# 5-AZA-2′-DEOXYCYTIDINE AS A POTENTIAL THERAPEUTIC OPTION IN THE TREATMENT OF HEAD AND NECK SQUAMOUS CELL CARCINOMA (HNSCC)

by

**Hannah Jessica Eadie**

A thesis submitted to the University of Birmingham for the degree

of

DOCTOR OF PHILOSOPHY

School of Dentistry

College of Medicine and Dentistry

University of Birmingham

August 2018

# Abstract

Head and neck squamous cell carcinoma (HNSCC) is the 6th most common cancer worldwide. Epigenetic mechanisms including changes in DNA methylation (5mC) and hydroxymethylation (5hmC) are major contributors to HNSCC progression and can be altered by the use of epigenetic therapies such as the DNA demethylating agent, 5-aza-2'-deoxycytidine (DAC). Here, the potential use of DAC in the treatment of HNSCC was investigated and a subset of HNSCC cell lines determined to respond to nanomolar concentrations of the drug. Furthermore, the efficacy of DAC was significantly increased by co-treatment with the common analgesic, paracetamol, providing an appealing new therapeutic option.

Genome-wide sequencing methods were used to compare the distribution of 5mC and 5hmC between normal oral keratinocytes and an HNSCC cell line and investigate how this is altered by DAC. In addition to causing global DNA demethylation, DAC partially restores 5mC and 5hmC patterns to those of normal oral keratinocytes. In normal cells, 5mC and 5hmC mark clusters of Alu elements in gene rich regions; this is dramatically reduced in HNSCC. Treatment with DAC restores DNA modifications to Alu elements, leading to repression of Alu RNA. Furthermore, DAC caused demethylation and increased expression of endogenous retroviruses, prompting a viral mimicry response.

# Dedication

This dissertation is dedicated to my family for their unwavering help and support: To my husband, for believing in me even when I didn't believe in myself and to my parents for a lifetime of encouragement.

# Acknowledgments

I would like to acknowledge and thank a number of people who have not only helped me with the preparation of this thesis but who have also supported me throughout my PhD.

Firstly, I would like to thank my primary supervisor, Dr Malgorzata Wiench, for all the knowledge, guidance and encouragement she has given me in the past four years and for helping this PhD come to fruition. I would also like to acknowledge the help of my secondary supervisors, Professor Paul Cooper and Dr Ben Scheven, in preparing this manuscript and supporting me through the course of my PhD.

Secondly, I must recognize the invaluable contribution of Dr Sam Clockie and Dr Wayne Croft in the hMeDIP and MeDIP data analysis, without whose help this thesis would not have been possible.

I would like to thank Dr Nikolaos Batis for sharing his knowledge on pharmacology and synergy and Dr Farhat Khanim for providing me with the drug repurposing panel and giving me guidance along the way.

Finally, I would like to thank the other PhD students in the lab, whose friendship, support and understanding helped to make my PhD more enjoyable.

# Table of Contents

List of Figures ..... 1

List of Tables ..... 8

List of Supplementary Data ..... 9

List of Abbreviations ..... 10

**Chapter 1. Introduction ..... 12**

1.1. Head and neck squamous cell carcinoma (HNSCC) ..... 12

*1.1.1. Occurrence and aetiology ..... 12*

*1.1.2. Genetic landscape of HNSCC ..... 12*

*1.1.3. Current therapies ..... 15*

1.2. Epigenetics ..... 17

*1.2.1. Introduction to epigenetics ..... 17*

*1.2.2. Gene structure and chromatin ..... 18*

1.3. DNA methylation ..... 22

*1.3.1. CpG islands (CGIs) ..... 23*

*1.3.2. DNA methyltransferases (DNMTs) ..... 23*

*1.3.3. Transcriptional silencing by 5mC ..... 26*

*1.3.4. Genome wide 5mC patterns ..... 28*

1.4. DNA hydroxymethylation ..... 30

*1.4.1. TET enzymes ..... 31*

*1.4.2. 5hmC function as opposed to 5mC ..... 32*

*1.4.3. DNA demethylation ..... 37*

1.5. Transposable elements ..... 39

*1.5.1. Transposable element families ..... 39*

*1.5.2. Epigenetic regulation of transposable elements ..... 42*

1.6. Epigenetic alterations in cancer ..... 44
*1.6.1. Genome wide hypomethylation ..... 46*
*1.6.2. CpG island hypermethylation ..... 46*
*1.6.3. CpG island hypermethylation in HNSCC ..... 48*
*1.6.4. Tumour suppressor potential of TET proteins ..... 48*
*1.6.5. 5hmC depletion and diagnostic potential ..... 49*
*1.6.6. Loss of epigenetic control of transposable elements in cancer cells ..... 52*
1.7. Epigenetic therapies in cancer ..... 53
*1.7.1 DNMTi ..... 54*
*1.7.2 Current usage and limitations of 5-aza-2′-deoxycytidine (DAC) ..... 54*
*1.7.3 Potential efficacy of DAC in solid tumours ..... 55*
1.8. Drug Repurposing ..... 58
1.9. Aims and Objectives ..... 61

**Chapter 2. Methods ..... 62**

2.1. Cell lines and culture ..... 62
*2.1.1. Cell lines ..... 62*
*2.1.2. Cell maintenance ..... 63*
2.2. Drug Treatments ..... 63
*2.2.1. 5-Aza-2′-deoxycytidine treatments ..... 63*
*2.2.2. CellTiter Blue Cell Viability assay ..... 64*
*2.2.3. Dose response curves ..... 64*
*2.2.4. DAC sensitising assay ..... 64*
*2.2.5. Drug treatments ..... 65*
*2.2.6. Determining synergy ..... 65*
2.3. DNA extraction and dot blotting of DNA ..... 66
*2.3.1. DNA extraction ..... 66*

*2.3.2. Dot blotting* ..... *66*

*2.3.3. Antibody incubation* ..... *67*

*2.3.4. Analysis* ..... *67*

2.4. Protein extraction and Western blotting ..... 68

*2.4.1. Protein extraction* ..... *68*

*2.4.3. SDS-PAGE* ..... *69*

*2.4.4. Transfer of proteins from gel to membrane* ..... *69*

*2.4.5. Antibody incubation* ..... *70*

*2.4.6. ImageJ Analysis* ..... *71*

2.5. Giemsa and Jenner Staining ..... 73

*2.5.1. Cell plating and fixation* ..... *73*

*2.5.2. Giemsa-Jenner staining* ..... *73*

2.6. Immunofluorescence ..... 73

*2.6.1. Cell fixation* ..... *73*

*2.6.2. Antibody incubation* ..... *74*

*2.6.3. Confocal microscopy* ..... *74*

*2.6.4. Immunofluorescent analysis* ..... *75*

2.7. Apoptosis and Necrosis detection ..... 75

*2.7.1. Annexin V and Propidium Iodide staining* ..... *75*

*2.7.2. Fluorescence-activated cell sorting (FACS)* ..... *76*

2.8. Cell cycle analysis ..... 76

2.9. Prostaglandin $E_2$ ($PGE_2$) enzyme-linked immunosorbent assay (ELISA) ..... 77

2.10. Determining Glutathione concentration ..... 77

2.11. Quantitative Reverse Transcriptase Polymerase Chain Reaction (qRT-PCR) Analysis ..... 78

*2.11.1. RNA extraction* ..... *78*

*2.11.2. cDNA synthesis* ..... *79*

*2.11.3. cDNA quantification* ..... *79*

*2.11.4. Quantitative reverse transcriptase PCR (qRT-PCR)* ..... *79*

*2.11.5. Primer design* ..... *80*

2.12. RNA sequencing and analysis ..... 82

2.13. Methylated and hydroxymethylated DNA Immunoprecipitation (MeDIP and hMeDIP) ..... 84

*2.13.1. Cell treatment* ..... *84*

*2.13.2. DNA extraction* ..... *84*

*2.13.3. Immunoprecipitation* ..... *85*

*2.13.4. Recovery of 5mC enriched DNA* ..... *86*

*2.13.5. Hydroxymethylated DNA Immunoprecipitation (hMeDIP)* ..... *88*

*2.13.6. (h)MeDIP-qPCR* ..... *88*

2.14. Illumina Library Preparation and Sequencing of (h)MeDIP samples ..... 90

*2.14.1. NEBNext ChIP-Seq Library prep Master Mix Set* ..... *90*

*2.14.2. Kapa Hyper Library Preparation* ..... *92*

*2.14.3. Library Fragment Size Determination* ..... *94*

*2.14.4. Library DNA Quantification* ..... *95*

*2.14.5. Illumina Sequencing* ..... *96*

*2.14.6. (h)MeDIP Sequencing Data Analysis.* ..... *96*

**Chapter 3. DAC as a potential therapeutic in the treatment of HNSCC .101**

3.1. Mechanism of action of DAC ..... 101

*3.1.1. DNA demethylation* ..... *101*

*3.1.2. DNA damage* ..... *102*

3.2. Mechanisms of resistance to DAC treatment ..... 105

3.3. Aims and objectives ..... 107

3.4. Cell viability assay ..... 108
3.5. The efficacy of DAC treatment varies in HNSCC cell lines ..... 110
3.6. A drug repurposing screen identifies paracetamol, valproic acid and zinc acetate as being able to sensitise HNSCC cells to DAC treatment ..... 119
3.7. DAC and paracetamol work in synergy to reduce viability in HNSCC cells ..... 123
3.8. Combined treatment with DAC and paracetamol slows down proliferation in VU40T cells ..... 128
3.9. Combined treatment with DAC and paracetamol alters the transcriptional profile of DAC-responsive cells ..... 133
3.10. Discussion and conclusion ..... 138

**Chapter 4. Mechanisms underlying the synergistic relationship between DAC and paracetamol ..... 141**

4.1. Paracetamol ..... 141
4.2. The mechanism of action of paracetamol ..... 142
*4.2.1. Inhibition of the cyclooxygenase pathway ..... 142*
*4.2.2. Inhibition of putative cyclooxygenase 3 ..... 145*
*4.2.3. Conversion of Paracetamol to N-arachidonoylphenolamine (AM404) ..... 145*
4.3. Paracetamol overdose ..... 148
4.4. Paracetamol and cancer ..... 151
*4.4.1. Inhibition of the cyclooxygenase pathway ..... 151*
*4.4.2. High dose paracetamol as a cancer therapeutic ..... 152*
4.5. Aims and objectives ..... 154
4.6. DAC alters the COX pathway ..... 155
4.7. Genes involved in the conversion of paracetamol to AM404 are not expressed in HNSCC cells ..... 164

4.8. Combined treatment with DAC and paracetamol may mimic paracetamol overdose ..... 166
4.9. Paracetamol treatment leads to a decrease in DNA methylation ..... 172
4.10. Discussion and conclusion ..... 177

**Chapter 5. HNSCC is characterised by global changes in the distribution of DNA methylation and hydroxymethylation ..... 182**

5.1 Methods of determining the distribution of 5mC ..... 182
5.2 Methods of determining the distribution of 5hmC ..... 183
5.3 Potential regulatory functions of Alu elements ..... 187
*5.3.1. Genomic Alu elements ..... 187*
*5.3.2. Embedded Alu RNA ..... 188*
*5.3.3. Free Alu RNA ..... 188*
5.4 Aims and Objectives ..... 190
5.5 Methylated and hydroxymethylated DNA immunoprecipitation (MeDIP and hMeDIP) and sequencing ..... 191
5.6 DNA hydroxymethylation is found in closer proximity to transcription start sites (TSS) and associated with a higher CpG content than DNA methylation ..... 193
5.7 5hmC-enriched regions are associated with enhancers ..... 197
5.8 Regions enriched in DNA methylation and hydroxymethylation were primarily found at Alu elements ..... 199
5.9 DNA methylation and hydroxymethylation at Alu elements occurs alongside but does not overlap regulatory elements. ..... 203
5.10 The function of 5mC and 5hmC may differ within and outside of Alu elements 210
5.11 5hmC marked Alu elements are found in regions associated with active histone modifications ..... 212

5.12 DNA methylation and hydroxymethylation outside of Alu elements are associated with bivalent chromatin ..... 218
5.13 5hmC is associated with tissue specific enhancers ..... 220
5.14 DNA methylation is gained at promoters and CpG islands in an HNSCC cell line 222
5.15 Promoters that gain 5mC in HNSCC are associated with decreased gene expression ..... 226
5.16 In HNSCC cells Alu elements are depleted of 5mC and 5hmC ..... 229
5.17 Discussion and Conclusion ..... 232

**Chapter 6. DAC partially restores the distribution of 5mC and 5hmC in an HNSCC cell line to one characteristic of normal oral keratinocytes ..... 239**

6.1. Viral mimicry ..... 239
6.2. 5hmC and DAC treatment ..... 243
6.3. Aims and objectives ..... 244
6.4. MeDIP and hMeDIP of DAC treated cells ..... 245
6.5. DAC treatment alters the genomic distribution of 5mC and 5hmC in HNSCC cells 247
6.6. DAC treatment does not reactivate tumour suppressor genes commonly inactivated in HNSCC ..... 254
6.7. DAC treatment induces a viral mimicry response in VU40T cells ..... 256
6.8. In HNSCC cells 5mC and 5hmC are partially restored at Alu elements after DAC treatment ..... 264
6.9. DAC treatment does not alter the distribution of DNA methylation in HOK cells 271
6.10. Discussion and conclusion ..... 277

**Chapter 7. Discussion and Future Work ..... 284**

7.1. In normal oral keratinocytes Alu elements are enriched in DNA methylation and hydroxymethylation ..... 284
7.2. DAC as a potential therapeutic in HNSCC ..... 286
*7.2.1. DAC is effective in a subset of HNSCC cell lines ..... 286*
*7.2.2. DAC shifts the distribution of 5mC and 5hmC back towards that of normal cells 287*
*7.2.3. Primary oral keratinocytes are not responsive to DAC treatment ..... 288*
*7.2.4. The relationship between 5mC and 5hmC ..... 289*
7.3. DAC and Paracetamol work in synergy to reduce viability in HNSCC cell lines 290
7.4. Conclusion ..... 292
Supplementary Information ..... 294
References ..... 309

# List of Figures

Figure 1.1. Molecular alterations contributing to HNSCC development ..... 14

Figure 1.2. Waddington's Epigenetic landscape ..... 17

Figure 1.3. The structure of a typical human gene ..... 19

Figure 1.4: Packaging of DNA ..... 21

Figure 1.5. The structure of 5-methylcytosine ..... 23

Figure 1.6. The role of DNA methyltransferases in the establishment and maintenance of DNA methylation ..... 25

Figure 1.7. Epigenetic regulation of gene expression ..... 27

Figure 1.8: The DNA methylation landscape in mammalian cells ..... 29

Figure 1.9. The chemical structure of cytosine and cytosine modifications ..... 31

Figure 1.10. 5hmC genomic distribution ..... 34

Figure 1.11. 5hmC and gene expression ..... 36

Figure 1.12. DNA demethylation ..... 38

Figure 1.13. Repetitive DNA in the human genome sequence ..... 39

Figure 1.14. Human retrotransposons ..... 41

Figure 1.15. The mutagenic potential of retrotransposons ..... 43

Figure 1.16. Epigenetic alterations in cancer ..... 45

Figure 1.17. 5hmC and cancer ..... 51

Figure 1.18. Mechanisms of epigenetic gene silencing and the potential for intervention ..... 53

Figure 2.1: Steps involved in the Methylated DNA immunoprecipitation (MeDIP) .. 87

Figure 2.2. Overview of the steps of MeDIP and hMeDIP sequencing and analysis 100

Figure 3.1. The mechanism of action of DAC ..... 104

Figure 3.2. DAC transport and activation ..... 106

Figure 3.3. CellTiter Blue cell viability assay ..... 109
Figure 3.4. HNSCC cells show variable sensitivity to DAC treatment ..... 112
Figure 3.5. Methylation changes in response to DAC treatment ..... 113
Figure 3.6. Differential RNA expression between DAC responsive and unresponsive cell lines ..... 115
Figure 3.7. DAC responsive cell lines exhibit increased expression of *SLC15A2, SLC22A5 and SLC29A1* ..... 117
Figure 3.8. Enrichment of nucleobase-metabolism related genes in DAC-unresponsive cell lines ..... 118
Figure 3.9. A drug-repurposing screen identified paracetamol, zinc acetate and valproic acid as able to sensitise HNSCC cell lines to DAC treatment ..... 121
Figure 3.10. Paracetamol, zinc acetate and valproic acid are able to sensitise HNSCC cell lines to DAC treatment ..... 122
Figure 3.11. DAC and paracetamol treatment in DAC-responsive cell lines ..... 125
Figure 3.12. DAC and paracetamol work in synergy to reduce viability in HNSCC cells ..... 126
Figure 3.13. Morphology of DAC and Paracetamol treated VU40T cells ..... 127
Figure 3.14. Paracetamol does not increase cell death caused by DAC treatment ..... 130
Figure 3.15. Combined treatment with DAC and paracetamol reduces proliferation ..... 131
Figure 3.16. Cell cycle changes after treatment with DAC and paracetamol ..... 132
Figure 3.17. Treatment with DAC and paracetamol alters the transcriptional profile of VU40T cells ..... 135
Figure 3.18. Gene ontology analysis of DAC and paracetamol treated VU40T cells 136
Figure 3.19. Combined treatment with DAC and paracetamol decreases the expression of genes involved in the cell cycle, DNA replication and repair ..... 137

Figure 4.1. The cyclooxygenase pathway ..... 144
Figure 4.2. The conversion of paracetamol to AM404 ..... 147
Figure 4.3. The mechanism of paracetamol overdose ..... 150
Figure 4.4. The tumour suppressive effects of paracetamol ..... 153
Figure 4.5. DAC treatment increases the expression of COX-2 in DAC responsive cell lines ..... 158
Figure 4.6. DAC increases the levels of $PGE_2$ in DAC-responsive VU40T cells ..... 159
Figure 4.7. DAC increases the expression of the $PGE_2$ receptors ..... 160
Figure 4.8. DAC treatment alters the cyclooxygenase pathway ..... 161
Figure 4.9. Alterations in the COX-2/$PGE_2$ pathway in HNSCC ..... 162
Figure 4.10. Other COX inhibitors do not sensitise HNSCC cell lines to DAC treatment ..... 163
Figure 4.11. Enzymes involved in the conversion of paracetamol into AM404 are not altered by DAC treatment in VU40T cells ..... 165
Figure 4.12. DAC treatment increases the expression of CYP2E1 in DAC-responsive cells ..... 168
Figure 4.13. Combined treatment with DAC and paracetamol reduces GSH levels. 169
Figure 4.14. Combined treatment with DAC and paracetamol may be mimicking the effects of paracetamol overdose ..... 170
Figure 4.15. Disulfiram does not rescue VU40T cells from combined treatment with DAC and paracetamol ..... 171
Figure 4.16. Paracetamol does not increase DAC-induced DNA damage ..... 173
Figure 4.17. Paracetamol decreases DNA methylation ..... 174
Figure 4.18. Paracetamol reduces *DNMT* gene expression ..... 175
Figure 4.19. Valdecoxib does not alter the levels of DNA methylation ..... 176

Figure 4.20. DAC and paracetamol work in synergy to reduce viability in HNSCC cells ..... 181
Figure 5.1. Determining 5hmC at base pair resolution ..... 185
Figure 5.2. Genome wide affinity-based methods for 5hmC profiling ..... 186
Figure 5.3. The potential functions of Alu elements ..... 189
Figure 5.4. The genomic location of 5mC and 5hmC-enriched regions ..... 194
Figure 5.5. Differential characteristics of 5mC and 5hmC-enriched regions ..... 195
Figure 5.6. Relationship between 5mC and 5hmC and transcription start sites and transcription factor binding sites ..... 195
Figure 5.7. 5mC and 5hmC are underrepresented at CpG islands but enriched in CpG island shores ..... 196
Figure 5.8. Chromatin domains associated with 5mC and 5hmC-enriched regions. 198
Figure 5.9. MeDIP and hMeDIP peaks at repetitive DNA ..... 200
Figure 5.10. Differential enrichment of 5mC and 5hmC in Alu element families ..... 201
Figure 5.11. Enrichment of 5hmC at Alu element subfamilies ..... 202
Figure 5.12. 5mC and 5hmC mark Alu elements in gene-rich regions ..... 205
Figure 5.13. 5mC and 5hmC at Alu elements were excluded from the TSS ..... 206
Figure 5.14. 5hmC at Alu elements at DNase hypersensitive sites ..... 207
Figure 5.15. 5hmC at Alu elements is excluded from DNase hypersensitive sites ... 208
Figure 5.16. 5mC and 5hmC enriched regions are excluded from transcription factor binding sites ..... 209
Figure 5.17. Chromatin domains associated with 5mC and 5hmC at Alu elements. 211
Figure 5.18. Histone modifications associated with 5mC and 5hmC-enriched Alu elements ..... 214
Figure 5.19. Example of co-localisation of 5hmC-marked Alu elements and active histone modifications ..... 215

Figure 5.20. The relationship between 5mC and 5hmC-enriched regions and histone modification peaks ..... 216
Figure 5.21. 5mC and 5hmC at Alu elements occurs close to regulatory regions ..... 217
Figure 5.22. 5mC and 5hmC and bivalent histone modifications ..... 219
Figure 5.23. Gene ontology analysis of 5hmC-enriched enhancers ..... 221
Figure 5.24. The distribution of 5mC and 5hmC in HNSCC cells ..... 223
Figure 5.25. 5mC is gained in CpG islands in HNSCC ..... 224
Figure 5.26. Characteristics of 5mC and 5hmC-enriched regions in VU40T cells ..... 225
Figure 5.27. Gene ontology analysis of VU40T specific promoters ..... 227
Figure 5.28. Expression of genes with methylated promoters ..... 228
Figure 5.29. DNA methylation and hydroxymethylation of repetitive DNA in HNSCC cells ..... 230
Figure 5.30. Enrichment of 5mC and 5hmC across Alu elements families in normal oral keratinocytes and HNSCC cells ..... 231
Figure 5.31. 5mC and 5hmC changes in HNSCC ..... 237
Figure 5.32. DNA modifications in normal oral keratinocytes and HNSCC cells ..... 238
Figure 6.1. Viral mimicry ..... 242
Figure 6.2. Global levels of 5mC and 5hmC in DAC treated VU40T cells ..... 248
Figure 6.3. DNA methylation and hydroxymethylation changes in response to DAC ..... 249
Figure 6.4. DAC treatment increases *TET2* RNA ..... 250
Figure 6.5. In HNSCC DAC treatment shifts the distribution of 5mC and 5hmC back towards that of normal oral keratinocytes ..... 251
Figure 6.6. Differential enrichment of 5mC at CpG islands in DAC treated VU40T cells ..... 252
Figure 6.7. Characteristics of DAC treated VU40T cells ..... 253

Figure 6.8. A comparison of the expression and methylation changes in VU40T cells in response to DAC ..... 255
Figure 6.9. 5mC is lost from endogenous retroviruses (ERVs) after treatment with DAC ..... 257
Figure 6.10. Decreased methylation correlates with increased RNA at some ERVs ..... 258
Figure 6.11. DAC treatment increases the levels of double stranded RNA ..... 259
Figure 6.12. DAC treatment increases the expression of dsRNA and viral response genes ..... 261
Figure 6.13. In VU40T cells DAC induces a viral mimicry response ..... 262
Figure 6.14. Immune genes upregulated by DAC treatment cluster together in HNSCC patient samples ..... 263
Figure 6.15. DNA methylation and hydroxymethylation at repetitive DNA after treatment with DAC ..... 266
Figure 6.16. Changes in methylation and hydroxymethylation at repetitive DNA in VU40T cells after DAC treatment ..... 267
Figure 6.17. Changes in the distribution of 5mC and 5hmC over repetitive DNA in VU40T cells in response to DAC ..... 268
Figure 6.18. Differential enrichment of 5mC and 5hmC at Alu element families in VU40T cells treated with DAC ..... 269
Figure 6.19. In HNSCC, increased 5mC and 5hmC at Alu elements correlates with decreased Alu expression ..... 270
Figure 6.20. DAC treatment leads to minimal changes in 5mC in HOK cells ..... 272
Figure 6.21. The distribution of 5mC in HOK cells after DAC treatment ..... 273
Figure 6.22. Characteristics of HOK MeDIP peaks after DAC treatment ..... 274
Figure 6.23. DNA methylation of repetitive DNA in HOK cells after treatment with DAC ..... 275

Figure 6.24. DAC treatment alters the distribution of 5mC in VU40T cells, but not HOK cells.....276
Figure 6.25. The distribution of 5mC and 5hmC is altered in HNSCC, and partially repaired by treatment with DAC.....282
Figure 6.26. The distribution of 5mC and 5hmC in normal oral keratinocytes, HNSCC cells and HNSCC cells treated with DAC .....283
Figure 7.1. DAC as a potential therapeutic option in the treatment of HNSCC.....293

# List of Tables

Table 1.1. DAC therapeutic application in solid tumours ..... 57

Table 1.2. Drugs repurposed for cancer treatment ..... 60

Table 2.1. Details of the cell types used in the studies ..... 62

Table 2.2: Buffers used for dot blotting of DNA ..... 68

Table 2.3: Buffers used for Western blotting ..... 72

Table 2.4: Programme used for quantification RNA by qPCR ..... 80

Table 2.5: Primers used for qRT-PCR ..... 81

Table 2.6: Buffers used for MeDIP and hMeDIP ..... 86

Table 2.7: Primer sequences used for MeDIP and hMeDIP qPCR check ..... 89

Table 2.8: NEBNext library preparation PCR programme ..... 92

Table 2.9: KAPA Hyper Library preparation PCR programme ..... 93

Table 2.10: KAPA Library Quantification thermal cycling conditions ..... 96

Table 3.1. DAC responsive cell lines exhibit increased expression of *SLC29A1* ..... 116

Table 3.2. DAC and paracetamol work in synergy to reduce viability in HNSCC ..... 126

Table 5.1. MeDIP and hMeDIP output ..... 192

Table 6.1. MeDIP and hMeDIP output ..... 246

Table 6.2. DAC treatment is associated with an increased expression of genes involved in the immune system activation ..... 260

# List of Supplementary Data

Supplementary data 1. FCM1 drug library ..... 294

Supplementary data 2. RNA sequencing reads ..... 295

Supplementary data 3. MeDIP and hMeDIP sequenced reads ..... 296

Supplementary data 4. DNA dot blot analysis of 5mC levels in HNSCC cell lines .. 297

Supplementary data 5. Gene ontology terms associated with DAC responsive cell lines ..... 298

Supplementary data 6. Gene ontology terms associated with genes differentially expressed in DAC responsive versus DAC unresponsive HNSCC cell lines ..... 299

Supplementary data 7. Gene ontology terms associated with genes differentially expressed after DAC, paracetamol or combined treatment ..... 300

Supplementary data 8. Western blot analysis ..... 301

Supplementary data 9. DNA dot blot analysis of DAC, paracetamol and valdecoxib treated HNSCC cell lines ..... 302

Supplementary data 10. MeDIP and hMeDIP qRT-PCR check ..... 303

Supplementary data 11. Overlap between 5mC and 5hmC peaks and NHEK chromatin domains ..... 304

Supplementary data 12. Alu element subfamilies ..... 305

Supplementary data 13. Overlap between HOK 5mC and 5hmC Alu groups and NHEK chromatin domains ..... 306

Supplementary data 14. Overlap between HOK 5mC and 5hmC Alu groups and NHEK histone modifications ..... 307

Supplementary data 15. DNA dot blot analysis of 5mC and 5hmC of DAC treated VU40T cells ..... 308

# List of Abbreviations

| | |
|---|---|
| **Frequently Used** | |
| HNSCC | Head and neck squamous cell carcinoma |
| DAC | 5-Aza-2′-deoxycytidine |
| 5mC | 5-methylcytosine |
| 5hmC | 5-hydroxymethylcytosine |
| DNMT | DNA methyltransferase |
| Para | Paracetamol |
| hMeDIP | hydroxymethylated DNA immunoprecipitation |
| MeDIP | Methylated DNA immunoprecipitation |
| **Other** | |
| 5caC | 5-carboxylcytosine |
| 5fC | 5-formylcytosine |
| ABC | ATP-binding cassette transporters |
| ac | acetylation |
| AM404 | N-arachidonoylaminophenol |
| AML | acute myeloid leukemia |
| ATP | Adenosine triphosphate |
| AZA | 5-Azacytidine |
| BS-seq | Bisulfite sequencing |
| BSA | Bovine serum albumin |
| C | Cytosine |
| CGI | CpG island |
| ChIP | Chromatin Immunoprecipitation |
| CI | Combination Index |
| Cmax | maximum (or peak) serum concentration |
| CML | chromatin myeloid leukemia |
| COX | cyclooxygenase |
| DAPI | 4',6-diamidino-2-phenylindole |
| DHS | Dnase hypersensitive site |
| DMSO | Dimethyl sulfoxide |
| DNA | Deoxyribonucleic acid |
| DNMTi | DNA methyltransferase inhibitor |
| DRI | Dose reduction index |
| DSF | Disulfiram |
| dsRNA | Double stranded RNA |
| EDTA | Ethylenediaminetetraacetic acid |
| EMA | European Medicines Agency |
| ENCODE | Encyclopedia of DNA Elements |
| ERV | Endogenous retrovirus |
| ESCs | Embryonic stem cells |
| Fa | Fraction affected |
| FACS | Fluorescence-activated cell sorting |
| FBS | Fetal Bovine Serum |
| FDA | U S Food and Drug Administration |
| FPKM | Fragments Per Kilobase of transcript per Million mapped reads |

| | |
|---|---|
| GO | Gene Ontology |
| GREAT | Genomic Regions Enrichment of Annotations Tool |
| GSH | glutathione |
| H1/2A/2B/3/4 | Histone 1/2A/2B/3/4 |
| HCL | Hydrochloric acid |
| HDAC | Histone deacetylase |
| HDACi | Histone deacetylase inhibitor |
| HPV | Human papilloma virus |
| Ibup | Ibuprofen |
| IC50 | half maximal inhibitory concentration |
| IP | immunoprecipitation |
| ISG | interferon-stimulated genes |
| K | lysine |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| LINE | Long interspersed nuclear element |
| LTR | Long terminal repeat |
| MBP | Methyl binding protein |
| MDS | Myelodysplastic syndromes |
| me | methylation |
| MEF | mouse embryonic fibroblasts |
| NAPQI | N-Acetyl-p-benzoquinone imine |
| NHEK | Normal Human Epidermal Keratinocytes |
| NSAIDs | Non-steroidal anti-inflammatory drugs |
| NT | No Treatment |
| OSCC | Oral squamous cell carcinoma |
| oxBS-seq | Oxidative bisulfite sequencing |
| PBS | Phosphate buffered saline solution |
| PCR | Polymerase chain reaction |
| $PGE_2$ | prostaglandin $E_2$ |
| PI | Propidium Iodide |
| Pol | Polymerase |
| qRT-PCR | Quantitative Reverse Transcriptase Polymerase Chain Reaction |
| RNA | Ribonucleic acid |
| ROS | reactive oxygen species |
| RT | Room temperature |
| SEM | Standard error of the mean |
| SINE | Short interspersed nuclear element |
| SPRI | Solid Phase Reversible Immobilization |
| TCGA | The Cancer genome altas |
| TFBS | Transcription facor binding site |
| TSS | Transcription start site |
| Vald | Valdecoxib |
| VPA | valproic acid |
| α-HG | α-hydroxyglutarate |
| α-KG | α-ketoglutarate |

# Chapter 1. Introduction

## 1.1. Head and neck squamous cell carcinoma (HNSCC)

### 1.1.1. Occurrence and aetiology

Head and neck cancer is the 6$^{th}$ most common cancer worldwide with up to 650,000 new cases diagnosed each year [1, 2]. Most cases are squamous cell carcinomas, referred to as head and neck squamous cell carcinoma or HNSCC [3, 4]. This term refers to tumours that develop from more than 15 different anatomical subsites, however the majority of cases arise in the oral cavity or the oropharynx (oral squamous cell carcinomas (OSCC)) or the larynx (Laryngeal cancer) [4-6]. The most important risk factors for HNSCC are tobacco and alcohol consumption, with a synergistic relationship described between the two [1, 7]. More recently, cases of HNSCC have been attributed to infection with oncogenic forms of the human papilloma virus (HPV). These cancers are found to be restricted to the oral pharynx and could be responsible for more than 60% of HNSCC cases in Europe [1].

### 1.1.2. Genetic landscape of HNSCC

Cancer is a complex disease; an accumulation of genetic mutations and epigenetic alterations lead to changes in gene expression that drive the progressive transformation of a healthy cell into a malignant one [8]. The genetic and transcriptomic signatures differ greatly between different cancers. HNSCC, for example, is a very heterogeneous disease [9]. However, genetic analysis of HNSCC samples has identified several mutations that are associated with disease initiation

and progression (Figure 1.1) [8, 9]. Mutations in the tumour suppressor gene *TP53* are found in 46-73% of HNSCC cases [12]. This mutation often occurs early in HNSCC development but incidence increases as premalignant lesions progress to invasive carcinomas (Figure 1.1)[3, 10]. P53 is activated upon DNA damage and triggers a halt in the cell cycle to allow for DNA repair, or prompts apoptosis or senescence when repair is not possible [12, 13]. Therefore mutant P53 can lead to an accumulation of mutations and evasion of apoptosis [12, 13]. Upregulation of the oncogene *PIK3CA* (found in 34-56% of HNSCC) can control the tumour suppressor protein RB and activate oncogenes *MYC* and *NF-kB*, leading to cell cycle progression and survival (Figure 1.1) [8]. Epidermal growth factor receptor (EGFR) is involved in various pathways associated with growth and survival and is upregulated in 21% of HNSCC cases, leading to increased proliferation, metastasis, migration, angiogenesis and inhibition of apoptosis (Figure 1.1) [8, 11]. Numerous other pathways contribute to HNSCC development including promotion of angiogenesis through *VEGF* overexpression, mutations in the tumour suppressor gene *CDKN2A* and overexpression of immune-related genes such as cyclooxygenase-2 (COX-2 encoded for by *PTGS2*) (Figure 1.1) [3, 9, 11, 12]. Furthermore, HNSCC tumours show an average of 141 copy number alterations per tumour, including losses of chromosomal regions 3p, 8p, 9p and 18q (Figure 1.1) [7, 8].

While many pathways overlap, the genetic landscape of HNSCC differs between those associated with HPV infection, and those attributed to smoking or alcohol [8]. Almost all smoking-related HNSCC cases show loss or inactivation of *TP53* and *CDKN2A*, while HPV positive tumours rely on the activity of viral oncoproteins E6 and

E7 [8, 13]. E7 binds to and degrades RB causing increased proliferation, while E6 causes the degradation of P53 and activates telomerase activity allowing cells to evade apoptosis [13].

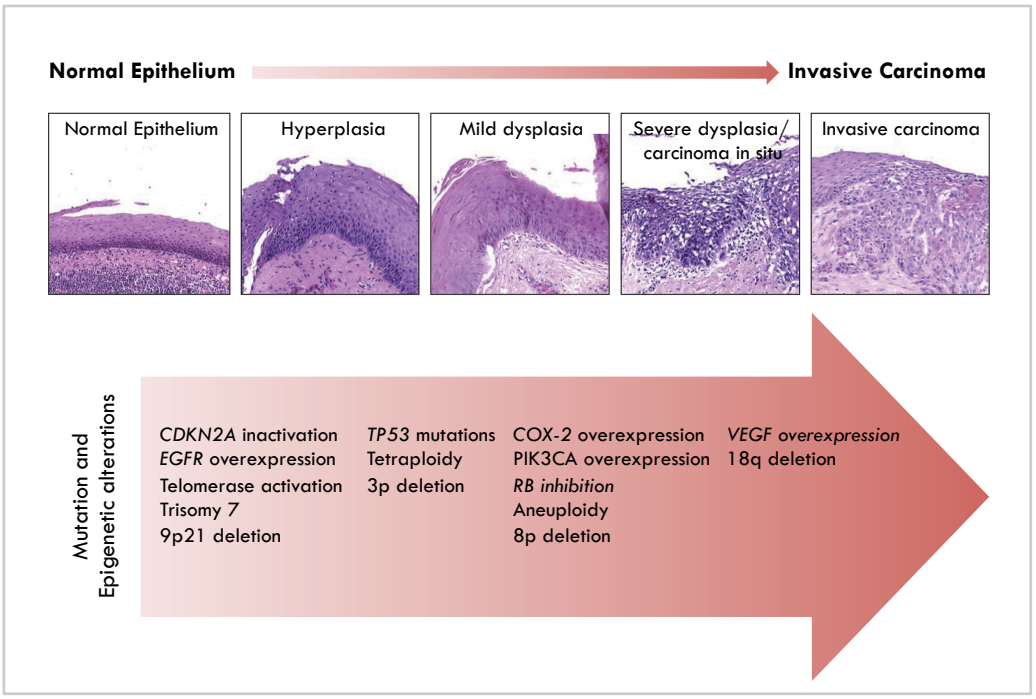



**Figure 1.1. Molecular alterations contributing to HNSCC development**

Mutations and epigenetic alterations contribute to the accumulation of transcriptional changes in HNSCC. In the top panel haematoxylin and eosin staining of tissue sections is used to represent the morphological changes that occur from normal epithelium to invasive carcinoma. The arrow shows the relationship between this progressive transformation and common mutations or gene expression changes known to contribute to HNSCC initiation and development. This figure was adapted from Argiris et al., 2008 [7].

### 1.1.3. Current therapies

When presented at an early stage (stage I-II) HNSCC treatment usually comprises a single agent, either surgery or radiotherapy [14]. However, approximately two-thirds of HNSCC patients present with late stage disease and a more complex response is required, often involving multiple different types of treatment [7, 14]. Traditionally this would be surgery followed by radiotherapy, however locoregional relapse and distant metastasis are common [15]. Furthermore the 5-year survival rate is less than 40% [15]. In 2004 two clinical trials described the benefits of combining radiotherapy with chemotherapy instead of surgery: One examined the effects of radiotherapy followed by cisplatin, while the other examined the concurrent use of the two treatments; both found a significant increase in progression free survival [14, 15]. Treatment of HNSCC is not currently based on molecular markers, instead the decisions are primarily based on the stage of disease at presentation and the age and fitness of the individual [14]. Due to the heterogeneity of HNSCC, clinically identical tumours can respond differently to the same treatment, therefore biological markers and targeted therapies are needed to improve the clinical outcomes [14].

Cetuximab is a monoclonal antibody targeted against EGFR, which, as described in section 1.1.2, is commonly upregulated in HNSCC. Its approval in 2006 was the first new HNSCC drug on the market for 50 years [16]. While cetuximab proved successful in preclinical models, it is only effective in a subset of HNSCC patients [16, 17]. Several possibilities have been proposed to explain resistance to cetuximab including overexpression of *EGFR* and the presence of an *EGFR* mutant that does not allow

ligand binding. Furthermore, the heterogeneity of HNSCC makes the response to cetuximab hard to predict [16-18]. Other EGFR inhibitors have been considered and while these have shown promise in preclinical studies, the current clinical data are unconvincing [16, 18].

Programmed cell death 1 (PD-1) and programmed cell death ligand 1 (PD-L1) are upregulated via various mechanisms in HNSCC, including activation by EGFR or PIK3CA [19]. The results of two clinical trials in 2016 led to USA Food and drug agency (FDA) approval for the PD-1/PD-L1 antagonists' pembrolizumab and nivolumab and the inclusion of these drugs into the standard therapeutic response to HNSCC [19]. DNA and RNA sequencing of HNSCC patients has uncovered other potential targets for molecular targeted therapy, including inhibition of VEGF, PI3K/mTOR and COX-2 [3, 8, 16]. Early results suggest these drugs may have some promise either as single agents or combined with conventional HNSCC treatment regimens [16].

Despite the advances in traditional therapies and the discovery of molecular targets, the survival rate for HNSCC has not significantly improved in recent decades [3]. There is a concern that the benefits of current treatment options have plateaued, therefore there is an urgent need for novel therapies [3]. HNSCC is characterised by epigenetic alterations as well as genetic ones, therefore epigenetic therapies provide a promising approach [2].

## 1.2. Epigenetics

### 1.2.1. Introduction to epigenetics

Since the late 1990s there has been a rapidly growing interest in the field of epigenetics [20]. The term was coined in 1942 by Conrad Waddington who described it as '*the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being*'; namely how a single genotype can give rise to complex phenotypes during development [21-23]. This can be summarised by his now famous drawing of the '*epigenetic landscape*'; a single cell faces several potential fates, the decision of which to go down is determined by epigenetics (Figure 1.2) [22]. The current definition of epigenetics is less well defined, and subject to some controversies [21]. It is a broad term that covers a range of mechanisms that regulate gene expression without altering the sequence of DNA [21, 22, 24].

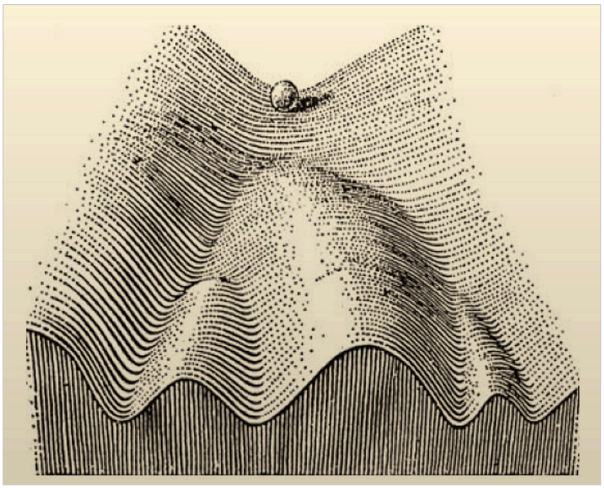


**Figure 1.2. Waddington's Epigenetic landscape**

This image was proposed by Conrad Waddington to represent embryonic development. The ball at the top represents a pluripotent cell and the valleys show the different pathways a single cell can go down to reach a differentiated state. Figure taken from of Goldberg *et al.,* 2007 [22].

### 1.2.2. Gene structure and chromatin

The central dogma of molecular biology states that genetic information stored as DNA is first transcribed into RNA and then translated into protein [25]. However only about 2% of the human genome is translated into proteins [26]. Eukaryotic genes are composed of introns and exons, where exons encode proteins, and introns, containing transposable elements and regulatory sequences, are spliced out of the RNA prior to translation. Promoters mark the start of gene transcription, where transcription factors and RNA polymerase bind to the DNA and transcribe RNA. Therefore, one mechanism of epigenetic regulation involves modifying access of transcription machinery to promoters. An additional level of complexity is added in the form of enhancers and insulators, these are DNA sequences situated at variable distances from promoters which are key to controlling gene expression and also subject to epigenetic control [27]. The structure of a typical human gene is outlined in Figure 1.3.

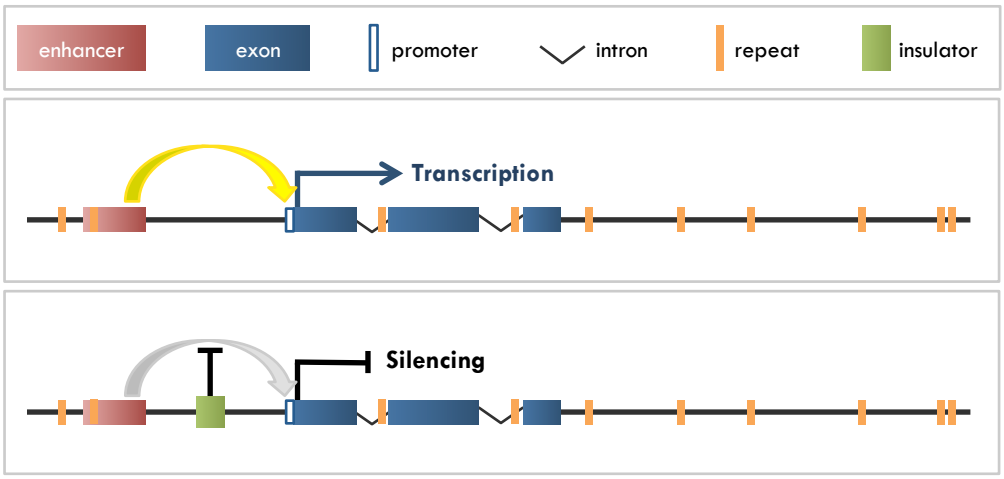



**Figure 1.3. The structure of a typical human gene**

An example of a typical human gene and how it can be regulated by distal regulatory elements is represented as a schematic representing. Enhancers promote expression of the promoter (top), while insulators can prevent the activity of enhancer elements (bottom).

DNA of eukaryotic cells is wrapped around proteins in a complex called chromatin [31]. Chromatin has two functions: to package DNA into the nucleus and to regulate gene expression [31]. The major unit of chromatin is the nucleosome, which consists of 165bp of DNA wrapped 1.67x around a core of histone proteins (duplexes of H2A, H2B, H3 and H4) [31]. Nucleosomes are separated by 10-80bp of DNA and are further packaged into a 30nm thick fibre with the help of the linker histone H1 (Figure 1.4) [28]. In its most compact form (termed heterochromatin) chromatin is inaccessible; therefore the positioning of nucleosomes can determine gene expression by regulating the access of transcription machinery to promoters and regulatory elements [28, 29]. This can be altered in several ways: complexes can remodel the chromatin in an ATP-dependent process; the core histones can be replaced with more permissive histone variants; and 'tails' which extend from the core histones can be modified [28]. Histone tails are subject to more than 100 post-translational modifications; some of these are associated with a permissive chromatin state (e.g. acetylation of lysine residues and methylation of lysine 4 (K4) of histone H3); and some are associated with heterochromatin and gene silencing (e.g. methylation of lysine 9 (K9) or lysine 27 (K27) of H3) [24, 30]. These cooperate with other epigenetic modifications (e.g. DNA methylation) to regulate differential expression of genes throughout development and between cell types [30].

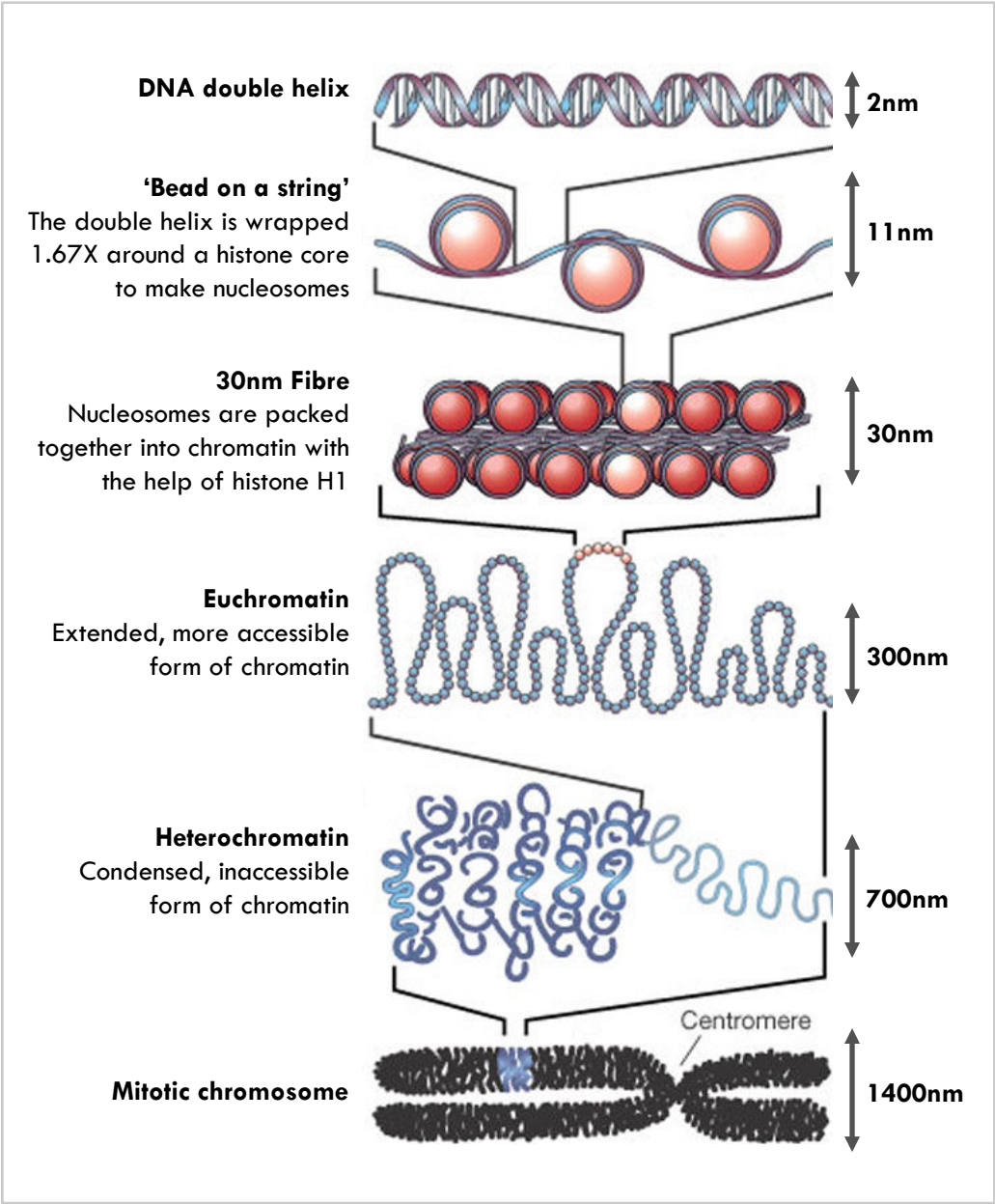



**Figure 1.4: Packaging of DNA**

The double helix of DNA is packaged into nucleosomes, which are then packed together to form a 30nm thick fibre. This fibre is condensed further to form the chromosomes. Some regions are more accessible and are referred to as euchromatin, while densely packed regions are referred to as heterochromatin. Figure adapted from Felsenfeld *et al.,* 2003 [28].

