"   "ACSL1"  "CHRNA1"
```

```
# Export the overlap list
```

Correlation of logFC for all genes found in DE in both tools ie overlap list

```
# Make a data frame of the fold change values
DE_fc <- data.frame(edgeR = DGE.results_edgeR[intsect$`edgeR:DESeq2`, ]$table$logFC,
    DESeq2 = DGE.results[intsect$`edgeR:DESeq2`, ]$log2FoldChange, row.names = intsec
t$`edgeR:DESeq2`)
# Visually check and see the estimated logFC of the different tools agree
# for teh DE genes
pairs(DE_fc, main = "Correlation of FC in edgeR and DESeq2")

# Make a data frame of the all the overlaps Write to csv
```

GSEA testing on EdgeR results

```
library(fgsea)
```

```
## Loading required package: Rcpp
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:matrixStats':
##
##     count
```

```
## The following object is masked from 'package:Biobase':
##
##     combine
```

```
## The following objects are masked from 'package:GenomicRanges':
##
##     intersect, setdiff, union
```

```
## The following object is masked from 'package:GenomeInfoDb':
##
##     intersect
```

```
## The following objects are masked from 'package:IRanges':
##
##     collapse, desc, intersect, setdiff, slice, union
```

```
## The following objects are masked from 'package:S4Vectors':
##
##     first, intersect, rename, setdiff, setequal, union
```

```
## The following objects are masked from 'package:BiocGenerics':
##
##     combine, intersect, setdiff, union
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

# Create ranks for all genes
DGE.results_edgeR.ord <- DGE.results_edgeR$table[order(-DGE.results_edgeR$table[,
    "logFC"]), ]
head(DGE.results_edgeR.ord, 20)

##                 logFC    logCPM        LR       PValue          FDR
## CHRNA1       2.827382 2.5270242 202.96144 4.716311e-46 1.040261e-42
## KRTAP3-1     2.540549 2.1676810 163.48980 1.955320e-37 2.352428e-34
## ITK          2.314698 2.5077353 173.51934 1.260526e-39 1.853534e-36
## NOD2         2.260122 2.7195717 161.26359 5.992054e-37 5.664203e-34
## SCG2         2.182944 3.3797584 235.25153 4.267028e-53 1.411746e-49
## SNX20        2.109074 2.0143537  66.20725 4.059167e-16 3.783030e-14
## SLC35D3      2.044643 2.1112858 110.71738 6.823699e-26 2.580138e-23
## THBS1        1.967704 4.0054960 263.61832 2.791441e-59 1.231398e-55
## JAKMIP2      1.922765 1.4287487  68.83978 1.067966e-16 1.095617e-14
## NGFR         1.849087 2.1948102 101.29491 7.925908e-24 2.558329e-21
## TNNC1        1.768917 0.9696215  34.48270 4.300565e-09 9.846658e-08
## RPA4         1.759168 0.8775354  38.91155 4.434486e-10 1.273015e-08
## SYTL5        1.734362 0.8169060  35.06925 3.181855e-09 7.506002e-08
## SESN3        1.582709 6.8355087 427.98844 4.453544e-95 5.893820e-91
## LYNX1        1.563797 1.1067949  40.47383 1.992677e-10 6.278830e-09
## GFRA3        1.553770 3.9609821 160.10925 1.070972e-36 8.858280e-34
## LOC101927272 1.535369 2.1407688  49.73681 1.758157e-12 8.192766e-11
## AKAP5        1.526360 0.8872967  31.38468 2.116446e-08 4.047550e-07
## GTF2H2B      1.512916 1.6795482  11.31768 7.677247e-04 3.740820e-03
## RGAG4        1.496195 1.3510352  40.29681 2.181654e-10 6.825535e-09
##                         gene
## CHRNA1               CHRNA1
## KRTAP3-1           KRTAP3-1
## ITK                     ITK
## NOD2                   NOD2
## SCG2                   SCG2
## SNX20                 SNX20
## SLC35D3             SLC35D3
## THBS1                 THBS1
## JAKMIP2             JAKMIP2
## NGFR                   NGFR
## TNNC1                 TNNC1
## RPA4                   RPA4
## SYTL5                 SYTL5
## SESN3                 SESN3
## LYNX1                 LYNX1
## GFRA3                 GFRA3
## LOC101927272 LOC101927272
## AKAP5                 AKAP5
## GTF2H2B             GTF2H2B
## RGAG4                 RGAG4

ranks <- DGE.results_edgeR.ord$logFC
names(ranks) <- rownames(DGE.results_edgeR.ord)
head(ranks, 20)