## 1.3. DNA methylation

DNA methylation is the addition of a methyl ($CH_3$) group to DNA. In 1975 two papers hypothesized that DNA methylation was an epigenetic mark that could be inherited across cellular generations [20, 31]. Arthur Riggs proposed that in mammalian DNA methylation would be found at symmetrical sites and that the enzyme responsible for adding the methyl moiety would have a preference for hemi-methylated DNA. Furthermore, he proposed that methylation of the 5' position of the pyrimidine ring of cytosine (5mC) would interfere with protein-DNA interactions [20]. Holliday and Pugh published similar theories the same year [31]. Three years later Adrian Bird confirmed that DNA methylation was primarily found as 5-methylcytosine (5mC) and almost always in the context of palindromic CpG dinucleotides (Figure 1.5) [32]. 5mC is now one of the most well characterised epigenetic modifications and is associated with transcriptional silencing and maintaining stability in the genome [30].

The methyl moiety of 5mC sits in the major groove of DNA. This is where transcription factors usually bind and, as predicted by Riggs in 1975, several proteins are unable to bind to the DNA in the presence of methylated CpGs [24, 30]. An example of this is at the imprint control region of *H19/Igf2*. The maternal locus is not methylated, allowing the transcriptional repressor CTCF to bind, prevent the activity of the *Igf2* enhancer and halt *Igf2* expression. However, at the paternal locus, the region is methylated so CTCF cannot bind and *Igf2* is expressed [20, 24].

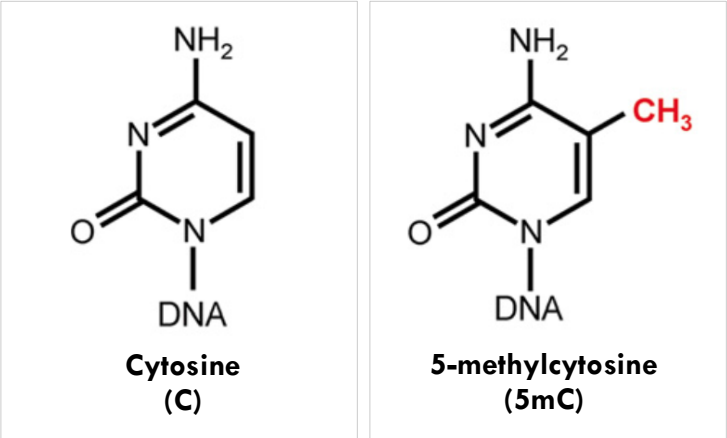



**Figure 1.5. The structure of 5-methylcytosine**

Left panel shows cytosine (C) and the right panel shows 5-methylcytosine (5mC).

### 1.3.1. CpG islands (CGIs)

CpG dinucleotides are underrepresented in the human genome, perhaps due to the spontaneous deamination of 5mC to thymine [33-35]. While mismatch repair machinery can recognise and repair the deamination of unmethylated cytosine to uracil; the deamination of methylated cytosine to thymine is not detected and hence the mutation is passed to future generations [24, 33, 35]. CpG islands (CGIs) are the exception; these regions have a ten times higher frequency of CpGs compared with the rest of the genome (1 per 10bp vs. 1 in 100bp). CGIs are located around the promoters of human genes and are unmethylated; perhaps explaining how such high densities of CpGs persisted throughout evolution [33, 34].

### 1.3.2. DNA methyltransferases (DNMTs)

The addition of a methyl group to DNA is catalysed by DNA methyltransferases (DNMT-1, -3A, and -3B) using S-adenosyl-L-methionine (SAM) as a methyl donor [34].

DNMT1, the maintenance methyltrasferase, has a preference for hemi-methylated DNA [20, 31, 36, 37]. After DNA replication the DNA is hemi-methylated, as the parental strand retains the previous pattern of 5mC but the daughter strand lacks the modification [38]. At the replication fork the UHRF1 protein preferentially binds to hemi-methylated DNA and recruits DNMT1 to copy the pattern of 5mC from the original strand onto the daughter strand; so 5mC is heritable across cellular generations (Figure 1.6)[37, 38].

DNMT3A and DNMT3B are *de novo* methyltransferases that do not differentiate between hemi-methylated or unmethylated DNA [39]. The catalytically inactive homologue, DNMT3L, aids *de novo* methylation during embryogenesis by increasing the ability of DNMT-3A and -3B to bind to the methyl donor, SAM [39]. Both DNMT3s are expressed highest in undifferentiated embryonic stem cells (ESCs), downregulated after differentiation and maintained at a low level in adult somatic tissue [40]. Therefore, it is proposed that the de novo methyltransferases establish the pattern of 5mC during development and this is then maintained by DNMT1 throughout adult life (Figure 1.6) [40, 41].

Initial experiments identified relatively low fidelity of DNMT1 with failure of 5mC maintenance estimated to occur at about 5% of CpG sites per cell division [36, 40]. Furthermore, CGI methylation is stably maintained in the absence of DNMT1 [42]. In a revised model of 5mC maintenance Jones *et al.,* proposed that at methylated CGIs and repetitive DNA where the density of CpGs makes 5mC maintenance more taxing, DNMT-3A and -3B associate with nucleosomes to help maintain the methylation pattern [43].

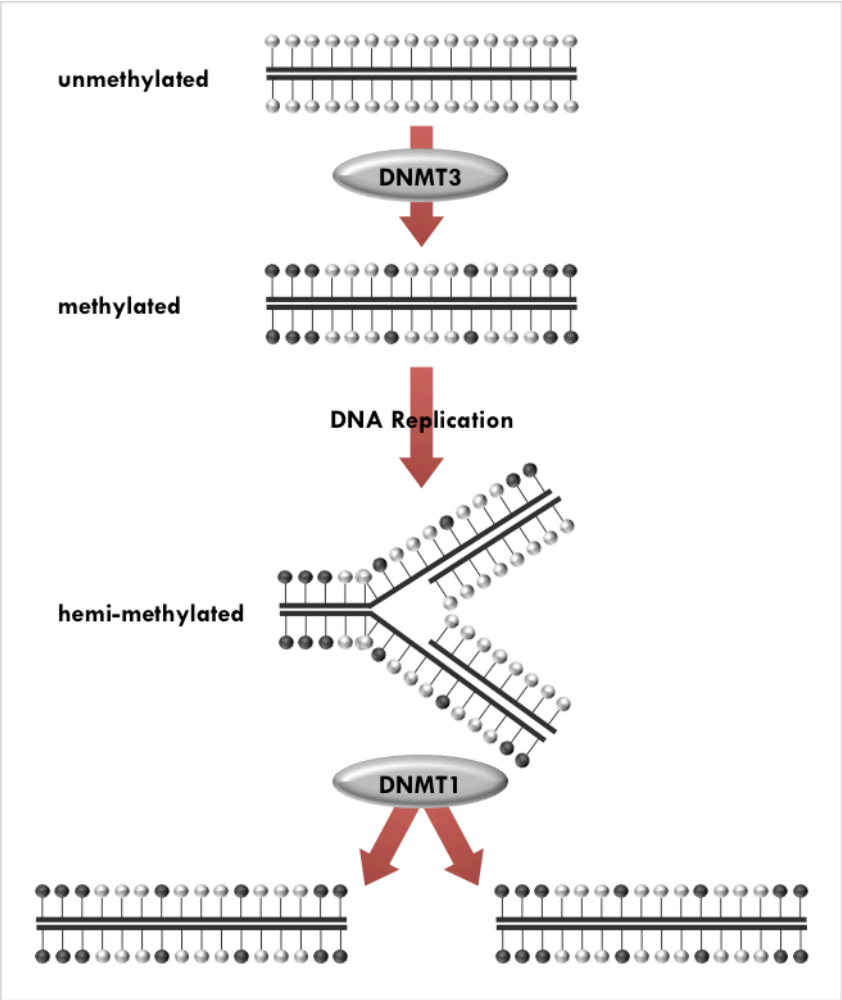



**Figure 1.6. The role of DNA methyltransferases in the establishment and maintenance of DNA methylation**

Unmethylated DNA is recognised and methylated by de novo methyltransferases DNMT3A and 3B. DNA replication results in two hemi-methylated strands of DNA. The maintenance methyltransferase DNMT1 copies the methyl mark onto the unmethylated strand so that the pattern of 5mC is maintained across cellular generations.

### 1.3.3. Transcriptional silencing by 5mC

Two mechanisms contribute to transcriptional silencing by DNA methylation: preventing access of transcription factors to DNA, as is the case for the imprint control region of *Igf2 (1.3)* and interaction with methyl-CpG binding domain proteins (MBPs) (MeCP2, MBD1, MBD2, MBD3, and MBD4) [30, 34, 44]. MBPs bind to 5mC and associate with transcriptional repressors such as histone deacetylases (HDACs) and chromatin remodelling complexes to promote transcriptional silencing [30, 44].

DNMTs and histone modifying enzymes cooperate to maintain the distribution of heterochromatin and euchromatin during DNA replication (Figure 1.7) [43]. Active and nucleosome-depleted regions (accessible chromatin) are characterised by unmethylated CGIs while inactive regions are maintained stably repressed by 5mC and H3K9me3 [27]. UHRF1 is required for DNMT1 localisation to replicating heterochromatic regions and has an affinity for the repressive histone modification, H3K9me [38]. DNMT-3A, -3B and -3L were found to directly bind to all four core histones [45]. DNMT3L can only bind to H3 tails that are not methylated at lysine 4. There it can activate DNMT -3A and –3B to *de novo* methylate the DNA [45]. Therefore, while UHRF1 may regulate a feedback loop to maintain 5mC and H3K9me3 at transcriptionally silenced regions, methylation of H3K4 prevents *de novo* DNA methylation at sites of permissive chromatin [38, 45]. Furthermore, in mESCs, SETD2-mediated tri-methylation of H3K36 has been shown to be involved in the selective recruitment of DNMT3B to the bodies of actively transcribed genes [46].

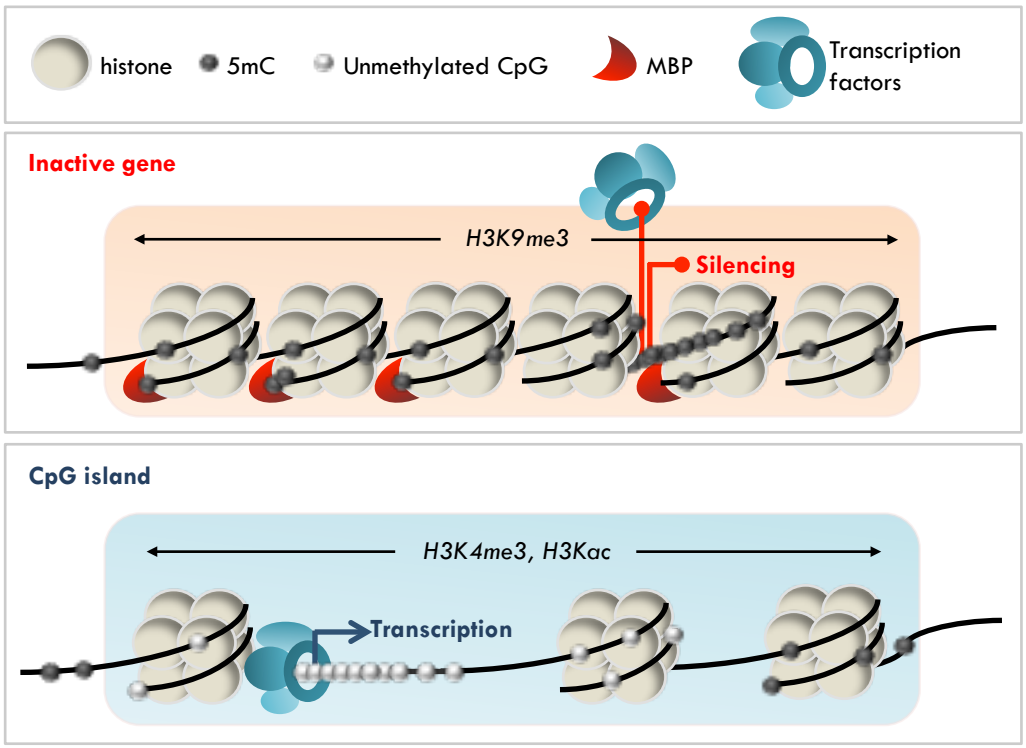



**Figure 1.7. Epigenetic regulation of gene expression**

**Top**: The maintenance of condensed heterochromatin is brought about by a combination of the H3K9me3 histone modification and DNA methylation. Methyl-binding domain proteins (MBPs) bind to 5mC and associate with transcriptional repressors. The presence of 5mC and heterochromatin prevents transcription factors from binding. **Bottom:** At active CpG islands and promoters 5mC is absent from the transcription start site but is present within the gene body to facilitate transcription. The histones are marked with H3K4me3 and H3Kac to maintain an open chromatin state.

### 1.3.4. Genome wide 5mC patterns

Most CGIs are unmethylated and this is understood to allow for transcription factor binding and expression of the corresponding gene. Methylation of CGIs is uncommon and tends to be associated with long term, stable silencing of gene expression [27]. A key example of this is X chromosome inactivation where DNA methylation interacts with other epigenetic mechanisms to stably silence one X chromosome in female cells and allow for dosage compensation [27]. Other examples include imprinted genes and gamete-specific genes that are not required in adult somatic cells [27]. Outside of CGIs the majority of CpGs in the genome are methylated [47]. 50-70% of the human genome is composed of transposable elements and other repetitive DNA [48, 49]. These are highly methylated to maintain genome stability [27]. Unlike promoters, gene bodies tend to be CpG poor and extensively methylated, and this corresponds with active transcription [27]. Gene bodies are rich in transposable elements and it is understood that DNA methylation is used to repress expression of these elements during gene transcription [27].

Enhancers and promoters of highly tissue-specific genes generally lack CGIs [33, 50]. DNA methylation at enhancers and CpG poor promoters is more variable and is often cell type-specific [50, 51]. Unlike most genomic regions, some enhancers show a partially methylated phenotype where a given CpG will be methylated in some cells of a tissue but not in others [27, 51]. This may suggest that enhancers are more prone to methylation changes than CpG rich regions.

The genomic positioning of 5mC and its relationship with CpG density is summarised in Figure 1.8.

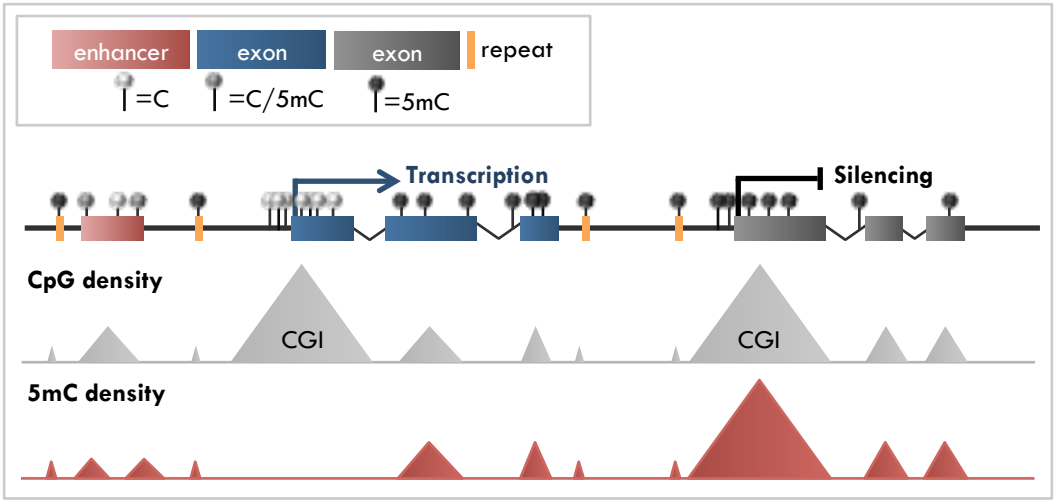



**Figure 1.8: The DNA methylation landscape in mammalian cells**

DNA methylation is located in low density genome wide, and at repeat sequences. CpG islands (CGI) are regions of high CpG density commonly found around transcription start sites. These are generally unmethylated, allowing the corresponding gene to be expressed. Methylation of CGIs leads to gene silencing. This figure shows the correlation between CpG density, 5mC and gene structure and expression (adapted from Baubec et al., 2014 [47]).

## 1.4. DNA hydroxymethylation

For many years 5mC was understood to be the only covalent modification to DNA. However, in 2009, an issue of Science was released in which two papers simultaneously reported the rediscovery of a modification to mammalian DNA called 5-hydroxymethylcytosine (5hmC) [50, 51]. 5hmC is the oxidised form of 5mC (Figure 1.9) [52]. The base was already known to exist in bacteriophages and in the 1970s 5hmC was found in the DNA of frog and rodent brains [52]. For many years this result could not be reproduced and hence it was not until 5hmC was rediscovered that its presence in mammalian DAC was accepted [52]. Kriaucionis and Heintz revealed 5hmC serendipitously when comparing the abundance of 5mC in Purkinje neurons and granule cells. Liquid chromatography mass spectrometry (LC-MS) determined the presence of an unexpected modified form of cytosine which was determined to be 5hmC [53]. Simultaneously, in a search for mammalian homologues of trypanosome proteins, JBP1 and JBP2, which are involved in the synthesis of a modified thymine base, Tahiliani and colleagues identified the TET enzymes [54]. The study determined that TET enzymes could convert 5mC into 5hmC in embryonic stem cells (ESCs) and that 5hmC was stable in mouse ESCs making up ~4% of all CpG dinucleotides [54].

Since then, research into the potential function of 5hmC has become a vastly expanding field of epigenetics with the abundance of the base now determined in the majority of mouse and human tissues [52, 53, 55-61]. The level of 5hmC varies greatly between tissues and stages of development and appears to be inversely correlated with proliferation [53, 54, 58, 62]. Furthermore, overexpression of TET

proteins can lead to further conversion of 5hmC into 5-formlycytosine (5fC) and 5-carboxylcytosine (5caC) [63]. While 5hmC is found in relatively high abundance, 5fC and 5caC are found at very low levels in mammalian DNA, 10-100 fold less than 5hmC, signifying rapid removal of 5fC and 5caC and a potential regulatory function of 5hmC [63].

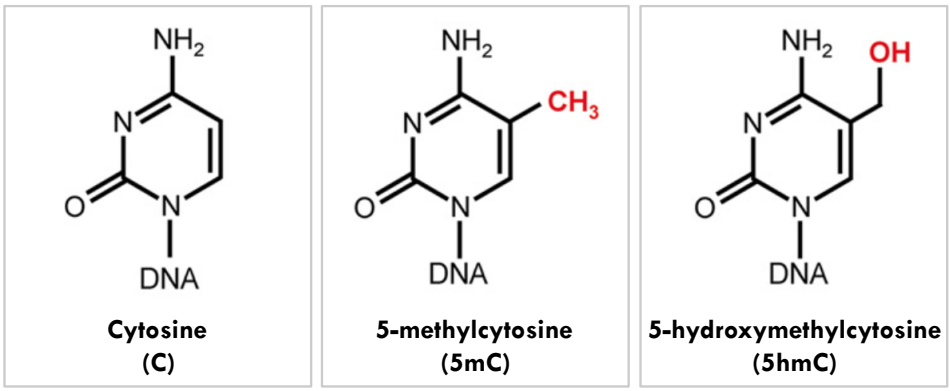



**Figure 1.9. The chemical structure of cytosine and cytosine modifications**

The addition of a methyl group to cytosine (C) (left) leads to 5-methylcytosine (5mC) (middle). This can then be oxidised to form 5-hydroxymethylcytosine (5hmC) (right).

### 1.4.1. TET enzymes

A family of Fe (II) and α-ketoglutarate (α-KG) -dependent dioxygenases called Ten-Eleven Translocation (TETs) catalyse the oxidation of 5mC into 5hmC [81]. There are three *TET* genes (*TET1, TET2* and *TET3*) in mammalian DNA [64]. Structurally, the enzymes are composed of a double-stranded beta-helix fold that contain metal binding residues and TET1 and TET3 contain a CxxC domain the recognises CpG-rich DNA; while TET2 is understood to partner with neighbouring CxxC containing protein, IDAX [65-67]. The TET1 CxxC domain has been shown to bind both modified

and unmodified CpGs [68, 69]. Knockdown of each TET enzyme in an embryonic carcinoma cell line determined differential effects of the three proteins [61]. While there was some overlap between knockdowns, TET1 depletion led to a 60% loss of global 5hmC whereas TET2 and TET3 depletion caused region specific 5hmC loss [61]. TET2 was associated with highly transcribed genes, enriched for active histone marks such as H3K4me3 where it is understood to be involved in removing 5hmC via conversion to 5caC and 5fC [61]. TET3 is primarily expressed in oocytes and zygotes where it is thought to be responsible for epigenetic reprogramming during mammalian development [61, 64]. TET1 associates with CGI promoters, potentially acting as a control mechanism against aberrant *de novo* methylation of active genes [27]. In cultured cells, *TET1* overexpression correlates with a decrease in 5mC and only a small increase in unmodified C; perhaps indicating its primary role in the conversion of 5mC to 5hmC while TET2 and TET3 further convert 5hmC into 5fC and 5caC [61, 64].

### 1.4.2. 5hmC function as opposed to 5mC

The most widely accepted function of 5hmC is as an intermediate in the DNA demethylation pathway. However, the levels of 5hmC are considerably higher than that of 5fC or 5caC and 5hmC has a distribution distinct from 5mC, suggesting the base may additionally function as an epigenetic modification [63, 70]. As discussed in section 1.3, the majority of 5mC is found outside of genes, associated with heterochromatin marks such as H3K9me3 [38, 47, 71]. However, 5hmC tends to be in gene-rich regions, at gene bodies and enhancers, and co-localises with active histone marks such as H3K4me2/3 (Figure 1.10, Figure 1.11)[52, 61].

Transcriptional silencing by 5mC is attributed to both preventing the access of transcription factors to DNA and interaction with methyl-binding proteins (MBPs) to promote heterochromatin formation [30, 34, 44]. An artificial system was developed to determine the relationship between 5hmC and transcription factor binding [72]. The presence of a single hemi-hydroxymethylated CpG or a single hemi-methylated CpG within a cAMP response element (CRE) was enough to reduce binding of the transcriptional activator, CRE binding protein (CREB), and cause a corresponding decrease in reporter gene expression. Therefore, a single 5hmC in the critical CpG is enough to reduce transcription factor binding [72]. In fact, a large number of proteins have a reduced affinity for DNA in the presence of either 5mC or 5hmC [73, 74]. Very few proteins bind to both DNA modifications and 5hmC impairs binding of a number of methyl-binding proteins [62, 73, 74]. These results suggest that 5hmC may prevent gene expression by inhibiting transcription factor binding; however, unlike 5mC, 5hmC does not colocalise with heterochromatin. Conversely, the presence of 5hmC in gene bodies and enhancers has been found to correlate with gene expression [4, 9, 21, 22]. TET2 knockdown correlates with loss of 5hmC from enhancers and reduced expression of adjacent genes [7, 22, 23]. In another study, high intragenic 5hmC correlated with organ-specific gene expression in mouse brain and liver cells [24].

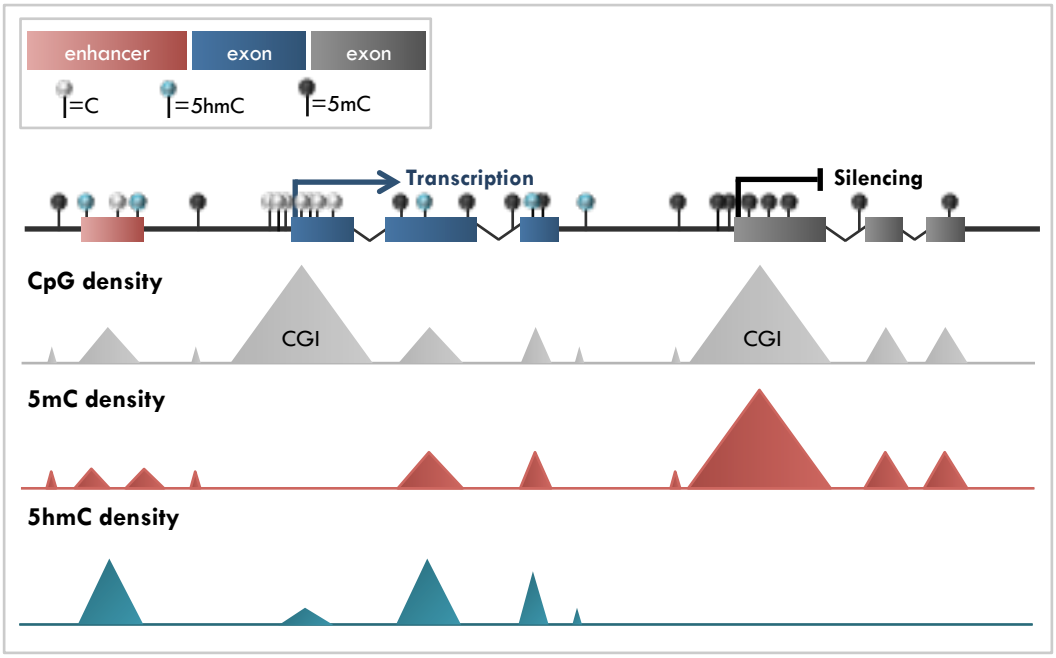



**Figure 1.10. 5hmC genomic distribution**

This figure represents the abundance of 5hmC and 5mC relative to CpG density, gene structure and transcriptional activity.

In summary, 5hmC plays a distinct role from 5mC in regulating gene expression. Unlike 5mC, 5hmC is not associated with heterochromatin marks such as H3K9me3. Instead 5hmC is enriched at enhancers and gene bodies of organ-specific genes; or found alongside bivalent chromatin marks to help maintain a poised chromatin state (Figure 1.10, Figure 1.11). These may be two distinct, but complementary, roles for 5hmC in preventing the formation of heterochromatin associated with 5mC (Figure 1.11)[72, 73].

The results of TET1 knockdown support this theory as while numerous genes are downregulated, a surprisingly large number are activated [64]. In ESCs these derepressed genes frequently correspond with sites of bivalent chromatin, where both activating (H3K4me3) and repressive (H3K27me3) marks are present [64]. Bivalent chromatin has been associated with a 'poised' chromatin state, generally associated with developmental genes that are primed to be activated or repressed depending on which differentiation pathway is pursued [64]. Therefore, the semi-repressive state of 5hmC may further contribute to this poised state of transcription (Figure 1.11).

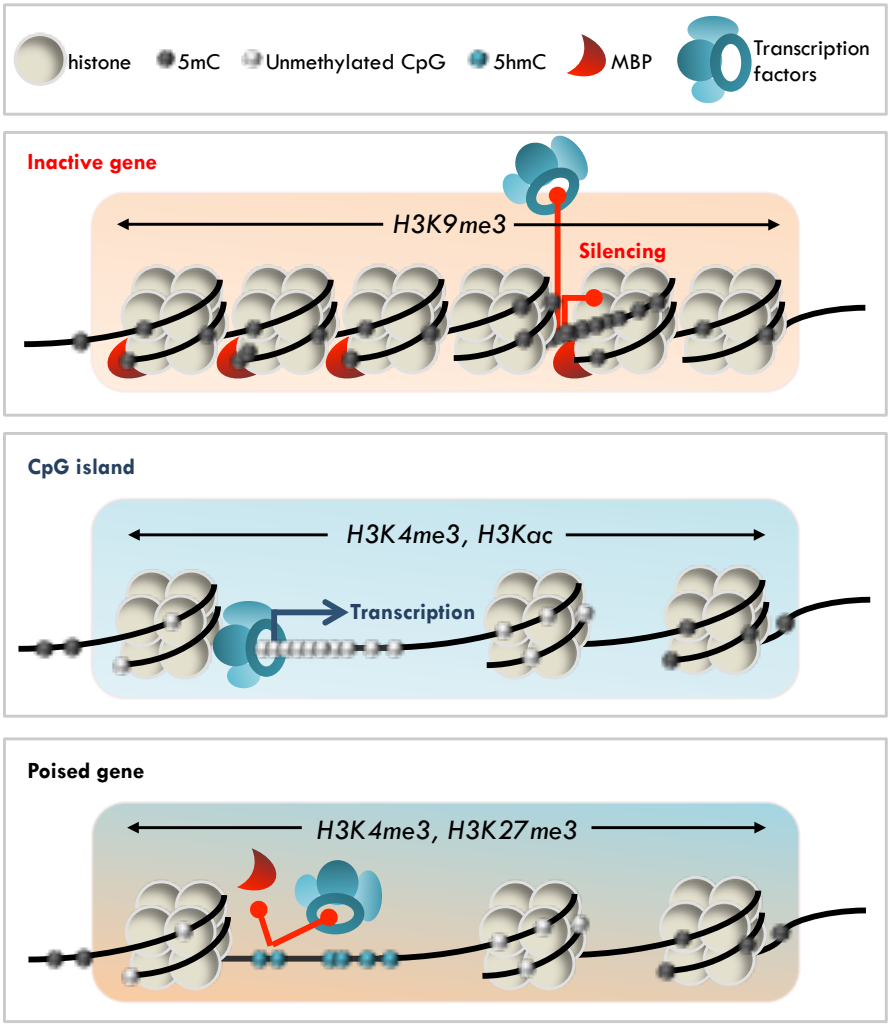



**Figure 1.11. 5hmC and gene expression**

**Top**: H3K9me3 and 5mC maintain heterochromatin. **Middle**: H3K4me3, H3 lysine acetylation and an absence of 5mC, mark active CpG islands. **Bottom**: 5hmC is found at poised genes, associated with the active histone mark H3K4me3 and the repressive histone mark H3K27me3. 5hmC inhibits the binding of both methyl-binding proteins (MBP) that associate with heterochromatin and transcription factors that promote gene expression. Therefore the gene is neither repressed nor activated, but remains in a 'poised' state.

### 1.4.3. DNA demethylation

DNA methylation can be lost from DNA passively via replication if DNMT1 fails to copy the methyl mark onto the daughter strand. This can happen when levels of DNMT1 are low and when 5mC has been oxidised to 5hmC, as DNMT1 poorly recognises 5hmC [54, 75]. DNA demethylation is also known to occur as an active process, independent of DNA replication [76]. For example, after fertilisation, the majority of 5mC is lost from the paternal genome before replication is initiated [64]. The discovery of 5hmC, 5fC and 5caC provided a major breakthrough in elucidating the mechanism of active demethylation as they are understood to represent the first steps in this pathway (Figure 1.12) [42]. The DNA repair protein thymine DNA glycosylase (TDG) excises 5fC and 5caC from the DNA and the base excision repair (BER) machinery replaces the missing base with an unmodified cytosine (Figure 1.12)[76, 77]. Accordingly, depletion of the TDG enzyme results in an accumulation of 5fC and 5caC [76, 77].

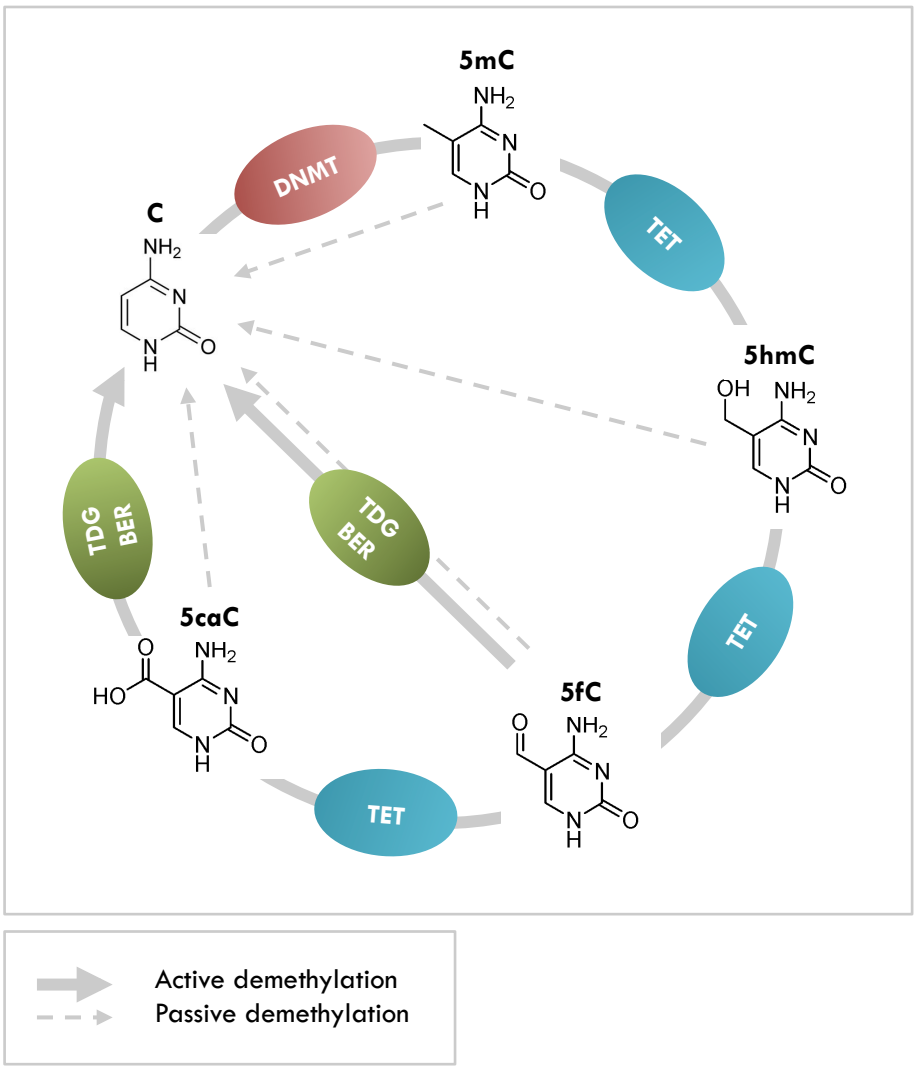



**Figure 1.12. DNA demethylation**

DNMT enzymes catalyse the addition of a methyl group to DNA, generating 5-methylcytosine (5mC). This can be oxidised by the TET enzymes to form 5-hydroxymethylcytosine (5hmC). 5hmC can be further oxidised by the TET enzymes into 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). 5mC, 5hmC, 5fC and 5caC can be lost passively via dilution during DNA replication. 5fC and 5caC are also subject to active removal by thymine DNA glycosylase (TDG), leaving a gap in the DNA that is repaired by the base excision repair (BER) pathway.

## 1.5. Transposable elements

Repetitive DNA, composed of tandem repeats and transposable elements, accounts for 50-70% of the human genome (Figure 1.13) [48, 49]. Tandem repeats are short non-coding sequences that are repeated head to tail [48]. They help maintain genome stability; at telomeres they protect chromosome ends and at centromeres they facilitate sister chromatid cohesion [48]. Transposable elements (transposons) are mobile DNA that have replicated and reinserted themselves across the human genome. Transposons have significantly contributed to human evolution, however the rate of transposition has been in steady decline in hominids since mammalian radiation [78, 79].

### 1.5.1. Transposable element families

Two classes of transposons exist: Class I, retrotransposons, that use an RNA intermediate to move about the genome and class II, DNA transposons, that excise and reintegrate themselves from one position to another [48].

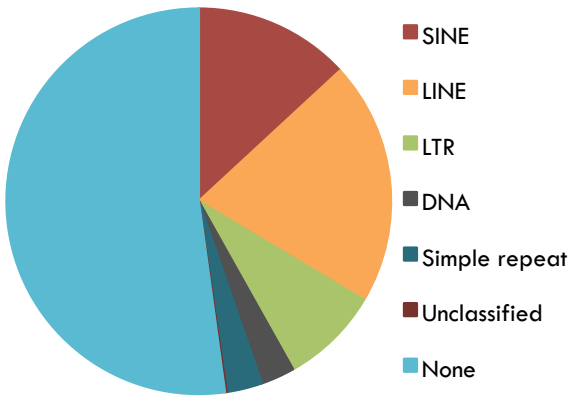



**Figure 1.13. Repetitive DNA in the human genome sequence**

The proportion of the human genome occupied by repetitive DNA classes. SINE= short interspersed repetitive element; LINE= long interspersed repetitive element; LTR= long terminal repeat; DNA=DNA transposon; simple repeat=short and tandem repeats [79].

### *Long terminal repeat (LTR) retrotransposons*

LTR retrotransposons are derived from ancient retroviral infections and contain long terminal repeats (LTRs) at the 3' and 5' ends [48]. LTRs are autonomous, containing the gag and pol genes that encode the protease, reverse transcriptase, RNase H and integrase required for retrotransposition (Figure 1.14) [79]. Endogenous retroviruses (ERVs) are the only active LTRs remaining in the mammalian genome [79].

### *Long interspersed nuclear elements (LINEs)*

LINEs are autonomous, class I retrotransposons, typically around 6kb long [79]. LINEs have an RNA polymerase II promoter and encode proteins ORF1 and ORF2, which facilitate their retrotransposition (Figure 1.14) [79, 80]. In humans the most prevalent LINE family is LINE-1, which accounts for ~17% of the human genome, only L1Hs remains active (Figure 1.14)[79].

### *Short interspersed nuclear elements (SINEs)*

SINEs are non-autonomous retrotransposons of 100-400bp in length [79]. SINEs are transcribed via RNA polymerase III and rely on LINE proteins to transpose [80]. The most prevalent human SINE family are the primate-specific Alu elements [79, 81]. These evolved from the 7SL RNA gene and make up 10-11% of the human genome [81]. Structurally Alu elements are ~280bp long, composed of two diverged dimers separated by an A-rich region, with a longer A-rich region at the 3' end (Figure 1.14) [81]. Alu elements continue to transpose in humans, with a new insertion understood to occur roughly 1 in every 20 births [81, 82]. Three subfamilies of Alu elements exist; Alu Y, the most recently evolved (youngest) and least prevalent; Alu

J, the oldest; and Alu S, middle-aged but most abundant [83]. Only Alu Y remains active (Figure 1.14) [80].

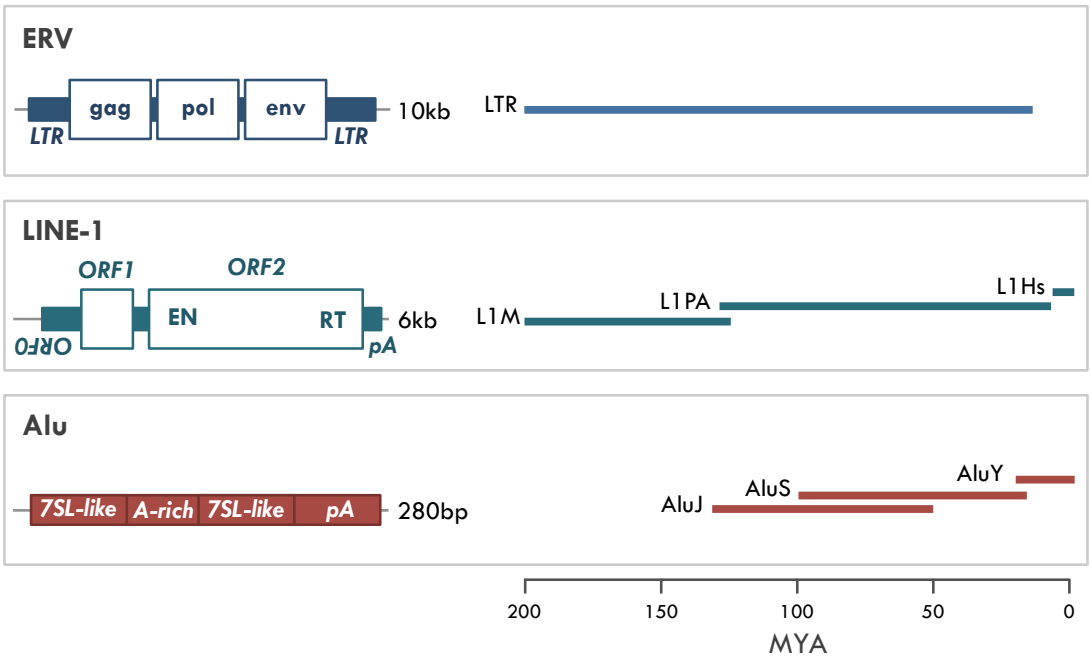



**Figure 1.14. Human retrotransposons**

The structure (**left**) and age (**right**) of human retrotransposons are represented as a schematic. **Top**: ERVs are LTR transposons, encoding the gag, pol and env genes required for their retrotransposition. The transposon is flanked by long terminal repeats (LTRs) [80]. **Middle**: LINE-1 is the most common LINE element in modern humans. ORF1 and ORF2 are the proteins required to facilitate retrotransposition. **Bottom**: the primate-speficic Alu elements make up the majority of human SINE elements. The retroelement encodes no proteins, but is made up of a dimer of 7SL-like sequences, separated by an A-rich region. A polyA tail at the 3'end facilitates its amplification. **MYA**= million years ago.

### 1.5.2. Epigenetic regulation of transposable elements

The insertion of transposons into DNA can directly cause mutations in protein coding regions, can influence splicing and can act as novel enhancers or promoters (Figure 1.15) [78, 80]. These alterations can be deleterious, therefore it has been proposed that epigenetic silencing evolved as means of repressing the expression of transposons [82]. Indeed, many transposons are enriched in 5mC and repressive histone modifications [82].

Alu elements contain 25% of all CpG dinucleotides in the human genome [81, 84]. This is distinct from other repeat sequences; LINEs are preferentially found within AT-rich regions and the proportion of CpGs found within LTRs correlates with their genomic coverage [84]. The CpGs within Alu elements are generally methylated. Primate-specific transposons show higher, more consistent levels of DNA methylation across tissues and between individuals [85, 86]. This may reflect the fact that these transposons are more likely to be active and therefore the need for silencing by repressive chromatin is greater [86]. Accordingly, the most recently evolved family of Alu elements, Alu Y, shows greater polymerase occupancy and greater DNA methylation than older families [86, 87].

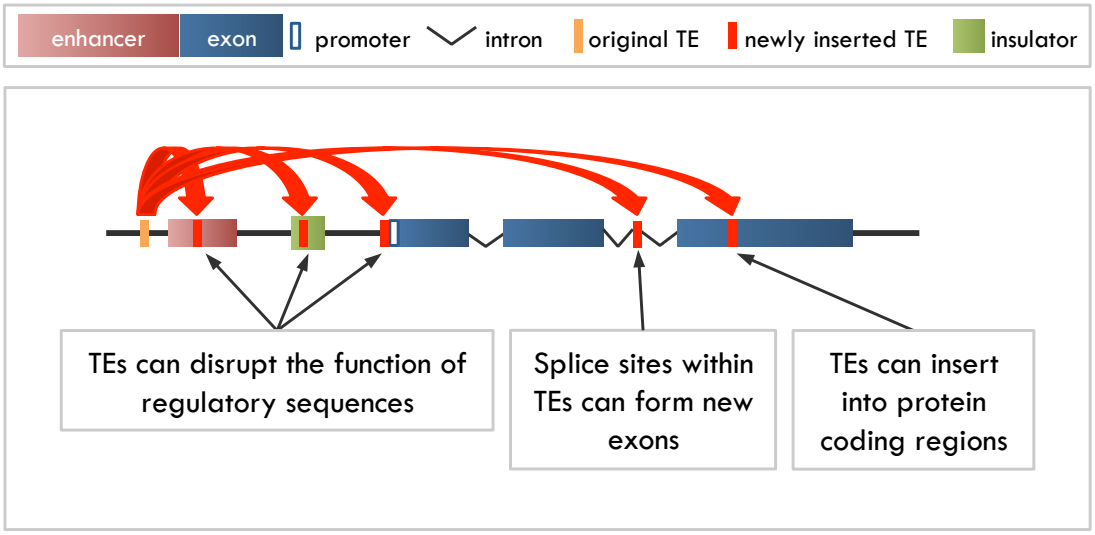



**Figure 1.15. The mutagenic potential of retrotransposons**

The insertion of retrotransposons into non-coding DNA can disrupt regulatory sequences such as enhancers or insulators, or can form new enhancers/ promoters; Alu elements can insert into gene bodies, providing new splice sites which can lead to exonisation; retrotransposons can insert into protein coding regions causing mutations, truncations and loss of function [80]. TE: transposable element

## 1.6. Epigenetic alterations in cancer

Epigenetic mechanisms tightly control gene expression to maintain differentiation of somatic cells. Disruptions to this control are a major contributor to cancer initiation and progression [88, 89]. Alterations to the pattern of 5mC are the most widely studied example of this [89]. Cancer cells exhibit both genome wide hypomethylation leading to genomic instability and hypermethylation of promoter CGIs leading to silencing of tumour suppressor genes (Figure 1.16) [34]. These 5mC changes are accompanied by changes in histone modifications and chromatin structure: global hypomethylation is accompanied by abnormal lysine acetylation and an open chromatin structure; while hypermethylated CGIs are associated with repressive histone marks and stable nucleosome positioning [90]. Furthermore, 5hmC is consistently reduced in many forms of cancer [91]. Epigenetic changes, such as alterations in 5mC and 5hmC, may play a seminal role in the earliest steps of cancer initiation; they tend to be early events in carcinogenesis and may initiate the expansion and evolution of pre-malignant cells into invasive carcinomas [88, 89, 92].

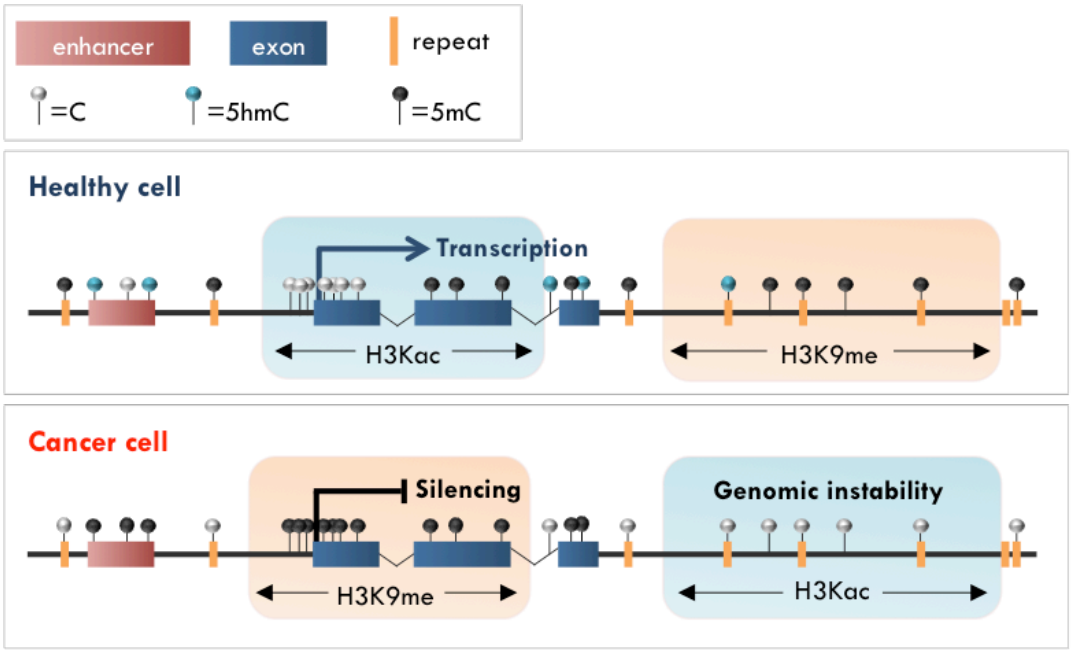



**Figure 1.16. Epigenetic alterations in cancer**

During cancer the pattern of 5mC, 5hmC and histone modifications are disrupted. **Top panel:** in healthy cells 5mC is found genome wide and across repeat sequences where it is associated with H3K9me3. CGIs are free of 5mC and associated with active histone marks such as acetylation of lysine residues. 5hmC is found at gene bodies and enhancers. **Bottom panel:** in cancer cells this pattern is reversed. A global decrease in 5mC and gain in H3Kac causes genomic instability while 5mC and repressive histone marks gained at CGIs leads to gene silencing. 5hmC is reduced globally in cancer.

### 1.6.1. Genome wide hypomethylation

Global loss of 5mC has been linked to cancer progression and metastasis [34]. This loss primarily occurs at CpG poor regions such as repetitive DNA, introns and gene deserts [2]. It increases genomic instability by loosening the chromatin around the bulk of the genome and allowing transcription of transposable elements [2]. In a mouse model of skin cancer global hypomethylation was observed as a two-stage process: first 25% of 5mC was lost during the transition from normal epithelium to benign papilloma cells; then a further 50% was lost when the cells transformed into a highly metastatic dedifferentiated morphology, suggesting that 5mC loss could be a marker for malignant transformation [93]. Hypomethylation of LINE1 repetitive elements have been identified as a putative marker for OSCC [2]. While hypomethylation primarily targets intergenic and intronic DNA, loss of methylation at genes usually silenced in the given tissue have also been described in HNSCC cells, including *PI3, AIM2, SPP1* and *ADGRES* [2, 94].

### 1.6.2. CpG island hypermethylation

Epigenetic silencing by 5mC of normally unmethylated CGI promoters is common to almost all types of cancer with 5-10% of all promoter CGIs methylated in cancer [90]. This can include the methylation of known tumour suppressor genes, such as those associated with familial forms of cancer such as *RB1, VHL, BRCA1* and *CDKN2A* [92]. Elevated levels of DNMT-1, -3A and -3B have been reported in cancers of the colon, prostate, breast and liver suggesting a potential mechanism for the hypermethylation observed [39]. In a 'two hit' model of cancer initiation, silencing by 5mC can often provide the second 'hit' to a somatic or hereditary mutation in a

tumour suppressor gene [92]. Interestingly, there is also a subset of genes including the repair gene *MGMT*, the cell cycle regulator *CDKN2B,* and a gene encoding for the RAS binding protein, *RASSF1* that are commonly silenced by 5mC in tumour cells, but which are rarely mutated [92]. Interestingly, CGI methylation changes in cancer do not always correspond with gene expression changes. Instead, many of the genes methylated in cancer were already silenced in the normal tissue [95]. In these situations, the addition of the methyl group may act by stabilising the gene silencing and therefore preventing differentiation or reactivation of the gene in the later stages of tumour progression [95].

Epigenetic silencing of tumour suppressor genes can occur at any point during carcinogenesis, however it commonly occurs during the very early stages and may predispose cells to genetic abnormalities that advance the neoplastic process [88]. For example silencing of the cell cycle regulator *CDKN2A* could allow cells to evade apoptosis and silencing of the DNA repair gene *MGMT* may promote DNA damage [2, 88, 96]. The chemical structure of 5mC can also increase mutation rates. The mark is more sensitive to UV radiation and as described in section 1.3.1, transition of 5mC→ T often eludes DNA repair machinery. Nearly 50% of the inactivating point mutations in the tumour suppressor gene *TP53* occur at 5mC sites [90, 92, 96]. Therefore, hypermethylation of CGIs in premalignant cells can promote cancer formation by creating mutational hotspots, repressing transcription of DNA repair machinery and silencing key tumour suppressor genes [90, 92, 96].

### 1.6.3. CpG island hypermethylation in HNSCC

Hypermethylated CGIs have been proposed as potential biomarkers of HNSCC [97]. Panels of several tumour suppressor genes were found to be more sensitive as diagnostic markers than single gene tests [97]. For example, in an analysis of 95 HNSCC patients 55% showed CGI hypermethylation in the promoter of at least one of *CDKN2A, MGMT, GSTP1* and *DAPK1* [98]. Another panel, consisting of *CDKN2A, DAP-Kinase, CDH1* and *RASSF1,* was able to identify 60% of HNSCC cases, with 5mC at CGIs increasing with tumour size [99]. The use of a larger panel of 16 tumour suppressor genes found that 5mC of CGIs was high in poorly differentiated tumours and increased in advanced tumours [100]. These results, and others, suggest that hypermethylation of CGIs may provide some diagnostic power for HNSCC, but that a well-defined panel of tumour suppressor genes must first be decided upon [97-102].

### 1.6.4. Tumour suppressor potential of TET proteins

Prior to the rediscovery of 5hmC in 2009, the TET proteins were already known to be associated with cancer, with reduced expression of all three TET genes identified in different cancer types [54, 57, 91, 103]. TET1 has been identified as a fusion partner of the MLL protein in rare cases of AML, and is frequently downregulated in solid tumours [54, 104, 105]. TET2 is inactive in about 15% of all myeloid cancers; including 22% of AML [91]. In pre-leukemic haematopoietic cells *TET2* loss contributes to leukemic transformation by deregulating a large number of genes involved in tumourigenesis [106]. Furthermore, overexpression of *TET2* in melanoma cells results in a re-establishment of 5hmC levels and smaller tumours developing

upon injection into nude mice [59]. Therefore TET2 can be considered a tumour suppressor protein.

The TET enzymes require the cofactor α-KG in order to convert 5mC to 5hmC [52]. This is produced during the Krebs cycle through the decarboxylation of isocitrate by IDH (-1, -2, -3) [107]. A mutant form of *IDH* is found in AML, gliomas and some other cancers, this produces α-hydroxyglutarate (α-HG) in place of α-KG [91]. α-HG is an oncometabolite that is structurally similar to α-KG and can bind and inactivate TET proteins in place of α-KG (Figure 1.17) [62, 108, 109]. *IDH* mutations are mutually exclusive with those of *TET2* loss [109].

### 1.6.5. 5hmC depletion and diagnostic potential

In a meta-analysis of published papers examining the relationship between 5hmC and cancer, Chen and colleagues determined a significant association between 5hmC loss and cancer progression [55]. The study included 1736 patients with different cancer types and determined that low to non-existent 5hmC levels significantly correlated with lymph node metastasis, advanced TNM stage and a poor prognosis, suggesting that 5hmC levels may function as a diagnostic and prognostic marker [55]. Of particular interest to this project is the inverse correlation between 5hmC and oral squamous cell carcinoma (OSCC) [57, 110]. Contrary to the metaanalysis, high 5hmC levels in OSCC are associated with aggressive tumour features and an unfavourable prognosis [55, 110]. This could be accounted for by differences in sample size, with only 95 OSCC patients examined in the OSCC study, or it may correspond to the variation in 5hmC levels found in healthy tissues [55, 110]. Either

way, further work is required to determine the prognostic potential of 5hmC in OSCC and other forms of HNSCC [55, 110].

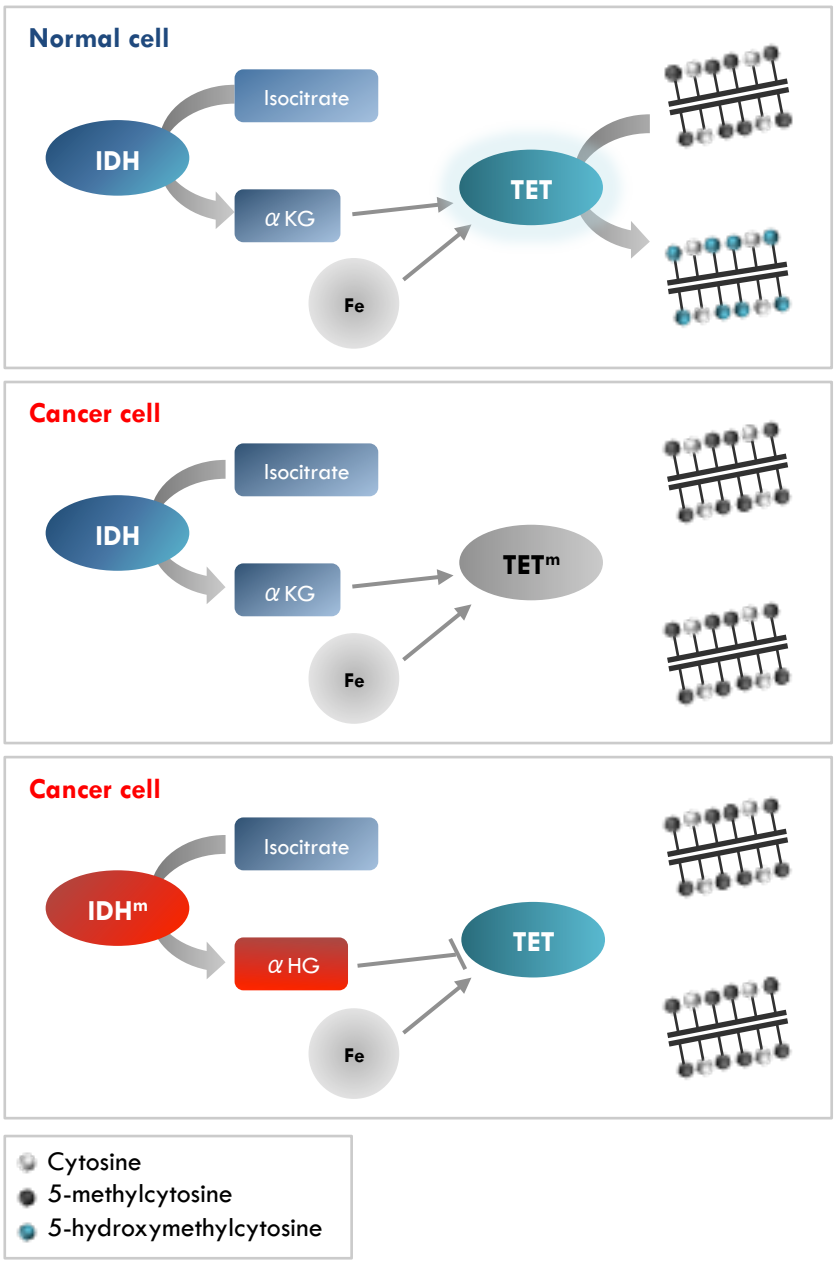



**Figure 1.17. 5hmC and cancer**

**Top**: In healthy cells IDH produces α-KG from isocitrate. TET proteins use α-KG and Fe(II) as cofactors to convert 5mC into 5hmC. **Bottom two panels:** 5hmC is reduced in cancer. **Middle**: TET is often mutated in leukemias, so 5hmC is not formed. **Bottom**: In cancer, a mutant form of IDH is often found that converts isocitrate to α-HG instead of α-KG. α-HG can bind to and inactivate TET proteins so no 5hmC is produced.

### 1.6.6. Loss of epigenetic control of transposable elements in cancer cells

Genome wide hypomethylation in cancer predominantly occurs at repeat sequences and is associated with decondensed chromatin; transposable elements are then free to transpose, causing mutations that further promote tumour progression (examples of this are outlined in (Figure 1.15) [78, 84, 111, 112]. A comparison of non-small cell lung cancer with paired adjacent tissue found a significant reduction in 5mC at LINE-1 and Alu elements and determined that this was associated with genomic instability [113]. Similar correlations have been found in other cancers including mucoepidermoid carcinoma, colorectal and prostate cancers [87, 111, 114, 115]. In HNSCC, primate specific repeats that are highly methylated in normal tissues are subject to a more substantial reduction in methylation than older families [86]. This may be selected for during cancer progression due to the higher mutagenic potential of these active elements [86]. Furthermore, overexpression of the LINE-1 protein, ORF1p is found in 90% of breast, ovarian and pancreatic cancer and 50-60% of oesophageal and colon cancers [80]. Therefore a clear correlation exists between the loss of epigenetic control of transposable elements and cancer progression [80].

## 1.7. Epigenetic therapies in cancer

Unlike genetic changes, epigenetic changes are reversible, making them a promising therapeutic target. The two main mechanisms for reversal of epigenetic silencing are inhibition of DNMT enzymes and inhibition of histone deacetylases (HDACs) (Figure 1.18)[2]. Currently two DNMT inhibitors (DNMTi) and four HDAC inhibitors (HDACi) have FDA approval, while the European Medicines Agency (EMA) has only approved the DNMTis [2, 116].

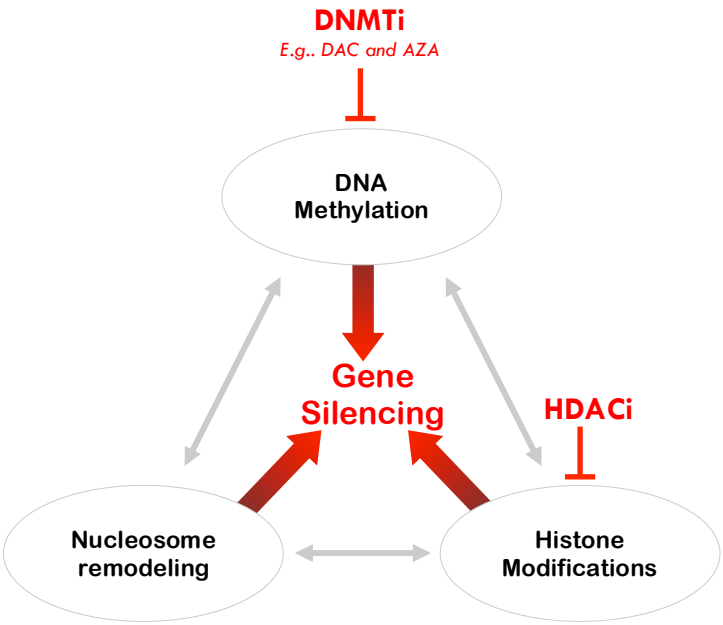



**Figure 1.18. Mechanisms of epigenetic gene silencing and the potential for intervention**

In normal and cancer cells the relationship between 5mC, histone modifications and nucleosome positioning maintains gene silencing. Therefore, inhibiting DNMTs or HDACs may provide a potential mechanism to reactivate epigenetically silenced tumour suppressor genes. Figure adapted from Jones and Baylin, 2007 [96].

### 1.7.1 DNMTi

5-Aza-2'-deoxycytidine (Decitabine or DAC) and 5-azacytidine (AZA) are cytosine analogues that act as DNMTis and DNA demethylating agents by incorporating into DNA in place of cytosine [117]. In preclinical models DAC is estimated to be 10 times more potent than AZA [118]. This is because DAC has a deoxyribonucleoside structure, while AZA has a ribonucleoside structure and 80-90% of AZA is incorporated into RNA instead of DNA [75, 118]. Elevated levels of the DNMT enzymes have been found in many cancers, therefore inhibition of DNMTs by DAC or AZA can reactivate silenced tumour suppressor genes and decondense the chromatin to induce cellular differentiation and stop tumour growth [39, 119].

High doses of DAC and AZA were initially tested as cytotoxic chemotherapy agents in the 1960s [75]. Several years later their DNA demethylating ability was determined and the two drugs received FDA approval for the treatment of myelodysplastic syndrome (MDS) in 2006 and 2004 [118, 120, 121]. Further success has been found in their use in the treatment of other haematological malignancies such as acute and chronic myeloid leukaemia (AML and CML) with DAC receiving EMA approval for use in AML in 2012 [116, 122, 123].

### 1.7.2 Current usage and limitations of 5-aza-2′-deoxycytidine (DAC)

As a single agent the response rate to DAC in the treatment of AML is reported as being between 33-89%, with hypomethylation correlating with patient response [124, 125]. Myelosuppression is the most widely recognised side effect to DAC treatment with neutropenia and thrombocytopenia common in MDS and AML

patients [118, 122, 126]. High doses of DAC induce DNA damage and cytotoxicity [71]. However, it has been observed that low dose treatments are just as effective [90, 122, 127]. *In vitro* work by Tsai *et al.,* demonstrated that low dose DAC treatment resulted in decreased promoter methylation, subsequent gene re-expression and the inhibition of subpopulations of both primary leukaemia and cultured breast cancer cells [117].

### 1.7.3 Potential efficacy of DAC in solid tumours

Epigenetic therapies only have FDA and EMA approval for the treatment of haematological malignancies [120]. However, preclinical studies suggest that DAC and AZA treatment may be transferable to solid tumours [117, 120, 121]. Cultured and primary breast tumour epithelial cells were subject to transient, clinically relevant, nanomolar doses of DAC and AZA and an anti-tumour 'memory' response was observed. This included widespread gene expression changes, inhibition of subpopulations of cancer stem-like cells and significantly reduced xenograft formation when injected into NOD/SCID mice [117].

In a review of published records of clinical activity, DAC was found to exhibit 35% response rate in breast cancer, 34% in colorectal, 36% in pancreatic and only 9% in HNSCC [125]. However, HNSCC patients make up a very small proportion (only 29 cases) of all patients treated with DAC [125]. Indeed, the only clinical trial involving DAC treatment in HNSCC was a phase II study in 1987 for a variety of solid tumours that determined that at the dose and schedule examined DAC was devoid of anti-tumour activity [128]. However, this was performed decades before the completion

of the successful phase III clinical trial of DAC treatment for MDS where a much lower dose was used [118, 128].

In the treatment of solid tumours, DAC, and other epigenetic agents, have displayed minimal efficacy when administered alone, however combined treatments may hold more promise [120, 121]. In HNSCC cell lines treatment with DAC plus the HDACi valproic acid (VPA) and all-trans-retinoic-acid (ATRA) was more effective than individual therapies [129]. Similarly, the results of a phase 1 clinical trial of DAC plus the HDACi vorinostat in solid tumours demonstrated anti-tumour activity with good tolerability and prolonged stable disease [120]. Early clinical trials have shown that DAC may sensitise solid tumours to chemotherapeutic agents such as cisplatin and carboplatin [120, 130]. In HNSCC, preclinical studies have shown that pre-treatment with DAC can restore sensitivity to cisplatin by altering 5mC and gene expression towards a cisplatin-sensitive profile [131].

In conclusion, clinical data has shown that when administered alone DAC has poor anti-tumour activity in solid tumours, however when combined with chemotherapy or other epigenetic agents, patient outcomes may improve [120, 121]. A huge variation exists in the dosing regime employed by different studies; this must be optimised to allow for accurate comparison [120, 121]. In regards to HNSCC, more research is needed to determine the potential efficacy of DAC treatment, particularly at low doses and when combined with other agents.

### Table 1.1. DAC therapeutic application in solid tumours

Clinical trials investigating the use of DAC in the treatment of cancer: either as a single agent (top), or combined with epigenetic (middle) or chemotherapeutic (bottom) drugs [120, 121].

| Study | Patient Cohort | Clinical Phase | No. Patients | Treatment Regime | Patient outcome |
|---|---|---|---|---|---|
| Abele (1987) [128] | Mixed solid tumours inc. HNSCC | II | 101 | **DAC**: 75 mg/m2 × 3 doses, 7h intervals | No anti-tumour activity |
| Clavel (1992) [132] | Nonsemino-matous testicular cancer | II | 15 | **DAC**: 75 mg/m2 1h IV infusion on day 1 with intervals of 7h every 5 weeks | No anti-tumour activity |
| Momparler (1997) [133] | Metastatic lung cancer | II | 15 | **DAC**: single 8h IV infusion of 200–660 mg/m2 | Potential anti-tumour activity |
| Stathis (2011) [134] | Advanced Solid Tumours and Non-Hodgkin's Lymphoma | II | 43 | **DAC**:10 mg/m2/day on days 1 to 5 and **Vorinostat:** 200 mg twice a day on days 6 to 12 | Anti-tumour activity and good tolerability |
| Chu (2013) [135] | Non-small-cell lung cancer | I | 8 | **DAC**: 5–15 mg/m2/day for 10 days. **VPA:** 10–20 mg/kg/day for days 5–21 q 28 days | Neurological toxicity Increase in foetal Hb levels in all patients |
| Schwartsmann (2000) [136] | Inoperable Non-small-cell lung cancer | II | 14 | **DAC**: 45- 120 mg/m2 **Cisplatin**: 33 mg/m2 for 3 consecutive days every 21 days | Insignificant anti-tumour activity |
| Pohlmann (2002) [137] | Advanced solid tumours | II | 25 | **DAC**: 50 mg/m2 **Cisplatin:** 40 mg/m2 for 3 consecutive days every 21 days IV | Moderate anti-tumour activity; significant hematologic toxicity |
| Matei (2012) [138] | Ovarian Cancer | | 17 | **DAC**: 10 mg/m2/day for 5 days **Carboplatin**: 5 day | 35% response rate 10.2 months progression-free survival |
| Tawbi (2013) [139] | Refractory Melanoma | I-II | 35 | **DAC**: 0.15 mg/kg IV daily × 5 days/wk for 2wks **Temozolomide** orally 75 mg/m2 daily for wks 2–5 of a 6wk cycle | Excellent pharmacokinetics and pharmacodynamics |
| Fang (2010) [140] | Ovarian | 1 | 10 | **DAC**: 10–20 mg/m2/day days 1–5 28 days **Carboplatin**: 5 day | Gene-specific DNA methylation. Toxicity: myelosuppression, nausea, fatigue global |
| Glasspool (2014) [141] | Ovarian | | 15 | **DAC:** IV 45 mg/m2 day 1 **Carboplatin** 6 day | DAC appears to reduce the efficacy of carboplatin |

## 1.8. Drug Repurposing

The costs and time demands of drug discovery are considerable, with only a small percentage of discovered drugs ever making it to the market [142, 143]. Furthermore, the costs are increasing; a study in 2003 estimated that the average research and development cost per drug was $802 million; $484 million higher than a similar study in 1999 [144]. Concurrently, the number of newly approved drugs is in decline, both in Europe and America [145-147]. Therefore, a novel strategy of drug development is required.

Drug repurposing (repositioning, reprofiling, redeployment) is the use of drugs (licensed or not) for indications that are different to their original indication or therapeutic purpose [147]. The risks associated with drug development are reduced as repurposed drugs have often been through several stages of clinical and technical development, providing information on the safety and pharmacokinetic profiles of the drug, as well as on issues such as bulk manufacturing and environmental toxicity [147]. While this does not always result in avoiding Phase I trials, it informs on the dosing structure and schedule and tends to result in a shorter development time for repurposed drugs (3-12 years) compared with *de novo* drugs (10-17 years) [148, 149]. Furthermore, repurposed drugs have a higher success rate in clinical trials [142]. Many repurposed drugs were found serendipitously [145, 146]. Examples include viagra, initially developed to treat hypertension and angina, but discovered to be effective against erectile dysfunction; buproprion, designed to treat depression, but found to aid in the cessation of smoking; and the type 2 diabetes drug, metformin, that was discovered to prevent cancer [145, 146, 148]. The

majority of available drugs are thought to possess secondary indications, either through off-target effects of the drug, or due distinct diseases originating from the same disrupted pathways [142, 150]. Nowadays the wealth of knowledge on the structure of drugs and drug targets allows for potential repurposing projects to be determined *in silico*, further reducing the failure rate of the repurposed compound [145].

A key issue with drug repurposing is the difference in patient compliance between therapeutic areas. Side effects that are tolerable for life threatening diseases are unacceptable for most conditions. As a result, oncology drugs are rarely repurposed [149]. Similarly, due to the severity of the disease, cancer is an ideal candidate for drug repurposing (Table 1.2).

**Table 1.2. Drugs repurposed for cancer treatment**

Table adapted from [142, 145].

| Drug | Primary Indication | Primary Target | Secondary Indication(s) | Cancer inhibitory effects |
|---|---|---|---|---|
| Thalidomide | Morning sickness | TNF-α | Leprosy<br>FDA approved for multiple myeloma | Inhibits angiogenesis and modulates inflammatory pathways |
| Valproic acid | Convulsions and migraines | HDACs | Phase II trials for a variety of solid and haematological malignancies | Inhibits survival, invasion, angiogenesis and metastasis. Modulates inflammatory pathways |
| Aspirin | Analgesic and antipyretic | COX-1/-2 | Decreased risk of numerous cancer types | Anti-inflammatory |
| Celecoxib | Rheumatoid arthritis and osteoarthritis | COX-2 | FDA approval for the prevention of polyps in cases of familial adenomatous polyposis. | Anti- inflammatory and anti-proliferative |
| Metformin | Type II diabetes | AMPK | Phase III trials for prostate, breast, endometrial and pancreatic cancer | Inhibits mTOR pathway, oxidative phosphorylation and the citric acid cycle leading to reduced proliferation |
| Raloxifene | Osteoporosis in menopausal and post-menopausal women | ER | FDA approved for breast cancer prevention | Regulates transcription of estrogen response elements |

## 1.9. Aims and Objectives

The primary aim of this thesis was to investigate the relationship between DNA modifications and HNSCC and to examine the efficacy of the DNA demethylating agent, DAC, in HNSCC treatment. This can be split into five objectives:

- Examine the potential efficacy of DAC treatment on HNSCC cell lines
- Utilise a drug repurposing screen to determine whether efficacy can be increased through combination therapy
    - Investigate the mechanisms behind combination therapies
- Compare the distribution of 5mC and 5hmC in normal oral keratinocytes and HNSCC cells
- Investigate effect of treatment with DAC on 5mC and 5hmC distribution in an HNSCC cell line.

# Chapter 2. Methods

## 2.1. Cell lines and culture

### 2.1.1. Cell lines

Three human HPV negative HNSCC cell lines (VU40T, SCC040, HN12); one human HPV positive HNSCC cell line (UDSCC2); and primary human oral keratinocyte cells (HOK) were used in this work. Details are provided in Table 2.1. SCC040 and VU40T cell lines were a gift from our collaborator Dr S. Roberts (School of Cancer Sciences; University of Birmingham): who obtained SCC040 cells from the German Culture Collection, DSMZ (#ACC660) and VU40T from Prof H. Joenje (VU University Medical Centre, Amsterdam). HN12 and UDSCC2 cells were obtained from Dr J. S. Gutkind (NIDCR, NIH, Bethesda). Primary human oral keratinocyte (HOK) cells, isolated from fetal oral mucosa were purchased from Caltag Medsystems.

**Table 2.1. Details of the cell types used in the studies**

| Cell Line | Site of Origin | Gender | Age | Stage | Smoking /Alcohol | HPV Status | Reference |
|---|---|---|---|---|---|---|---|
| **VU40T** | Tongue | Female | 65 | T3N0 | -/ - | - | [151] |
| **SCC040** | Tongue | Male | 50 | T2N2 | -/+ | - | [152] |
| **UDSCC2** | Hypopharynx | Male | 58 | T1N2 | N/A | + | [153] |
| **HN12** | Metastasized from base of tongue to lymph node | Female | N/A | T4N1 | N/A | - | [154] |
| **HOK** | Oral mucosa | Unknown | 20w gestation | - | N/A | N/A | ScienceCell (no: 2610) |

### 2.1.2. Cell maintenance

HNSCC cultures were maintained in Dulbecco's Modified Eagles Medium (Sigma-Aldrich) supplemented with 10% fetal bovine serum (FBS) (Sigma-Aldrich), 100U/ml penicillin-streptomycin (Life Technologies), 4mM L-glutamine, 1X non-essential amino acids (Life Technologies) and 1mM sodium pyruvate (Life Technologies).

HOKs were cultured in complete oral keratinocyte medium (SC-2611), containing amino acids, vitamins, hormones and growth factors, optimised for the growth of normal human keratinocytes and supplemented with 1% penicillin-streptomycin solution and 1% oral keratinocyte growth supplement (all purchased from ScienceCell).

All cell lines and treatments were maintained at 37°C in a humidified 5% $CO_2$ incubator (Sanyo).

To split HNSCC cell lines for maintenance or subculture the media were removed and cells washed once in phosphate buffered saline solution (PBS) (Sigma-Aldrich). Following this, cells were incubated with 0.25% Trypsin (Sigma-Aldrich) for 5min or until the cells begin to round up and detach as determined by inverted light microscope. The trypsin was inactivated with an equal volume of complete media.

## 2.2. Drug Treatments

### 2.2.1. 5-Aza-2′-deoxycytidine treatments

5-Aza-2'-deoxycytidine (DAC) was purchased from Sigma-Aldrich and dissolved in 50% acetic acid stock solution, aliquoted and stored at -80°C. HNSCC cells were

treated with a range of concentrations of DAC or an equivalent volume of the vehicle, 50% acetic acid. Unless otherwise stated, treatments were 96h with a change of treatments and media after 48h.

### 2.2.2. CellTiter Blue Cell Viability assay

Cell viability experiments were performed in a 96-well format. Cells were grown in 200µl media and viability was determined by incubation with 20µl of the CellTiter Blue Cell Viability reagent (Promega) for the final 8h of treatment. The reagent contains the compound resazurin can be absorbed at a wavelength 600nm. In viable cells resazurin is reduced to resorufin, which can be absorbed at 570nm. Therefore, absorbance at 570nm is indicative of cell viability. After incubation, absorbance was measured on a Synergy HT plate reader (BioTek). For each well the 600nm value was subtracted from the 570nm value. For all viability experiments the cell free control (blank) was subtracted from each of the sample readings to correct for background signal from the media. Triplicate wells were used and all treated wells were normalised to the vehicle only control.

### 2.2.3. Dose response curves

Sigma plot software (Systat Software Inc.) was used to generate sigmoidal, 4-parameter dose response curves after drug titrations.

### 2.2.4. DAC sensitising assay

A panel of 100 drugs (FMC1), dissolved in 100% DMSO, ethanol or water (Supplementary data 1) were a gift from Dr Farhat Khanim (School of Biosciences, University of Birmingham, UK) and administered at the Cmax concentrations [155].

The concentration of vehicle was always <0.1%. The sensitizing assay was performed blind with controls hidden within the panel. HNSCC cells were subject to 96h treatment with each drug either in the presence of 500nM DAC or an equivalent volume of the vehicle, 50% acetic acid. The assay was undertaken in a 96-well format. Each plate contained a series of controls: no cells; cells with vehicles only; cells with 500nM DAC only and a positive control of 10µM DAC. Triplicate wells were prepared and the drugs and media were refreshed at 48h into the treatment.

The assay was performed in triplicate and SPSS software (IBM Analytics) used to perform a paired t-test analysis with Bonferroni correction for each of the 'DAC plus drug' samples relative to the 'drug only' and the 'DAC only' samples.

### 2.2.5. Drug treatments

Paracetamol, zinc acetate, valdecoxib and disulfiram were purchased from Sigma-Aldrich. These were dissolved in 100% DMSO at 5000X the working concentrations, aliquoted and stored at -20°C.

### 2.2.6. Determining synergy

The Chou-Talalay method was employed to determine synergy [156]. Three 96 well plates were set-up with constant-ratio titrations (0.125X, 0.25X, 0.5X, 1X, 2X, 4X, 8X): one with titrations of 500nM DAC (0.0625µM, 0.125µM, 0.25µM, 0.5µM, 1µM, 2µM, 4µM); one with titrations of 132.3µM paracetamol (16.5µM, 33.1µM, 66.15µM, 132.3µM, 264.6µM, 529.2µM, 1058.4µM) and one with matched titrations of DAC plus paracetamol. Cultures were treated for 96h, following which viability was determined using as described in section 2.2.2. The CompuSyn software (CompuSyn

Inc.) was used to calculate combination index (CI) and dose reduction index (DRI) values from this data [156].

## 2.3. DNA extraction and dot blotting of DNA

### 2.3.1. DNA extraction

DNA was extracted using the DNeasy blood and tissue kit (Qiagen) as per the manufacturer's instructions. Cultures were harvested and centrifuged at 400xg to form a cell pellet and washed once in PBS. The pellet was incubated in buffer AL with proteinase K at 56°C for 10min to lyse the cells and digest the proteins. One volume of ethanol was added to precipitate the DNA and the sample was loaded onto a spin column and centrifuged at 14,000xg for 1min to bind the DNA to the silica-based membrane. This was washed with buffers AW1 and AW2 and the DNA eluted with 50-200µl of AE buffer. DNA concentration was determined using a Nanodrop spectrophotometer (Thermo Fisher Scientific).

### 2.3.2. Dot blotting

Dot blotting of DNA was performed as described in the Current Protocols in Molecular Biology using the buffers outlined below (Table 2.2) [157]. DNA (1µg, 0.5µg, 0.25µg or 0.125µg) was combined with 0.4M NaOH and 10mM EDTA in a total volume of 50µl. This mixture was heated at 100°C for 10min to denature the DNA then quenched on ice for 5min. Nytran Supercharge Positively Charged Nylon membrane (Fisher Scientific) and Whatman filter paper were soaked in $dH_2O$ prior to use. The 96-well dot blotting apparatus (BioRad) was assembled with the membrane placed on a piece of Whatman filter paper and sandwiched within the apparatus.

Wells were washed with $dH_2O$ and drained by vacuum, the samples loaded and let to vacuum-drain, followed by a wash with 100µl 0.4M NaOH. The membrane was briefly soaked in 2X SSC buffer and allowed to dry, suspended, at room temperature (RT) for 2h.

### 2.3.3. Antibody incubation

Dried membranes were blocked in 5% low-fat milk (Marvel) in PBS at RT for 1h and incubated with primary antibodies (5mC: Cell Signalling (28692S), 1:750; 5hmC: Active Motif 1:5000) in 1% bovine serum albumin (BSA) at 4°C overnight.

The following day membranes were washed in low fat milk and incubated with secondary antibodies (anti-rabbit-IgG-HRP (Santa Cruz sc-2004), 1:5000 dilution) for 2h at 4°C. Membranes were then washed in PBS and treated with a 1:1 ratio of Amersham ECL Western Blotting Reagent (GE Healthcare) and Supersignal West Femto reagent (Thermo Scientific) for 2min and placed in an autoradiography cassette. In a dark room the membrane was exposed to Amersham ECL chemiluminescence film (GE Healthcare) and the film developed on a photon imaging system (PROTEC). To stain for total DNA, the blots were incubated in 0.04% methylene blue for 20min or until the dots became visible. Blots were washed for 10min in 20% ethanol to remove background staining and photographs taken.

### 2.3.4. Analysis

Each experiment was performed in triplicate. ImageJ software was used to quantify the dot blots. Lanes were selected using the 'gels' sections within the 'analyse' function. This results in a line graph depicting the intensity of any bands or dots

within the lanes. The wand tool was then used to select the relevant peaks and the 'label peaks' function used to determine the relative intensity of each dot. 5mC or 5hmC dots were normalised to the corresponding methylene blue dot.

**Table 2.2: Buffers used for dot blotting of DNA**

| Buffer | Component | Company |
|---|---|---|
| 20X SCC Buffer | 0.3M trisodium citrate | Sigma-Aldrich |
| | 3M sodium chloride | Sigma-Aldrich |
| | $dH_2O$ | |
| 1% BSA | 1% Bovine serum albumin | Sigma-Aldrich |
| | PBS | Sigma-Aldrich |
| 0.04% Methylene Blue | 0.04% Methylene blue | Sigma-Aldrich |
| | 0.05M Sodium acetate | Sigma-Aldrich |

## 2.4. Protein extraction and Western blotting

The buffers used for this protocol are outlined in Table 2.3.

### 2.4.1. Protein extraction

HNSCC cells were washed twice in cold PBS and incubated with RIPA buffer containing 1X cOmplete EDTA-free protease inhibitors (Roche) for 45min at 4°C to lyse the cells. Centrifugation at 14,000xg for 10min pelleted the cellular debris, which was discarded, and the supernatant containing the proteins was retained.

### 2.4.2. SDS-PAGE

Prior to gel electrophoresis, 20µl of protein was combined with 4X loading buffer and denatured by incubation at 95°C for 5min.

A 10% SDS-PAGE resolving gel (Table 2.3) was prepared and loaded into a mini-Protean 3 BioRad apparatus. The top of the gel was covered with isopropanol (Sigma Aldrich) to prevent evaporation. Once the gel was set the isopropanol was removed and a stacking gel (Table 2.3) prepared and loaded with a 1mM comb. The stacking gel concentrates the sample into a tight band before it enters the resolving gel. The gel was transferred into a Mini Protean3 Cell (BioRad) tank filled with running buffer. 20µl of sample or 5µl of PageRuler Prestained Protein ladder (Thermo Fisher Scientific) were loaded to each well and the gel electrophoresed for 1h at 200V at RT.

### 2.4.3. Transfer of proteins from gel to membrane

Filter paper and sponges were soaked in transfer buffer for 10min. Nitrocellulose membrane was soaked briefly in 100% methanol (Sigma-Aldrich), washed in $dH_2O$ and soaked in transfer buffer for 10min.

When electrophoresis was complete the gel was removed from the glass casing and the transfer apparatus assembled (BioRad). The gel and membrane were placed side-by-side encased in soaked filter paper and sponge and held together in the transfer cassette. This assembly was transferred to a transfer tank filled with transfer buffer and an ice pack to ensure the gel remained cool. The transfer was electrophoresed for 1h at 100V.

Once the transfer was complete the gel was washed briefly in 1X PBST (Table 2.3) and stained with Ponceau S (Sigma-Aldrich) to visually ensure the transfer was successful. Ponceau S was removed by a 5min wash step in 1X PBST. The gel was also stained in Coomassie Blue solution (BioRad) overnight followed by 4h of washing in Coomassie destain solution (Table 2.3) to visualise the proteins remaining on the gel. Both were used to visually determine the relative amounts of protein loaded.

### 2.4.4. Antibody incubation

The membrane was blocked in 20% milk in PBST for 1h at RT then washed 3X 5min in PBST. The membrane was incubated with the primary antibody as described below:

- COX-2 (Abcam ab15191, 1:1000): 1h at RT in 1% BSA in PBST
- CYP2E1 (Abcam ab28146, 1:2500): overnight at 4°C in 5% milk in PBST
- Lamin A/C (Santa Cruz sc-20681, 1:10,000): 1h at RT in 5% milk in PBST

Following primary antibody incubation membranes were washed 3X 5min in PBST and incubated with the secondary antibody (anti-rabbit-IgG-HRP (Santa Cruz sc-2004), 1:5000 dilution) for 2h at RT. The membranes were washed 3X 15min in PBST. The EZ-ECL Chemiluminescence Detection Kit for HRP (Biological Industries) was prepared as described by the manufacturer and incubated with the membrane for 1min prior to placing the membrane into an autoradiography cassette. In a dark room, an X-ray film (HyperfilmTM, Amersham Biosciences) was placed into the cassette for an appropriate amount of time. Films were then developed on a photon imaging system (PROTEC).

Each membrane was subject to three antibodies: CYP2E1, COX-2 and the control Lamin A/C. After one antibody was developed the membranes were washed briefly in 1X PBST, then washed with stripping buffer (Table 2.3) for 2 x 7.5min. Membranes were washed briefly in 1X PBST and the antibody incubation step repeated as described above.

### 2.4.5. ImageJ Analysis

Image J was used to quantify the intensity of each band on the western blot as described for dot blots in section 2.3.4. COX-2 or CYP2E1 bands were normalised to the corresponding lamin bands.

**Table 2.3: Buffers used for Western blotting**

| Buffer | Components | Supplier |
|---|---|---|
| RIPA Buffer | 150mM sodium chloride | Sigma-Aldrich |
| | 0.5% Sodium deoxycholate | Sigma-Aldrich |
| | 0.1% Sodium dodecyl sulfate | Sigma-Aldrich |
| | 50mM Tris | Sigma-Aldrich |
| 4X Loading Buffer | 4% SDS, | Sigma-Aldrich |
| | 0.125 M Tris HCl | Sigma-Aldrich |
| | 20% glycerol | Sigma-Aldrich |
| | 10% 2-mercaptoethanol | Sigma-Aldrich |
| | 0.004% bromophenol blue | Sigma-Aldrich |
| 10% resolving gel | 37.5:1 acylamide:bisacrylamide | Protogel |
| | 375mM Tris-HCl pH 8.8 | Sigma-Aldrich |
| | 0.1% w/v SDS | Sigma-Aldrich |
| | 0.1% w/v N,N,N',N'-tetramethylethylenediamine (TEMED) | Sigma-Aldrich |
| | 0.1% w/v ammonium persulphate (APS) | Sigma-Aldrich |
| Stacking gel | 37.5:1 acylamide:bisacrylamide | Protogel |
| | 375 mM Tris-HCl pH 6.8 | Sigma-Aldrich |
| | 0.1% w/v SDS | Sigma-Aldrich |
| | 0.1% w/v TEMED | Sigma-Aldrich |
| | 0.1% w/v APS | Sigma-Aldrich |
| Running Buffer | 25 mM Tris-HCl pH 7.5 | Sigma-Aldrich |
| | 2 mM glycine | Thermo Fisher Scientific |
| | 10% SDS | Sigma-Aldrich |
| 1X PBST | 1X Phosphate Buffered Saline | Sigma-Aldrich |
| | 0.1% Tween-20 | Sigma-Aldrich |
| Transfer Buffer | 100 mM Tris-HCl pH 7.2 | Sigma-Aldrich |
| | 200 mM glycine | Sigma-Aldrich |
| | 0.05% SDS | Sigma-Aldrich |
| | 20% methanol | Sigma-Aldrich |
| Stripping Buffer | 1.5% w/v glycine | Sigma-Aldrich |
| | 0.1% SDS | Sigma-Aldrich |
| | 1% Tween20 | Sigma-Aldrich |
| | Adjust to pH2.2 with HCL | |
| Coomassie Destain | 30% Methanol | VWR |
| | 10% acetic acid | VWR |

## 2.5. Giemsa and Jenner Staining

### 2.5.1. Cell plating and fixation

Glass coverslips were incubated in methanol (Sigma-Aldrich) for at least 24h; individual coverslips were placed in each well of a 12-well plate (Corning) and allowed to dry. VU40T cells were plated onto the coverslips and treated with 500nM DAC +/- 132.3µM paracetamol, or vehicle only control. After treatment, the media was removed and the cells were washed with PBS and fixed for 15min in methanol at RT. The methanol was removed and the cells allowed to air-dry at RT overnight.

### 2.5.2. Giemsa-Jenner staining

The protocol was performed as described previously [155]. Fixed cells were incubated with Jenner stain (VWR) diluted 1:3 in 1mM sodium phosphate buffer pH5.6 for 5min at RT followed by repeated washing in $dH_2O$ to remove stain from the wells. The cells were then incubated with Giemsa stain (VWR), diluted 1:20 in 1mM sodium phosphate pH5.6 for 10min at RT, followed by repeated washes in $dH_2O$. The coverslips were left to dry at RT overnight prior to mounting onto slides using DePex (CellPath).

## 2.6. Immunofluorescence

### 2.6.1. Cell fixation

Immunofluorescence was performed as described in Roulois *et al.,* 2015 [158]. Cells were grown on methanol-sterilised coverslips, washed twice in cold PBS, blotted on paper to remove any excess liquid and fixed with ice cold methanol for 15min at -

20°C. The methanol was removed by aspiration and replaced with excess 100% acetate for 1min at -20°C to permeabilise the nuclei. The acetate was removed and the coverslips allowed to dry briefly at -20°C.

### 2.6.2. Antibody incubation

Coverslips were washed three times in cold PBS and blocked with saturation buffer (1% BSA in PBS) for 1h at 4°C. Coverslips were incubated overnight at 4°C with primary antibodies (dsRNA (Scicons: J2) diluted 1:1000; Ki67 (Abcam: ab15580) 1:1000; γH2AX (Abcam) 1:1000) diluted in 1% BSA in PBS. Wells were washed 3X 5min in PBS and incubated with the corresponding secondary antibody (goat anti-mouse Alexa Fluor 488 at 1:2000 dilution for J2; donkey anti-rabbit Alexa Fluor 488 at 1:250 for Ki67 and γH2AX) for 1h at room temperature, protected from light. Wells were washed 3X 5min in PBS. Coverslips were allowed to dry and mounted onto slides using ProLong Gold Antifade Mountant with DAPI (4',6-diamidino-2-phenylindole) (Fisher Scientific). The slides were stored in the dark at 4°C for at least 72h.

### 2.6.3. Confocal microscopy

Immunofluorescence slides were imaged on a Zeiss 780 Zen confocal microscope. The microscope settings were adjusted for each antibody and then maintained across all slides within the experiment. Each slide was manually focused and immunofluorescent images captured at 10X and 40X magnification. The 'averaging' function was used to remove background noise. This takes four stacked images and produces an average.

### 2.6.4. Immunofluorescent analysis

The ImageJ software was used to analyse confocal images as is described below.

For immunofluorescence images of dsRNA, the staining intensity was determined by converting all images to greyscale. The background fluorescence was subtracted and the intensity of DAPI and dsRNA determined. For each image J2 dsRNA levels were normalised against DAPI.

To determine the number of Ki67 positive nuclei in each image first the colour channels were split and the 'threshold' function used so that nuclei appeared as distinct dots. The 'analyse particle function' was then used to count the number of nuclei in each image. The number of Ki67-stained nuclei were normalised against the number of DAPI-stained nuclei.

For γH2AX the number of foci per nuclei was counted by hand while imaging across multiple focal planes. For each biological replicate 100 nuclei were counted.

## 2.7. Apoptosis and Necrosis detection

### 2.7.1. Annexin V and Propidium Iodide staining

The Annexin V Apoptosis Detection APC Kit (eBioscience) was used to stain HNSCC cells with annexin and propidium iodide prior to Fluorescence-activated cell sorting (FACS). HNSCC cells were washed in PBS, trypsinised and counted. Cells were resuspended in 1X Binding Buffer at a concentration of $5x10^6$ cells/ml. The kit was used as per manufacturers recommendations with 100µl of cell suspension incubated with 5µl of fluorochrome-conjugated annexin V for 15min at RT. Following

this the cells were washed in 1X Binding Buffer and resuspended in 500µl 1X Binding Buffer plus 5µL of Propidium Iodide (PI) Staining Solution and analysed by immediately on a Cyan B FACS analyser (Beckman Coulter).

### 2.7.2. Fluorescence-activated cell sorting (FACS)

The pulse width, forward and side scatter of unstained cells were used to discard doublets and debris by selecting the population to be analysed. Necrotic cells stain only with PI; early apoptotic cells stain only with annexin V, late apoptotic cells stain with both and viable cells stain with neither. When annexin V is plotted against PI the four populations can be separated into quadrants and the number of events in each quadrant counted.

## 2.8. Cell cycle analysis

VU40T cells were harvested, washed and centrifuged to form a pellet, then fixed by adding ice-cold 70% ethanol drop-wise whilst vortexing, followed by a 30min incubation at 4°C. The pellet was washed twice in PBS, treated with RNase (Roche) to remove RNA, resuspended in 50µg/µl propidium iodide (PI) (Invitrogen) and analysed on a Cyan B FACS analyser (Beckman Coulter).

Doublets and debris were removed as described in section 2.7.2. PI emission is directly proportional to the amount of DNA within each cell and therefore can be used to determine the cell cycle stage. A histogram of PI emission versus cell count, determined by FACS analysis, was used to determine the distribution of cells within each stage of the cell cycle by gating the appropriate peaks.

## 2.9. Prostaglandin $E_2$ ($PGE_2$) enzyme-linked immunosorbent assay (ELISA)

The levels of $PGE_2$ in the media of VU40T and SCC040 cells was determined using the $PGE_2$ ELISA kit (Abcam: ab133055) as the manufacturer recommends. Briefly, 100µl sample or standard was added to the antibody bound wells followed by 50µl Assay Buffer and 50µl $PGE_2$-AP conjugate, then 50µl Anti-$PGE_2$ antibody. The plate was incubated for 2h at RT and washed 3 times in 1X Wash Buffer. The plate was then incubated with 200µl p-Nitrophenyl Phosphate (pNpp) substrate solution for 1h at 37°C. 50µl Stop solution was added and absorbance read at 405nm with correction at 590nm. The assay was performed in 96-well format with duplicate sample wells and with the controls described by the manufacturer: blank wells containing only the pNpp substrate solution; total activity wells containing pure, unwashed $PGE_2$-AP conjugate; and non-specific binding wells containing no sample or standard. The absorbance values of the standards of known concentration were used to calculate the concentrations of the sample wells.

## 2.10. Determining Glutathione concentration

The levels of glutathione were measured using the GSH-Glo$^{TM}$ Glutathione Assay kit (Promega) as described by the manufacturer. Briefly VU40T cells were grown on a white 96 well plate (Greiner Bio-One) treated with DAC +/- paracetamol, or a vehicle only control for 96h. As a control, 1mM paracetamol was included. All treatments were performed in triplicate. Following treatment the media was removed from the wells. GSH-Glo (1X) reagent was made by combining luciferin-NT substrate with

Glutathione S-transferase and GSH-Glo reaction buffer at a ratio of 1:1:10. Wells were incubated with 100µl of this mixture for 30min at RT. In the presence of glutathione the Glutathione S-Transferase can convert Luciferin-NT into Luciferin. Wells were incubated with 100µl of reconstituted Luciferin Detection Reagent at RT for 15min and luminescence recorded. The luminescence signal is relative to the amount of glutathione in each sample. As the concentration of glutathione is relative to cell density all treated wells were normalised against the corresponding CellTiter Blue cell viability results (section 2.2.2). For each experiment a standard curve was generated spanning 0-5µM glutathione using the 5mM glutathione provided by the kit.

## 2.11. Quantitative Reverse Transcriptase Polymerase Chain Reaction (qRT-PCR) Analysis

### 2.11.1. RNA extraction

RNA was extracted using the RNeasy Mini Kit (Qiagen) as the manufacturer recommends. In brief, 500µl of the lysis buffer RLT with freshly added 5µl 2-mercaptoethanol was added to the washed well of a 6-well plate, allowed to digest for 5min and homogenized using a 1mm needle and syringe. The homogenates were mixed with one volume of 70% ethanol and transferred to an RNeasy spin column. RNA was captured on the membrane by centrifugation at 10,000xg for 15secs. The membrane was washed in RW1 buffer, incubated in DNase (RNase-free DNase I (Qiagen)) to digest DNA contaminants and washed in RW1 again. This was washed twice in RPE buffer and RNA eluted with 40µl of nuclease free $H_2O$.

### 2.11.2. cDNA synthesis

RNA was converted into cDNA using the iScript cDNA Synthesis kit (Bio-Rad) as described in the protocol. The reaction mix of 1µg RNA, 5X iScript reaction buffer and iScript reverse transcriptase, adjusted to a total volume of 20µl with nuclease free water, was incubated at 25°C for 5min, 46°C for 20min and finally 95°C for 1min. The cDNA was purified using QiaQuick PCR purification kit (Qiagen).