##       CHRNA1     KRTAP3-1          ITK         NOD2         SCG2
##     2.827382     2.540549     2.314698     2.260122     2.182944
##        SNX20      SLC35D3        THBS1      JAKMIP2         NGFR
##     2.109074     2.044643     1.967704     1.922765     1.849087
##        TNNC1         RPA4        SYTL5        SESN3        LYNX1
##     1.768917     1.759168     1.734362     1.582709     1.563797
```

```
##        GFRA3 LOC101927272        AKAP5      GTF2H2B        RGAG4
##    1.553770     1.535369     1.526360     1.512916     1.496195

# plot(ranks)
barplot(ranks, ylab = "logFC", xlab = "Genes", horiz = F, axisnames = F, main = "Nrf2
K.O ranked gene list by logFC",
    sub = "N=13,234")

# Download the gmt hallmarks file from the Broad Institute website and save
# to R
pathways.hallmark <- gmtPathways("/Users/socathail/Documents/R/RNAseq/h.all.v6.2.symb
ols.gmt")
# Run Enrichments
fgseasRes <- fgsea(pathways = pathways.hallmark, stats = ranks, nperm = 1000)
```

Repeat for KEGG analysis

Download the gmt KEGG file from the Broad Institute website and save to R

pathways.hallmark.K<-

gmtPathways("/Users/socathail/Documents/R/RNAseq/c2.cp.kegg.v6.2.symbols.gmt")

Plot enrichments and stacked barplots using ggplot2

###END###

Duplicate code was used for the analysis of the HCT116 isogenic data. Not shown for brevity.

9.3 APPENDIX 2: List of probes from microarray datasets in Chapter 5

Supplementary Table 3. The list of Affymetrix U133 v2.0 probes used in the analysis and the genes to which they map

Probe ID	Gene	Probe ID	Gene	Probe ID	Gene
1565778_at	ABCA8	211424_x_at	METTL7A	1568574_x_at	SPP1
204719_at	ABCA8	244376_at	METTL7A	209875_s_at	SPP1
220518_at	ABI3BP	207761_s_at	METTL7A	218863_s_at	TNS1
223395_at	ABI3BP	209703_x_at	METTL7A	221246_x_at	TNS1
1559077_at	ABI3BP	217739_s_at	NAMPT	214840_at	TOM1L2
202952_s_at	ADAM12	1555167_s_at	NAMPT	226198_at	TOM1L2
204943_at	ADAM12	217738_at	NAMPT	237109_at	TOM1L2
215613_at	ADAM12	243296_at	NAMPT	209890_at	TSPAN5
213790_at	ADAM12	202174_s_at	PCM1	225387_at	TSPAN5
226777_at	ADAM12	209996_x_at	PCM1	225388_at	TSPAN5
208214_at	ADRB1	209997_x_at	PCM1	224674_at	TTYH3
229277_at	ADRB1	214937_x_at	PCM1	204620_s_at	VCAN
229309_at	ADRB1	214118_x_at	PCM1	215646_s_at	VCAN
217626_at	AKR1C1	228905_at	PCM1	204619_s_at	VCAN
205608_s_at	ANGPT1	211668_s_at	PLAU	211571_s_at	VCAN
1552939_at	ANGPT1	205479_s_at	PLAU	221731_x_at	VCAN
205609_at	ANGPT1	203896_s_at	PLCB4	223845_at	VSIG10
229308_at	ANKRD29	203895_at	PLCB4		
238332_at	ANKRD29	207388_s_at	PTGES		
228471_at	ANKRD44	210367_s_at	PTGES		
233361_at	ANKRD44	1558115_at	RECK		
226641_at	ANKRD44	1558116_x_at	RECK		