### 2.11.3. cDNA quantification

The concentration of cDNA was determined using Qubit dsDNA High Sensitivity Assay kit (Thermo Fisher Scientific). Qubit dsDNA HS reagent was diluted 1:200 in Qubit dsDNA HS buffer. In Qubit microtubes, 199µl of the reagent was combined with 1µl of DNA sample, or 190µl with 10µl of the standards provided. These were incubated for 2min at RT to allow the fluorescent dye to bind the DNA. The dye only emits a signal when bound to DNA. A Qubit 3.0 Fluorometer (Thermo Fisher Scientific) was used to record this and the signals of the standards used to quantify the DNA in each sample.

### 2.11.4. Quantitative reverse transcriptase PCR (qRT-PCR)

qRT-PCR was performed on the Roche LightCycler 480 II using the programme outlined in Table 2.4. The reaction mix comprised 1µl of cDNA, 0.075µl of 100µM forward and reverse primers, 6µl LightCycler 480 SYBR Green Master Mix and adjusted to 12µl with nuclease free $H_2O$. Standards were made from a mixture of cDNA samples, STD1: undiluted, STD2: 10X dilution, STD 3: 100X dilution.

The Roche LightCycler records the number of cycles taken for a sample to cross the threshold fluorescent level (Crossing point, or Cp value). The software utilises the Cp values of the standards to determine the relative RNA concentration in each sample. This was normalised to the cDNA concentration (2.11.3). SPSS software was used to calculate paired t-tests with Bonferroni correction to determine whether treated samples showed a significant difference in RNA compared with untreated controls.

The program outlined in Table 2.4 was used to generate a melting curve for each set of amplicons to determine specificity.

**Table 2.4: Programme used for quantification RNA by qPCR**

| Cycle Step | Temperature (˚C) | Time (Seconds) | |
|---|---|---|---|
| Preincubation | 95 | 300 | |
| Amplification | 95<br>60<br>72 | 10<br>10<br>10 | x45 Cycles |
| Melting Curve | 95<br>65 | 5<br>60 | |
| Cooling | 40 | 30 | |

### 2.11.5. Primer design

Primers for qRT-PCR were designed using the Primer3 software with RNA sequences downloaded from the UCSC genome browser [159]. Primers were designed to produce a product between 90-150bp and where possible, all RNA primers were designed to cross a large intron, as control against DNA contaminants.

**Table 2.5: Primers used for qRT-PCR**

| GENE SYMBOL | FORWARD (5'-3') | REVERSE (5'-3') | Product size |
|---|---|---|---|
| ***TET1*** | TCATGGGTGTCCAATTGCTA | GATGAGCACCACCATCACAG | 121 |
| ***TET2*** | GGACATGATCCAGGAAGAGC | CCCTCAACATGGTTGGTTCT | 118 |
| ***GAPDH*** | CCTGGCCAAGGTCATCCAT | AGGGGCCA TCCACAGTCTT | 98 |
| ***DNMT1*** | GAGCCACAGATGCTGACAAA | GACACAGGTGACCGTGCTTA | 133 |
| ***DNMT3A*** | AAGGAGGAGCGCCAAGAG | GGATGGGGACTTGGAGATCA | 129 |
| ***DNMT3B*** | GGGAGGTGTCCAGTCTGCTA | GGCTTTCTGAACGAGTCCTG | 129 |
| ***PTGS1 (COX-1)*** | CGTCCCGCACCCCAGCAG | CTGGGTCCGCGAGCAGGA | 104 |
| ***PTGS2 (COX-2)*** | TCATCATCAGCGCCCTCAA | GCTCGTTCACAGCCTTCATG | 93 |
| ***PTGS1 (COX-3)*** | CGTCCCGCACCCCAGCAG | GCTGAGCCTGGCATTCAAGG | 104 |
| ***PTGER1 (EP-1)*** | GCCAGCTTGTCGGTATCATG | CTGCAGGGAGGTAGAGCTC | 101 |
| ***PTGER2 (EP-2)*** | AAGCTGTGGTCAAGGCTACA | GCCAAGTACCATGCTCACTG | 97 |
| ***PTGER3 (EP-3)*** | GGATCATGTGCGTGCTGTC | TGTGTCTTGCAGTGCTCAAC | 97 |
| ***PTGER4 (EP-4)*** | TGCTCATCTGCTCCATCCC | ATTCGGATGGCCTGCAAATC | 107 |
| ***FAAH*** | CTGCTCTGGACTTGAATGCC | CCCCAAAGTAGCCCCTGTAA | 108 |
| ***TRPV1*** | TATCACCATCCAGAGGCCAG | ACTCCTGCGATCATAGAGCC | 109 |
| ***CYP2E1*** | CGGAACTATGGGATGGGGAA | CGGAAGAGGATGTCGGCTAT | 158 |
| ***IFN$\beta$1*** | TCACTGTGCCTGGACCATAG | AGCAATTGTCCAGTCCCAGA | 124 |
| ***IRF7*** | GGAGGCCCAAGGAGAAGAG | TATCCAGGGAAGACACACCC | 102 |
| ***TLR3*** | ACAGCATCAAAAGAAGCAGAAA | AGCTTGTTGAACTGCATGATGT | 104 |
| ***IFIH1 (MDA5)*** | GTTGGACTCGGGAATTCGTG | CAAACGATGGAGAGGGCAAG | 102 |
| ***MAVS*** | CAGCAAGAGACCAGGATCGA | GCCGCTGAAGGGTATTGAAG | 90 |

## 2.12. RNA sequencing and analysis

RNA was extracted from VU40T cells as described in section 2.11.1 and quantified using a Nanodrop. For each cell type of condition three RNA samples were pooled in equal amounts and a single library was made using the TruSeq Stranded mRNA Library Prep kit. Qubit fluorometric quantitation (Thermo fisher scientific) was used to quantify the pool and the integrity of the RNA (RIN number) was determined by Tape station (Agilent). All samples included in this work had a RIN number of 10. Libraries were loaded onto an Illumina NextSeq 500 and sequenced in paired-end mode at 2x75 bases in the Genomics Birmingham facility.

The analysis was performed using Galaxy [160]. Reads were aligned to the genome (hg19) using HiSAT (Galaxy version 2.0.3) using default parameters. Total and aligned reads for each sample are outlined in Supplementary data 2. HiSAT was set to search for at most 5 distinct, primary alignments for each read and to terminate the search either when this was reached or when no further alignments were detected. Gene expression was quantified using StringTie (Galaxy version1.3.3), specifying a minimum transcript length of 200 nucleotides, with GENCODE v19 as the reference genome [161, 162].

Comparison between treatment groups was performed using DeSeq2 (Galaxy version 2.11.39) [163]. Prior to this HiSAT bam files were converted into a count file using htseq-counts (Galaxy Version 0.6.1) using Gencode v19 as the reference. Those with a log2 fold change above 1 or below -1 were subject to gene ontology (GO) analysis using the Gene Ontology Consortium software [164] followed by removal of redundant terms using Revigo [165].

Cuffcompare (Galaxy version 2.2.1.0) was used to determine expression of transposable element using a transposable element reference sequence (rmsk, downloaded from UCSC, created using RepeatMasker programme) for comparison.

Genes or transposons with an FPKM more than 1 were considered expressed and used in further analysis. Venn diagrams were generated using the Venny software [166] and heatmaps generated using ClustVis [167].

## 2.13. Methylated and hydroxymethylated DNA Immunoprecipitation (MeDIP and hMeDIP)

### 2.13.1. Cell treatment

VU40T cells were treated with 100nM or 1µM DAC or an equivalent volume of 50% acetic acid for 96h in media containing charcoal stripped FBS (Gibco by Life Technologies) in place normal the FBS described in section 2.1.2. Charcoal stripped serum has reduced levels of lipid-based materials such as some hormones and growth factors. This provides a more defined background to test the molecular effects of drug treatment. Previous observations from the lab have suggested that charcoal stripped serum does not alter the growth of HNSCC cell lines and can amplify the effect of DAC treatment on gene expression. The media was changed daily and experiments performed in duplicate.

### 2.13.2. DNA extraction

MeDIP was performed as described previously and is summarised in Figure 2.1 using the buffers described in Table 2.6 [168]. Briefly, cells were harvested by cell scraping, homogenized with a 1.1x 40mm needle and syringe (Neolus) and centrifuged at 400xg for 10min. The cell pellet was resuspended in 300µl lysis buffer with 10µl 20mg/ml Proteinase K solution (Thermo Fisher Scientific) for 6h at 55°C. An equal volume of Phenol:Chloroform:Isoamyl alcohol 25:24:1 (PCIA) (Sigma-Aldrich) was added to the sample, mixed and centrifuged for 5min at 14,000xg. The top phase was collected, mixed with an equal volume of PCIA and centrifuged at 14,000xg for 5min. The top phase was retained and DNA precipitated using 100% ethanol and

75mM sodium acetate (Sigma-Aldrich). Centrifugation at 14,000xg was used to pellet the DNA, which was washed once in 70% ethanol and allowed to dry at room temperature. Pellets were resuspended in 100μl TE buffer and treated with 20μg/ml RNase A (Roche) for 1h at 37°C. DNA was quantified on a NanoDrop.

### 2.13.3. Immunoprecipitation

DNA (30μg) was sonicated using an EpiShear multisample sonicator (Active Motif) at 80% amplitude for 10 cycles of 30s 'on', 30s 'off' to generate fragments of 300-1000bp. For each immunoprecipitation 5-10μg of sonicated DNA was used, the amount was kept constant within each set of experiments. An input of 10μl of each sample was retained. A "no antibody" control sample was compiled from a mixture of all samples. The DNA was diluted in 1X TE buffer and denatured at 100°C for 10min then immediately quenched on ice. 10X IP buffer was added to give a final concentration of 1X. Prior to immunoprecipitation (IP) sheep anti-mouse IgG Dynabeads (M-280 Invitrogen) were washed three times in 0.1% BSA (Sigma-Aldrich) in PBS using a magnetic rack. To reduce nonspecific binding each sample was pre-cleared by incubation with 60μl of beads for 1h, rotating at 4°C, after which the sample was retained and the Dynabeads discarded. The samples were then incubated with 5μl of anti-5mC antibody (Eurogentec) for 2h, rotating at 4°C. To capture the antibody bound DNA 60μl of pre-washed beads were added and the sample incubated at 4°C for 2h.

### 2.13.4. Recovery of 5mC enriched DNA

The Dynabeads, now bound to 5mC enriched DNA, were washed three times in 1X IP buffer and incubated at 50°C overnight, shaking, in proteinase K digestion buffer + 7μg proteinase K (Ambion) to detach the DNA from the beads. A magnetic rack was used to collect the Dynabeads that were discarded and the supernatant, now containing the 5mC enriched DNA was retained. The DNA was extracted using PCIA extraction as described in section 2.13.2 and precipitated in 100% ethanol, 400mM NaCl and 2μl 4mg/ml glycogen (Thermo Fisher Scientific). The final 5mC enriched DNA pellet was resuspended in 30μl 1X TE.

**Table 2.6: Buffers used for MeDIP and hMeDIP**

All buffers were filter sterilized using a 0.22μm filter (Millex). Components were purchased from Sigma-Aldrich.

| Buffer | Components |
|---|---|
| **Lysis Buffer** | 20mM Tris ph 8.0<br>4mM EDTA<br>20mM NaCl<br>1% SDS |
| **0.1% BSA in PBS** | 9ml PBS<br>1ml BSA (10mg/ml stock) |
| **Proteinase K Digestion Buffer** | 50mM Tris pH8.0<br>10mM EDTA<br>0.5% SDS |
| **TE buffer** | 1M Tris-HCl<br>0.5M EDTA |
| **IP Buffer** | 100mM Na-Phosphate<br>1.4M NaCl<br>0.5% Triton X-100 |
| **100mM Na-Phosphate** | 39ml 2M monobasic sodium phosphate (276g/L)<br>61ml 2M dibasic sodium phosphate (284g/L)<br>1000ml $dH_2O$ |

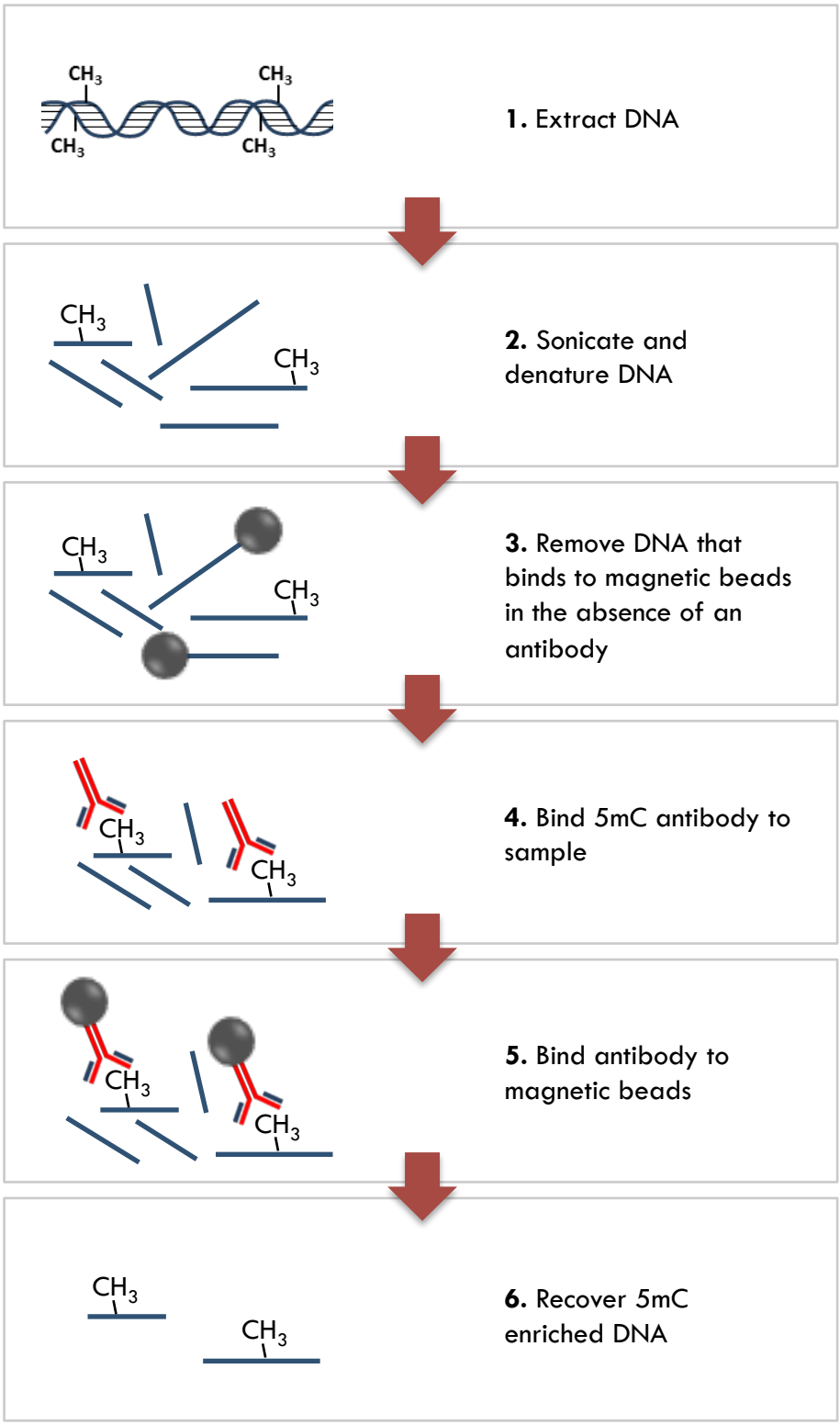



**Figure 2.1: Steps involved in the Methylated DNA immunoprecipitation (MeDIP)**

### 2.13.5. Hydroxymethylated DNA Immunoprecipitation (hMeDIP)

To determine the distribution of 5hmC in VU40T cells hMeDIP experiments were performed using the same protocol as described for MeDIP. To enrich for 5hmC marked DNA an anti-5hmC antibody (Active Motif: 33791) was used along with anti-rabbit IgG Dynabeads (M-280, Invitrogen).

### 2.13.6. (h)MeDIP-qPCR

To determine the efficacy of the MeDIP and hMeDIP immunoprecipitation, a qPCR was performed. Primers against *RARβ* promoter (methylated, positive control) and *KRT6A* (unmethylated, negative control) were used to test the MeDIP efficacy (Table 2.7). Previous hMeDIP sequencing results were used to design primers for regions positive (*BPIFB2*) or negative for 5hmC (Table 2.7). The qPCR was performed as described in Section 2.11.4. (h)MeDIP samples, input DNA and no antibody controls were diluted 1:4. A set of three standards was made from the diluted input sample at 100%, 10% and 1%. In the analysis the (h)MeDIP sample was normalised against the input sample to determine the percentage of total DNA immunoprecipitated by the 5(h)mC antibody and this was compared against the no antibody control.

**Table 2.7: Primer sequences used for MeDIP and hMeDIP qPCR check**

| Amplified Region | | Forward (5'-3') | Reverse (5'-3') | Product Size (bp) |
|---|---|---|---|---|
| Genomic Coordinates | Function | | | |
| chr20:31,613,004-31,613,142 | hMeDIP +ve control | TAGGTGGATGCATGGGTGAA | CACCCACCCATCAATCCGT | 139 |
| chr7:4899107-4899223 | hMeDIP –ve control | TGAGGGGTACGGGACTAGAT | ACAGTACCTCTAGCTGTCACA | 117 |
| chr3:25469696-25469800 (*RARβ*) | MeDIP +ve control | TAGACCCTCCTGCCTCTGAA | GGTAGGGTTCACCGAAAGT | 105 |
| chr12:52887101-52887235 (*KRT6A*) | MeDIP –ve control | GCTGGAAGGCAGGAGAATT | GAAGGTGAGCTTGCAGGTTG | 135 |

## 2.14. Illumina Library Preparation and Sequencing of (h)MeDIP samples

### 2.14.1. NEBNext ChIP-Seq Library prep Master Mix Set

VU40T (h)MeDIP libraries were prepared using the NEBNext ChIP-Seq Library prep Master Mix Set for Illumina using Multiplex oligonucleotides for Illumina (New England Biolabs). The total volume of hMeDIP or 50ng of MeDIP were used for library preparation.

#### *2.14.1.1. End repair of sonicated DNA fragments*

The DNA was incubated with NEBNext End Repair Enzyme Mix and Buffer for 30min at 20°C to repair the ends of the sonicated DNA fragments.

#### *2.14.1.2. Solid Phase Reversible Immobilisation (SPRI) clean up*

The product was cleaned up by SPRI using Agencourt AMPure XP reagent (Beckman Coulter). This process involves incubating the DNA with AMPure XP magnetic beads for 5min. A magnetic rack is then used to capture the DNA bound beads, which are washed twice in 80% ethanol and allowed to dry briefly at RT. DNA is recovered by incubation with 0.1X TE for 2min, following which the magnetic beads are captured and the supernatant retained.

#### *2.14.1.3. dA-tailing and adapter ligation*

To add dA tails to the end-repaired DNA, Klenow fragment plus dA-tailing Reaction Buffer were added to the sample and incubated at 37°C for 30min. The samples were cleaned up using the SPRI method (Section 2.14.1.2). Quick T4 DNA Ligase,

NEBNext Adaptor and Quick Ligation Reaction Buffer were added and the sample incubated at 20°C for 15min to ligate the illumina adaptor to the DNA. The NEBNext adapter contains a stem loop structure with a uracil base at the centre of the ring. USER enzyme (3µl) was added and the sample incubated at 37°C for 15min to excise the DNA at this uracil.

#### *2.14.1.4. Dual-SPRI size selection*

Dual-SPRI size selection was performed to select for smaller fragments of adapter ligated DNA (~300bp) which are suitable for Illumina sequencing. The library was incubated with 0.9X AMPure XP beads at RT for 5min. AMPure XP beads, bound to fragments over 300bp, were captured with a magnetic rack and discarded. The supernatant, containing smaller fragments, was transferred to a new tube and incubated with 0.2X AMPure XP beads for 5min. This captured DNA of the correct size on the beads, which were retained, while the supernatant was discarded. Beads were washed twice in 80% ethanol and DNA eluted into 0.1X TE buffer.

#### *2.14.1.5. Sample indexing and amplification*

Distinct indexes must be attached to each sample so that sample-specific reads can be separated in the analysis when multiple samples are sequenced on the same flow cell. A PCR step was used to enrich the adapter ligated DNA and attach these indexes to the samples. This contained NEBNext High-Fidelity Master Mix, Universal PCR Primer and one of 12 indexes from the NEBNext® Multiplex Oligos for Illumina (E7335S)(different for each sample). The PCR protocol used is summarized in Table

2.8 below. After final SPRI clean up (Section 2.14.1.2), the library DNA was eluted in 15µl TE buffer and stored at -80°C prior to use.

qPCR using (h)MeDIP positive and negative control primers (Section 2.13.6) was used to confirm the quality of the libraries prior to sequencing.

**Table 2.8: NEBNext library preparation PCR programme**

| | Temperature | Time (Seconds) | |
|---|---|---|---|
| **Initial denaturation** | 98 | 30 | |
| **Denaturation** | 98 | 10 | X15 |
| **Annealing** | 65 | 30 | |
| **Extension** | 72 | 30 | |
| **Final Extension** | 72 | 5mins | |
| | 4 | Hold | |

## 2.14.2. Kapa Hyper Library Preparation

The Kapa Hyper library preparation kit (KapaBiosystems) was used to generate Illumina libraries from HOK (h)MeDIP experiments. This kit claims a greater yield from lesser quantities of DNA and therefore was better suited for experiments involving primary cells. The protocol is similar to the one described in 2.14.1, however the adapter ligation and indexing are applied in a single step, and fewer SPRI clean up steps are applied, therefore the loss of material during library preparation is reduced.

### *2.14.2.1. End repair, dA-tailing and adapter ligation*

The library preparation was performed as the manufacturer recommends. Approximately 50ng of (h)MeDIP DNA was used for each library. In a single step

fragmented DNA was end repaired and A-tailed by incubating the DNA with End Repair and A-tailing Buffer and Enzyme Mix at 20°C for 30min then 65°C for 30min. Illumina-indexed adapters were then ligated to the fragmented DNA by incubating the A-tailed DNA fragments with DNA Ligase, Ligation Buffer and distinct indexed adapters for 1h at 20°C.

The ligated product was cleaned up by SPRI as described in section 2.14.1.2. After adapter ligation this was performed twice to remove adapter dimers.

#### *2.14.2.2. Library amplification*

The adapter-ligated library was amplified with KAPA HiFi HotStart ReadyMix and KAPA Library Amplification Primer Mix using the PCR programme outlined in

Table 2.9: KAPA Hyper Library preparation PCR programme. A further SPRI clean up (section 2.14.1.2) was performed.

**Table 2.9: KAPA Hyper Library preparation PCR programme**

| | Temperature | Time (Seconds) | |
|---|---|---|---|
| **Initial denaturation** | 98 | 45 | |
| **Denaturation** | 98 | 15 | X14 |
| **Annealing** | 60 | 30 | |
| **Extension** | 72 | 30 | |
| **Final Extension** | 72 | 5mins | |
| | 4 | Hold | |

#### *2.14.2.3. Dual-SPRI size selection*

To select for adapter ligated DNA fragments of 250-450bp in size, dual-SPRI size selection was performed. The library was incubated with 0.6X AMPure XP reagent for 10min at RT. The magnetic beads containing the larger fragments were captured and discarded and the supernatant retained. The supernatant was then incubated with 0.2X AMPure XP beads at room temperature for 10min. A magnetic rack was used to capture the beads and the supernatant, containing fragments smaller than required, was discarded. Beads were washed twice in 80% ethanol, allowed to dry briefly and the library DNA was eluted in 10mM Tris HCl.

### 2.14.3. Library Fragment Size Determination

The average size of the DNA fragments in the library were determined on a 2100 BioAnalyzer using High Sensitivity DNA analysis kit (Agilent) as per the manufactures instructions. The BioAnalyser chip is equivalent to a small, very sensitive gel, where the fragment sizes of the samples can be determined by the rate at which they move through the gel relative to the provided ladder.

#### *2.14.3.1 Preparing the gel mix*

The gel mix was prepared prior to use by adding 15µl High Sensitivity DNA dye concentrate to one vial of High Sensitivity DNA gel matrix. This was vortexed to mix, transferred to a spin filter (provided) and centrifuged at 2240xg for 10min.

#### *2.14.3.2 Set up chip priming station*

To set up the chip priming station a 1ml syringe was inserted into the luer lock adapter and screwed tightly onto the priming station. The clip lever was then

released and the luer slid down to the lowest position. A high sensitivity DNA chip was then placed on the priming station and the plunger of the syringe brought to the 1ml position.

#### *2.14.3.3 Loading the gel-dye mix and samples*

9µl of gel-dye mix was loaded into the well indicated by the protocol. Following this the chip priming station was closed and the plunger of the syringe pushed down. After 1min the plunger was released, and slowly pulled back to the 1ml position. This procedure allows a thin layer of gel to be spread across the base of the chip. In other wells indicated by the protocol, 9µl of gel-dye mix was loaded.

To each sample and ladder well, 5µl of marker (provided) was loaded. Following this 1µl of ladder (provided) was added to the ladder well and 1µl sample to each of the sample wells. The chip was vortexed at 2400rpm for 1min and was immediately read on the 2100 BioAnalyser. The BioAnalyser images the bands on the gel and displays them as a graph. The size of the fragments in the sample was determined relative to the ladder.

### 2.14.4. Library DNA Quantification

The SYBR FAST LightCycler 480 qPCR Kit (KAPA) was used as per the manufacturer's instructions to quantify the total amount of adapter-ligated DNA in each library. Libraries were first diluted to 1:1000 and 1:5000. For each reaction 6µl KAPA SYBR FAST qPCR Master Mix containing Primer Premix was combined with 2µl diluted library or standard and 2µl nuclease free water. These were analysed on the Roche Light Cycler 480 II using the PCR programme detailed in Table 2.10. The kit provides

a set of 6 standards ranging from 0.0002pM to 20pM and the concentration of the libraries was determined from this. The size-adjusted concentration can be calculated by multiplying the concentration by (452 (average fragment length of the standards)/ (average fragment length of a sample).

**Table 2.10: KAPA Library Quantification thermal cycling conditions**

| Step | Temperature (°C) | Time | |
|---|---|---|---|
| Initial activation/denaturation | 95 | 5 min | |
| Denaturation | 95 | 30 sec | 35 cycles |
| Annealing/extension/data acquisition | 60 | 45 sec | |

### 2.14.5. Illumina Sequencing

(h)MeDIP Libraries to be analysed on one flow cell were pooled in equimolar quantities to a total concentration of 2nM and sequenced on a HiSeq 2500, using V3 SBS Chemistry (1x51 cycles).

### 2.14.6. (h)MeDIP Sequencing Data Analysis.

#### *2.14.6.1. MeDIP and hMeDIP alignment*

MeDIP and hMeDIP sequencing data were aligned to the human genome (build hg19) using Bowtie2 (v2-2.2.6), and with '*very-sensitive-local*' alignment option (example command line: *bowtie2-2.2.6/bowtie2-align-s --wrapper basic-0 -p 2 --very-sensitive-local -x ucsc.hg19.fa –reads sample.fastq.gz*). Reads overlapping the Encode blacklisted regions were removed using bedtools (v2.22.0). MeDIP and hMeDIP sequencing alignment was performed by Dr S. Clokie (West Midlands

Regional Genetics Laboratory, Birmingham Women's Hospital, UK). The resulting sequenced and aligned reads are outlined in Supplementary data 3.

#### *2.14.6.2. MeDIP peak calling and annotation*

For each sample peaks were called from MeDIP aligned reads using MACS v2.1.1 [169]. To increase peak-calling specificity and better eliminate background bias, a non-MeDIP sample (input DNA) was used as the control. Candidate peaks were called relative to genomic background and filtered using a minimum FDR cut-off of 0.1 to call significant regions. Blacklisted regions were removed.

Peaks from different samples were merged using bedtools to obtain an aggregated list of all identified peaks. Upon merging, peaks within 150bp are counted as the same peak and the midpoint taken as the peak summit. For each individual sample, HOMER v4.8 [170] was used to annotate peaks and calculate the normalized read density (RPM) averaged over a 400bp window centred on the peak summits taken from the aggregated peaks list. MeDIP peak calling and annotation is attributed to Dr W. Croft (University of Birmingham, UK).

#### *2.14.6.3. hMeDIP peak calling and annotation*

For hMeDIP analysis the Homer suite (v4.8) [170] was used to call peaks and perform annotation of the peaks using default settings. The command '*makeTagDirectory VU40T-Aza-unique-tbp1-frag350 hMeDIP.merged.hg19.brm.bam -checkGC -tbp 1 -fragLength 350 -genome hg19 -i input.hg19.brm.bam'* was used to ensure uniquely mapped reads are retained, the background signal is removed (to improve sensitivity) and the expected fragment size (350bp) utilised. The –tbp 1 setting on

Homer was used to count only unique alignments. Peaks from different experimental conditions were merged using the '*mergePeaks*' program from Homer. It produced a list of peaks occurring in at least one of the conditions as well as a summary statistics of the intersect between different input lists.

Peaks were annotated with counts, CpG and CG%, detailed annotations, distance to TSS and closest gene determined using the following command from Homer: *'annotatePeaks.pl merged-peaks.txt hg19 -size 350 -CpG -go output-go-results -annStats annotation-stats.txt -d /sample1 /sample2 /etc > annotation-results.txt'*. hMeDIP peak calling and annotation was performed by Dr S. Clokie.

MeDIP and hMeDIP peaks were annotated independently for each replicate and sample. In the case of the VU40T samples, where replicates were available, analysis was performed on each replicate and the mean calculated. When comparisons between treatments or cell lines were performed (i.e. when using the merged file) all peaks from all replicates were considered.

#### *2.14.6.4. Comparison of hMeDIP peaks with published datasets*

The UCSC table browser [171] was used to transfer ENCODE datasets for histone modification, transcription factor binding sites, DNase hypersensitive sites and chromatin states for primary Normal Human Epidermal Keratinocytes (NHEK) [172, 173] and CpG islands to Galaxy [160].

On Galaxy the BedTools '*intersect intervals'* (Galaxy version 2.27.0.1) function was used with the default settings to determine how many hMeDIP peaks overlap with NHEK datasets [174, 175]. Using the default settings, this analysis counts any overlap

between the two data sets and will only count one incident per MeDIP or hMeDIP peak.

Using the default settings on Galaxy [160] the '*computeMatrix*' tool was used to align hMeDIP peaks to NHEK datasets and the '*plotProfile*' tool was used to visualise these [176]. These were performed using 'reference point' mode, using the centre of the regions as the reference point. Relative enrichment was calculated by separating each region into 50bp bins and showing the average overlap within each bin relative to the overlap across the region of interest.

### *Gene Ontology*

The online software GREAT [177] was used to assign genes to genomic co-ordinates and perform gene ontology on these.

### *Enrichment of 5hmC at Alu element subfamilies*

The repeatmasker bed file that contains annotated Alu elements was downloaded from UCSC. HOK hMeDIP reads were counted and the read density over each Alu element subfamily was calculated using Homer histogram [170]. Histograms were scaled using a *'-size given'* option of 4000 (+/- 2kb of centre of the Alu element) and a '*-hist'* option of 10 (bins of 10bp). This scales the histograms individually across the 4kb region for each Alu element; regions are separated into bins of 10bp and scaled based on the total number of reads aligned to the 4kb region. These individual Alu histograms were collated together to create a matrix suitable for plotting. Dr S. Clokie conducted this analysis.

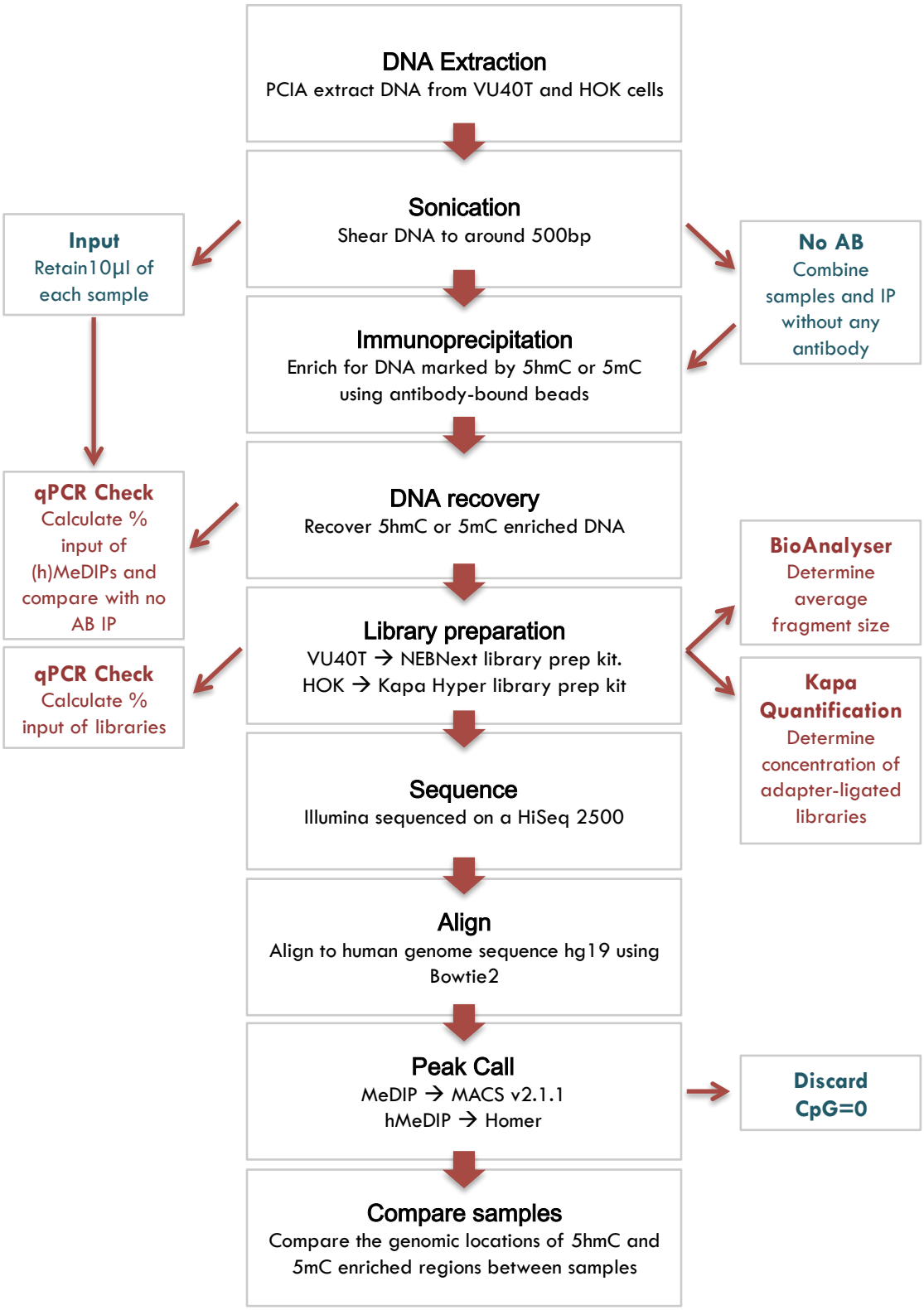



**Figure 2.2. Overview of the steps of MeDIP and hMeDIP sequencing and analysis**

Quality control steps are shown in red and anything removed from the samples shown in blue.

# Chapter 3. DAC as a potential therapeutic in the treatment of HNSCC

The DNA demethylating agent DAC, has FDA and EMA approval for the treatment of some haematological malignancies. Progress is hindered by toxicities at high doses. However, it has been observed that low dose treatments are effective and that efficacy might be transferable to solid tumours [90, 122, 127]. Like other cancers, HNSCC is characterised by promoter methylation of tumour suppressor genes, therefore, this chapter aims to examine the potential efficacy of DAC, either alone, or combined with other drugs, in the treatment of HNSCC [97].

## 3.1. Mechanism of action of DAC

### 3.1.1. DNA demethylation

DAC is a cytosine analogue that is recognised by the DNA replication machinery as the equivalent to cytosine and incorporated into DNA in its place [75]. One mechanism by which DAC decreases 5mC levels is via this incorporation, as unlike cytosine, DAC cannot be methylated and so the mark is diluted out of the DNA through subsequent rounds of replication. Similar mechanisms have been proposed for AZA and other nucleoside analogues [75, 178]. However, the degree of hypomethylation observed after nucleoside analogue treatment exceeds the levels

of analogues found in the DNA [178]. Furthermore, DNMT1 was co-purified with DNA in cells treated with DAC [179]. Once DAC is incorporated into the DNA, DNMT1 recognises and binds to DAC as it would to cytosine [75]. This involves covalent bond formation between the DNMT enzyme and the carbon-6 atom on the pyrimidine ring. When bound to cytosine this bond is resolved by beta-elimination through the carbon-5 atom. However, DAC has a nitrogen atom at this position and so the covalent bond cannot be resolved [75]. DNMT becomes trapped on the DNA, preventing it from functioning as a methyltransferase. Furthermore, the adducts formed from this trapping trigger DNA damage signalling pathways resulting in the eventual degradation of DNMT [75]. The 5mC mark is then diluted out through subsequent rounds of replication (Figure 3.1).

Demethylation by DNMTi treatment can reactivate epigenetically silenced tumour suppressor genes. Additionally, at low doses, DAC and AZA demethylate endogenous retroviruses (ERVs), allowing them to be expressed as double-stranded RNA (dsRNA) [158, 180]. The ERV dsRNA activates the immune system to produce an antiviral response, eventually leading to apoptosis [158, 180]. This mechanism of action of low dose DNMTi is referred to as viral mimicry and is discussed in more detail in (Chapter 6) [158, 180].

### 3.1.2. DNA damage

In addition to demethylating the DNA, DAC causes DNA damage [181]. This is probably due to adducts in the DNA formed by DNMT trapping (Figure 3.1) [75]. Cells with reduced levels of DNMTs were significantly more resistant to the toxic effects of DAC than wild type cells; the induction of γH2AX (a marker of DNA damage) by DAC

was not observed in *DNMT1* knockout cells. After incorporation into DNA, DAC traps DNMT1, causing DNA adducts which trigger cell cycle arrest and the inhibition of proliferation via activation of the G1 checkpoint protein p53 [75, 181, 182].

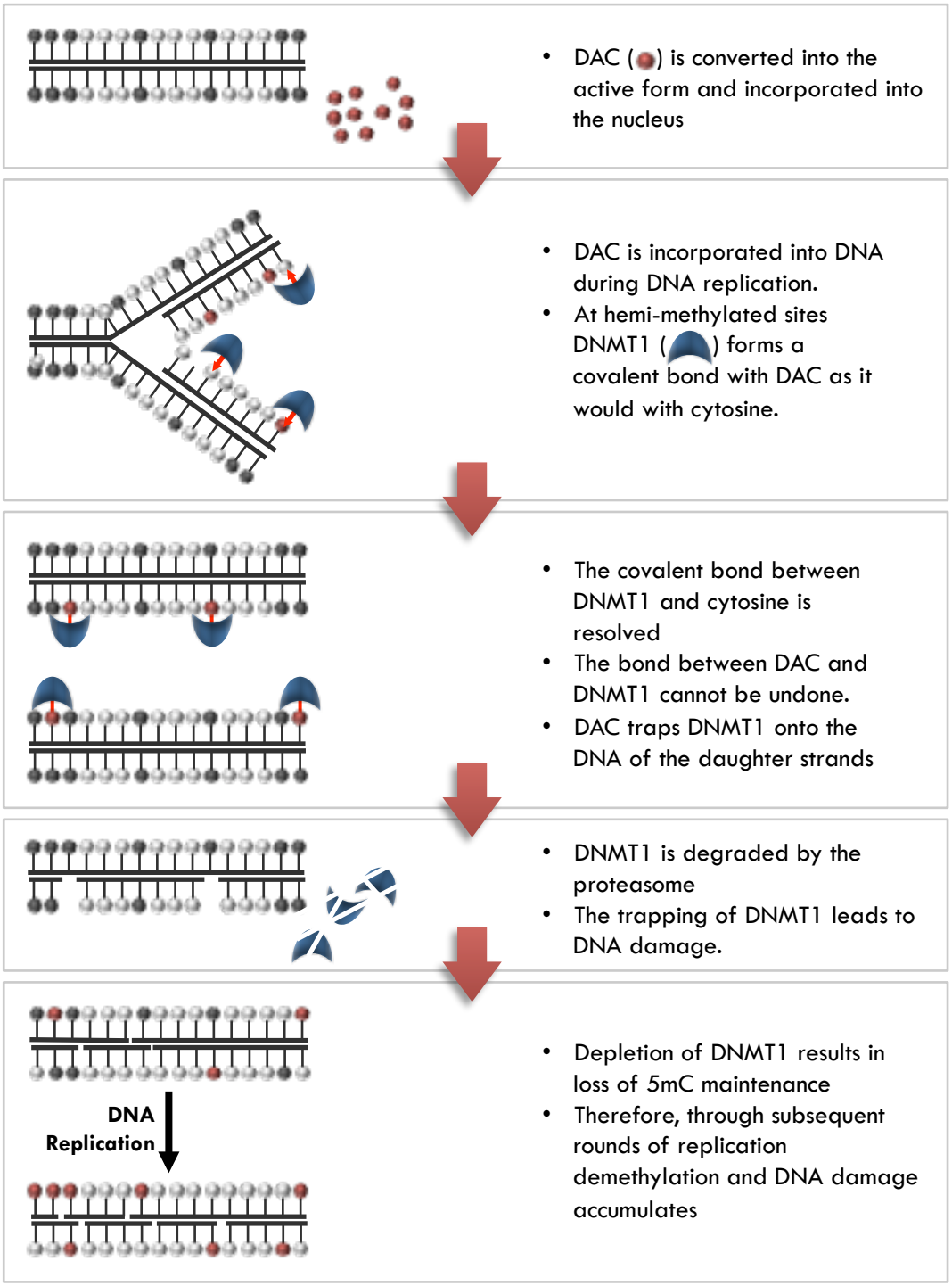



**Figure 3.1. The mechanism of action of DAC**

DAC is incorporated into DNA in place of cytosine. It then covalently traps the DNMT enzyme. This causes adducts in the DNA that lead to DNA damage. Additionally the DNMT enzyme is degraded, causing less DNMT available to remethylate the DNA after replication. Subsequent rounds of replication lead to genome-wide hypomethylation and the accumulation of DNA damage. Adapted from [75].

## 3.2. Mechanisms of resistance to DAC treatment

A subset of MDS patients displays resistance to DAC in the clinic [183, 184]. Moreover, when the response to DAC treatment was examined in a panel of cancer cell lines the concentration required to inhibit cell growth by 50% ($IC_{50}$) varied 1000-fold between the cell lines [185]. This variation correlated with global 5mC changes after DAC but not with baseline 5mC levels [185]. Therefore efficacy is most likely related to uptake, retention or activation of the drug. In humans the transport of nucleosides across membranes entails: equilibrative uniporters that transport nucleosides down a concentration gradient (members of the *SLC29A* gene family such as hENT1 and hENT2); concentrating transporters that transport nucleosides against a concentration gradient (*SLC28A* gene family, hCNT1 and hCNT3 proteins); organic cationic/anionic transporters that use facilitated transport to promote the uptake of nucleosides (*SLC22A* gene family, OCT and OAT proteins); and ATP-dependent exporters which use ATP to export nucleosides out of the cell (ABC family) (Figure 3.2). Nucleoside analogues, such as DAC, harness these transporters to gain access into the cell [186]. Accordingly, response to DAC treatment inversely correlates with *hENT1* expression in cancer cell lines and MDS patients [183-185].

Once in the cell, DAC is converted into the active form (5-aza-2'-deoxycytidine-triphosphate) by deoxycytidine kinase (DCK), while cytidine deaminase (CDA) can inactivate the drug (to 5-aza-2'-deoxyuridine) [75]. Therefore, the ratio of CDA/DCK may correlate with response to treatment (Figure 3.2) [184, 185]. Indeed, a three-fold higher CDA/DCK ratio was found in DAC non-responsive MDS patients [184].

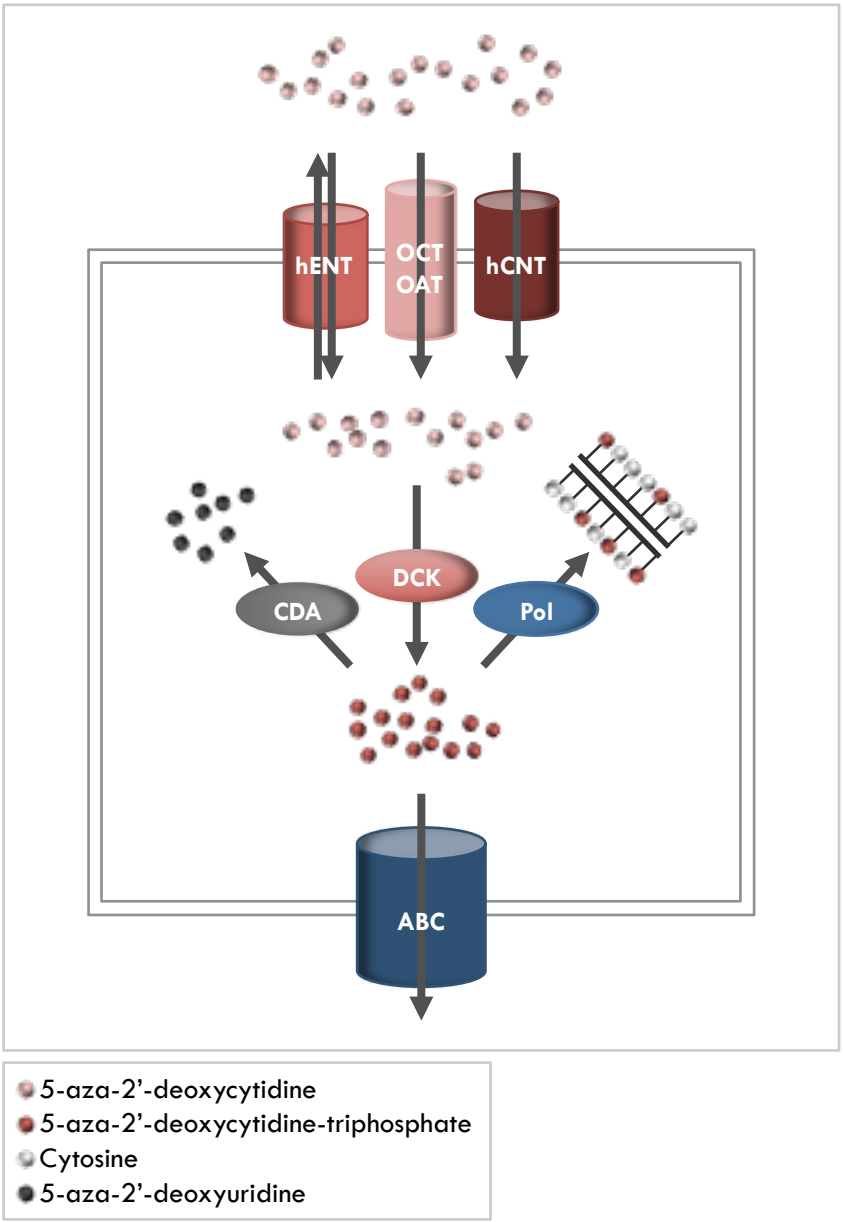



**Figure 3.2. DAC transport and activation**

DAC can be transported into the cell by equilibrative uniporters (hENTs); concentrating transporters (hCNTs); organic cationic/anionic transporters (OCT and OATs). Deoxycytidine kinase (DCK) can then convert DAC into the active form that can be incorporated into DNA by polymerase (Pol) during DNA replication. Cytidine deaminase (CDA) can inactivate DAC and ATP-dependent exporters (ABC family) can export DAC out of the cell. Adapted from [75].

## 3.3. Aims and objectives

In this chapter the DNA demethylating agent, DAC, is studied for the potential use in the treatment of HNSCC and a drug-repurposing screen is used to determine whether efficacy can be increased through combination therapy. These aims are explored through the following objectives:

- The response to DAC is examined in four HNSCC cell lines and in primary human oral keratinocytes
    - Differences in the response to DAC are investigated through expression of *DNMT*s, extent of DNA demethylation and RNA-sequencing analysis.
- A panel of 100 off-patent drugs is screened for DAC-sensitising qualities and the top hits further investigated
    - The Chou-Talalay method is used to determine whether the two drugs are working in synergy
    - Giesma / Jenner staining, FACS analysis, Ki67 staining and RNA sequencing were used to investigate the mechanism behind combination therapy.

## 3.4. Cell viability assay

In this report the CellTiter Blue cell viability assay was used to determine changes in viability of HNSCC cell lines in response to drug treatments. Viable cells convert resazurin into resorufin. Therefore, the amount of each compound in a sample gives a relative measure of viable cell density. This provides a means of high throughput drug testing. However, it is important to note, that metabolic changes could also alter the results.

To validate the efficacy of this assay the HNSCC cell line, VU40T was plated at an increasing cell density followed by measurement using the CellTiter Blue assay. The density of VU40T cells corresponds with an increasing pink colour and an increasing absorbance at 570nM, indicating conversion from resazurin to resorufin (Figure 3.3). The addition of 500nM DAC to the treatment does not alter the curve, suggesting that 24h treatment with DAC has no effect on viability at this concentration and that the drug does not interfere with the assay; therefore this reagent can be used to measure the effect of DAC on the viability of HNSCC cells (Figure 3.3). Both curves begin to plateau at ~$1x10^5$ cells / well suggesting that the accuracy of the reagent will decrease when cell density is too high. Therefore all subsequent experiments were performed at less than 80% confluency at the time of data collection.

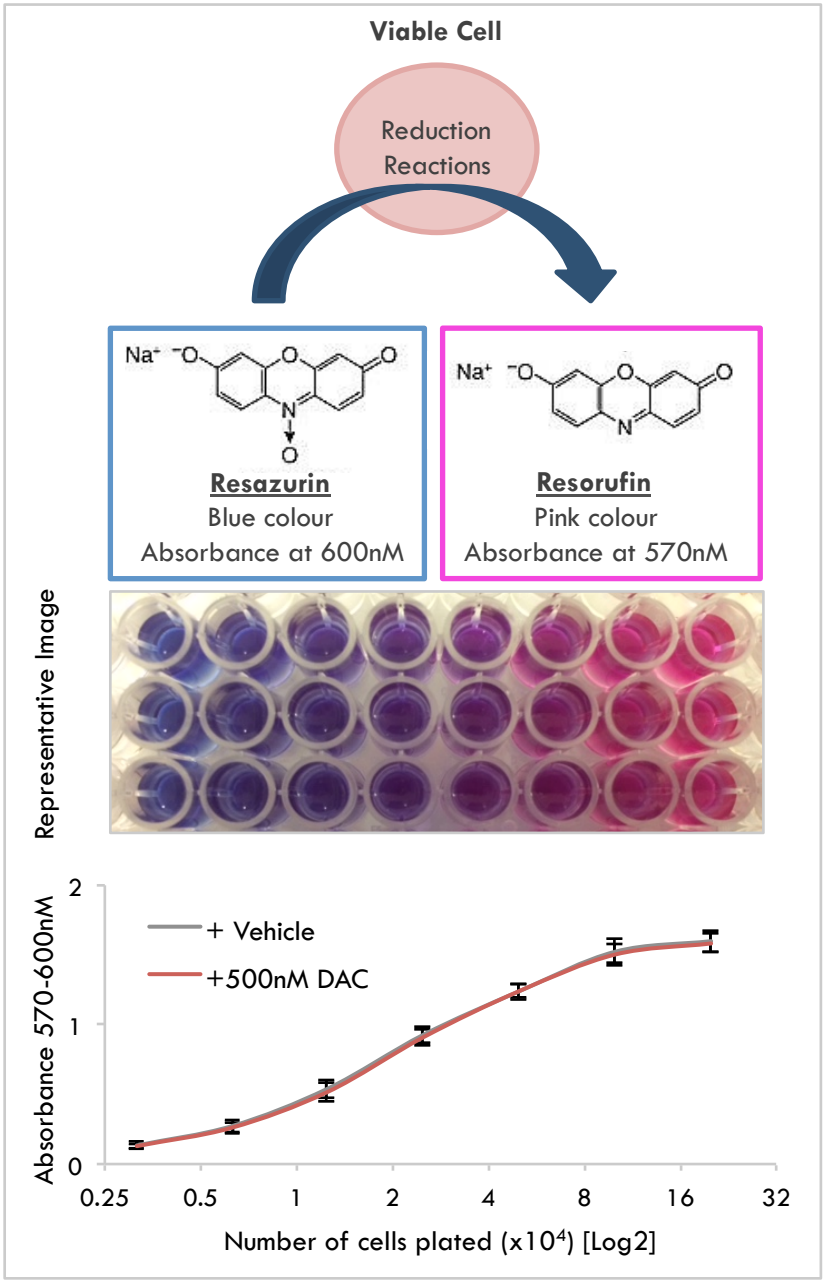



**Figure 3.3. CellTiter Blue cell viability assay**

To test the efficacy of the cell viability experiment (Section 2.2.2) CellTiter Blue reagent was added to a titration of VU40T cells +/- 500nM DAC cultured for 24h. The reagent contains the compound resazurin that is visible as a blue colour in the media and can be read at an absorbance of 600nM. Viable cells convert this into resorufin, which is pink and the absorbance can be read spectrophotometrically at 570nM. The middle panel shows a representative image of the colour change. The bottom panel shows the corresponding change in absorbance. Results are a mean of 3 experiments, and error bars show SEM.

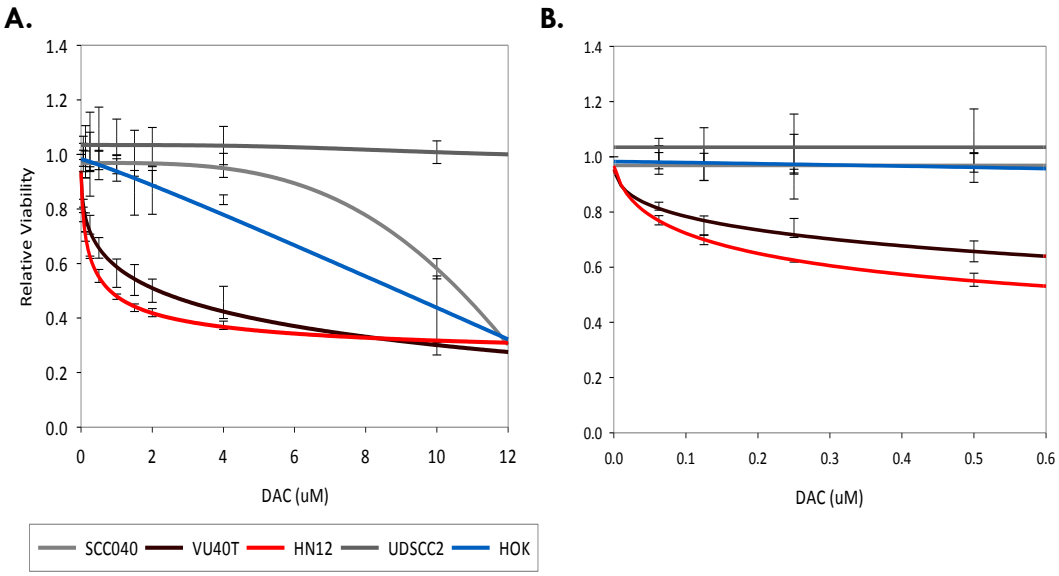
## 3.5. The efficacy of DAC treatment varies in HNSCC cell lines

To determine the efficacy of DAC treatment in HNSCC a panel of relevant cell lines (HNSCC cell lines VU40T, HN12, UDSCC2, SCC040 and primary human oral keratinocyte cells, HOK) were subject to a titration of DAC (0.0625, 0.125, 0.25, 0.5, 1, 1.5, 2, 4, 10μM), following which viability was determined (Figure 3.4). The four HNSCC cell lines stratified into two distinct groups: those that respond to nanomolar doses of DAC (VU40T and HN12) and those that showed minimal response to treatment (SCC040 and UDSCC2) (Figure 3.4). According to the EMA datasheet, in patients with AML or MDS utilising a 5-day regimen of 20mg/$m^2$ DAC, the maximum plasma concentration (Cmax) was determined to be 107ng/ml, equivalent to 469nM [187]. Therefore, in order to remain clinically relevant the dose of DAC used in cell culture cannot exceed 500nM. The two responsive cell lines, HN12 and VU40T, showed an initial response from as little at 50nM and a drop in viability of roughly 40% at 500nM DAC (Figure 3.4). However, UDSCC2 and SCC040 did not respond to clinically relevant doses; UDSCC2 did not respond to DAC at any concentration tested, while SCC040 only responded to the clinically irrelevant dose of 10μM (Figure 3.4). None of the cell lines tested showed $IC_{50} \leq$ 500nM DAC. Primary human oral keratinocyte cells, HOK, were used to represent non-cancerous tissue. These grouped with the non-responsive HNSCC cells, only responding to clinically irrelevant concentrations of the drug, above 4μM (Figure 3.4).

DAC treatment exerts its effects via two main mechanisms: global DNA demethylation and DNA damage. Both of these are dependent upon DAC incorporating into the DNA and trapping the DNMT enzymes [75]. Therefore the

basal levels of the DNMT enzymes could correspond to efficacy of the drug. qRT-PCR analysis was used to determine basal RNA levels for *DNMT-1, -3A,* and *-3B* (Figure 3.5). In general, HNSCC cell lines showed higher levels of *DNMTs* than primary oral keratinocytes. However, while *DNMT* expression varied between cell lines, this did not correspond with sensitivity to DAC.

The global level of DNA methylation after 96h of 500nM DAC treatment was determined using a DNA dot blot. Encouragingly, 5mC changes corresponded with the sensitivity to DAC previously described. DAC-responsive cell lines showed a significant decrease in 5mC after DAC treatment, while unresponsive cell lines did not (Figure 3.5, Supplementary data 4). There was no significant difference in the basal levels of 5mC between the cell lines. These data would indicate that the efficacy of DAC treatment is dependent upon how DAC is activated, incorporated or retained within the cell.




**C.**

| | SCC040 | VU40T | HN12 | UDSCC2 | HOK |
|---|---|---|---|---|---|
| **IC50 ($\mu$M)** | 10.61 | 2.17 | 0.81 | N/A | 8.93 |

**Figure 3.4. HNSCC cells show variable sensitivity to DAC treatment**

A DAC titration was undertaken over a 96h period on 4 HNSCC cell lines and primary oral keratinocytes (HOK). Viability was determined as described in section 2.2.2. The results show differing efficacy of the drug between cell lines with HN12 and VU40T showing a rapid reduction in viability while SCC040 cells appeared resistant to concentrations of DAC up to 4µM and UDSCC2 was resistant to up to 10µM. **A.** All concentrations applied. **B.** Clinically relevant concentrations only. Results, shown relative to the vehicle only control, are mean + SEM (n=3) **C.** The concentration of DAC required to reduce growth by 50% (IC50) values calculated from the curves in **A.** N/A: not determined.

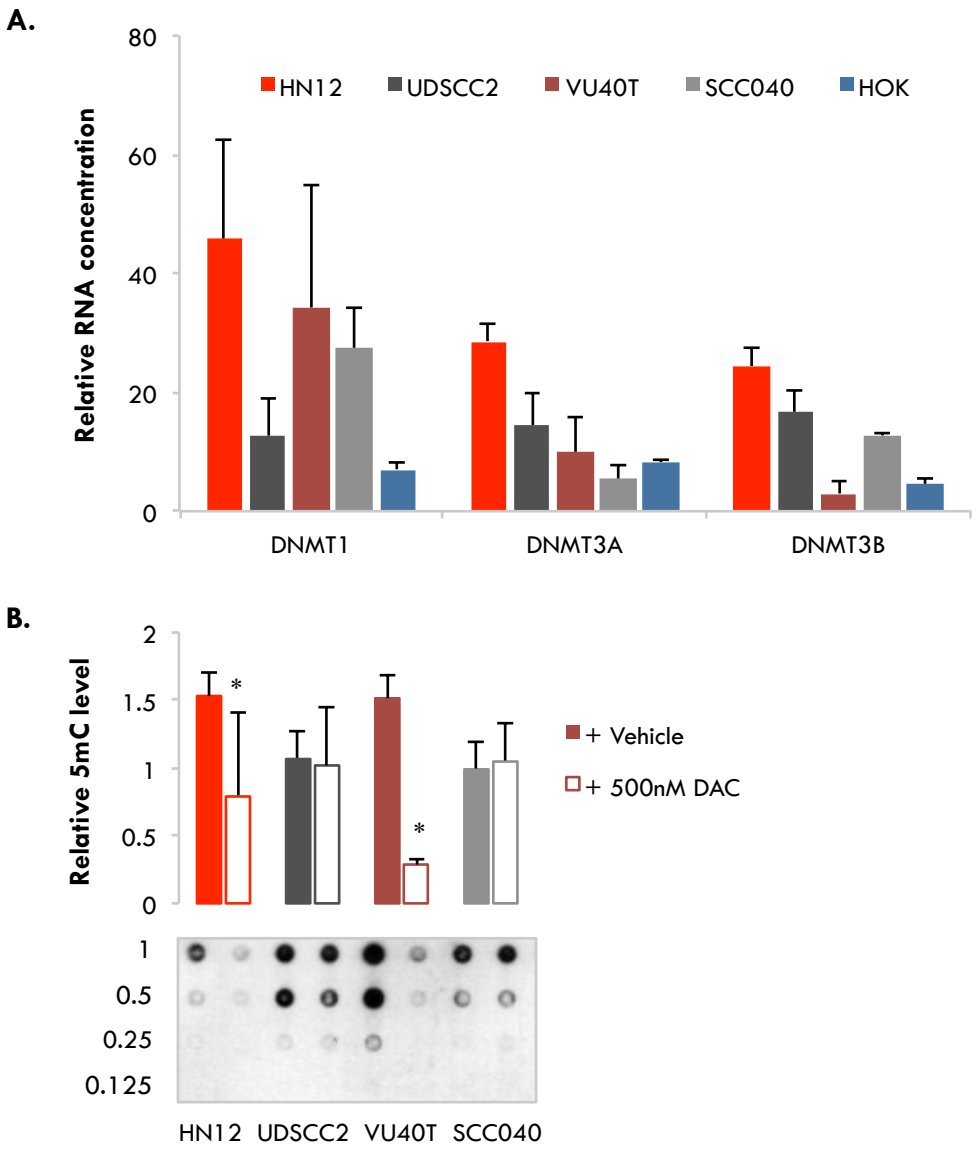



**Figure 3.5. Methylation changes in response to DAC treatment**

**A.** The relative mRNA levels of DNMT enzymes in HNSCC cell lines and HOK primary cells was determined by qRT-PCR analysis. **B.** DNA dot blot showing global DNA methylation changes after 96h treatment with 500nM DAC. Total DNA was determined by staining blots with methylene blue. The intensity of each 5mC blots was normalised to the corresponding methylene blue dot. The blot shown is representative while the graph displays the mean + SEM (n=3). Additional blots shown in Supplementary data 4. A paired t-test was used to determine significance between untreated and treated samples where * $p<0.05$.

RNA sequencing was performed on DAC-responsive (HN12, VU40T) and unresponsive (SCC040, UDSCC2) cell lines to examine whether differences in the transcriptional profile could account for the response to DAC. The four cell lines were initially compared in an unbiased manner: a Venn diagram was used to determine the whether any genes were jointly expressed in the responsive cell lines and lacking in the unresponsive cell lines (Figure 3.6). To do this, a threshold FPKM value of 1 was used to discard lowly expressed genes from the comparison. The majority of genes (5591) are expressed in all HNSCC cell lines while 255 genes were expressed only in DAC-responsive cell lines (HN12 and VU40T) and 309 genes – in DAC-resistant cells (UDSCC2 and SCC040). These groups were subject to gene ontology (GO) analysis for biological processes using the analysis software DAVID [188]. The GO term 'drug transport' was enriched in the DAC-responsive group and included the genes *SLC15A2 and SLC22A5* (Supplementary data 5). To further examine this result, the RNA levels (FPKM values) of a user selected list of genes reported to be involved in nucleoside transport and activation was determined [75] (Table 3.1). The majority of these genes had FPKM values less than 1 and therefore were not regarded as expressed. However, the equilibrative nucleoside transporter *SLC29A1* was expressed in all cell lines, and had higher FPKM values in the DAC-responsive cell lines. qRT-PCR results confirmed higher expression of the drug transporters *SLC22A5* and *SLC15A2* and the nucleoside transporter *SLC29A1* in the DAC responsive cell lines over the unresponsive cells, with significantly higher *SLC15A2* expression in HN12 cells and *SLC29A1* significantly higher in VU40T cells (Figure 3.7).

The comparison software DeSeq2 was then used to compare the DAC-responsive cell lines with the unresponsive cell lines. 2274 genes had higher expression in the DAC-responsive cell lines, while 2881 genes were associated with the DAC-unresponsive group. The results were subject to gene ontology (GO) analysis and interestingly, GO terms associated with the regulation of metabolism of nucleobase containing compounds were enriched in the DAC-unresponsive group (Figure 3.8, Supplementary data 5).

In conclusion, DAC-responsive cell lines are characterised by increased expression of genes involved in drug transport and DAC-unresponsive cell lines are characterised by genes involved in the metabolism of nucleobase containing compounds (Figure 3.6, Figure 3.8). Therefore, the differential response to DAC treatment could be attributed to either or both of these results.

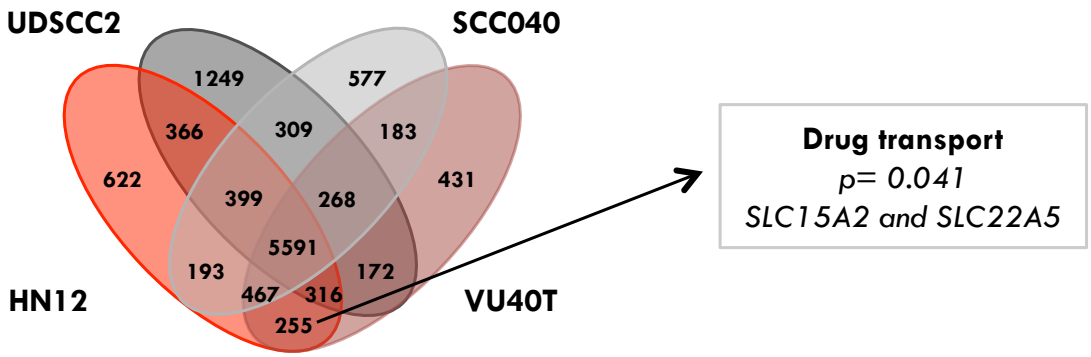



**Figure 3.6. Differential RNA expression between DAC responsive and unresponsive cell lines**

RNA sequencing of HNSCC cell lines. RNA was aligned using HiSAT2 and expressed genes determined using StringTie. All genes with an FPKM >1 for each cell line are compared using the Venny software [189]. The genes common to VU40T and HN12 cells were subject to gene ontology biological process analysis using DAVID [188].

**Table 3.1. DAC responsive cell lines exhibit increased expression of *SLC29A1***

The FPKM values for a user defined list of genes with a putative involvement in DAC transport or activation from the transcriptome of HNSCC cell lines. Genes with an FPKM<1 not considered as expressed.

| Gene | UDSCC2 | HN12 | SCC040 | VU40T | Function |
|---|---|---|---|---|---|
| SLC22A1 | 0.01 | 0.01 | 0.05 | 0.02 | Organic cationic/anionic transporters (OCT/OAT): promote the uptake of nucleoside analogues |
| SLC22A2 | 0 | 0 | 0 | 0 | |
| SLC22A3 | 0 | 0.03 | 0.06 | 0.62 | |
| SLC22A4 | 0.46 | 0.72 | 0.5 | 0.19 | |
| SLC22A6 | 0 | 0 | 0 | 0 | |
| SLC22A7 | 0 | 0 | 0 | 0 | |
| SLC22A8 | 0 | 0 | 0 | 0 | |
| SLC28A1 | 0 | 0 | 0 | 0 | Concentrating transporters (CNT): transport nucleoside analogues across a concentration gradient |
| SLC28A2 | 0 | 0 | 0 | 0 | |
| SLC28A3 | 0 | 0 | 0 | 0 | |
| SLC29A1 | 2.36 | 20.49 | 13.69 | 34.00 | Equilibrium nucleoside transporters (ENT): facilitate the transport of nucleoside analogues down the concentration gradient |
| SLC29A2 | 0.18 | 0.15 | 0.80 | 0.05 | |
| ABCC10 | 2.33 | 1.81 | 7.77 | 3.68 | ATP dependent exporters – can remove nucleosides from the cell |
| ABCC11 | 0.04 | 0.01 | 0.01 | 0.03 | |
| ABCC4 | 0 | 0 | 0 | 0 | |
| ABCC5 | 0.01 | 0.01 | 0.02 | 0 | |
| DCK | 31.82 | 28.84 | 20.43 | 35.73 | Coverts DAC into the active form |
| CDA | 0.02 | 13.75 | 9.60 | 1.26 | Catalyses the deamination of cytidine and deoxycytidine |

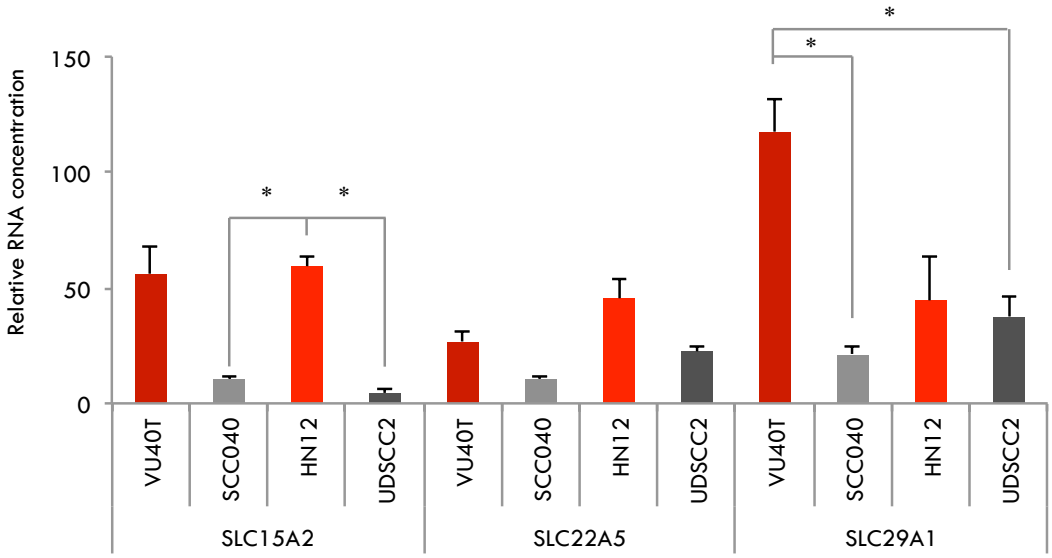
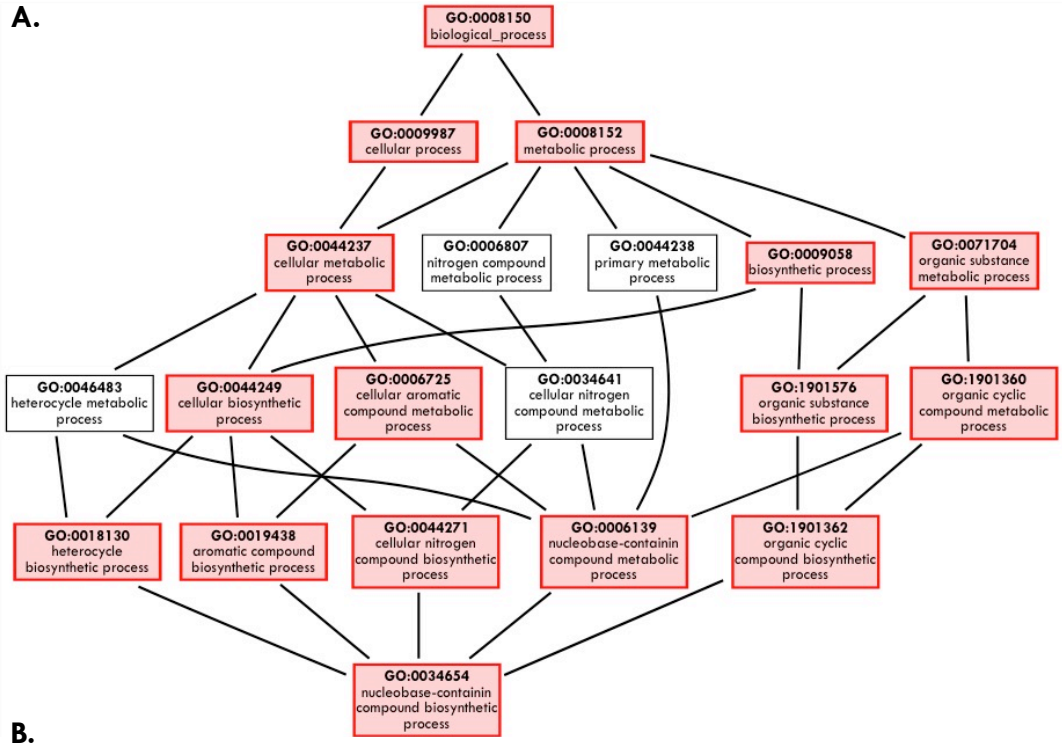



**Figure 3.7. DAC responsive cell lines exhibit increased expression of *SLC15A2, SLC22A5 and SLC29A1***

qRT-PCR analysis of drug transporters. Results are mean + SEM (n=3) and significance was determined using the SPSS software to calculate independent t-tests with Bonferroni correction where *p<0.05

**A.**




**B.**

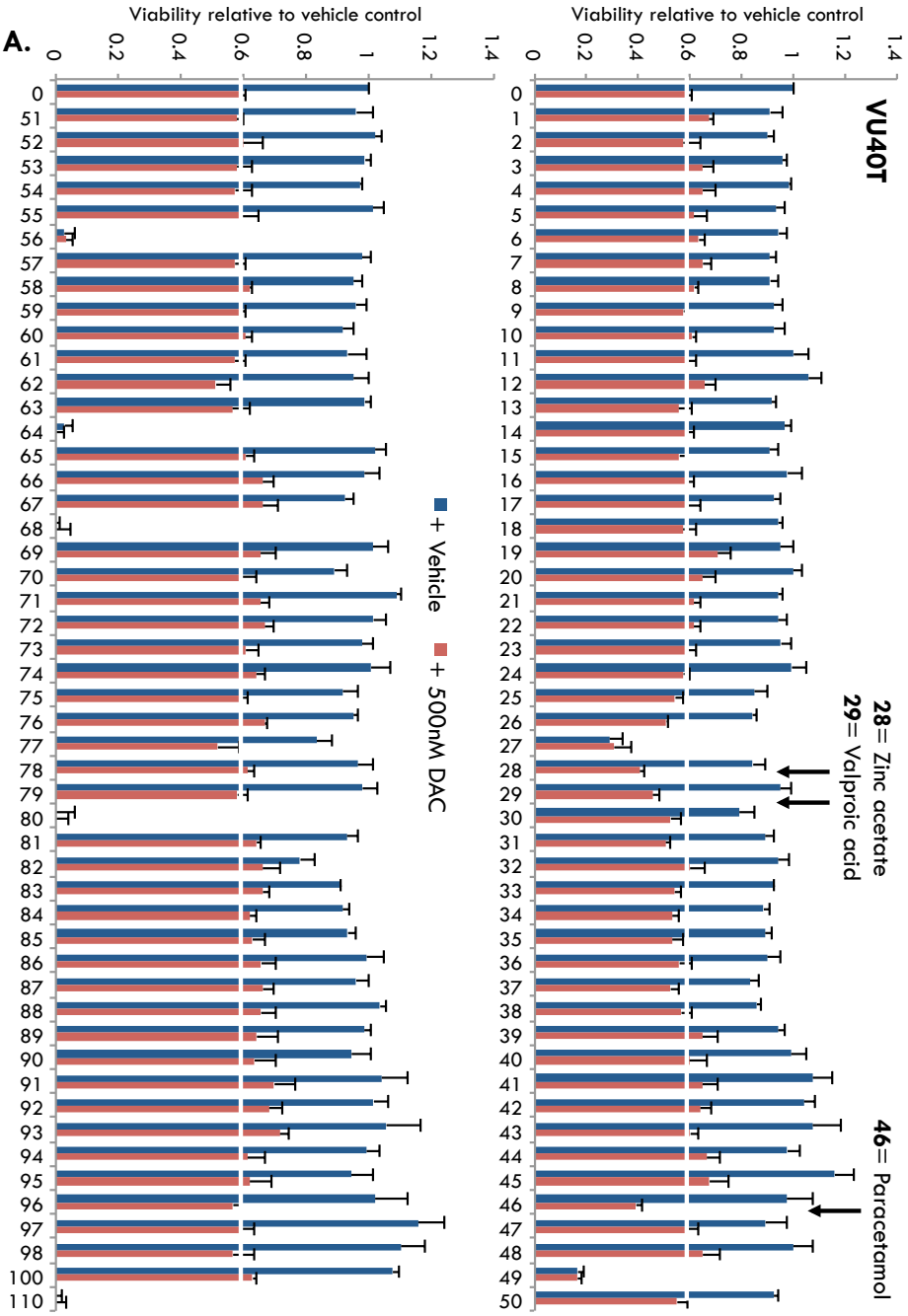
| Term ID | Description | log10 p-value |
|---|---|---|
| GO:1901362 | organic cyclic compound biosynthetic process | -5.9586 |
| GO:0034654 | nucleobase-containing compound biosynthetic process | -5.4202 |
| GO:0018130 | heterocycle biosynthetic process | -5.3197 |
| GO:0019438 | aromatic compound biosynthetic process | -5.251 |
| GO:0019219 | regulation of nucleobase-containing compound metabolic process | -3.7011 |
| GO:1901360 | organic cyclic compound metabolic process | -3.1308 |
| GO:0044271 | cellular nitrogen compound biosynthetic process | -2.9469 |
| GO:0044249 | cellular biosynthetic process | -2.5045 |
| GO:0045934 | negative regulation of nucleobase-containing compound metabolic process | -2.4962 |
| GO:0006725 | cellular aromatic compound metabolic process | -2.3233 |
| GO:0006139 | nucleobase-containing compound metabolic process | -2.1858 |
| GO:1901576 | organic substance biosynthetic process | -2.1524 |
| GO:0046483 | heterocycle metabolic process | -2.0783 |

**Figure 3.8. Enrichment of nucleobase-metabolism related genes in DAC-unresponsive cell lines**

DeSeq2 was used to determine genes differentially expressed between the DAC-responsive (HN12, VU40T) and unresponsive (SCC040, UDSCC2) cell lines. Differentially expressed genes were subject to gene ontology (GO) analysis. **A.** The gene ontology consortium relationship map for GO term 0045934 (negative regulation of nucleobase containing compound). Outlined in red are GO terms significantly enriched ($p<0.05$) in the DAC-unresponsive group. **B.** The log10 p-values for the GO terms outlined in **A.**

## 3.6. A drug repurposing screen identifies paracetamol, valproic acid and zinc acetate as being able to sensitise HNSCC cells to DAC treatment

Data has indicated that the HNSCC cell lines examined are highly variable in their response to DAC, however they all have $IC_{50}$ values higher than the $C_{max}$ of ~500nM. A DAC sensitising assay was developed to determine whether the efficacy of DAC treatment in HNSCC could be increased through combination therapy. A screen was performed on DAC-responsive cell line, VU40T, and unresponsive cell line, SCC040 using a panel of 100 off-patent drugs (Supplementary data 1). Cells were treated with one of the panel of drugs, +/- 500nM DAC for 96h, following which viability was determined (Figure 3.9). For the unresponsive cell line, none of the 100 drugs tested were able to increase the efficacy of DAC treatment (Figure 3.9). However, for the DAC-responsive cell line, zinc acetate, valproic acid and paracetamol were able to increase the efficacy of DAC (Figure 3.9). This pattern was replicated when the other DAC-responsive and unresponsive cell lines were examined: HN12 cells were sensitized by paracetamol and zinc acetate; while UDSCC2 cells were not (Figure 3.10). HOK cells were highly sensitive to both zinc acetate and valproic acid but not paracetamol, and paracetamol was unable to increase the sensitivity of HOK cells to DAC treatment (Figure 3.10). Therefore further work focused on the relationship between DAC and paracetamol in HNSCC.


A.
Viability relative to vehicle control
0
0.2
0.4
0.6
0.8
1
1.2
1.4
+ Vehicle
+ 500nM DAC
VU40T
28= Zinc acetate
29= Valproic acid
46= Paracetamol

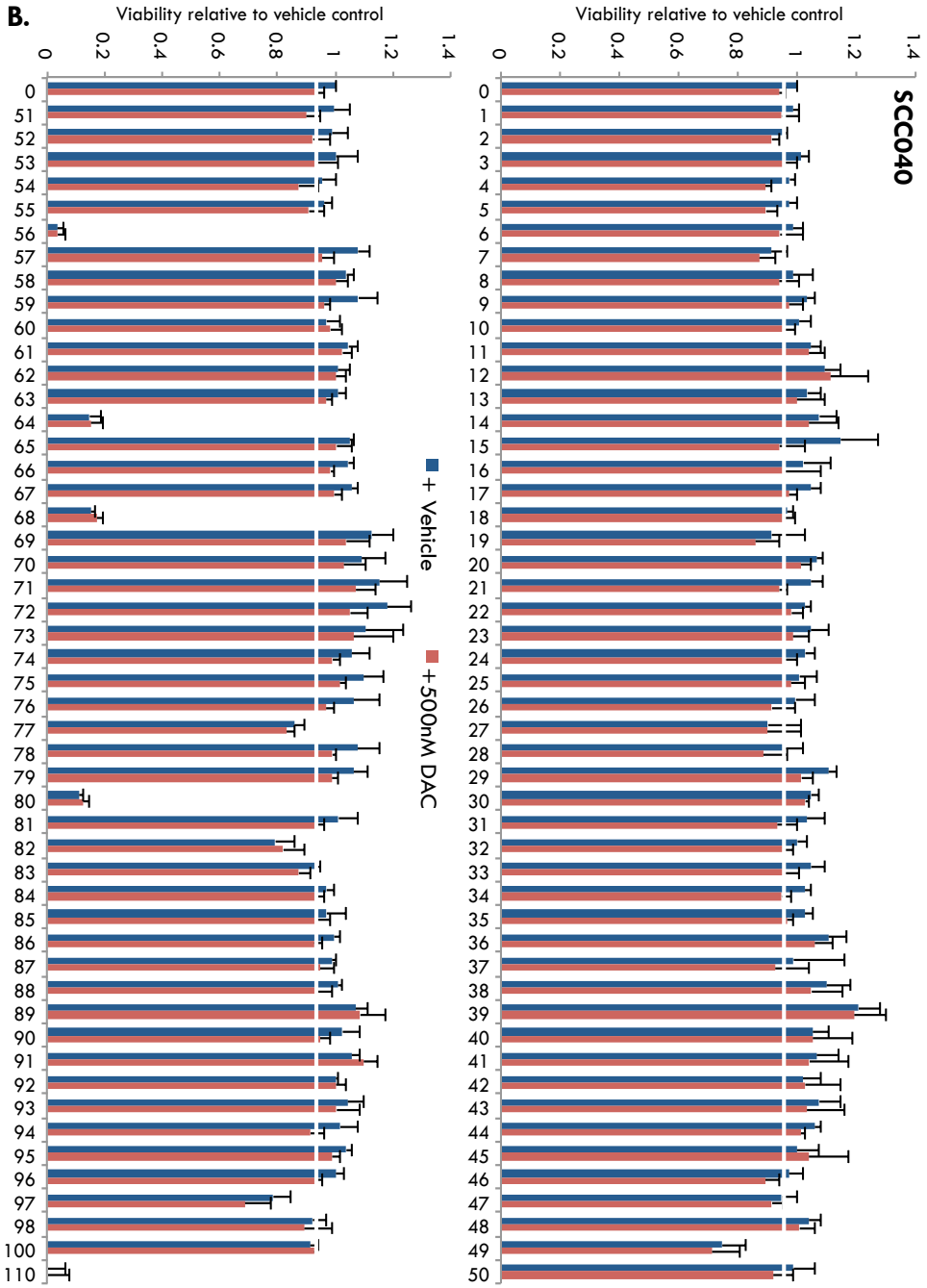
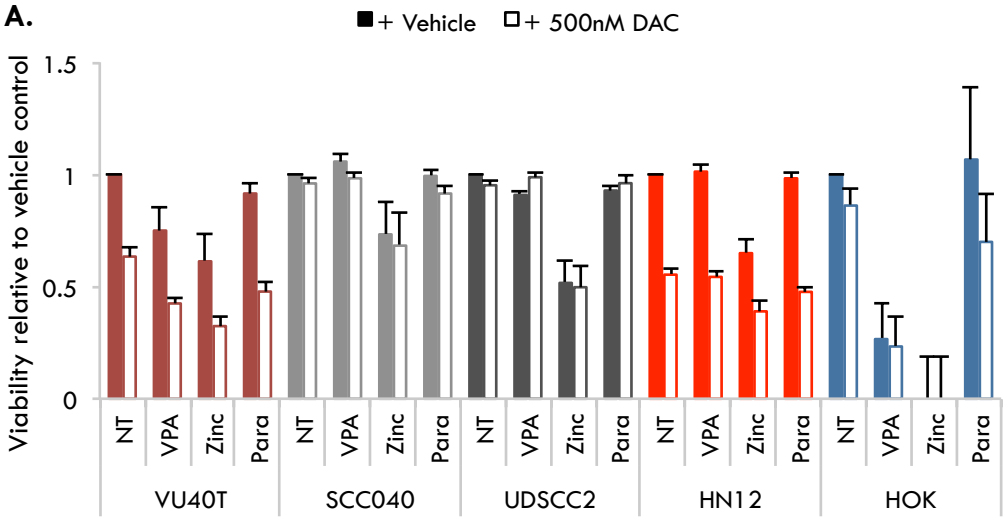



**Figure 3.9. A drug-repurposing screen identified paracetamol, zinc acetate and valproic acid as able to sensitise HNSCC cell lines to DAC treatment**

DAC responsive (**A.** VU40T) and unresponsive (**B.** SCC040) cell lines were treated with 500nM DAC +/- one of a panel of 100 drugs for 96h, following which viability was determined using the CellTiter Blue Cell Viability Assay (Section 2.2.2). Cells were determined as sensitising if the combined effect of the drug + DAC was greater than the effect of DAC alone (under the white line) and greater than the effect of the drug alone (red bar lower than grey bar). Results are shown relative to the vehicle only control (0) + SEM (n=3).

**A.**




**B.**

| Cell Line | Sensitivity to DAC Treatment (%) | Sensitising Drugs |
|---|---|---|
| **VU40T** | 36.63 (***) | Zinc Acetate (*) ,Valproic acid (*) ,Paracetamol(*) |
| **SCC040** | 4.021 (ns) | None |
| **HN12** | 44.63 (***) | Zinc Acetate(ns) , Paracetamol(**) |
| **UDSCC2** | 4.84 (ns) | None |
| **HOK** | 13.88 (ns) | None |

**Figure 3.10. Paracetamol, zinc acetate and valproic acid are able to sensitise HNSCC cell lines to DAC treatment**

**A.**HNSCC cell lines and primary oral keratinocyte cells were treated with drugs at the Cmax concentration either alone (filled bars) or in combination with 500nM DAC (empty bars) for 96h. Viability was determined and results are shown relative to the vehicle only control. Results are mean +/- SEM (n=3). NT= vehicle only; Zinc= zinc acetate; VPA= valproic acid; Para= paracetamol. **B.** The results from **A.** are displayed as a table. Significance was determined via paired t-test against the untreated control (sensitivity to DAC column) or the DAC only sample (sensitising drugs column). *** $p<0.001$, **$p<0.01$, *$p<0.05$.

## 3.7. DAC and paracetamol work in synergy to reduce viability in HNSCC cells

To determine whether DAC and paracetamol were working in synergy to reduce viability in HNSCC cells the Chou-Talalay method was employed [107]. The main benefit of this method is its simplicity [107]. Experimentally, in place of interrogating all possible dose combinations between the two drugs, only matched titrations of each drug alone and in combination are required. The results of the combined titration are treated equivalent to a third drug, thereby simplifying the experimental design and mathematical analysis [107]. DAC-responsive cells, HN12 and VU40T, were subject to matched titrations of DAC, paracetamol, or DAC + paracetamol at multiples of the Cmax (500nM for DAC, 132.3µM for paracetamol) for 96h, following which viability was determined (Figure 3.11). The CompuSyn software was used to determine synergy. Synergy is defined as when the combined effect of two drugs is greater than the sum of their individual effects, and it is measured by combination index (CI) values, where a CI value less than 1 denotes synergy. For HN12 cells the CI value was below 1 at all concentrations tested (Figure 3.12). For VU40T cells, at all bar the lowest concentrations, the CI values were below 1 (Figure 3.12). Therefore DAC and paracetamol are working in synergy in DAC-responsive HNSCC cells.

The CompuSyn software also calculates the dose reduction index (DRI). DRI values determine how many times less each drug can be used when used in combination (Table 3.2). For example, 2.26µM of DAC or 595µM of paracetamol is needed to achieve a 50% reduction in viability in VU40T cells, if used alone. These doses are considerably higher than those used in the clinic. However, almost five times less of

each drug can be used when combined, so toxic concentrations of each drug can be replaced with therapeutically relevant ones (Table 3.2). A similar result can be observed for HN12 cells. HN12 cells exhibit minimal response to paracetamol alone; this is mirrored by very high DRI values for paracetamol (Table 3.2, Figure 3.11).

To confirm the efficacy of the viability assay and to determine whether any morphological changes are observed following DAC and paracetamol treatment Giemsa/Jenner staining was performed on DAC-responsive VU40T cells. The cells were untreated (NT) or treated with 500nM DAC +/- 132.3μM paracetamol for 96h. Cells were collected and stained every 24h. Minimal change in cell morphology or density was observed after 48h of treatment. However, at 72h and 96h, DAC treated cells showed a reduction in cell number, in accordance with the viability results (Figure 3.4, Figure 3.13). After 96h, this decrease is amplified by the addition of paracetamol, although paracetamol alone caused no apparent change in cell density (Figure 3.13). The morphology of the cells was altered by DAC; the cells appeared larger, less uniform and less cobblestone-like. However, paracetamol did not visibly affect cell morphology (Figure 3.13).

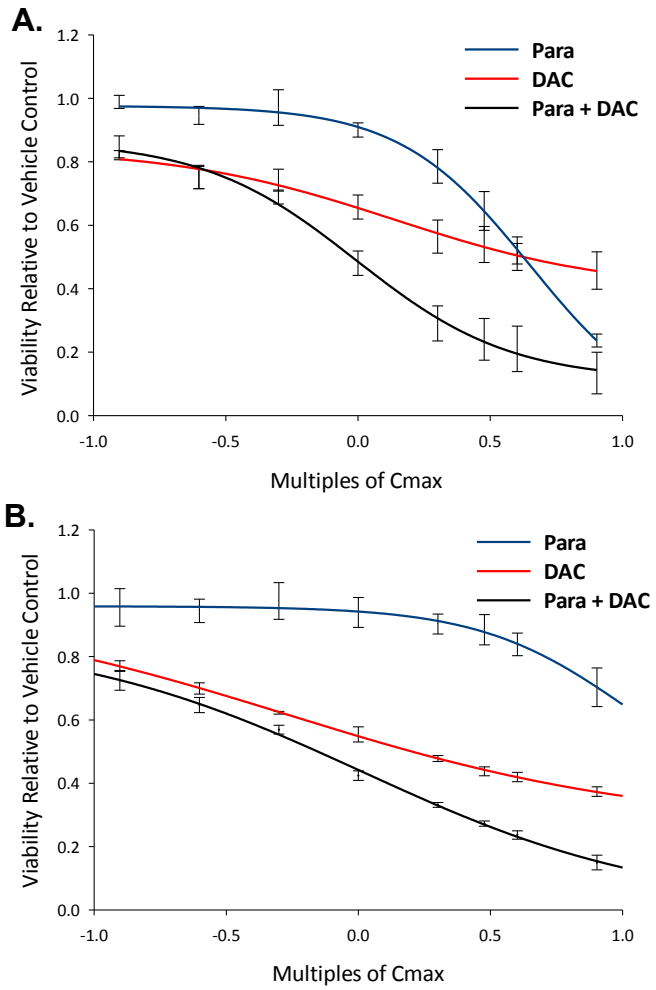



**Figure 3.11. DAC and paracetamol treatment in DAC-responsive cell lines**

**A.** VU40T cells were treated with matched titrations (0.125X, 0.25X, 0.5X, 1X, 2X, 4X and 8X) of DAC (at 500nM), paracetamol (at 132.3μM), or both for 96h, following which viability was determined. **B.** The same study was performed on HN12 cells. Mean +/- SEM (n=3).

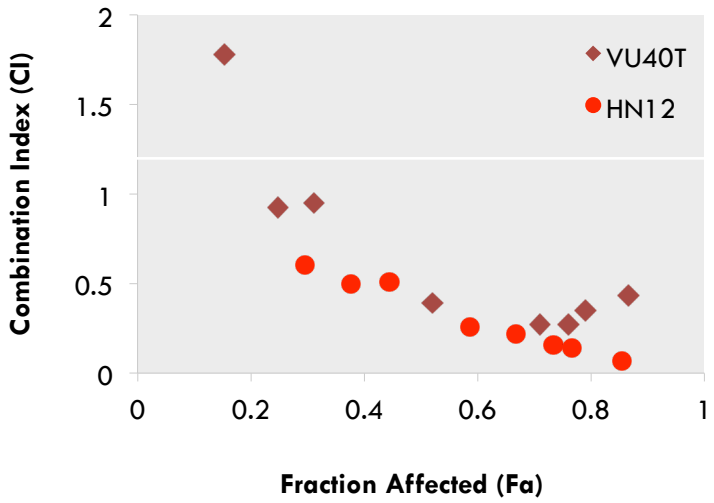



**Figure 3.12. DAC and paracetamol work in synergy to reduce viability in HNSCC cells**

The Chou-Talalay method was used to determine synergy from the results in Figure 3.11. Combination Index (CI) values are shown relative to the fraction of cells affected by treatment (Fa). Synergy is defined by a CI value less than 1, or below the white line.

| Fa (%) | VU40T | | | | HN12 | | | |
|---|---|---|---|---|---|---|---|---|
| | Dose (µM) | | DRI | | Dose (µM) | | DRI | |
| | DAC | Para | DAC | Para | DAC | Para | DAC | Para |
| **10** | 0.01 | 105 | 0.30 | 10.18 | 0.004 | 186 | 0.49 | 80 |
| **25** | 0.16 | 251 | 1.22 | 7.11 | 0.06 | 1962 | 1.19 | 147 |
| **50** | 2.26 | 595 | 4.99 | 4.97 | 0.85 | 20653 | 2.92 | 269 |
| **75** | 31 | 1414 | 20.35 | 3.47 | 11.92 | 217412 | 7.15 | 493 |
| **97** | 9300 | 9200 | 426 | 1.59 | 3630 | 35450000 | 49.5 | 1827 |

**Table 3.2. DAC and paracetamol work in synergy to reduce viability in HNSCC**

The CompuSyn software calculated dose reduction index (DRI) values for each drug using the titrations described in Figure 3.11. The table shows the concentrations of each drug required to affect the corresponding percentage of cells (Fa) when used individually and the number of times this dose can be reduced if used in combination (DRI).

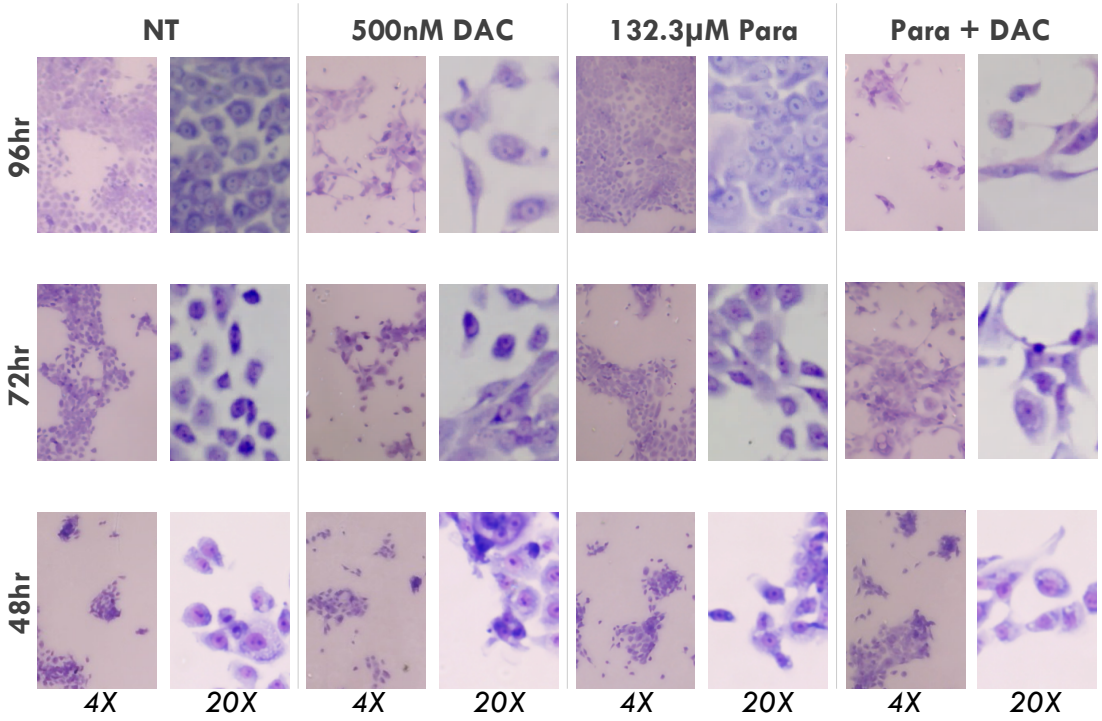



**Figure 3.13. Morphology of DAC and Paracetamol treated VU40T cells**

VU40T cells were treated with DAC +/- paracetamol for 96h. Cells were harvested at 48h, 72h and 96h, fixed in methanol and Giemsa/Jenner stained to visualise changes in morphology and density after each treatment. Representative images from 3 experiments. Microscope magnification is shown below images in italics.