235777_at	ANKRD44	216156_at	RECK		
235778_s_at	ANKRD44	205407_at	RECK		
205433_at	BCHE	218723_s_at	RGCC		
223484_at	C15orf48	239827_at	RGCC		
232458_at	COL3A1	201890_at	RRM2		
201852_x_at	COL3A1	209773_s_at	RRM2		
215076_s_at	COL3A1	216402_at	SEC14L4		
203325_s_at	COL5A1	207714_s_at	SERPINH1		

219232_s_at	EGLN3	209260_at	SFN		
222847_s_at	EGLN3	33322_i_at	SFN		
227771_at	LIFR	33323_r_at	SFN		
205876_at	LIFR	234948_at	SLC27A5		
229185_at	LIFR	203813_s_at	SLIT3		
208433_s_at	LRP8	216216_at	SLIT3		
205282_at	LRP8	203812_at	SLIT3		
228955_at	LRP8				

Supplementary Table 4. The list of Affymetrix Xcel probes used in the analysis and the genes to which they map

Probe ID	Gene	Probe ID	Gene
ADXECAD.4694_at	ABCA8	ADXEC.24209.C1_at	ANKRD44
ADXECAD.4694_x_at	ABCA8	ADXECADA.8703_at	ANKRD44
ADXECEMUTR.2314_at	ABCA8	ADXECAD.6885_at	ANKRD44
ADXEC.34095.C1_at	ABCA8	ADXEC.32445.C1_s_at	ANKRD44
ADXEC.34095.C1_x_at	ABCA8	ADXECEMUTR.1806_at	ANKRD44
ADXECAD.2828_at	ABI3BP	ADXECRS.10687_x_at	ANKRD44
ADXECADA.22250_at	ABI3BP	ADXEC.20524.C1_at	BCHE
ADXECADA.24689_s_at	ABI3BP	ADXEC.5762.C1_at	C15orf48
ADXECADA.13221_at	ADAM12	ADXEC.10342.C1_at	COL3A1
ADXECADA.16091_at	ADAM12	ADXEC.1138.C6_s_at	COL3A1
ADXECADA.19748_at	ADAM12	ADXECAD.16292_at	COL3A1
ADXECADA.22926_at	ADAM12	ADXECNTDJ.8556_at	COL3A1
ADXEC.10435.C1_at	ADAM12	ADXEC.1138.C8-a_s_at	COL3A1
ADXEC.24608.C1_at	ADAM12	ADXEC.1138.C8_s_at	COL3A1
ADXECAD.23192_s_at	ADRB1	ADXEC.2639.C2_s_at	COL5A1
ADXECRS.43563_s_at	ADRB1	ADXECEMUTR.7334_at	COL5A1
ADXEC.10477.C1_s_at	ADRB1	ADXEC.14436.C1_s_at	EGLN3
ADXECADA.7121_s_at	ADRB1	ADXEC.30477.C1-a_s_at	EGLN3
ADXECEMUTR.4343_at	ADRB1	ADXEC.30477.C1_at	EGLN3
ADXECEMUTR.4343_s_at	ADRB1	ADXEC.9508.C1_at	EGLN3
ADXEC.1215.C1_x_at	AKR1C1	ADXECNTDJ.1518_s_at	LIFR
ADXEC.34313.C1_s_at	ANGPT1	ADXECNTDJ.8595_at	LIFR
ADXECAD.17898_at	ANGPT1	ADXEC.28103.C1_s_at	LIFR
ADXECEMUTR.2049_at	ANGPT1	ADXECAD.15818_at	LIFR
ADXECEMUTR.2049_x_at	ANGPT1	ADXEC.17620.C1_at	LRP8
ADXECAD.22091_at	ANGPT1	ADXEC.17620.C1_x_at	LRP8
ADXECNTDJ.6462_s_at	ANGPT1	ADXECAD.26545_at	LRP8
ADXEC.28165.C1_at	ANGPT1	ADXECADA.3575_at	LRP8
ADXEC.11758.C1_at	ANKRD29	ADXEC.15165.C1_at	LRP8

ADXECAD.11878_s_at	ANKRD29	ADXEC.8514.C1_x_at	METTL7A
ADXPCEC.11748.C1_at	ANKRD29	ADXECADA.24486_at	METTL7A
ADXPCEC.11748.C1_x_at	ANKRD29	ADXECNTDJ.5826_s_at	METTL7A
ADXPCEC.865.C1_at	ANKRD29	ADXECADA.16813_at	METTL7A
ADXPCEC.865.C1_x_at	ANKRD29		
ADXEC.4289.C1_at	TOM1L2		
ADXEC.32336.C1_at	NAMPT	ADXEC.11904.C1_at	SLIT3
ADXECADA.10815_at	NAMPT	ADXEC.19191.C1_at	SLIT3
ADXEC.2660.C1-a_s_at	NAMPT	ADXEC.19191.C1_x_at	SLIT3
ADXEC.2660.C1_at	NAMPT	ADXEC.4613.C1_at	SLIT3
ADXECAD.1453_s_at	NAMPT	ADXECAD.25829_s_at	SLIT3