## 3.8. Combined treatment with DAC and paracetamol slows down proliferation in VU40T cells

Reduced viability can be caused due to several mechanisms, including; the rate of apoptosis and necrosis could be increased, or the rate of proliferation decreased. Annexin V and propidium iodine staining and FACS analysis was used to determine whether VU40T cells treated with DAC, paracetamol or both were undergoing apoptosis or necrosis (Figure 3.14). In this assay, annexin V is used as a marker of apoptosis, while cells will only stain with PI if the cell membrane is damaged, a feature of necrotic or late apoptotic cells. Therefore, live cells can be distinguished from necrotic cells which are positive for PI; apoptotic cells which are positive for annexin V and late apoptotic cells which are positive for both. A slight increase in apoptosis and necrosis could be observed from as early as 48h of DAC treatment, this effect was increased after 72h and 96h (Figure 3.14). Paracetamol alone had no effect on cell death and the addition of paracetamol to DAC treatment did not significantly increase apoptosis or necrosis (Figure 3.14).

To determine the relative rate of proliferation, VU40T cells were treated with DAC, paracetamol or both for 96h and immunostained for Ki67 (Figure 3.15). Proliferating cells show up as positively stained nuclei. The proportion of nuclei positive for Ki67 was calculated for each treatment (Figure 3.15). Treatment with DAC significantly decreased the proliferation levels after 24h; this was further decreased by the addition of paracetamol suggesting that the efficacy of combined treatment was due to a reduction in the rate of proliferation (Figure 3.15).

Reduced proliferation may indicate that the cell cycle has been stalled. Therefore propidium iodide staining and FACS analysis were used to determine whether VU40T cells treated with DAC and paracetamol were progressing through the cell cycle at the same rate as untreated cells (Figure 3.16). In this assay, cells were fixed in ethanol to cause cell cycle arrest, PI was then intercalated into the DNA, and FACS was used to separate the cells based on PI level, which is proportional to the total amount of DNA. This approach can be used to distinguish the stages of the cell cycle, as cells in G2/M have double the amount of DNA as those in the G0/G1 stage. Both paracetamol and combined treatment cause a decrease in the amount of cells in the G2/M stage of the cell cycle and a corresponding increase in G0/G1 suggesting that the cells are stalled at the G1 or S phase checkpoints (Figure 3.16). However, this result was not significant (Figure 3.16). DAC and combined treatment samples have an additional peak (sub G1) prior to G1/G0 that contains apoptotic cells, in agreement with the result previously described (Figure 3.14).

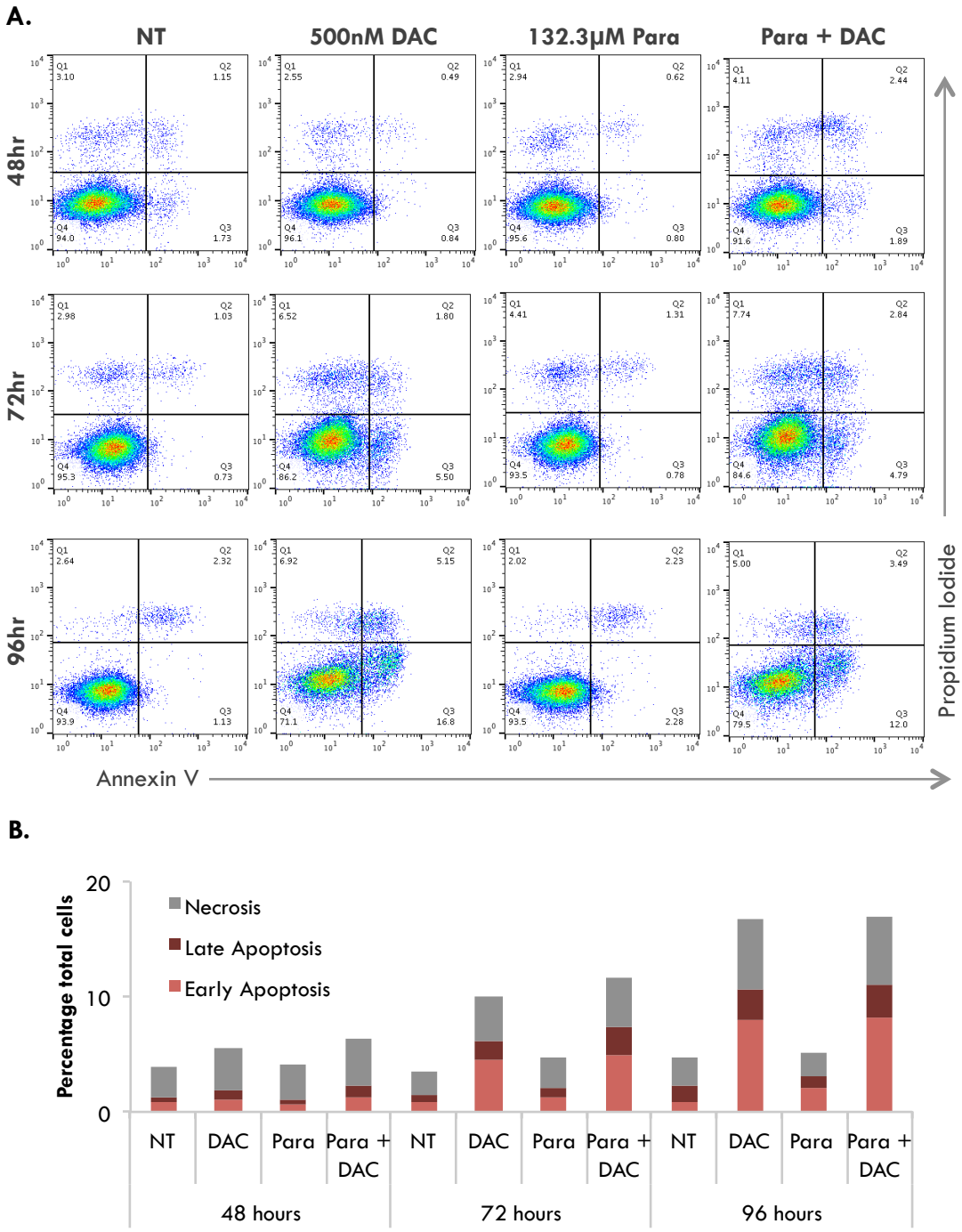



**Figure 3.14. Paracetamol does not increase cell death caused by DAC treatment**

VU40T cells were treated with DAC +/- paracetamol for 48, 72 or 96h and stained with annexin V and propidium iodide. Cells positive for each label were then counted using FACS. **A.** Representative plots show the cell populations at each time point. **B.** The mean proportion of cells from each quadrant in **A.** were calculated + SEM (n=3).

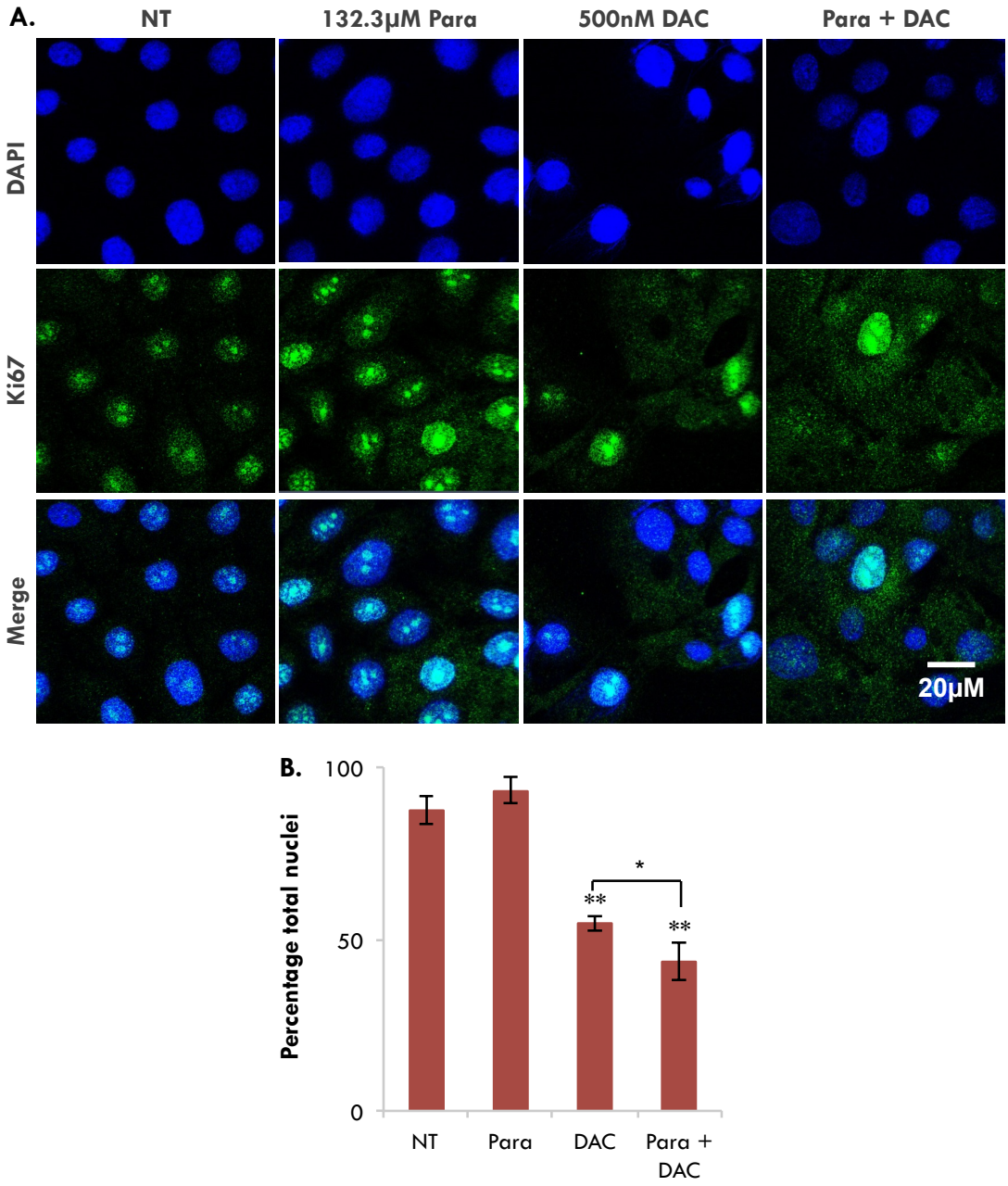



**Figure 3.15. Combined treatment with DAC and paracetamol reduces proliferation**

VU40T cells were treated with DAC +/- paracetamol for 96h, fixed and immunostained with Ki67. Cells were imaged on a Zeiss LSM 780 confocal microscope. **A.** Representative images of 3 experiments. **B.** ImageJ software was used to calculate the number of positively stained nuclei from **A.** Results are the mean + SEM (n=3). An average of 50 nuclei were counted per treatment per replicate. SPSS software was used to calculate a paired t-test for each treatment against the untreated samples, and the combined treatment against DAC alone. *p<0.05, ** p<0.01.

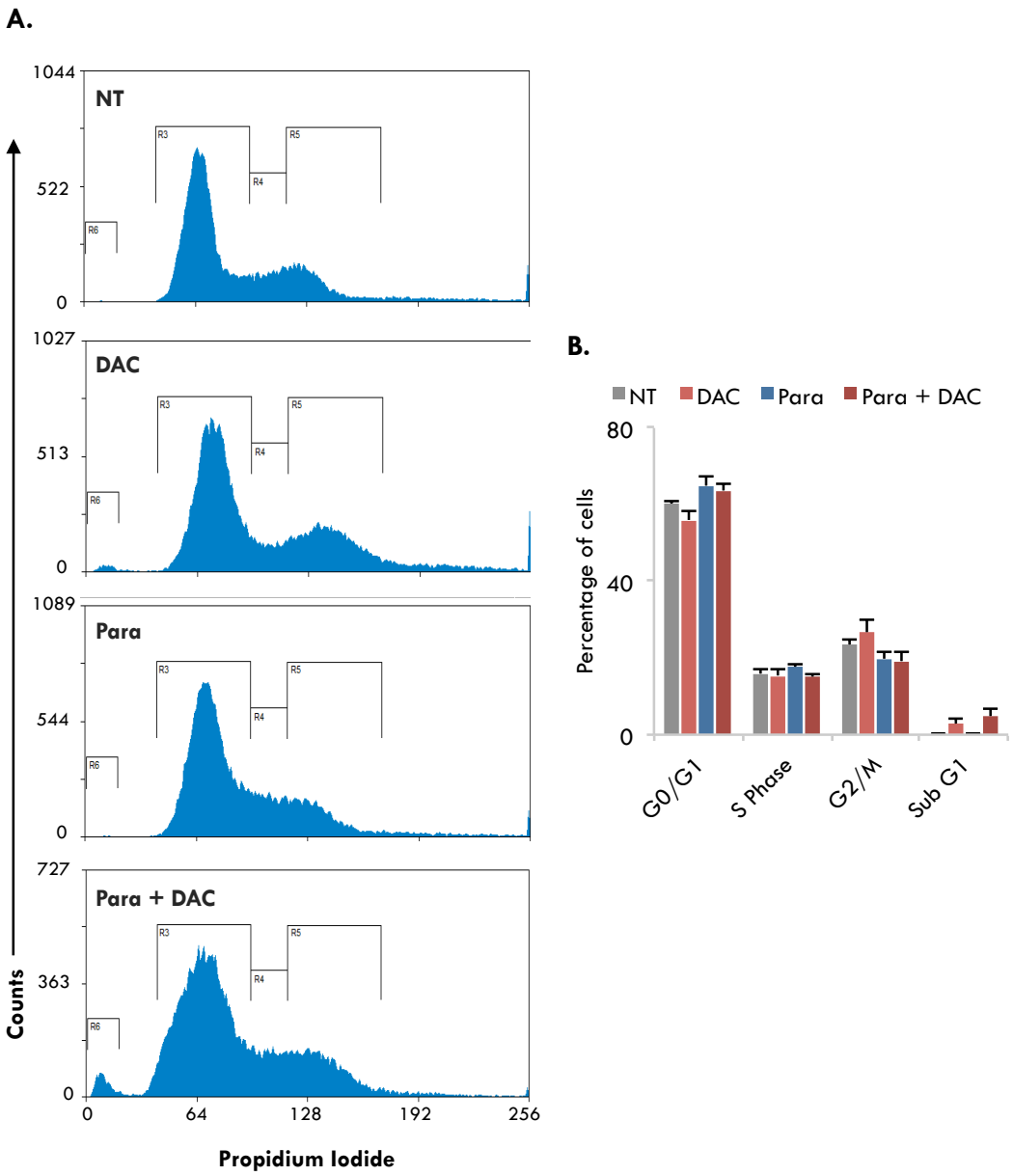



**Figure 3.16. Cell cycle changes after treatment with DAC and paracetamol**

VU40T cells were treated with DAC +/- paracetamol for 96h, fixed in ethanol and stained with PI prior to sorting by FACS. **A.** Representative histograms showing the levels of PI vs count. Peaks define populations of cells in each stage of the cell cycle and are gated as such. The peak at 64 represents G1/G0; the peak at 128 represents G2/M and in between is S phase. **B.** The average results from **A.** are shown as a bar graph (n=5 + SEM).

## 3.9. Combined treatment with DAC and paracetamol alters the transcriptional profile of DAC-responsive cells

RNA sequencing was performed on VU40T cells after 96h treatment with 500nM DAC +/- 132.3µM paracetamol to investigate the effect of treatment on the transcriptome (Figure 3.17-Figure 3.19). The majority (62.7%) of expressed genes were consistent between the treatment groups (Figure 3.17). DAC treatment induced the expression of a moderate amount of genes relative to control (11%) while fewer genes were activated by paracetamol (6.2%) and only 4.8% of genes were specific to the combined treatment group (Figure 3.17). These data suggest that any transcriptional changes in response to treatment are primarily due to DAC. Clustering of the most differentially expressed genes confirms this observation: untreated and paracetamol treated cells exhibit a similar expression profile which is distinct from DAC and DAC + Para treated cells (Figure 3.17).

Treated samples were compared against the vehicle-treated control using the DeSeq2 software (Figure 3.18). The results were divided into 6 groups: i) genes up-regulated by paracetamol treatment (n=151); ii) genes down-regulated by paracetamol treatment (n=127); iii) genes up-regulated by DAC treatment (n=658); iv) genes down-regulated by DAC treatment (n=293); v) genes up-regulated by combined treatment (n=915) and; vi) genes downregulated by combined treatment (n=721). Genes from each group were subject to GO analysis to determine whether they were enriched for any biological processes (Supplementary data 7). No GO terms were found to be significantly associated with the paracetamol down-regulated gene dataset. The majority of ontological groupings associated with

increased expression after paracetamol treatment associated with the production of ATP (Figure 3.18, Figure 3.19). Genes involved in the cell cycle, DNA replication and cellular metabolic processes were decreased by treatment with DAC, while anti-viral and immune related genes were increased (Figure 3.18, Figure 3.19). Genes down-regulated by combined treatment were enriched for similar terms down-regulated by DAC alone, however they had substantially lower p-values, suggesting that the addition of paracetamol to the treatment regime amplifies the reduced expression of cell cycle and DNA replication related genes (Figure 3.18, Figure 3.19).

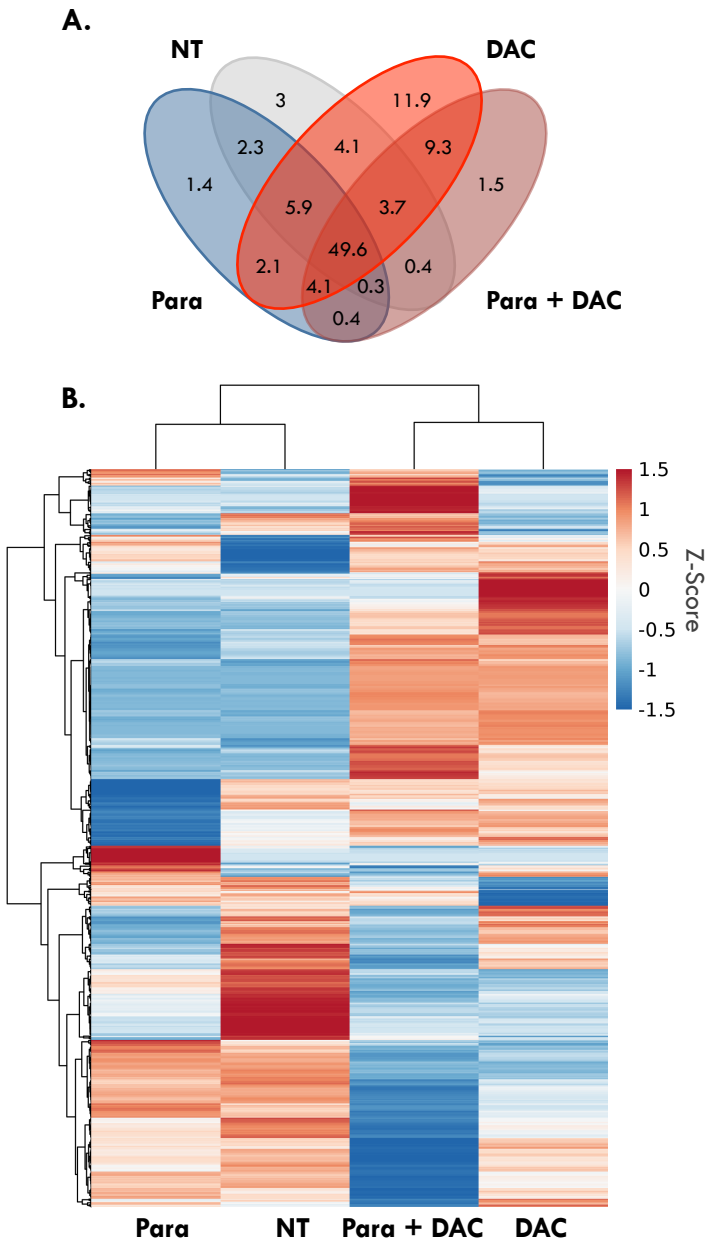



**Figure 3.17. Treatment with DAC and paracetamol alters the transcriptional profile of VU40T cells**

RNA was extracted and Illumina sequencing was performed in VU40T cell cultures treated with DAC +/- paracetamol for 96h. Sequences were aligned to the genome using HiSAT2. **A**. StringTie was used to determine which genes were expressed, and the top 25% for each treatment group are shown as a Venn diagram. Number represent percentage of total genes. **B**. All differentially expressed genes (STD>1) are shown as a heatmap generated using ClustVis software where both samples (columns) and genes (rows) are clustered using correlation distance and average linkage and unit variance scaling is applied..

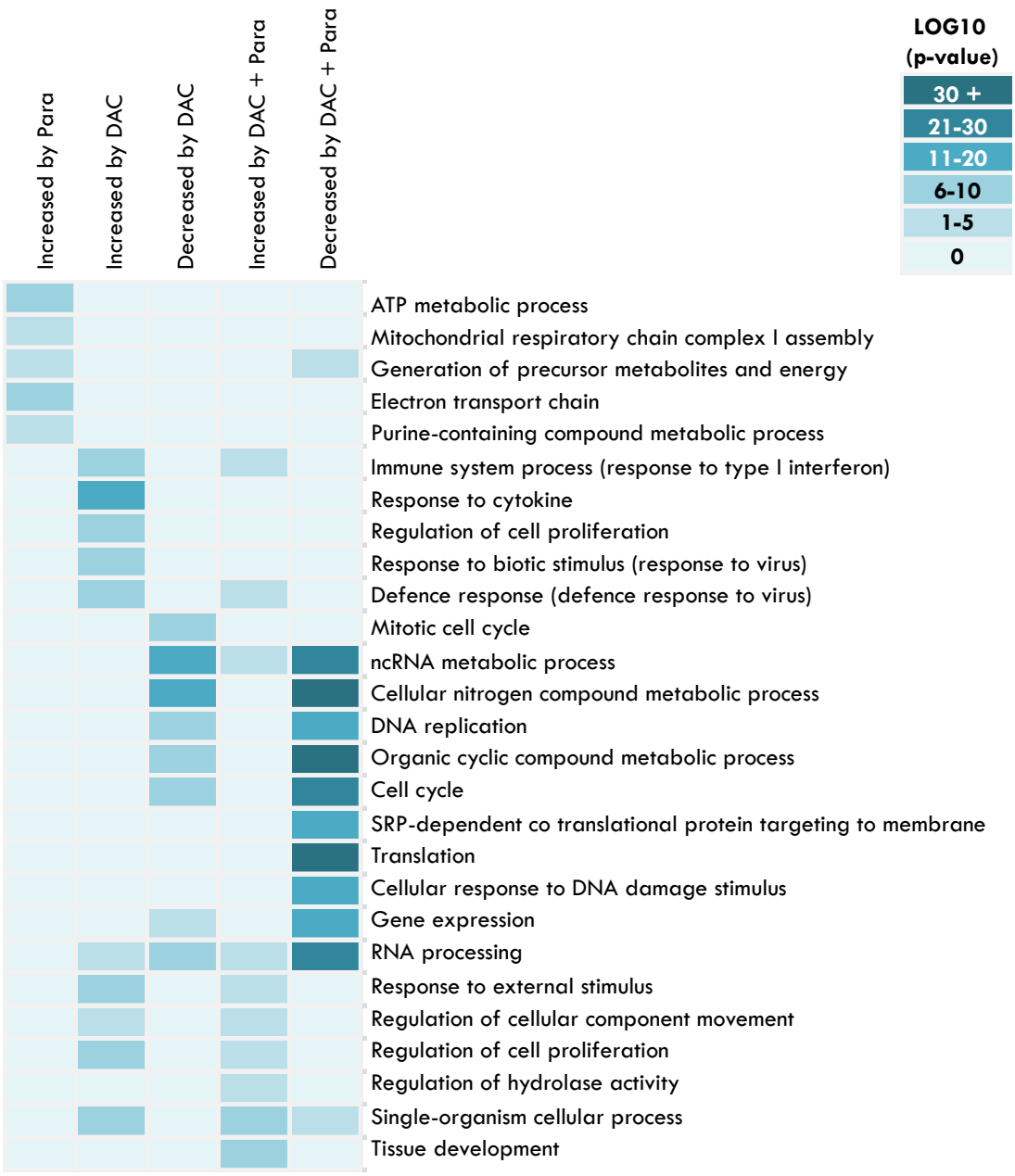



**Figure 3.18. Gene ontology analysis of DAC and paracetamol treated VU40T cells**

RNA was extracted and Illumina sequenced from VU40T cells treated with DAC +/- paracetamol for 96h. Sequences were aligned to the genome using HiSAT2. Differentially expressed genes for each treatment group were determined using DeSeq2. Those with a log2 fold change above 1 were subject to gene ontology (GO) analysis followed by removal of redundant terms using Revigo. The figure shows LOG10 (p-value) for selected GO terms.

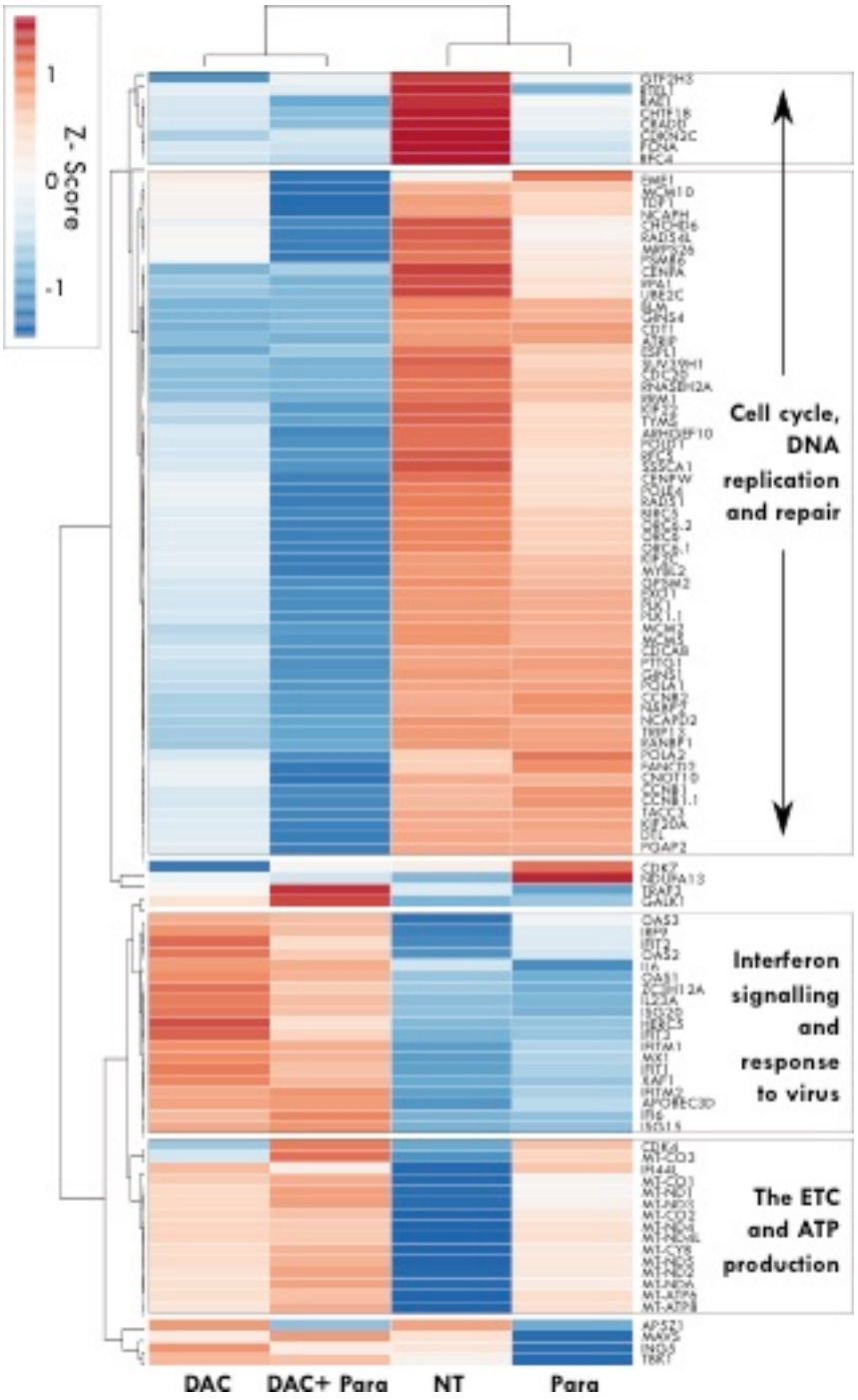



**Figure 3.19. Combined treatment with DAC and paracetamol decreases the expression of genes involved in the cell cycle, DNA replication and repair**

Genes from chosen ontological groupings (Figure 3.14) are displayed as a heatmap using the ClustVis software. Both rows and columns are clustered using correlation distance and average linkage. Unit variance scaling was applied.

## 3.10. Discussion and conclusion

Altered promoter methylation is understood to contribute to the initiation and development of HNSCC; therefore, the disease is a good candidate for DAC treatment [29, 100, 101]. However, this study has shown considerable variability in the response of HNSCC cell lines to DAC, with the cell lines stratifying into two distinct groups: those that did not respond to DAC and those that responded to nanomolar concentrations. The DNA methylation change in response to DAC in each cell line mirrors these results, suggesting that efficacy of the drug is directly proportional to the ability of DAC to demethylate the DNA. This would suggest that sensitivity is dependent upon the activation, incorporation or retention of the drug. Accordingly, the RNA expression of drug transport genes *SLC15A2, SLC22A5* and *SLC29A1* are higher in DAC-responsive cell lines than unresponsive cell lines. Follow up experiments are needed to confirm whether this could be contributing to the variable sensitivity observed. Interestingly, genes involved in the regulation of nucleobase-containing compound metabolism were enriched in the gene set associated with DAC-unresponsive cell lines suggesting that there may be a difference in the nucleotide turnover in DAC-unresponsive cells compared with responsive ones. This could contribute to the variable sensitivity observed as high dNTP pools can contribute to DAC resistance; therefore this is another area of future research [184].

To improve upon the response of HNSCC cell lines to DAC treatment a DAC sensitising assay was designed. The screen had no initial bias towards chemotherapeutic or epigenetic drugs and instead aimed to repurpose commonly

used, cost-effective and off-patent drugs for the treatment of HNSCC [155]. None of the 100 drugs tested were able to sensitise unresponsive cells to DAC treatment. This is in agreement with the theory previously proposed, and suggests that targeting the activation or uptake of DAC is the primary means of conferring sensitivity to unresponsive cases. In the DAC-responsive cell lines, valproic acid (VPA), zinc acetate and paracetamol were all able to increase sensitivity to DAC. The potential benefits of combining VPA with DAC have been described elsewhere, confirming the results of the sensitising assay employed here [190]. However, the relationship between DAC and paracetamol was chosen to be the focus of this work because normal human oral keratinocytes were highly sensitive to both zinc acetate and VPA treatment, while paracetamol had no effect on their viability.

Recent work suggests that, especially at low doses, DAC acts by inducing the expression of double stranded endogenous retroviruses that trigger an immune response [158, 180]. The RNA sequencing results described in this chapter are in keeping with this finding as DAC treatment leads to an upregulation of genes associated with the immune system and viral response. The addition of paracetamol to this treatment did not increase the expression of these genes; instead many genes exhibited lower expression levels. Therefore it does not appear likely that combined treatment is harnessing this mechanism. Instead the expression of genes downregulated by DAC is further dysregulated by the addition of paracetamol. These include genes involved in metabolism, DNA replication and the cell cycle. Accordingly cells treated with DAC and paracetamol show altered cell cycle profiles and reduced proliferation.

In conclusion, this chapter has outlined the potential use of DAC in the treatment of HNSCC and identified combination with paracetamol as a potential therapy. The two drugs were shown to work in synergy to reduce viability in DAC-responsive HNSCC cells, allowing toxic concentrations of each drug to be replaced by clinically relevant ones. While DAC alone can induce cell death, the addition of paracetamol significantly slows down the rate of proliferation. Therefore, this could have great therapeutic potential, as combined treatment would protect against cancer progression by slowing tumour growth while actively inducing apoptosis in the existing cancer cells.

# Chapter 4. Mechanisms underlying the synergistic relationship between DAC and paracetamol

## 4.1. Paracetamol

Paracetamol (or acetaminophen in the USA) is a popular analgesic and antipyretic with a similar spectrum of action to non-steroidal anti-inflammatory drugs (NSAIDs) [191]. It was first synthesised by Morse in 1878 during a search for derivatives of acetanilide. Acetanilide had been in use at the time as an antipyretic and analgesic, however it was found to have some serious side effects including severe methemoglobinemia, prompting a search for less toxic derivatives [192]. The two main outcomes of this search were paracetamol and phenacetin. Phenacetin was favoured and paracetamol set-aside for several decades [192]. In 1948, Brodie and Axelrod determined that the active metabolite in both phenacetin and acetanilide was paracetamol [193]. Furthermore the methemoglobinemia was caused by a separate metabolite, phenylhydroxylamine. This lead to marketing and widespread distribution of paracetamol; it became available in the USA in 1955 and in the UK the following year [192].

## 4.2. The mechanism of action of paracetamol

Interestingly, despite widespread usage and extensive history, the exact mechanism of action of paracetamol is still under debate. Upon searching the literature three main theories emerge to explain the analgesic effects of paracetamol: Inhibition of the cyclooxygenase (COX) pathway through COX-2; specific inhibition of COX-3; or conversion to the analgesic N-arachidonylphenolamine (AM404).

### 4.2.1. Inhibition of the cyclooxygenase pathway

The spectrum of action of paracetamol is similar to that of non steroidal anti-inflammatory drugs (NSAIDs), and as such it was initially classified as one [191]. NSAIDs act on the cyclooxygenase (COX) pathway, by inhibiting COX enzymes to prevent the conversion of arachidonic acid into prostaglandin $H_2$ ($PGH_2$). In the absence of COX inhibitors $PGH_2$ is rapidly converted into one of a number of effector prostaglandins (PG) or thromboxane (TXA) that exert their effects through G-protein coupled receptors (Figure 4.1). Two COX enzymes catalyse the conversion of arachidonic acid into $PGH_2$: COX-1 which is constitutively expressed; and COX-2 which is induced by a variety of growth factors [194]. While both COX enzymes have the ability to convert arachidonic acid into $PGH_2$, there is some preference for which downstream prostaglandins are produced, with $PGE_2$ and $PGI_2$ mainly derived from COX-2 [195, 196]. Paracetamol is thought to specifically inhibit COX-2 [191].

The efficacy of paracetamol can be inhibited by increasing the peroxide tone of the cell, either directly by the addition of hydroperoxides or indirectly, by increasing the concentration of arachidonic acid [197, 198]. When the levels of arachidonic acid

are low, prostaglandin production predominantly occurs via COX-2, and $PGH_2$ is converted into $PGE_2$ by membrane associated PGE synthases [191]. This is termed as delayed response. An immediate response takes effect when arachidonic acid levels are high, which utilises COX-1 and cytosolic PGE synthases [199]. High levels of COX enzymes or arachidonic acid can inhibit paracetamol; therefore it appears to be most potent on the delayed response pathway, perhaps explaining the preference of paracetamol for COX-2 inhibition [197].

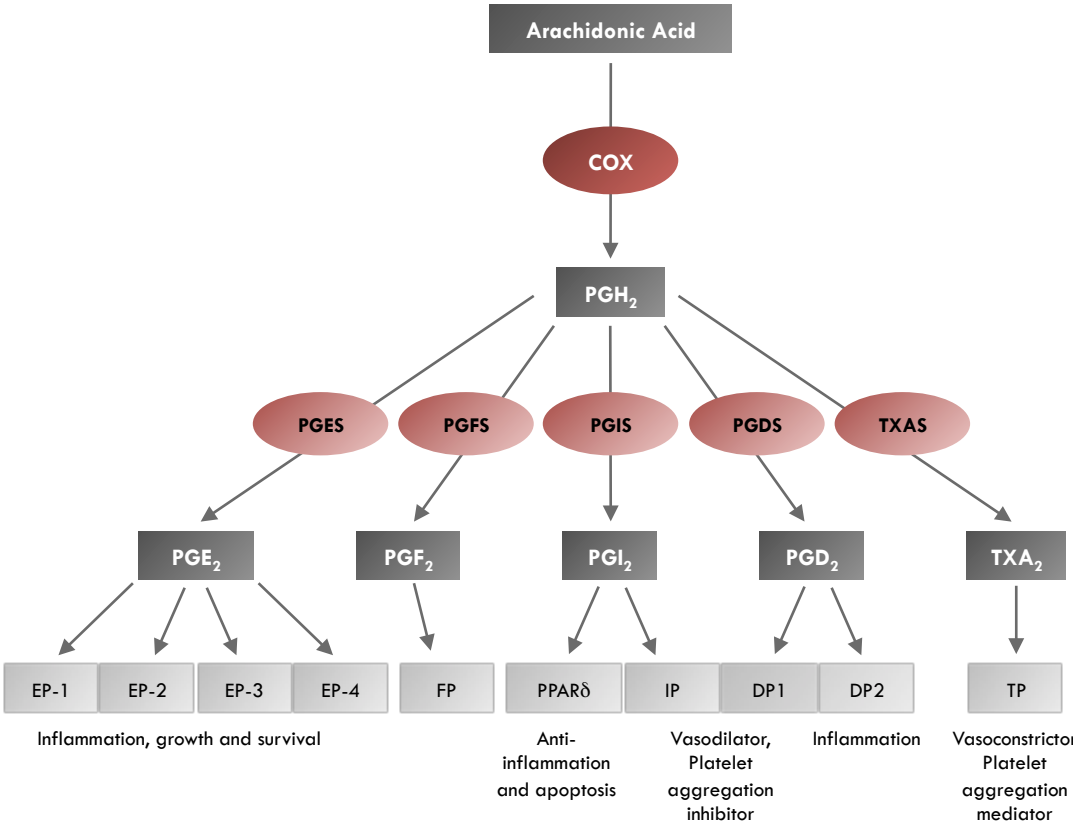



**Figure 4.1. The cyclooxygenase pathway**

Cyclooxygenase (COX) proteins (PTGS-1/2 genes) convert arachidonic acid into prostaglandin $H_2$ ($PGH_2$). This is then rapidly converted into one of a number of effector prostaglandins (PG-$E_2$, $F_2$, $I_2$, and $D_2$) or thromboxane (dark grey boxes). These exert their effects through G-protein coupled receptors (light grey boxes). EP-1/2/3/4: prostaglandin $E_2$ receptors 1/2/3/4 (PTGER-1/2/3/4 genes); FP: Prostaglandin $F_2$ receptor (PTGFR gene); PPARδ: peroxisome-proliferator-activated receptor δ; IP: prostaglandin $I_2$ receptor (PTGIR gene); DP-1/2 prostaglandin $D_2$ receptor-1/2 (PTGDR1/2 genes); TP: thromboxane receptor (TBXA2R gene). PGES= prostaglandin $E_2$ synthase (PTGES gene); PGFS: prostaglandin $F_2$ synthase (FAM213B gene); PGIS: prostaglandin $I_2$ synthase (PTGIS gene); PGDS: prostaglandin $D_2$ synthase (PTGDS gene); TXAS: thromboxane synthase (TBXAS1 gene). Figure adapted from Cebola et al., 2012 [194].

### 4.2.2. Inhibition of putative cyclooxygenase 3

In 2002, Simmons and colleagues discovered a splice variant of COX-1 (encoded for by the *PTGS1* gene), where the first intron is retained in the canine cerebral cortex [200]. They termed it COX-3, determined that it possessed cyclooxygenase activity and that some analgesic drugs including paracetamol, and the NSAID, ibuprofen preferentially inhibited COX-3 over COX-1 or COX-2 [200].

The first intron of the canine *PTGS1* gene is 90 nucleotides long, therefore the resulting mRNA remains in frame and a functional COX protein would result [200]. In humans and rodents the inclusion of intron 1 would lead to a frame shift, causing premature termination and a truncated, inactive protein [201]. Intron 1-retained *PTGS1* transcripts have been detected in a variety of human tissues. This primarily results in truncated proteins, however single base deletions can prevent premature termination from occurring and result in active proteins [202]. When cloned into insect cells these full length COX-3 proteins have the ability to convert arachidonic acid into $PGF_2$, albeit at a lower efficiency than COX-1. However the truncated form of human COX-3 is by far the most frequent and while human COX-3 was able to be inhibited by paracetamol, this was equal to the inhibition of COX-1 [202].

### 4.2.3. Conversion of Paracetamol to N-arachidonoylphenolamine (AM404)

Zygmunt Hogestatt and colleagues (2005), noted that there was a structural similarity between paracetamol and the metabolite N-arachidonoylphenolamine (AM404) [203]. AM404 is one of a group of bioactive N-acylamines that can produce analgesia in various animal tests [192]. The group found that, following

deacetylation to its primary amine, p-aminophenol, paracetamol is conjugated with arachidonic acid to form AM404 [203] (Figure 4.2). Oral administration of paracetamol and microinjection of AM404 produced comparable antinociception. The conversion of paracetamol to AM404 relies on the fatty acid amide hydrolase enzyme (FAAH) as no AM404 is produced in *FAAH*$^{-/-}$ mice [203].

The analgesic effects of AM404 can be attributed to two pathways: AM404 can inhibit the activity of COX-1 and COX-2 and reduce $PGE_2$ levels; and AM404 can indirectly activate the cannabinoid receptor $CB_1$ [192, 204]. Common to both is that AM404 activates TRPV1 [192, 203, 205]. TRPV1 is a receptor involved in the response to changes in body temperature, and can provide nociception. TRPV1 agonists and antagonists can be used to achieve analgesia and while paracetamol lacks TRPV1 agonist/antagonist activity, AM404 is a potent activator [203]. Furthermore the antinociceptive effect of paracetamol is lacking in *TRPV1* $^{-/-}$ mice [205].

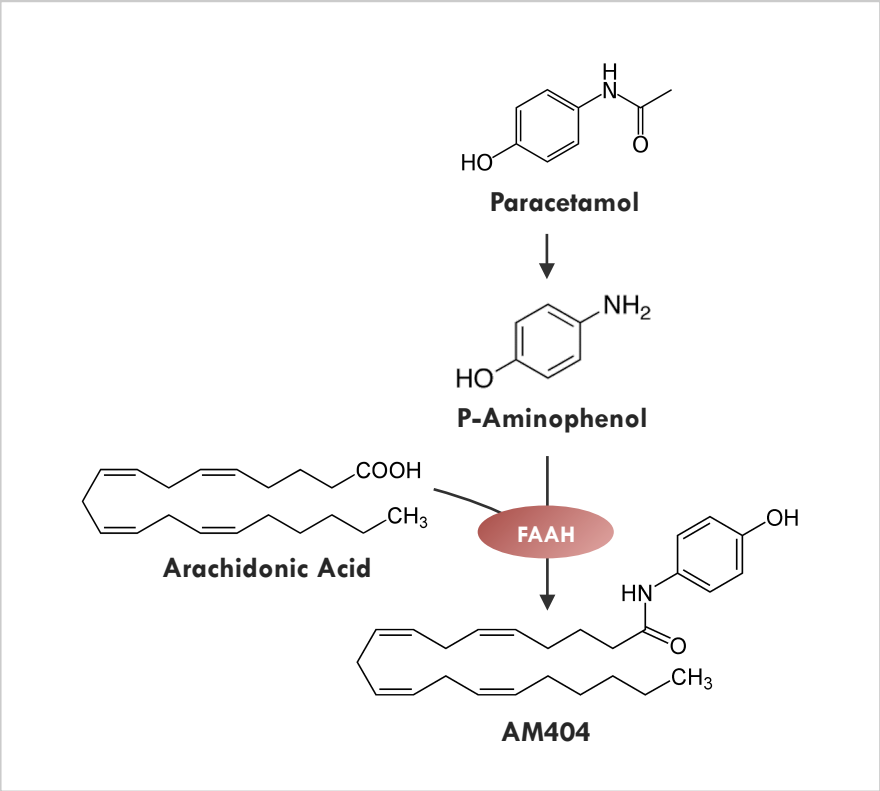



**Figure 4.2. The conversion of paracetamol to AM404**

Paracetamol is deacetylated to p-aminophenol that is then conjugated by FAAH enzyme to arachidonic acid to form N-arachidonoylphenolamine (AM404) [203].

In conclusion, the exact mechanism by which paracetamol produces analgesia remains unclear. The most widely discussed theory that paracetamol inhibits the COX-2 enzyme has some convincing results, however inconsistencies exist between the spectrum of action of paracetamol, known COX-2 inhibitors and NSAIDs [205]. The discovery of an active splice variant of *PTGS1* (COX-1) in canine brain was thought to overcome these discrepancies; yet this result was not reproduced in human cells [200-202]. The recent finding that paracetamol can be converted into the analgesic metabolite AM404 is an interesting proposition, however, to date only one paper has directly shown this conversion takes place [203]. While circumstantial evidence from decades of common usage has confirmed the efficacy and tolerability of paracetamol no mechanism proposed to date can definitively explain how it works.

## 4.3. Paracetamol overdose

In the UK and US, the most common cause of acute liver failure is paracetamol overdose [193]. This tends to be caused by excessive self-medication and lack of awareness of the presence of paracetamol in common cold and flu medicines. Furthermore toxicity can be increased by chronic alcohol use and malnutrition [192, 193]. A small portion of all paracetamol consumed is converted into N-Acetyl-p-benzoquinone imine (NAPQI) by cytochrome p450 (CYP450) [206]. At therapeutic doses of paracetamol NAPQI is detoxified by irreversibly binding to reduced glutathione (GSH). When paracetamol is taken in excess, NAPQI accumulates, GSH depletes and NAPQI can covalently bind to cellular proteins [207]. This leads to mitochondrial dysfunction, decreased ATP production and cell death (Figure

4.3)[193, 207]. The antidote to paracetamol overdose is treatment with N-acetylcysteine (NAC) [208]. NAC replenishes the GSH levels and prevents NAPQI accumulation [209].

Typical signs of apoptosis and necrosis are observed after paracetamol overdose. In the livers of wild-type mice high dose paracetamol causes severe tissue destruction and damaged, necrotic hepatocytes and the cells from paracetamol treated rats show increased necrosis [206, 207, 210]. Paracetamol overdose leads to increased phosphorylation of anti-apoptotic proteins Bcl 2 and Bcl xl, induces the expression of the pro-apoptotic protein Bim and causes translocation of pro-apoptotic protein Bax to the mitochondria [206, 207, 210]. After a toxic dose of paracetamol, sustained activation of JNK was observed and treatment with the JNK inhibitor SP600125 provided protection against paracetamol overdose [206]. This protection was not accompanied by a change in the extent of GSH depletion [206]. Therefore, after paracetamol overdose GSH levels deplete and NAPQI accumulates leading to activation of the apoptosis and necrosis pathways via JNK (Figure 4.3).

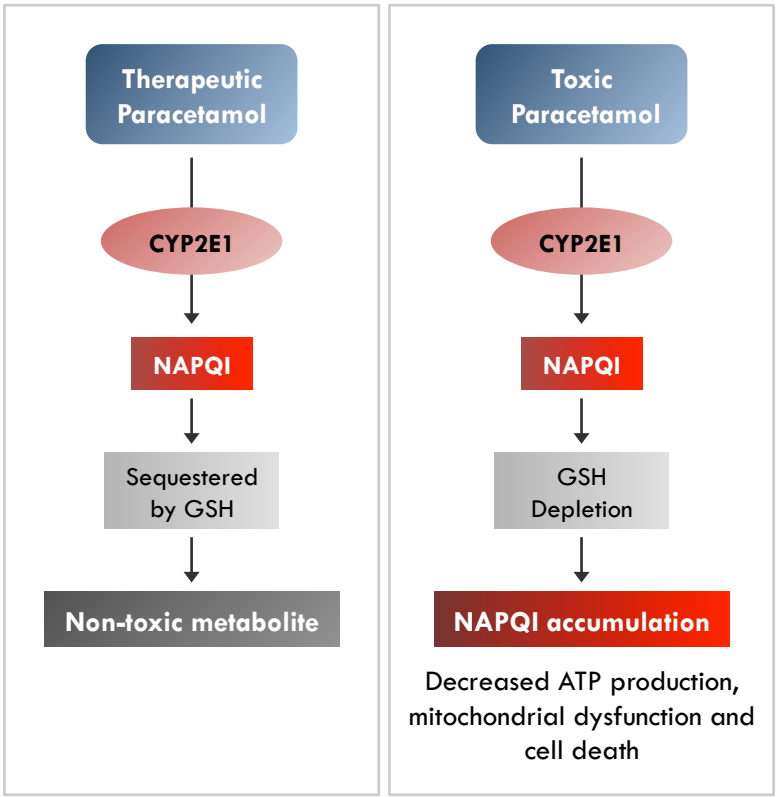



**Figure 4.3. The mechanism of paracetamol overdose**

**Left**: 5% of all paracetamol consumed is converted into the toxic metabolite NAPQI. This is sequestered by GSH to form a non-toxic metabolite. **Right**: During overdose the levels of NAPQI overwhelm the GSH stores and NAPQI can accumulate. NAPQI then binds to cellular proteins leading to cell death.

## 4.4. Paracetamol and cancer

The long term or frequent use of the NSAID, aspirin, is associated with a reduced risk of cancer [211-213]. Many studies have tried to expand upon this to determine whether a similar association exists for non-aspirin NSAIDs and paracetamol. The results are varied between different cancer types. In the non-melanoma skin cancers such as squamous cell carcinoma (SCC) and basal cell carcinoma (BCC) NSAIDs use was associated with a reduced risk of SCC and paracetamol use was associated with a reduced risk of both SCC and BCC [214]. A similar correlation was determined for prostate cancer [215]. Therefore, paracetamol may have tumour suppressive effects: the potential mechanisms to explain this are discussed below.

### 4.4.1. Inhibition of the cyclooxygenase pathway

The COX-2/$PGE_2$ pathway is associated with inflammation, growth and survival and is thought to play an important role in the 'inflammogenesis of cancer' [194]. COX-2, encoded by gene *PTGS2*, is overexpressed in many solid tumours, including colorectal, liver, pancreatic, breast, lung and head and neck [216]. COX-2 is overrepresented in both premalignant lesions and invasive HNSCC [216]. This tends to be associated with elevated levels of *PTGS2* and a poor prognosis. In healthy cells the pro-inflammatory prostaglandin, $PGE_2$, is balanced out by the other prostaglandins; when COX-2 expression is deregulated the balance shifts in favour of $PGE_2$ leading to chronic inflammation and cancer [194]. The COX-2/$PGE_2$/EP pathway is thought to be able to supress the activity of dendritic cells (DCs), natural killer (NK)-cells and T-cells, repressing type 1 immunity while promoting type 2 and allowing evasion of the anti-tumour immune response [196]. Furthermore COX-2

overexpression was found to be associated with reduced response to immunotherapy [196]. Accordingly the use of selective COX-2 inhibitors, and NSAIDs, could significantly reduce the risk of cancer and have been shown to overcome tumour immune evasion [196].

### 4.4.2. High dose paracetamol as a cancer therapeutic

Several studies have examined the link between high dose paracetamol, either alone or in combination with other drugs, and cancer cell death. A phase 1 trial on 19 patients with advanced cancer examined the effect of high dose paracetamol treatment with NAC rescue, and determined a partial response rate of 15.8% [209]. In neuroblastoma cells, staurosporine alone can induce apoptosis, when combined with high dose paracetamol staurosporine-induced cell death is increased [217]. Furthermore combined treatment is accompanied by a reduction in GSH [217]. Similarly, in ovarian cancer cells high dose paracetamol enhances the cytotoxic effects of cisplatin and paclitaxel by reducing GSH levels, increasing ROS and reducing mitochondrial membrane potential [218]. A similar result was observed in liver cancer cells [208]. Therefore, high dose paracetamol can potentiate the efficacy of common chemotherapeutics by inducing apoptosis and necrosis via NAPQI.

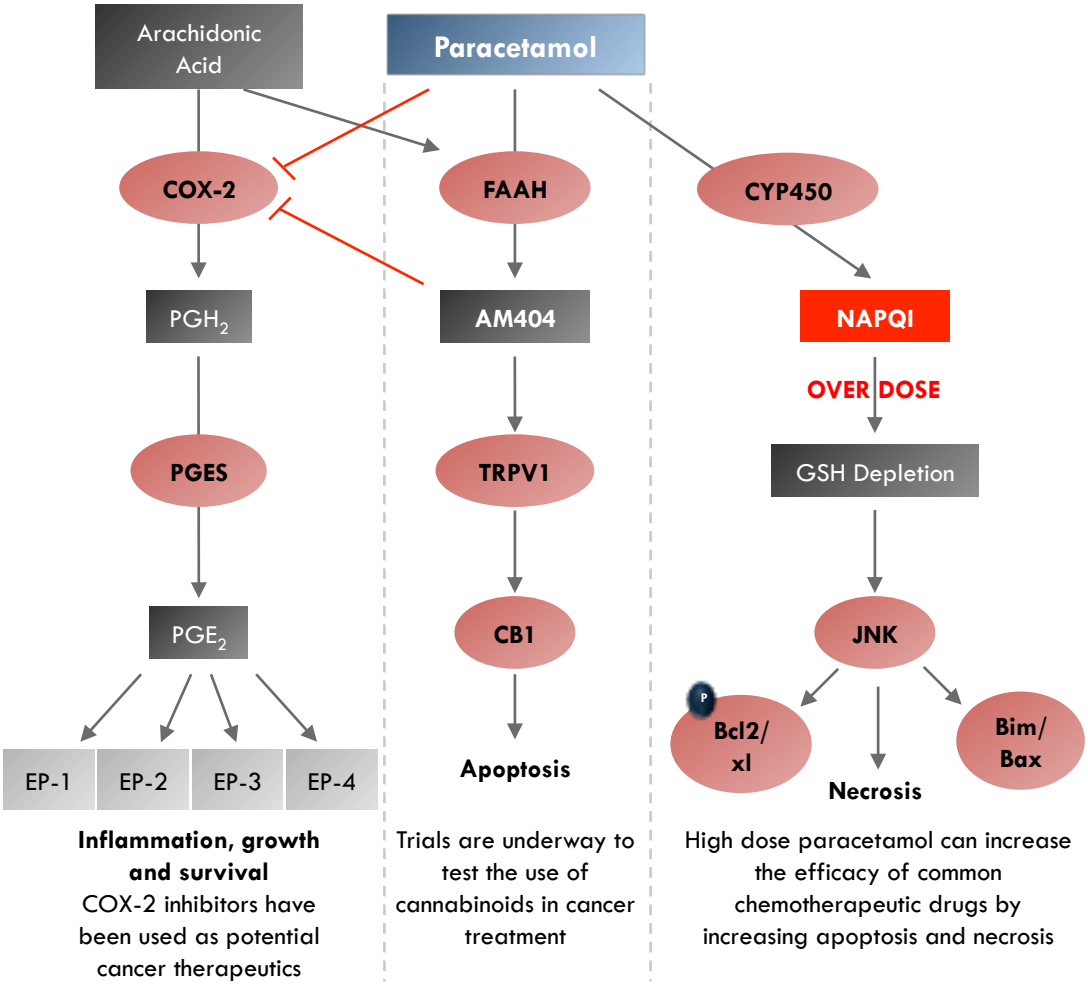



**Figure 4.4. The tumour suppressive effects of paracetamol**

**Left**: In many cancers COX-2 is overexpressed, leading to an increase in the pro-inflammatory prostaglandin PGE2 levels and contributing the 'inflammogenesis of cancer'. Paracetamol has been proposed to inhibit COX-2, causing a reduction of PGE2 levels. **Middle**: FAAH is thought to convert paracetamol to the metabolite AM404. This indirectly activates the cannabinoid receptor, CB1. Cannabinoids and similarly acting drugs are potential cancer therapeutics, thought to induce apoptosis via CB1. **Right**: A small amount of all paracetamol is converted by CYP450 into the metabolite NAPQI; this is usually detoxified by GSH; however when paracetamol is taken in excess GSH levels deplete and NAPQI accumulates leading to activation of the apoptosis and necrosis pathways via JNK. High dose paracetamol, either alone, or in combination with other drugs has the ability to inhibit cancer cell growth via this pathway.

## 4.5. Aims and objectives

The previous chapter has described a synergistic relationship between DAC and paracetamol that can reduce viability in HNSCC cell lines. This could indicate one of two mechanistic possibilities: either DAC is altering the mechanism of action of paracetamol or paracetamol is increasing the rate of DNA damage or DNA demethylation caused by DAC. The aim of this chapter was to examine each of these possibilities. This will be done via three main objectives:

- Examine the effect of DAC treatment on the cyclooxygenase pathway;
- Determine the effect of combined treatment with DAC and paracetamol on the paracetamol overdose pathway;
- Investigate the effect of paracetamol on DNA damage and DNA methylation levels in HNSCC.

## 4.6. DAC alters the COX pathway

Paracetamol is understood to act on the cyclooxygenase pathway. Therefore, HNSCC cells were treated with a titration of DAC with or without 132.3µM paracetamol for 96h and the relative RNA levels of the COX enzymes (encoded for by genes *PTGS1* and *PTGS2*) determined. *PTGS1* expression was not significantly changed by DAC and/or paracetamol treatment in all cell lines examined. A significant, dose-dependent increase in *PTGS2* gene expression was observed after DAC treatment in DAC-responsive HNSCC cell lines, HN12 and VU40T, while a similar pattern was not observed for the DAC-unresponsive cell lines, UDSCC2 and SCC040 (Figure 4.5). Similarly the protein levels of COX-2 were increased after DAC treatment in the DAC responsive cell line, but were in fact minimally decreased in the unresponsive cell line (Figure 4.5). Paracetamol alone caused no significant change to COX-2 protein or RNA expression and the addition of paracetamol to 500nM DAC did not alter the DAC-increased COX-2 RNA and protein levels described (Figure 4.5).

*COX-3* is a splice variant of *PTGS1* that has been proposed to have cyclooxygenase function [200]. RNA sequencing of the HNSCC cell lines shows that in untreated cells this variant is not expressed, and neither DAC nor paracetamol can induce expression. This was confirmed by very high crossing point (CP) values during qPCR analysis.

The level of prostaglandin $E_2$, one of the downstream products of COX-2 activity, was determined for VU40T and SCC040 cells. Treatment with paracetamol or the specific COX-2 inhibitor, valdecoxib caused a small, but insignificant decrease in $PGE_2$ in DAC-responsive cells, VU40T (Figure 4.6). After DAC treatment, a significant increase in

$PGE_2$ was observed only in the DAC-responsive cell line, VU40T. This directly reflects the COX-2 expression results previously described (Figure 4.5, Figure 4.6). In VU40T cells the increased $PGE_2$ observed after DAC was reversed by the addition of paracetamol or valdecoxib, back to the level of untreated cells (Figure 4.6).

$PGE_2$ exhorts its effects through G-protein coupled receptors (EP-1-4, encoded by *PTGER-1-4* genes) [194]. DAC-responsive cell lines showed increased RNA expression of these receptors after DAC treatment (Figure 4.7). For VU40T cells, a significant, dose-dependent increase in *PTGER1* was observed after treatment with DAC; while in HN12 cells *PTGER2* RNA levels were significantly, and *PTGER3* insignificantly, upregulated from 100nM DAC (Figure 4.7). DAC-unresponsive cell lines, UDSCC2 and SCC040, did not show increased *PTGER* expression.

The results thus far suggested that the efficacy of combined treatment may be dependent upon upregulation of the COX-2/$PGE_2$/EP pathway by DAC treatment (Figure 4.8). In HNSCC, increased expression of COX-2 correlates with tumour stage and a worse clinical outcome [216, 219-221]. Indeed, using the analysis software cBioPortal, a comparison of The Cancer Genome Atlas data from 504 HNSCC tumours identified amplifications in all steps in the COX-2/$PGE_2$ pathway; with *PTGS2* (*COX-2*) alterations generally associated with high levels of smoking and alcohol, a middle to high tumour T grade at diagnosis and a poor survival (Figure 4.9) [8, 222]. A Kaplan-Meier plot shows a trend of lower survival for patients with alterations in these genes, however the association was not significant (Figure 4.9) [8, 222]. Therefore upregulation of this pathway may be a negative result of DAC treatment. Paracetamol can then rescue this via inhibition of COX-2, as seen by the restored

$PGE_2$ levels after treatment with DAC and paracetamol (Figure 4.6). To test this theory, the ability of other COX inhibitors to sensitize HNSCCs to DAC treatment was examined (Figure 4.10). Ibuprofen is a generic COX inhibitor. Unlike paracetamol, the addition of ibuprofen at the Cmax concentration did not increase sensitivity to 500nM DAC in any of the cell lines tested (Figure 4.10). Interestingly, treatment with ibuprofen increased viability in HNSCC cells, but not normal oral keratinocytes and partially protected HN12 cells from DAC (Figure 4.10). Furthermore, the specific COX-2 inhibitor valdecoxib was unable to sensitise HNSCC cells to DAC at a range of concentrations (Figure 4.10).

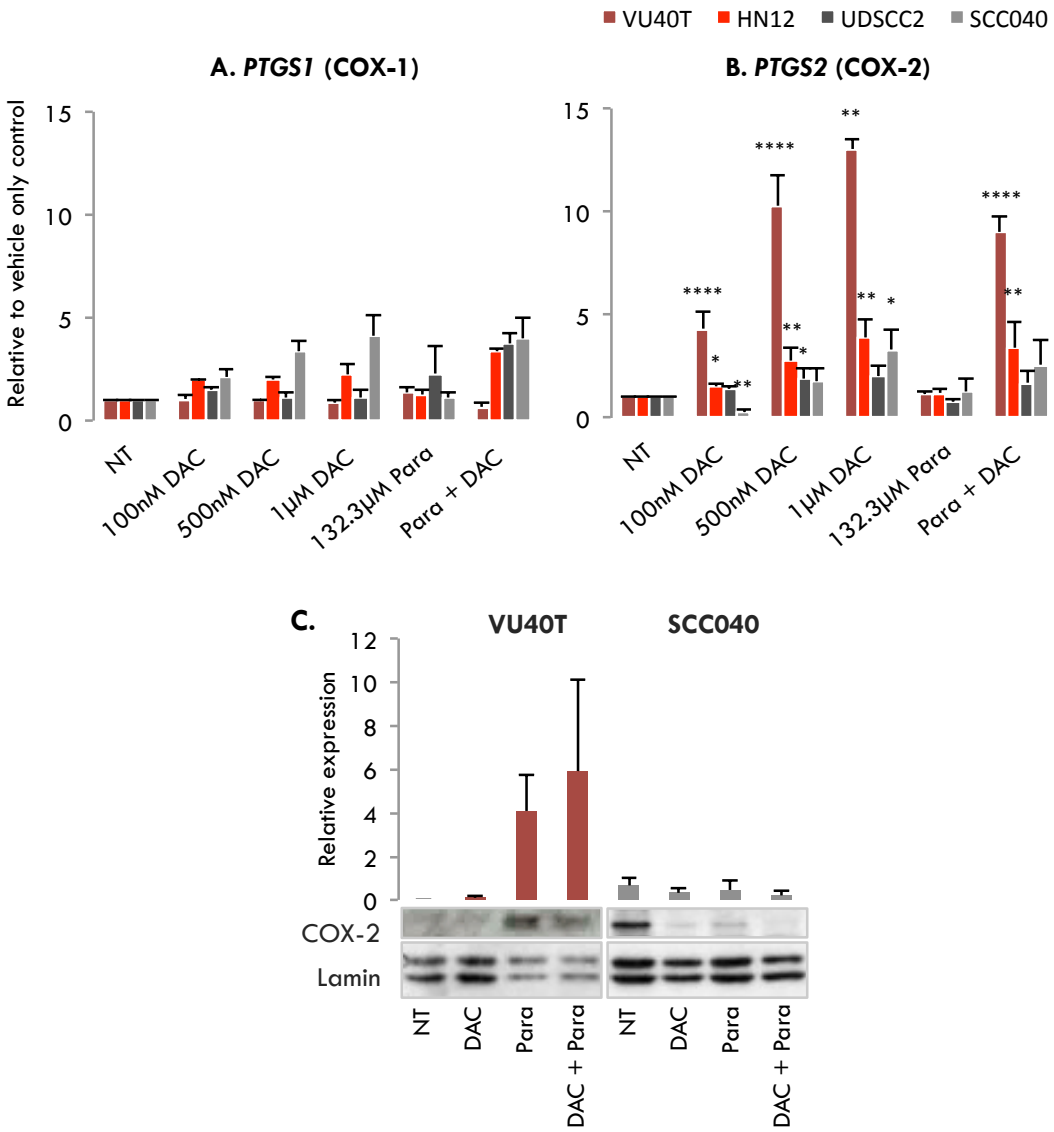



**Figure 4.5. DAC treatment increases the expression of COX-2 in DAC responsive cell lines**

HNSCC cell lines were treated with 100nM, 500nM and 1µM DAC, 132.3µM paracetamol or 500nM DAC + 132.3µM paracetamol for 96h following which RNA was extracted and qRT-PCR analysis of *PTGS1* (COX-1) (**A**) and *PTGS2* (COX-2) (**B**) performed. **C**. Proteins were extracted from VU40T and SCC040 cells and immunoblotted for COX-2. Additional blots are shown in Supplementary data 6. **A-C:** DAC-responsive cell lines are shown in red, while unresponsive cell lines are displayed in grey. Results display COX-2 normalised against the corresponding lamin band and shown relative to SCC040 NT. Results display mean + SEM (n=3). SPSS software was used to calculate paired t-tests with Bonferroni correction where *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

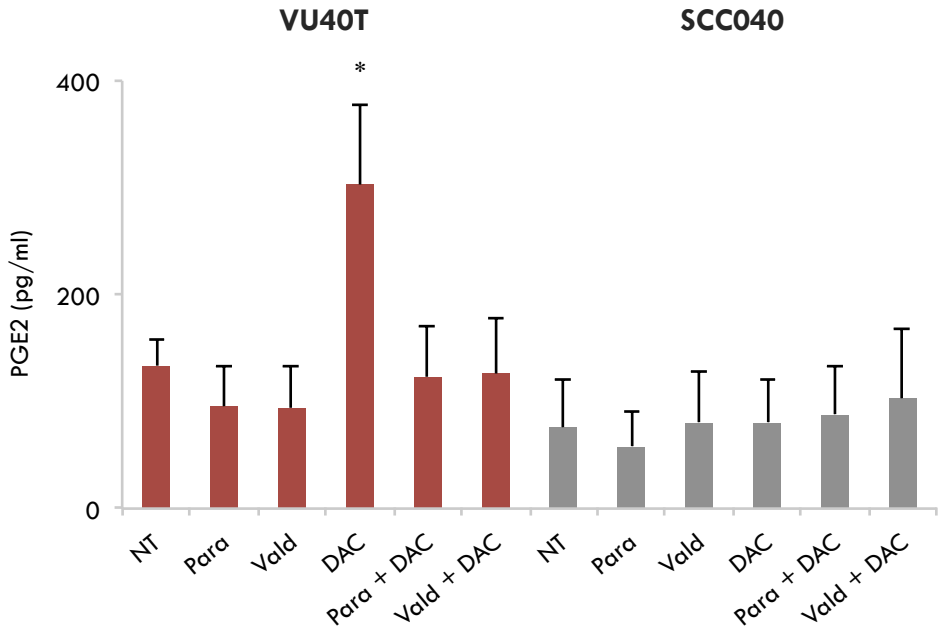



**Figure 4.6. DAC increases the levels of $PGE_2$ in DAC-responsive VU40T cells**

The $PGE_2$ levels in the media of VU40T and SCC040 cells after 96h treatment with combinations of 500nM DAC, 132.3μM paracetamol or 10μM valdecoxib were determined by ELISA. Results are the mean + SEM (n=3). A paired t-test with Bonferroni correction was used to determine significance compared to the untreated control where *p=0.05.

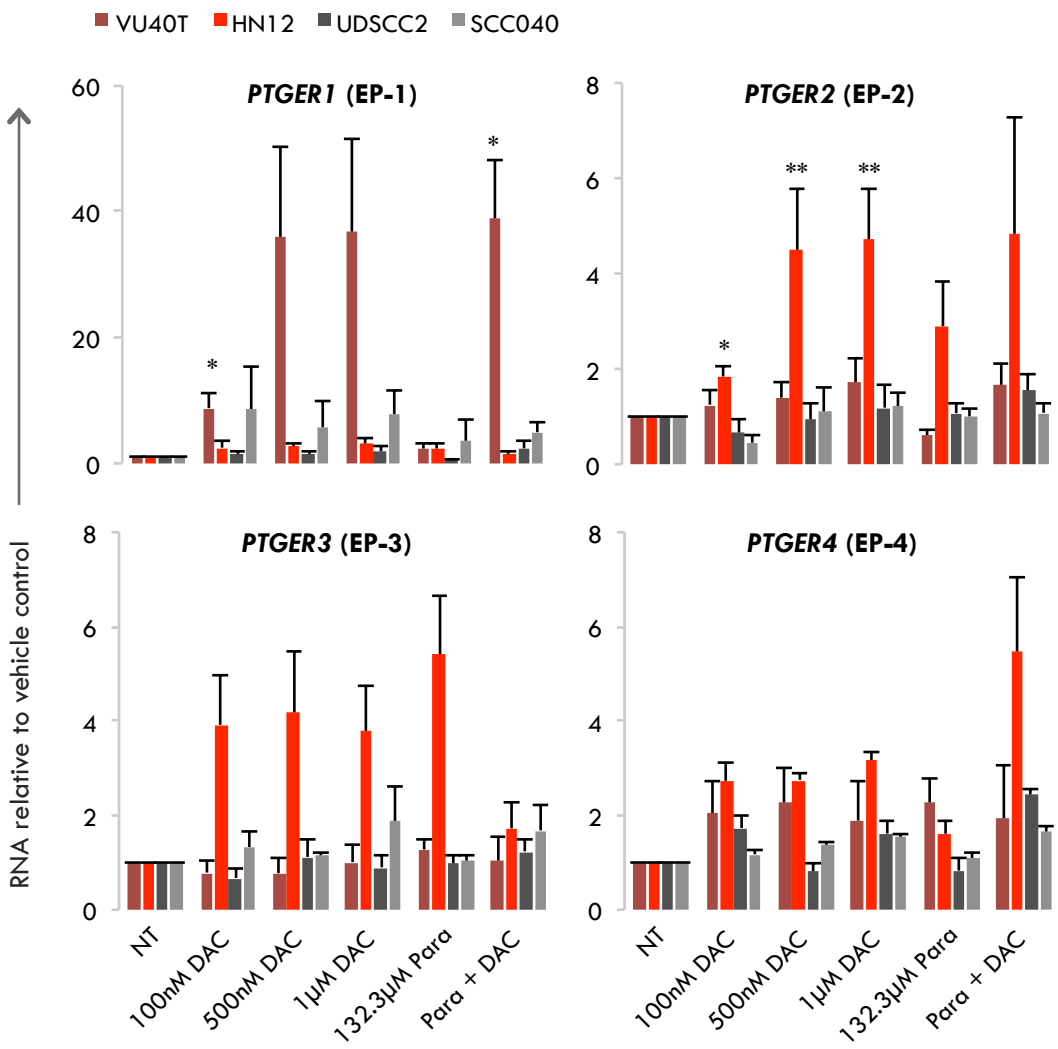



**Figure 4.7. DAC increases the expression of the $PGE_2$ receptors**

HNSCC cells were treated with a titration of DAC +/- 132.3μM paracetamol for 96h following which RNA was extracted. qRT-PCR analysis was performed for $PGE_2$ receptors EP-1-4 encoded by *PTGER-1-4*. Red bars denote DAC-responsive cell lines while DAC-unresponsive cells are shown in grey. Results are mean + SEM (n=3). Significance was determined using a paired t-test with Bonferroni correction where *p>0.05, **p>0.01

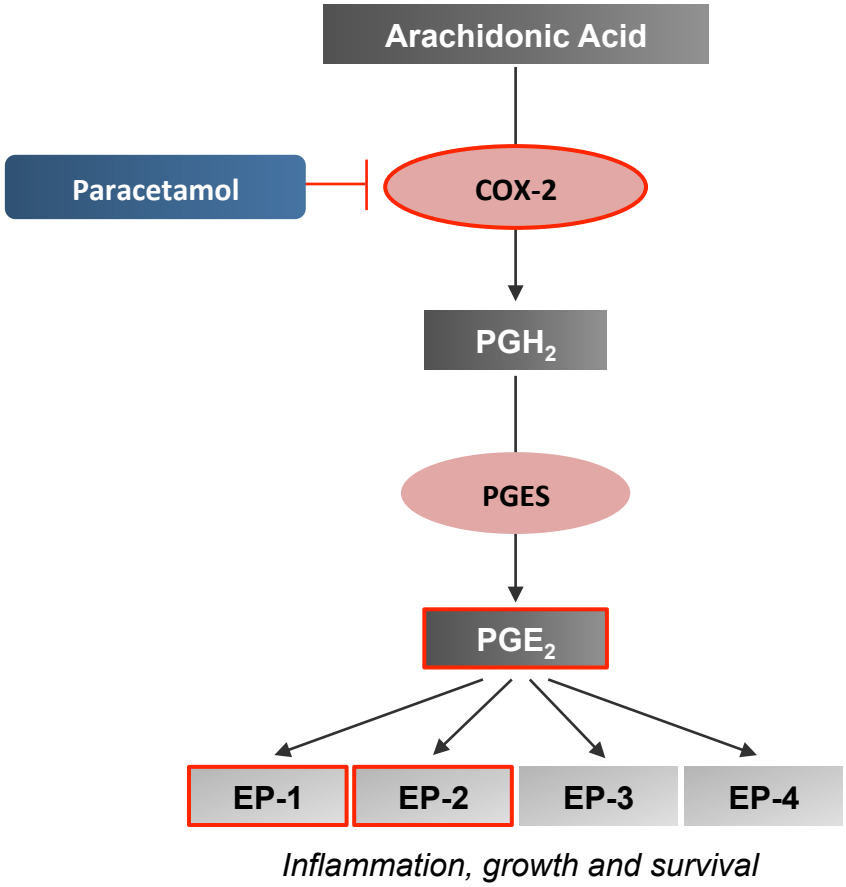



**Figure 4.8. DAC treatment alters the cyclooxygenase pathway**

Schematic representing the COX-2→$PGE_2$ pathway. COX-2 converts arachidonic acid into $PGH_2$ which is then rapidly converted into one of a number of effector prostaglandins. PGES converts $PGH_2$ into $PGE_2$ that is then recognized by one of four G-protein coupled receptors (EP-1-4). The COX-2→$PGE_2$ pathway is generally associated with inflammation, growth and survival. Aspects upregulated by DAC treatment are outlined in red. Paracetamol inhibits the activity of COX-2.

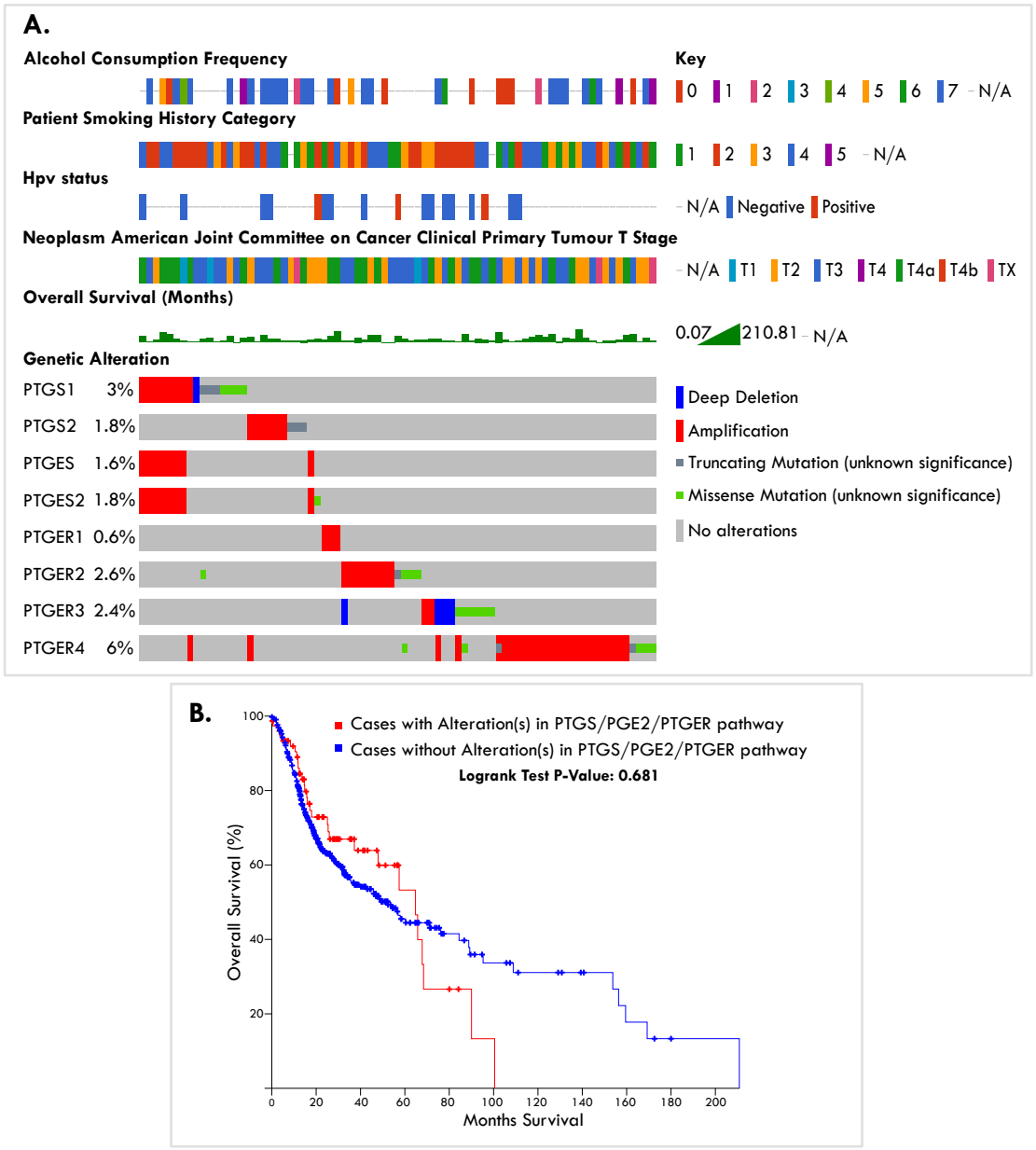



**Figure 4.9. Alterations in the COX-2/$PGE_2$ pathway in HNSCC**

The cBioPortal online software was used to visualise genetic alterations in the *PTGS/PTGES/PTGER* pathway in 504 HNSCC samples analysed by the cancer genome atlas [8, 222]. **A.** Alterations in *PTGS-1-2/PTGES-2/PTGER-1-4* occur in 15% of HNSCC cases (only altered cases are shown = 76/ 504 cases). Deep deletion indicates a deep genetic loss, for example a homozygous deletion **B.** Specific comparison of HNSCC cases with COX-2/$PGE_2$ pathway alterations. The level of alcohol consumption and smoking consumption; HPV status; tumour T stage; and overall survival time are shown above genetic alterations in COX/$PGE_2$ pathway genes. Each box corresponds with a single patient. **C.** Kaplein-Meier curve for HNSCC samples with or without alterations in *PTGS-1-2/PTGES-2/PTGER-1-4* genes as determined by cBioPortal. All images adapted from cBioPortal [222].

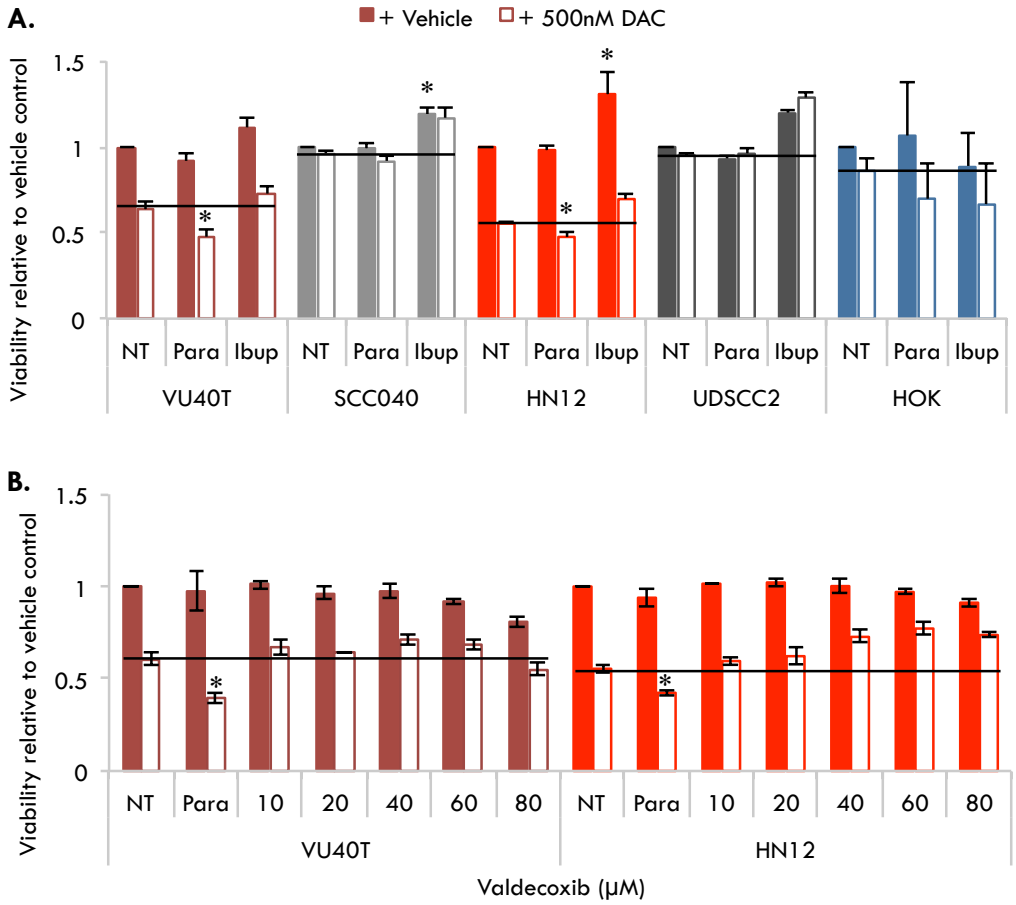



**Figure 4.10. Other COX inhibitors do not sensitise HNSCC cell lines to DAC treatment**

**A**. Four HNSCC cell lines and primary human oral keratinocytes (HOK) were subject to 96h treatment with 193.9μM ibuprofen (Ibup) or 132.3μM paracetamol (Para) +/- 500nM DAC following which viability was determined**. B.** DAC responsive HNSCC cell lines, VU40T and HN12, were subject to 96h treatment with a titration of Valdecoxib +/- 500nM DAC following which viability was determined. **A-B**. Results are mean + SEM (n=3), shown relative to the vehicle only control (NT). A paired t-test with Bonferroni correction was used to determine significance against the vehicle only or DAC only sample where *p<0.05.

## 4.7. Genes involved in the conversion of paracetamol to AM404 are not expressed in HNSCC cells

In recent years an alternate mechanism, involving the conversion of paracetamol into the metabolite AM404 by the enzyme FAAH and subsequent activation of TRPV1 and CB1, has been proposed to explain the analgesic effects of paracetamol [192, 203, 204] (Figure 4.11*A*). RNA sequencing of the HNSCC cell line showed little to no expression of these three factors, suggesting that this is not a mechanism employed in HNSCC cells (Figure 4.11*B*). Furthermore, while *FAAH* and *TRPV1* RNA could be detected by qRT-PCR, the levels were not altered by treatment with DAC +/- paracetamol (Figure 4.11C).

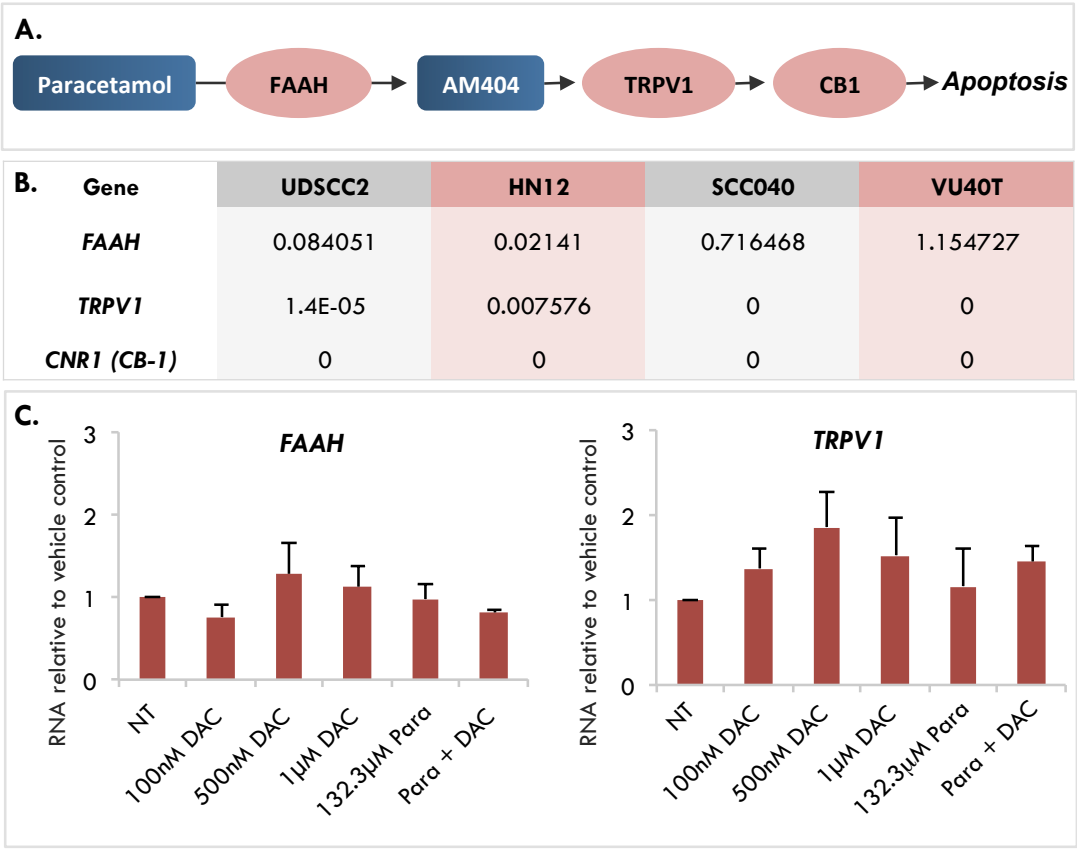



**Figure 4.11. Enzymes involved in the conversion of paracetamol into AM404 are not altered by DAC treatment in VU40T cells**

**A.** Schematic showing the potential conversion of paracetamol into the metabolite AM404. The FAAH enzyme is responsible for converting paracetamol into AM404, which leads to the activation of TRPV1 and CB1 and may trigger apoptosis. **B.** RNA sequencing of untreated HNSCC cell lines. FPKM values for genes outlined in **A. C.** VU40T cells were treated with a range of concentrations of DAC +/- 132.3μM paracetamol for 96h. qRT-PCR was performed for AM404 related genes *FAAH* and *TRPV1*. Results are mean + SEM (n=3).

## 4.8. Combined treatment with DAC and paracetamol may mimic paracetamol overdose

The potential use of paracetamol in the treatment of established cancers has been previously discussed [208, 217, 218]. These papers use relatively high doses of the drug and rely on the toxic effects of its metabolite, NAPQI, to induce cell death [208, 217, 218]. In contrast, the current project has described clinically relevant concentrations of the drug; 10-20 times lower than those previously described [208, 217, 218]. The CYP450 enzyme, CYP2E1, is understood to be primarily responsible for the conversion of paracetamol into NAPQI [223]. Interestingly, in DAC-responsive, VU40T cells, the RNA levels of *CYP2E1* were significantly and dose-dependently increased by DAC treatment, while unresponsive cell lines showed no change (Figure 4.12). This was unaltered by the addition of paracetamol (Figure 4.12). The protein levels were also increased by DAC in VU40T cells (Figure 4.12). The RNA levels of *CYP2E1* increased in DAC responsive HN12 cells after combined treatment with DAC and paracetamol, however this was not significant (Figure 4.12).

These results suggest that DAC might cause an increased conversion of paracetamol into NAPQI and mimic the effects of paracetamol overdose at clinically relevant concentrations. Due to the highly reactive and transient nature of NAPQI it is very difficult to measure in whole cell systems [224, 225]. Instead, in DAC-responsive VU40T cells, the level of glutathione (GSH) was used as a measure of paracetamol toxicity (Figure 4.13). 1mM of paracetamol was used as a control and, as expected, the levels of GSH decreased significantly at this dose (Figure 4.13). When administered alone 500nM DAC and 132.3µM paracetamol caused minimal change

to the concentration of GSH (Figure 4.13). However, when the two drugs were combined, a significant reduction in GSH was observed, to even lower than that detected after high dose (1mM) paracetamol treatment (Figure 4.13). These data support the theory that DAC is increasing the conversion of paracetamol into NAPQI and thereby mimicking the effects of paracetamol overdose (Figure 4.14).

Disulfiram (DSF) is a CYP450 inhibitor, known to inhibit CYP2E1. VU40T cells were treated with 500nM DAC and 132.3µM paracetamol for 96h plus 5, 10 or 50nM DSF for the final 24h or 48h of treatment (Figure 4.15). None of the concentrations of DSF tested were able to rescue VU40T cells from the reduced viability caused by DAC and paracetamol co-treatment, however, the levels of GSH after DSF rescue were not examined (Figure 4.15).

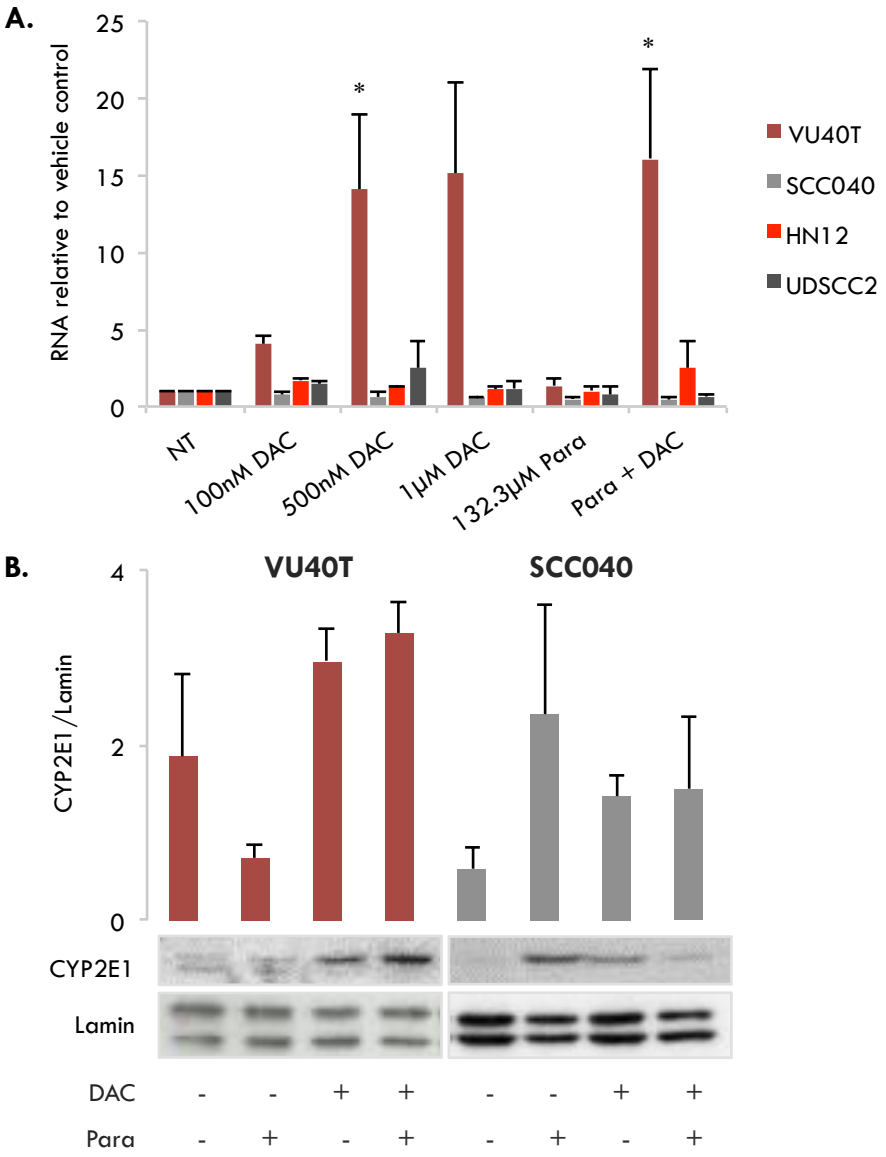



**Figure 4.12. DAC treatment increases the expression of CYP2E1 in DAC-responsive cells**

Four HNSCC cell lines were treated with 100nM, 500nM or 1µM DAC, 132.3µM paracetamol (Para), 500nM DAC + 132.3µM paracetamol (Para+ DAC), or an equivalent volume of vehicle (NT) for 96h, following which RNA and proteins were extracted. **A.** qRT-PCR was performed to determine the relative RNA levels of *CYP2E1*. **B.** Proteins extracted from VU40T and SCC040 cells were immunoblotted for CYP2E1. Representative blots from 3 independent experiments for CYP2E1 and lamin A/C control are shown with all blots shown in Supplementary data 6. Graph shows the mean intensity of the CYP2E1 bar relative to the corresponding Lamin A/C bars. **A-B.** Results are mean + SEM (n=3). Significance was determined using a paired t-test with Bonferroni correction where *p<0.05.

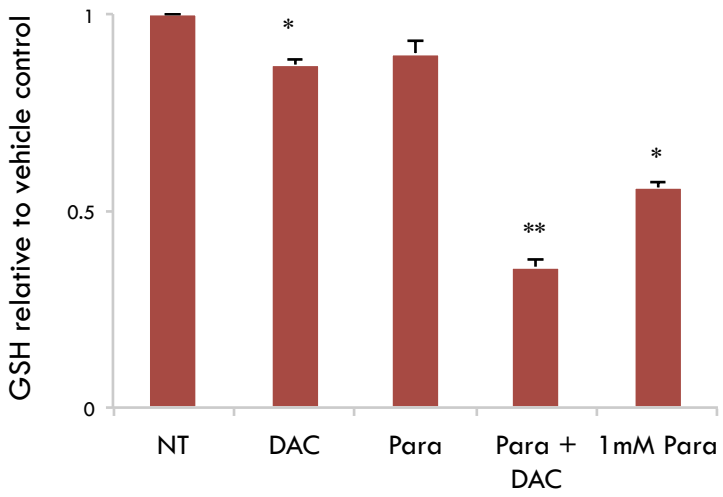



**Figure 4.13. Combined treatment with DAC and paracetamol reduces GSH levels.**

VU40T cells were treated with 132.3µM paracetamol +/- 500nM DAC for 96h following which glutathione (GSH) levels were determined using the GSH-glo kit. 1mM paracetamol was used as a high dose comparison. Results are mean (n=3) + SEM. Significance was determined using a paired t-test with Bonferroni correction where *p<0.05 and **p<0.01.

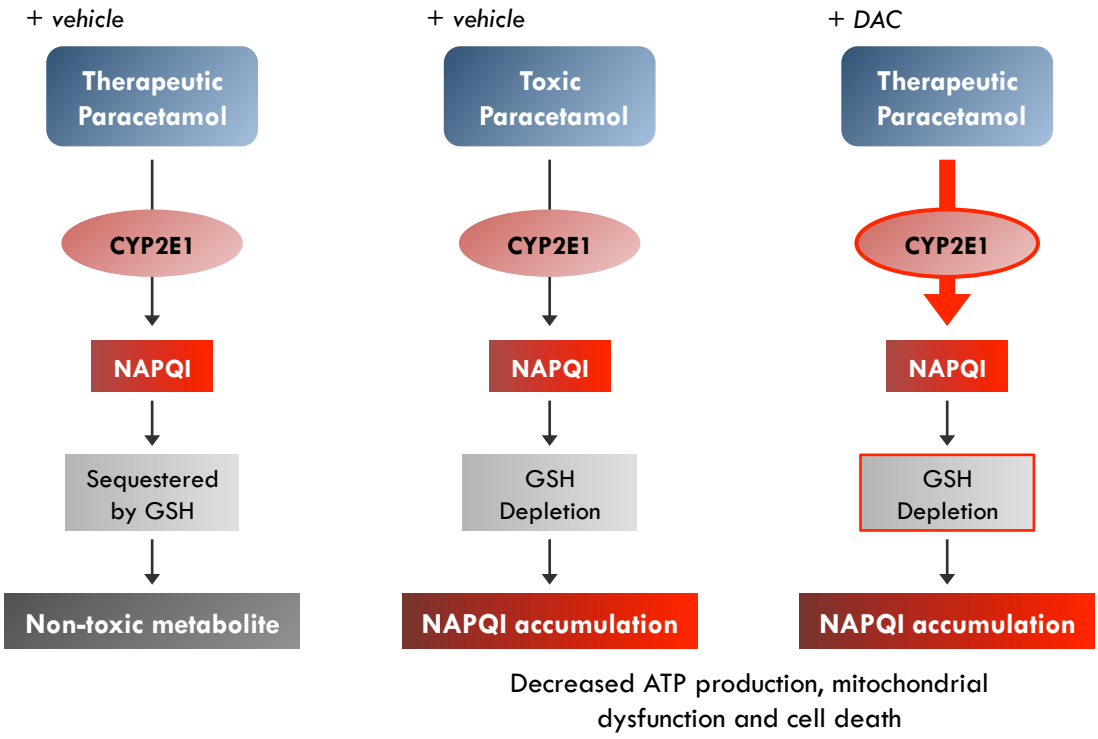



**Figure 4.14. Combined treatment with DAC and paracetamol may be mimicking the effects of paracetamol overdose**

Schematic representing the paracetamol overdose pathway and the potential mechanism by which DAC may be mimicking it. **Left.** A small amount of all paracetamol consumed is converted into the toxic metabolite NAPQI by CYP450 enzymes, primarily CYP2E1. This is usually detoxified by binding to GSH [226]. **Middle.** When paracetamol is taken in excess the GSH levels deplete, NAPQI accumulates, binds to cellular proteins and promotes decreased ATP production, mitochondrial dysfunction and cell death [206]. **Right**. DAC increases the expression of CYP2E1, potentially leading to increased conversion of paracetamol into NAPQI causing GSH depletion at therapeutic concentrations of paracetamol.