ADXECAD.24382_s_at	NAMPT	ADXECADA.16195_x_at	SLIT3
ADXEC.10262.C1_at	PCM1	ADXECAD.5047_at	SPP1
ADXEC.10262.C1_s_at	PCM1	ADXECAD.23641_s_at	SPP1
ADXEC.23999.C1_s_at	PCM1	ADXOCEC.14560.C1_at	SPP1
ADXEC.10262.C2_s_at	PCM1	ADXOCEC.14560.C1_x_at	SPP1
ADXEC.25350.C1_at	PCM1	ADXEC.26864.C1_at	TNS1
ADXEC.4679.C2-a_s_at	PLAU	ADXECAD.24662_at	TNS1
ADXECAD.24282_at	PLAU	ADXECEMUTR.2541_s_at	TNS1
ADXEC.21964.C2_at	PLCB4	ADXEC.3298.C1_at	TNS1
ADXECAD.27780_at	PLCB4	ADXECADA.10891_at	TOM1L2
ADXECAD.27780_s_at	PLCB4	ADXEC.13429.C1_at	TOM1L2
ADXECRS.4084_s_at	PLCB4	ADXEC.4289.C1_at	TOM1L2
ADXEC.21189.C1_s_at	PTGES	ADXEC.21443.C1_at	TSPAN5
ADXECADA.20482_at	PTGES	ADXEC.34129.C1_at	TSPAN5
ADXECADA.20482_x_at	PTGES	ADXECAD.10916_at	TSPAN5
ADXEC.10125.C1_at	RECK	ADXECAD.13537_at	TSPAN5
ADXEC.10125.C1_s_at	RECK	ADXEC.15643.C1_at	TSPAN5
ADXECAD.2455_s_at	RECK	ADXEC.8833.C1_s_at	TSPAN5
ADXECAD.11511_at	RECK	ADXEC.4902.C1_at	TTYH3
ADXEC.25901.C1_at	RGCC	ADXEC.4902.C1_x_at	TTYH3
ADXEC.10181.C1_s_at	RGCC	ADXEC.4902.C2_s_at	TTYH3

ADXEC.3228.C1_s_at	RRM2	ADXEC.25748.C1_at	VCAN
ADXECRS.44245_at	RRM2	ADXEC.4587.C1-a_s_at	VCAN
ADXECNTDJ.6564_at	SEC14L4	ADXEC.4587.C1_s_at	VCAN
ADXECEMUTR.6644_at	SERPINH1	ADXECAD.17188_at	VCAN
ADXECNTDJ.6701_s_at	SERPINH1	ADXECAD.17188_s_at	VCAN
ADXEC.1556.C1_at	SFN	ADXECAD.21682_s_at	VCAN
ADXEC.13004.C1_at	SLC27A5	ADXECEMUTR.4185_s_at	VCAN
ADXEC.13004.C1_s_at	SLC27A5	ADXECNTDJ.1389_at	VCAN
ADXEC.13004.C1_x_at	SLC27A5	ADXECRS.7864_s_at	VCAN
ADXECADA.15636_at	SLC27A5	ADXECADA.19309_at	VSIG10
ADXECNTDJ.6726_at	SLC27A5		

Supplementary Table 5. The list of Agilent Human 4 x 44k v2 array probes used in the analysis and the genes to which they map

Probe ID	Gene	Probe ID	Gene
A_23_P118615	ABCA8	A_33_P3364864	NAMPT
A_33_P3342305	ABCA8	A_33_P3364869	NAMPT
A_23_P218858	ABI3BP	A_33_P3375934	NAMPT
A_33_P3411975	ABI3BP	A_23_P82950	PCM1
A_33_P3411980	ABI3BP	A_24_P555510	PCM1
A_23_P350512	ADAM12	A_23_P24104	PLAU
A_33_P3256848	ADAM12	A_33_P3306146	PLAU
A_33_P3310929	ADAM12	A_33_P3260634	PLCB4
A_33_P3310189	ADRB1	A_24_P403417	PTGES
A_24_P152968	AKR1C1	A_23_P83028	RECK
A_23_P216023	ANGPT1	A_33_P3269899	RECK
A_23_P412577	ANKRD29	A_24_P225616	RRM2
A_23_P209347	ANKRD44	A_33_P3389286	SFN
A_24_P128713	ANKRD44	A_23_P4611	SLC27A5
A_23_P212050	BCHE	A_23_P58588	SLIT3
A_23_P26024	C15orf48	A_33_P3391496	SLIT3
A_24_P935491	COL3A1	A_23_P7313	SPP1
A_23_P83818	COL5A1	A_24_P105733	TNS1
A_33_P3383351	COL5A1	A_33_P3209491	TNS1
A_33_P3256952	EGLN3	A_33_P3738458	TNS1
A_23_P200222	LRP8	A_23_P141180	TOM1L2
A_23_P415021	METTL7A	A_33_P3325349	TSPAN5
		A_23_P331895	TTYH3
		A_23_P144959	VCAN

This page is the final page of the thesis