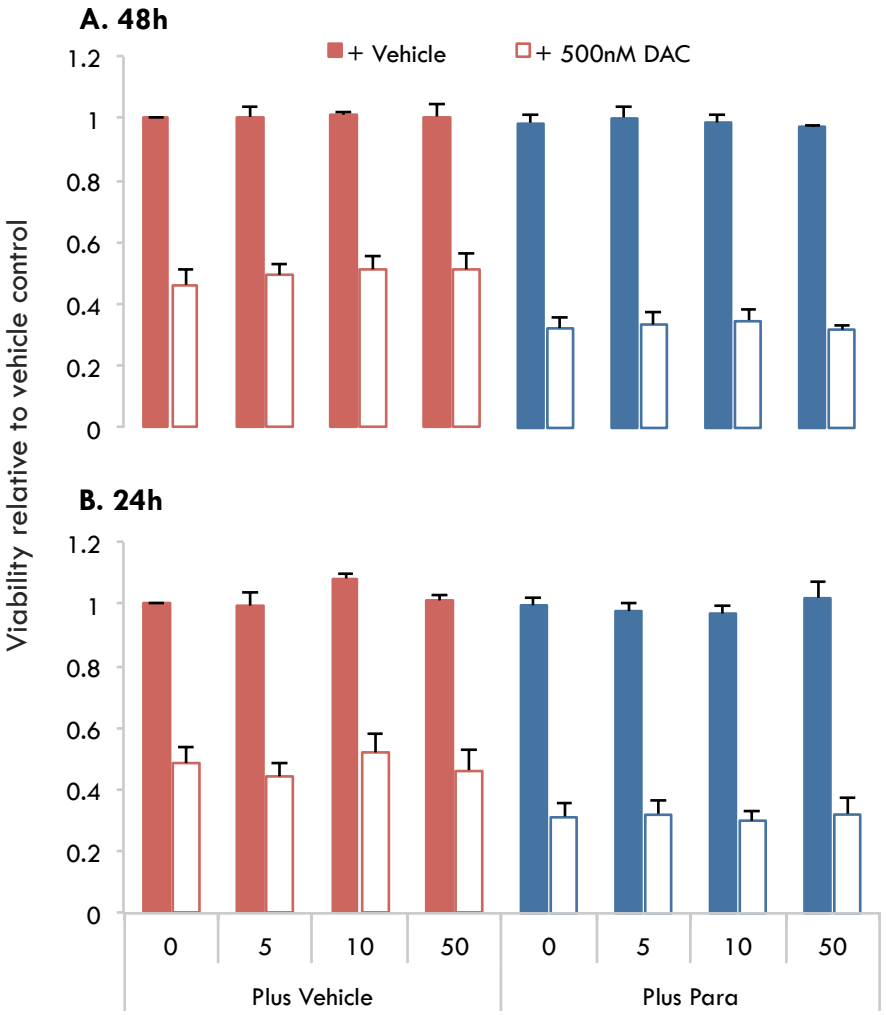



**Figure 4.15. Disulfiram does not rescue VU40T cells from combined treatment with DAC and paracetamol**

VU40T cells were treated with 500nM DAC +/- 132.3µM paracetamol for 96h, following which media was removed and replaced with varying concentrations of the CYP2E1 inhibitor disulfiram (DSF) for **A.** 24h or **B.** 48h. Viability was determined using CellTiter Blue reagent. Results are mean + SEM (n=3) and are shown relative to the vehicle only control.

## 4.9. Paracetamol treatment leads to a decrease in DNA methylation

The efficacy of DAC treatment can be attributed to two main mechanisms: increased DNA damage and reduced DNA methylation. Therefore the effect of paracetamol on each process was examined. Visible γH2AX foci within the nuclei of cells can be used as a measure of DNA damage (Figure 4.16). As expected, 500nM DAC increased the number of cells with γH2AX foci, with significantly more cells found to have 10 or more foci (Figure 4.16). Paracetamol did not increase the number of γH2AX foci and the addition of paracetamol to DAC treatment also did not cause any significant increase (Figure 4.16).

DAC causes a reduction in DNA methylation in DAC-responsive VU40T cells but not DAC-unresponsive SCC040 cells (Figure 4.17). Interestingly, paracetamol treatment causes a relatively small, but significant, drop in DNA methylation in VU40T, but not SCC040 cells (Figure 4.17, Supplementary data 9). This can be attributed to a corresponding decrease in the RNA expression of *DNMT-1, -3A* and *-3B* in VU40T, but not SCC040 cells (Figure 4.18). For *DNMT1* this was maintained, and increased slightly by the addition of DAC to the treatment (Figure 4.18). Interestingly, treatment with the specific COX-2 inhibitor valdecoxib did not give the same result (Figure 4.19). In VU40T cells valdecoxib did not decrease 5mC or the RNA levels of the three *DNMT* enzymes (Figure 4.19).

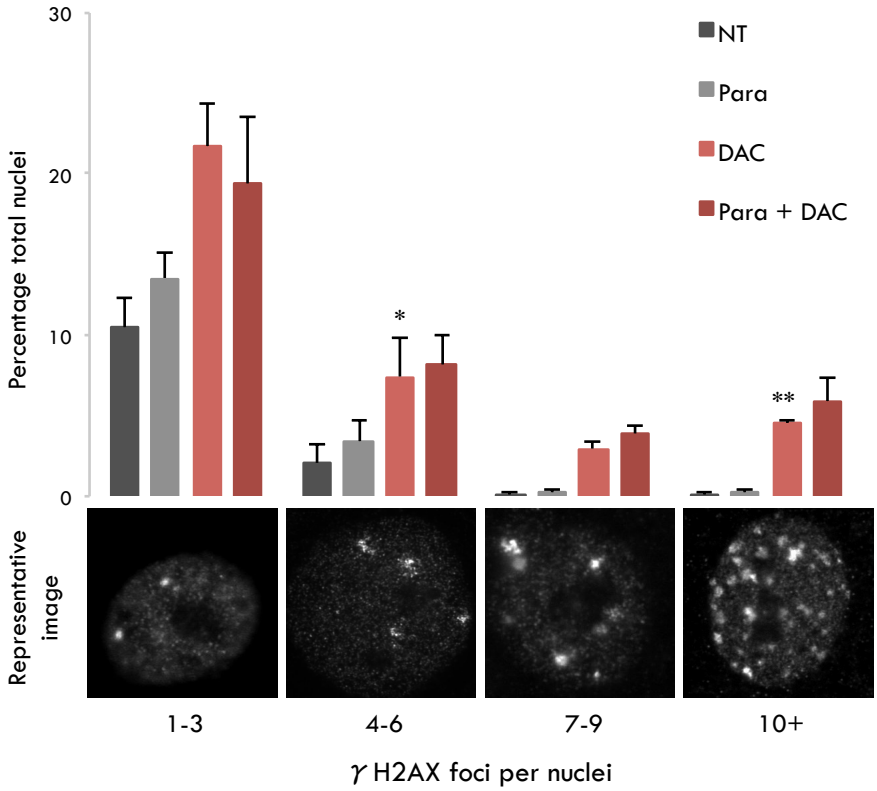



**Figure 4.16. Paracetamol does not increase DAC-induced DNA damage**

VU40T cells were grown on coverslips and treated with 500nM DAC +/- 132.3µM paracetamol for 96hrs. Cells were fixed in methanol and immunofluorescent staining for γH2AX performed. For each biological replicate 100 cells were counted. Results are mean +SEM (n=3) and are shown as a percentage of total cells counted. Significance was determined using a paired t-test with Bonferroni correction against the untreated control where *p<0.05 and **p<0.01. NT= vehicle only. Example cells for each γH2AX group are shown below.

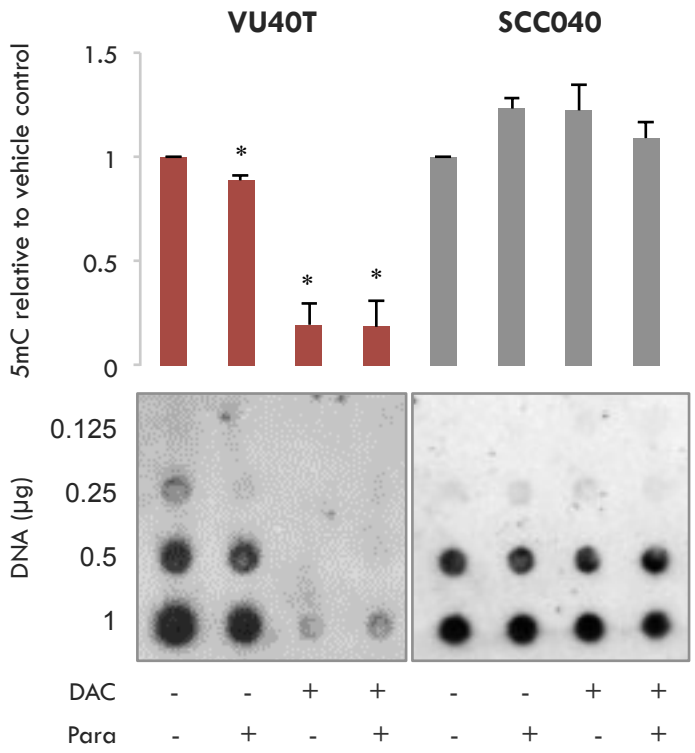



**Figure 4.17. Paracetamol decreases DNA methylation**

VU40T and SCC040 cells were treated with 500nM DAC +/- 132.3µM paracetamol for 96h following which DNA was extracted and immunoblotting for 5mC performed. Blots were incubated in methylene blue to stain for total DNA. The graph shows the 5mC blot normalised against the corresponding methylene blue dot and shown relative to the vehicle control. Results are mean (n=3) + SEM. Significance was determined using a paired t-test where *p<0.05. Representative blots are shown below and additional blots shown in Supplementary data 7.

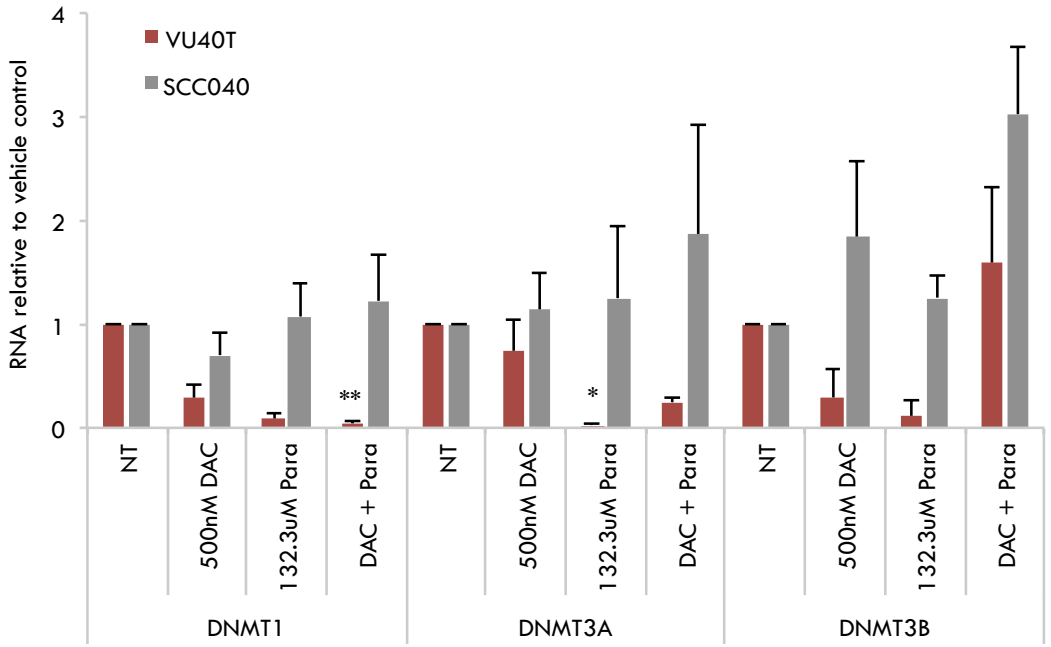



**Figure 4.18. Paracetamol reduces *DNMT* gene expression**

VU40T and SCC040 cells were treated with 500nM DAC +/- 132.3µM paracetamol for 96h following which RNA was extracted and qRT-PCR performed for *DNMT1, -3A* and *-3B*. Results are mean + SEM (n=3). Significance was determined using a paired t-test with Bonferroni correction where *p< 0.05 and **p<0.01.

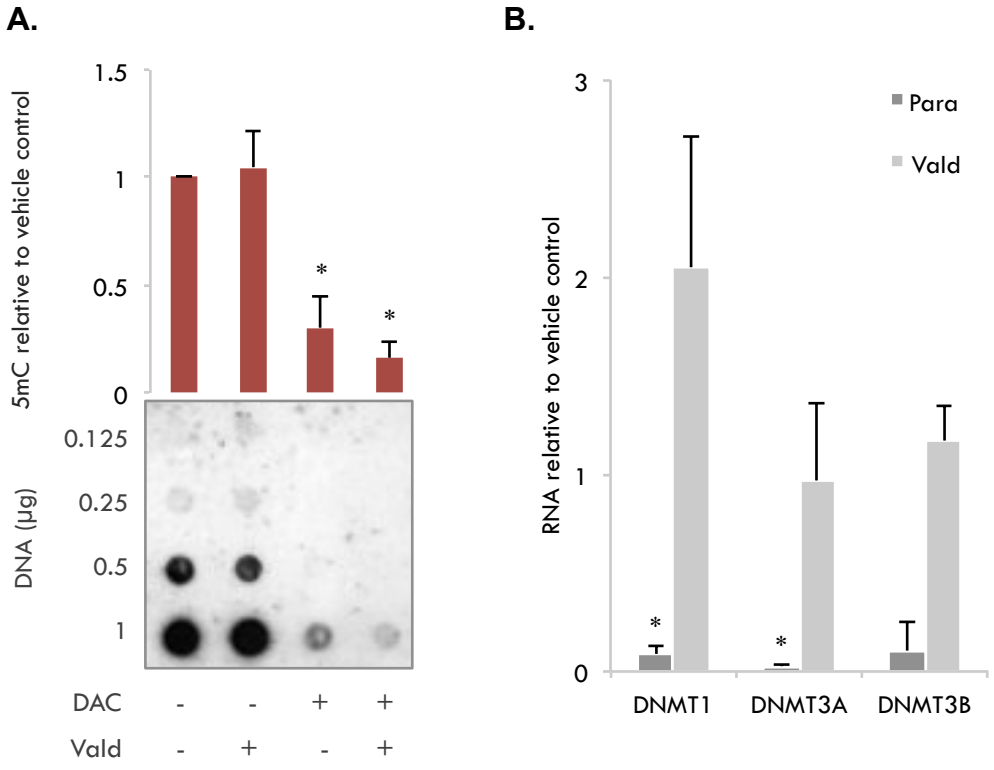



**Figure 4.19. Valdecoxib does not alter the levels of DNA methylation**

VU40T cells were treated with 500nM DAC +/- 10mM valdecoxib (vald) or 132.3µM paracetamol (para) for 96h. **A.** DNA was extracted and immunoblotting for 5mC performed. Methylene blue was used to stain for total DNA and results are shown normalised against the methylene blue loading control relative to the vehicle only sample. A representative blot is shown below and additional blots shown in Supplementary data 7. **B.** RNA was extracted and qRT-PCR performed for *DNMT-1, -3A and -3B*. Results are mean + SEM (n=3) and shown relative to the vehicle only control. Significance was determined using a paired t-test with Bonferroni correction where *p<0.05.

## 4.10. Discussion and conclusion

The aim of this chapter was to determine whether the synergistic relationship between DAC and paracetamol could be attributed to paracetamol altering the mechanism of action of DAC or DAC altering the mechanism of action of paracetamol. Interestingly, both outcomes were found to be true, suggesting that the reduction in viability described is brought upon by a combinatorial effect from the disruption of several pathways.

The most commonly held theory on the mechanism of action of paracetamol involves inhibition of the COX pathway, primarily through COX-2 [191, 195]. However, some authors contend that paracetamol inhibits a splice variant of COX-1, referred to as COX-3 [200]. In the system described here, this possibility is unlikely as the isoform is found primarily in canine cells, and as expected, was not present in the HNSCC cell lines used [201]. Instead, DAC treatment resulted in a significant increase in COX-2 expression only in DAC-responsive cell lines. This was accompanied by an increase in the downstream product, $PGE_2$. This pathway is understood to play a pivotal role in the 'inflammogenesis of cancer' with COX-2 inhibitors already being considered as HNSCC therapeutics [194, 219-221]. Therefore the increase in this pathway by DAC treatment should be pro-tumourigenic. DAC demethylates DNA in an indiscriminate and global manner and hence it is unsurprising that some cancer-promoting pathways are also activated by treatment. However, the efficacy of the drug demonstrates that the balance between the activation of cancer promoting genes and the activation of tumour suppressor genes is in favour of the latter. Therefore, while DAC may be inadvertently promoting

cancer survival through COX-2/$PGE_2$ pathway promotion, this is overpowered by the reactivation of silenced tumour suppressor genes. The addition of paracetamol to the treatment may then remove the cancer-promoting effects of COX-2 pathway promotion by reducing production of $PGE_2$ and hence, as described, an increase in cancer inhibition will be observed. Recently, the efficacy of DAC in cancer therapeutics has been attributed to an indirect activation of the interferon (IFN) pathway [158, 180]. This could account for the increased COX-2 expression observed as interferon treatment has been shown to increase COX-2 expression. Indeed, a COX-2 inhibitor was shown to work in synergy with IFN-β to induce apoptosis in human hepatoma cell lines and mouse models [227]. However, this cannot fully account for the synergistic relationship between DAC and paracetamol as neither the COX-2 specific inhibitor valdecoxib, nor the COX-generic inhibitor ibuprofen were able to produce the same result.

Several studies have described the potential benefits of high dose paracetamol treatment in cancer, either alone, or in combination with common chemotherapeutics [208, 209, 217, 218]. The toxicity of paracetamol at this dose is dependent upon the conversion of paracetamol into the toxic metabolite NAPQI [206]. While NAPQI is usually detoxified by glutathione (GSH), when paracetamol is taken in excess GSH is depleted allowing NAPQI to accumulate and bind to other cellular proteins. This can include covalent binding to mitochondrial proteins leading to decreased production of ATP and eventual cell death [193, 207, 228]. In the current study, the dose of paracetamol used was not sufficient to overwhelm GSH stores alone. However, DAC caused an increase in expression of the enzyme *CYP2E1*,

understood to be primarily responsible for the conversion of paracetamol into NAPQI. This combination results in a depletion of GSH greater than that observed from high dose paracetamol, suggesting that combined treatment with DAC and paracetamol may be mimicking the effects of paracetamol overdose. Interestingly, treatment with the CYP2E1 inhibitor, DSF could not rescue DAC-responsive HNSCC cells from combined treatment, again suggesting that this mechanism cannot fully account for the synergistic relationship between DAC and paracetamol. This result could be investigated further by determining whether GSH levels after DAC and paracetamol co-treatment are rescued by DSF, and whether the addition of GSH to the treatment regime eliminated the synergistic effect of DAC and paracetamol.

Furthermore, in DAC-responsive HNSCC cells, paracetamol reduced 5mC via reduced expression of the *DNMT* enzymes. This result is unlikely to be attributed to inhibition of the COX pathway, as the COX-2 inhibitor valdecoxib was unable to alter 5mC levels. Unlike paracetamol, valdecoxib does not have a synergistic relationship with DAC. Therefore the combined effect of DAC and paracetamol could partially be attributed to the DNA demethylation brought about by paracetamol. As discussed in chapter 3, paracetamol treatment causes a slight stalling of the cell cycle at G1/G0. Interestingly, expression of both DNMT1 and DNMT3b is significantly downregulated in G1/G0, and a corresponding decreased in 5mC is observed [229]. This could partially account for the reduction in 5mC described here [229].

Examination of the putative pathways that could contribute to the synergistic relationship between DAC and paracetamol did not produce a clear outcome. Instead, alterations in several pathways were found: DAC increased the COX-2/$PGE_2$

pathway; combined treatment mimics the effects of paracetamol overdose; and paracetamol can decrease DNA methylation (Figure 4.20). Comparable experiments using other COX inhibitors were not successful; neither was rescue experiments using an inhibitor of CYP450 enzymes. This suggests that no individual pathway can fully account for synergistic relationship between the two drugs, and instead suggests that alterations in all three pathways cooperate to reduce viability (Figure 4.20).

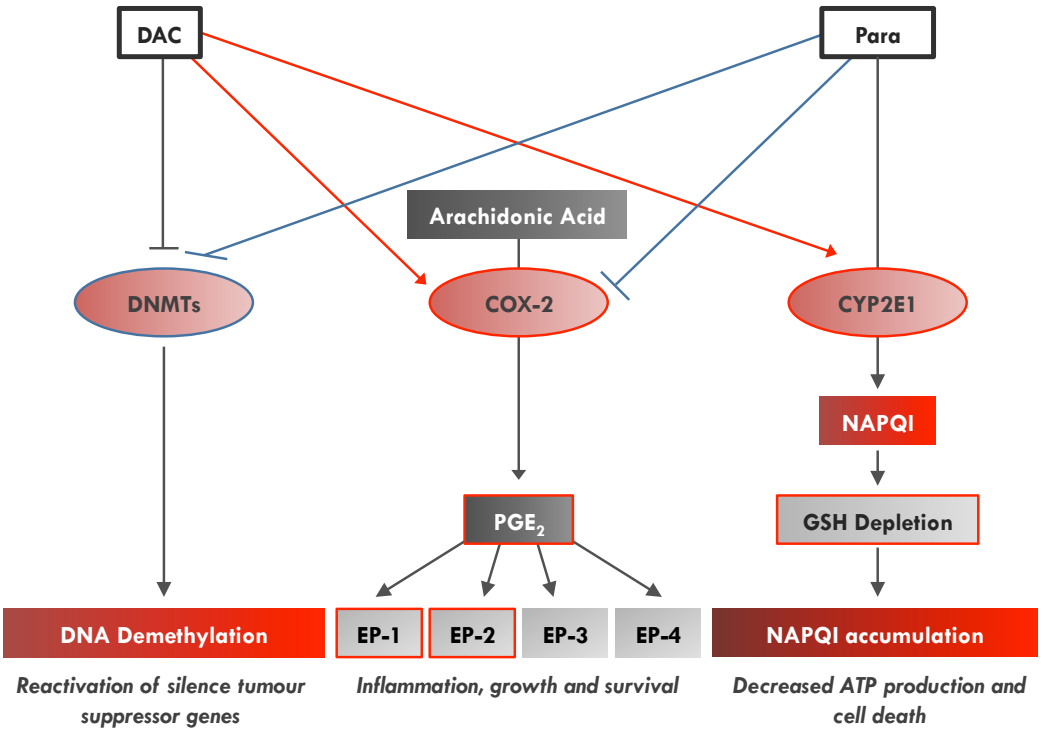



**Figure 4.20. DAC and paracetamol work in synergy to reduce viability in HNSCC cells**

Schematic representing the potential mechanisms utilised by combined treatment with DAC and paracetamol to reduce viability in HNSCC cells. **Left.** DAC demethylates DNA by inhibiting DNMT enzymes. In DAC responsive cells paracetamol reduces the expression of DNMTs leading to reduced DNA methylation. **Middle.** Arachidonic acid is converted into $PGH_2$ by COX-2 which is preferentially converted into $PGE_2$. $PGE_2$ can then activate EP receptors 1-4. DAC treatment increases the levels of COX-2, $PGE_2$ and EP receptors -1 and -2. Increases in this pathway by DAC may promote inflammation, growth and survival, however paracetamol can inhibit COX-2 activity and therefore protect from this. **Right.** Paracetamol is converted into the toxic metabolite NAPQI by CYP2E1. This is usually detoxified by glutathione, however if high dose paracetamol is present, GSH depletes and NAPQI accumulates leading to decreased ATP production and cell death. DAC increases the expression of CYP2E1 leading to GSH depletion after combined treatment with DAC and paracetamol.

# Chapter 5. HNSCC is characterised by global changes in the distribution of DNA methylation and hydroxymethylation

In previous chapters a novel therapeutic option was described for the treatment of HNSCC, where the DNA demethylating agent, DAC, was combined with the analgesic, paracetamol. The focus of this thesis will now shift towards investigating epigenetic changes in HNSCC in response to DAC treatment. However, this chapter will first investigate differences in the distribution of DNA methylation and hydroxymethylation between normal oral keratinocytes and HNSCC cells. The work highlights the central role of DNA modifications in the control of Alu elements and how this is dramatically altered in HNSCC.

## 5.1 Methods of determining the distribution of 5mC

The 'gold standard' of 5mC sequencing is bisulfite sequencing (BS-seq). In this technique, DNA is treated with sodium bisulfite to convert unmethylated cytosines into uracil, while methylated cytosines are protected. Subsequent sequencing and comparison of the bisulfite-converted sample with the reference genome can determine the position of 5mC at base pair resolution [230, 231]. The main

drawback to this technique is the sequencing depth required [230]. The associated costs have lead to the development of enrichment-based techniques [230]. Reduced representation bisulfite sequencing (RRBS) utilises restriction enzymes to select for CpG-rich regions of the genome prior to BS-seq [232]. This provides base pair resolution of 5mC but leaves CpG-poor regions uncharacterised [230]. Other enrichment based techniques do not provide base pair resolution; instead the DNA is sheared and fragments containing 5mC selected and sequenced, this is compared to the background or input DNA sequence and results in 'peaks' in the DNA sequence where 5mC is enriched. Methylation restriction enzyme sequencing (MRE-seq) utilises methylation-sensitive restriction enzymes to select for 5mC sites; MBD-seq uses the methyl binding domain of MeCP2 to capture 5mC; methylated DNA immunoprecipitation (MeDIP) uses an antibody against 5mC to specifically immunoprecipitate 5mC- containing fragments of DNA [168, 230]. Of the 5mC-enrichment methods currently available MeDIP provides the best coverage of repetitive elements over other 5mC sequencing methods and allows small quantities of input DNA to be used [230].

## 5.2 Methods of determining the distribution of 5hmC

The discovery of oxidised derivatives of 5mC brought technical challenges in determining the genomic distribution of these bases. Many single base resolution 5mC detection techniques are unable to distinguish between 5mC and 5hmC: Bisulfite sequencing cannot distinguish between 5mC and 5hmC, or between unmodified C and 5fC or 5caC, and some methylation-sensitive restriction enzymes such as *HpaII* and *McrBC* cannot reliably distinguish between 5mC and 5hmC [233].

Therefore, many of the previously published reports of 5mC distribution at single base resolution will represent 5hmC and 5mC combined [233]. Furthermore, due to the relatively low abundance of these bases in the genome, sensitive detection techniques are required.

Two methods have been proposed to overcome the limitations of conventional bisulfite sequencing: TET-assisted BS-seq (TAB-seq) and oxidative bisulfite sequencing (oxBS-seq) (Figure 5.1) [234, 235]. Comparing these techniques with conventional BS-seq allows 5hmC and 5mC to be determined at base pair resolution. However, the sequencing costs are high as both a bisufite converted and an oxBS or TAB sample need to be sequenced for each experiment [230].

Affinity-based techniques can be used to determine the distribution of 5hmC across the genome by directly (hMeDIP) or indirectly (Anti-CMS, JBP-1, GLIM, hMe-Seal) enriching for 5hmC-marked DNA [236, 237] (see Figure 5.2). Hydroxymethylated DNA immunoprecipitation (hMeDIP) involves the direct enrichment of 5hmC using an antibody against the modified base [237]. The main benefit of this technique over other 5hmC enrichment methods is its simplicity; the DNA is not damaged or lost in chemical reactions and there are fewer variables that may differ between experiments [236]. Indeed, a comparison of hMeDIP with oxBS-seq found a good correlation between the two techniques even if hMeDIP is known to bias towards CpG rich regions [231].

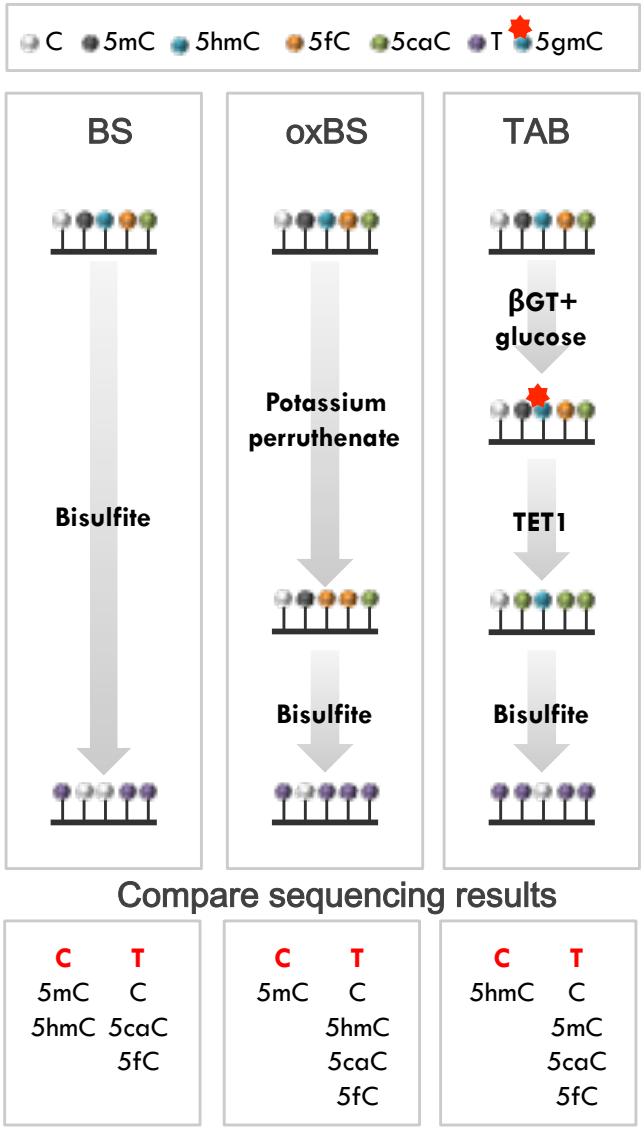



**Figure 5.1. Determining 5hmC at base pair resolution**

Conventional BS-seq cannot determine the difference between 5mC and 5hmC or 5fC, 5caC and unmodified C. TAB-seq involves tagging 5hmC with a glucose molecule using β-glucosyltransferase (βGT), followed by conversion of 5mC and 5fC into 5caC by recombinant TET1. After BS-seq 5hmC is recognised as C while all other cytosines are recognised as T. Oxidative bisulfite conversion (oxBS-seq) involves the conversion of 5hmC into 5fC by potassium perruthenate. This is then bisulfite converted and read as C, while 5mC is detected as T. After sequencing the results of BS converted DNA are compared with oxBS or TAB converted DNA. The bottom panels shows what the results from sequencing converted DNA (in red) represent in terms of modified bases (underneath in black). Figure adapted from [231].

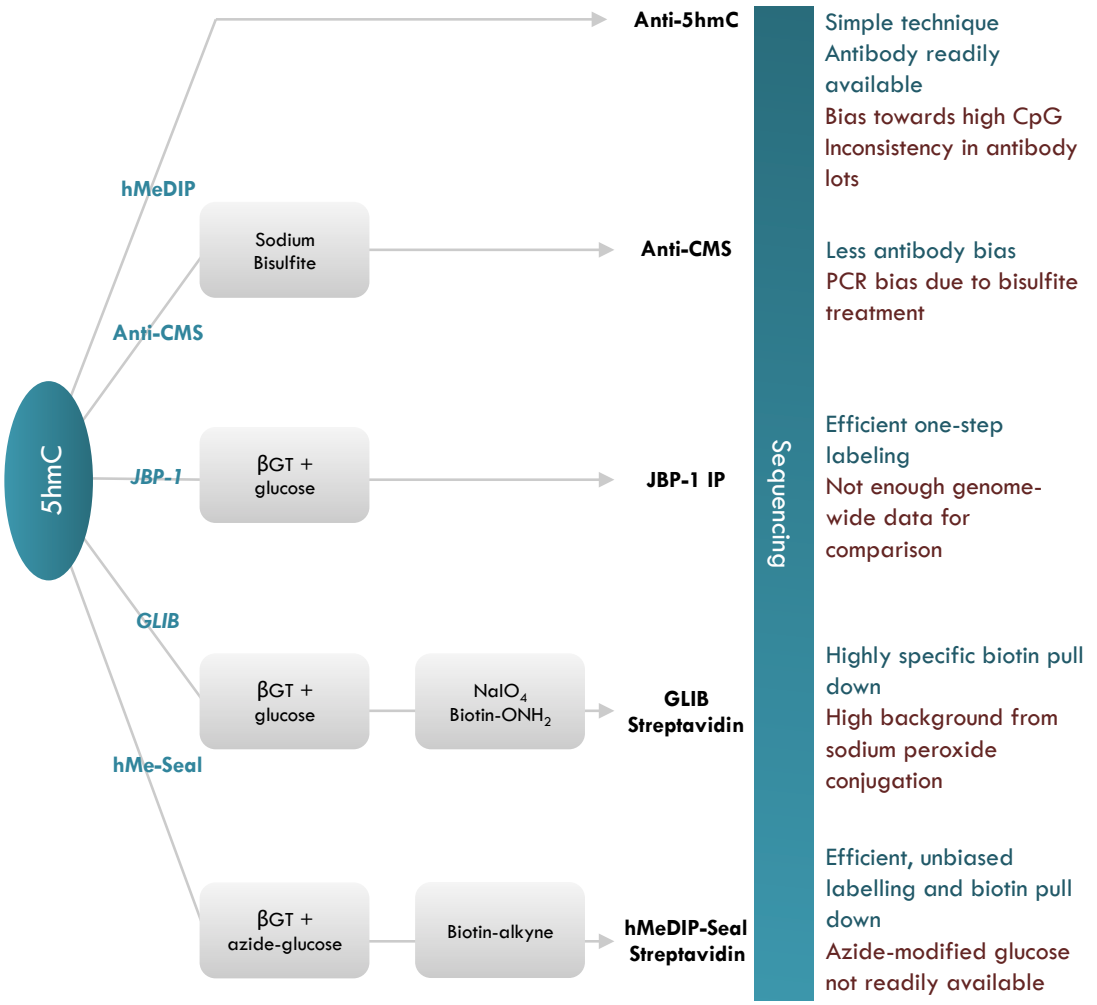



**Figure 5.2. Genome wide affinity-based methods for 5hmC profiling**

In **hMeDIP**, an anti-5hmC antibody is used to immunoprecipitate DNA containing 5hmC. **Anti-CMS** involves first sodium bisulfite converting 5hmC into cytosine 5-methylenesulfonate (CMS) and then using an antibody against CMS to enrich for 5hmC-marked DNA. The **JBP-1** technique involves first conjugating a glucose molecule onto 5hmC using the T4 bacteriophage enzyme β-glucosyltransferase (βGT). The J-binding protein 1 (JBP-1) can then be used to enrich for 5gmC-marked DNA. The **GLIB** technique utilises the same first step, however this is followed by treatment with sodium periodate to generate reactive aldehyde groups, which can be biotinylated using an aldehyde-reactive hydroxylamine-biotin. The biotin-conjugated modified 5hmC can be pulled down with streptavidin beads. **hMe-Seal** entails the addition of an azide-glucose molecule to 5hmC using βGT, followed by addition of a biotin group and 5hmC-enriched DNA can be pulled down by streptavidin beads. The advantages of each technique are shown alongside in **blue** and the disadvantages in **red.** Figure adapted from [236].

## 5.3 Potential regulatory functions of Alu elements

In humans the majority of repetitive DNA is made up of LINE-1 and Alu elements [79]. Alu elements require active LINE-1 elements in order to transpose, as the ORF2p protein is essential for Alu retrotransposition [79]. However, the distribution of LINE-1 and Alu elements is distinct; Alu elements are enriched at CG-rich DNA and commonly found in euchromatic regions, while LINE elements have a preference for AT-rich DNA [79, 111]. Interesting, when comparing the distribution of Alu subfamilies, it was determined that younger Alu elements were more frequently found in AT-rich DNA, while older Alu elements were enriched in DNA with a high CG% [79]. CG-rich DNA accumulates around genes and regulatory regions [79]. Alu elements are disproportionately found in intragenic regions, particularly introns, with at least three quarters of human genes thought to be associated with Alu elements [81, 83, 238]. Therefore the presence of Alu elements within genes and regulatory regions has been selected for throughout evolution, suggesting that they may have a regulatory role [79].

### 5.3.1. Genomic Alu elements

Alu elements resemble enhancers; they are enriched in CpGs and contain numerous transcription factor binding sites [84, 238]. The similarity of Alu elements to enhancers increases with evolutionary age from the young, AluYa5, to the old, AluJo, with many middle-aged Alu elements appearing as proto-enhancers [66]. This corresponds with an increasing enrichment of active histone marks and suggests a gradual domestication of Alu elements towards enhancers [66]. Additionally, Alu

elements induce the formation of stable nucleosomes that are positioned in phase around the repeat [239].

### 5.3.2. Embedded Alu RNA

RNA Pol II inadvertently transcribes the majority of Alu elements with an average of 10 embedded Alu elements per pre-mRNA molecule [84, 240]. Inverted Alu repeats can form dsRNA duplexes that are subject to hyper-editing by the ADAR enzymes [240-242]. ADARs deaminate adenosine (A) in RNA to form inosine (I) to aid innate immune sensors in discriminating between host and viral dsRNA [240].

Additionally, in humans, embedded Alu elements are understood to contribute 64% of new exons [82, 243]. Alu exonisation is associated with reduced expression of the corresponding gene as the transcripts can be targeted by nonsense-mediated decay [244].

### 5.3.3. Free Alu RNA

The transcription of independent Alu elements by RNA Pol III can be induced by a variety of stresses such as heat shock, toxins or viruses [242]. These transcripts are mainly Alu S and Alu J and do not encode the sequences required for their own retrotransposition [245]. Instead the Alu RNA molecule binds to RNA Pol II and disrupts the interaction with DNA, slowing transcription [82].

The potential functions of Alu elements are outlined in Figure 5.3.

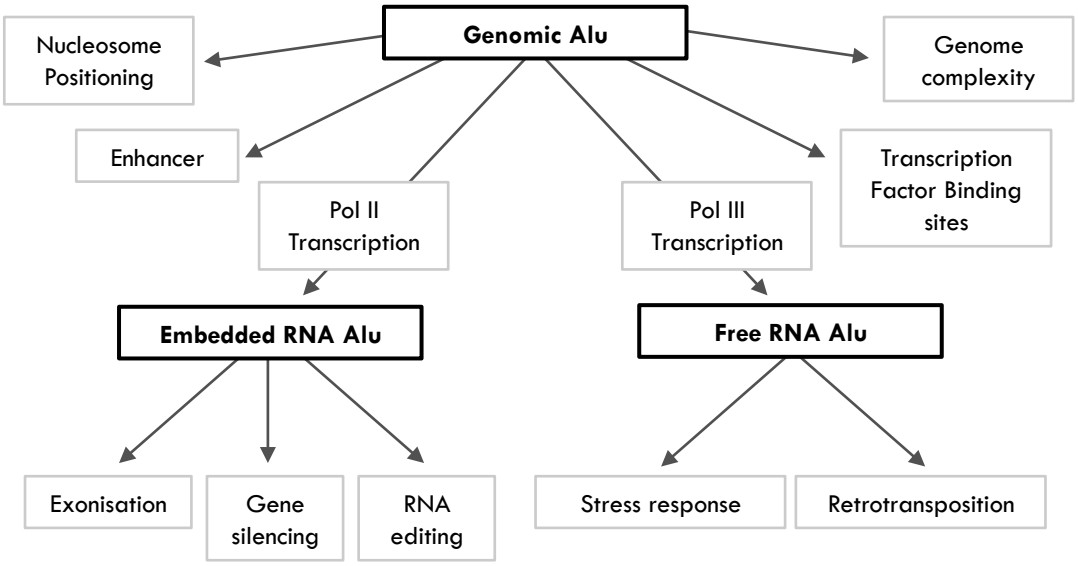



**Figure 5.3. The potential functions of Alu elements**

Alu elements have been proposed to be involved in a number of host mechanisms, either within the host DNA; embedded in mRNA when transcribed by Pol II; or as free Alu elements when transcribed by Pol III.

## 5.4 Aims and Objectives

HNSCC is characterised by genome wide changes in the distribution of DNA methylation and hydroxymethylation [55, 110, 246]. Methylation is gained at promoters and CpG islands, and lost genome wide and over repetitive DNA [11]. However, less is known about 5hmC alterations in HNSCC. In this chapter, MeDIP and hMeDIP sequencing will be used to compare the genomic distribution of 5mC and 5hmC in normal oral keratinocytes (HOK) with an HNSCC cell line. The analysis will focus on the relationship between these modifications and repetitive DNA, particularly Alu elements, and the effect of altered promoter methylation in HNSCC.

## 5.5 Methylated and hydroxymethylated DNA immunoprecipitation (MeDIP and hMeDIP) and sequencing

The MeDIP and hMeDIP methods were used to determine the distribution of 5mC and 5hmC in normal human oral keratinocytes (HOK) and the HNSCC cell line, VU40T. For VU40T cells, experiments were performed and sequenced in duplicate for MeDIP, and triplicate for hMeDIP and the efficacy of each experiment determined by qRT-PCR prior to library preparation (Supplementary data 8). For HOK cells the (h)MeDIP experiments were performed in triplicate and pooled prior to sequencing. Regions enriched for 5mC or 5hmC (peaks) were identified for each cell type and treatment using Homer and MACS for hMeDIP and MeDIP, respectively (Table 5.1). Alignment and peak calling of MeDIP and hMeDIP data can be attributed to Dr W. Croft and Dr S. Clokie respectively. Only uniquely mapped reads were considered for all (h)MeDIP data. For VU40T cells where the replicates were used, the peak should be present in at least one of the repeats to be included. The antibody will only specifically recognise 5(h)mC in the context of CpG dinucleotides; therefore, any peaks without CpGs were removed and only the retained peaks were included in subsequent analyses (Table 5.1). The peaks were then annotated using Homer '*annotatepeaks'*. In order to compare the 5mC or 5hmC distribution between cell types the 'merged peaks' file was created using Homer '*mergepeaks'*. It consists of peaks present in at least one sample and takes into account the overlap between them. To visualize the data in the UCSC genome browser bigwig files were created

for genome coverage representation and bed files for peaks' localization (including merged peaks file).

**Table 5.1. MeDIP and hMeDIP output**

The total number of 5mC and 5hmC enriched regions ('peaks') obtained for each MeDIP and hMeDIP experiment. **First row**: before filtering; **second row**: the number of peaks removed due to lack of CpGs; **third row**: the total number of peaks retained for further analysis; **forth row:** the percentage of peaks retained for further analysis. MeDIP peaks were identified using MACS by Dr W. Croft, while hMeDIP peaks were identified using Homer by Dr S. Clokie.

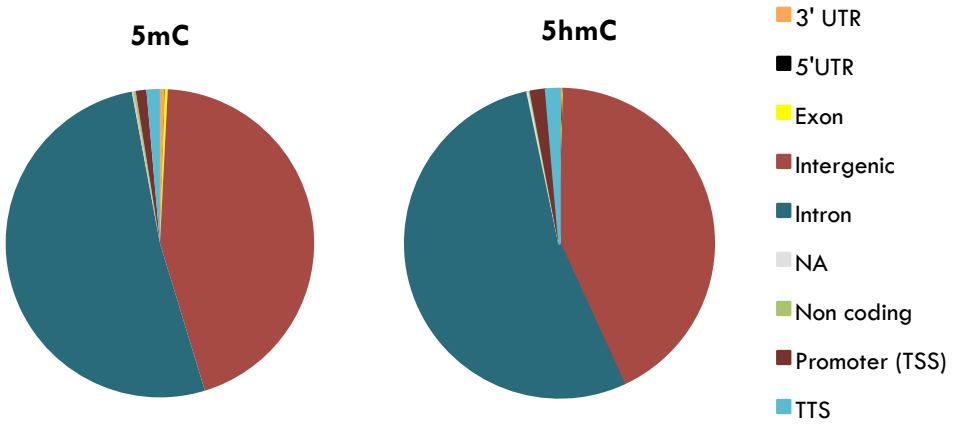
| | **MeDIP** | | **hMeDIP** | |
|---|---|---|---|---|
| | **HOK** | **VU40T** | **HOK** | **VU40T** |
| Total Peaks | 547344 | 512502 | 26678 | 41562 |
| CpG =0 | 1839 | 1822 | 2358 | 10481 |
| No. Peaks retained | 545505 | 510680 | 24320 | 31081 |
| Percentage peaks retained | 99.66 | 99.64 | 91.16 | 74.78 |

## 5.6 DNA hydroxymethylation is found in closer proximity to transcription start sites (TSS) and associated with a higher CpG content than DNA methylation

To determine the distribution of 5mC and 5hmC in normal human oral keratinocytes (HOK cells) the MeDIP and hMeDIP peaks were separated by genomic annotations. The vast majority of both 5mC and 5hmC-enriched regions were found within introns and intergenic regions, promoters only accounted for 1.11% of 5mC peaks and 1.57% of 5hmC peaks and exons represented less than 1% of both (Figure 5.4). In this regard, the distribution of 5mC and 5hmC is very similar. However, 5hmC-enriched regions are found closer to the transcription start site and were associated with a higher CpG content than 5mC-enriched regions (Figure 5.5).

5mC and 5hmC-enriched regions were then aligned to transcription start sites (TSS) and transcription factor binding sites (TFBS) (Figure 5.6). Both 5mC and 5hmC were exempt from the TSS and the centre of transcription factor binding sites and were found at high levels surrounding TFBS' (Figure 5.6). Similarly, both 5mC and 5hmC were underrepresented in CpG islands (CGI), but enriched in the 2kb region either side, termed CGI shores (Figure 5.7). This result was stronger for 5hmC, with 5hmC-enriched regions highly underrepresented in CGI, and highly overrepresented in CGI shores (Figure 5.7).

A.




B.

| | 5mC | | 5hmC | |
|---|---|---|---|---|
| | No. Peaks | % Peaks | No. Peaks | % Peaks |
| **3' UTR** | 2732 | 0.50 | 100 | 0.41 |
| **5' UTR** | 163 | 0.03 | 7 | 0.03 |
| **Exon** | 1543 | 0.28 | 12 | 0.05 |
| **Intergenic** | 242606 | 44.47 | 10381 | 42.69 |
| **Intron** | 282498 | 51.79 | 13011 | 53.51 |
| **NA** | 921 | 0.17 | 55 | 0.23 |
| **non coding** | 1328 | 0.24 | 31 | 0.13 |
| **Promoter-TSS** | 6071 | 1.11 | 382 | 1.57 |
| **TTS** | 7643 | 1.40 | 338 | 1.39 |

**Figure 5.4. The genomic location of 5mC and 5hmC-enriched regions**

MeDIP and hMeDIP sequencing peaks were assigned annotations. MeDIP peaks were identified and assigned using MACS by Dr W. Croft, while hMeDIP peaks were identified and assigned using Homer by Dr S. Clokie.The genomic location of 5mC and 5hmC-enriched regions is shown as **A**. a pie chart and **B**. a table.

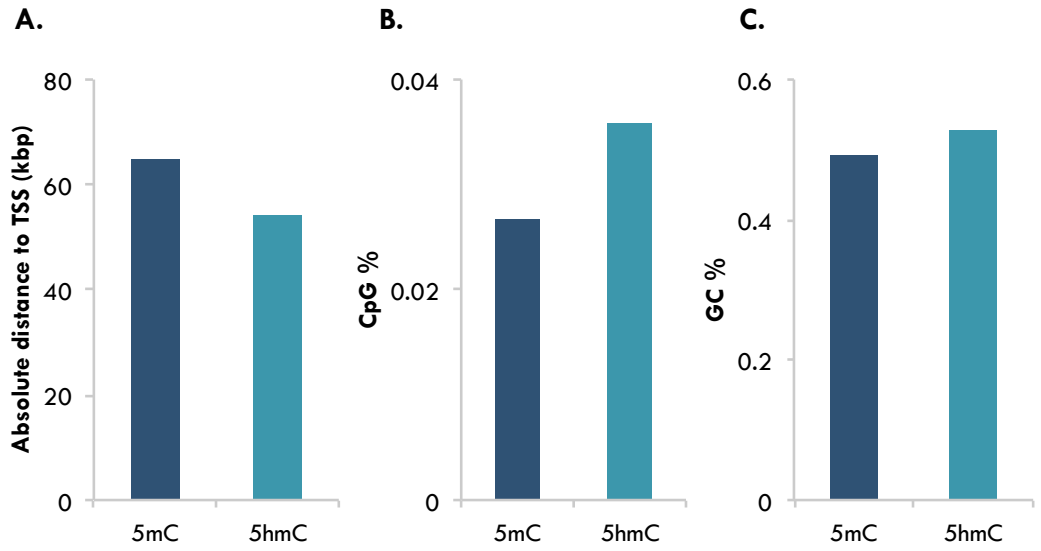



**Figure 5.5. Differential characteristics of 5mC and 5hmC-enriched regions**

Using the peaks defined for each MeDIP and hMeDIP sequencing samples the mean absolute distance to the transcription start site (TSS) (**A**.), CpG content (**B.**) and GC% (**C.**) were calculated.

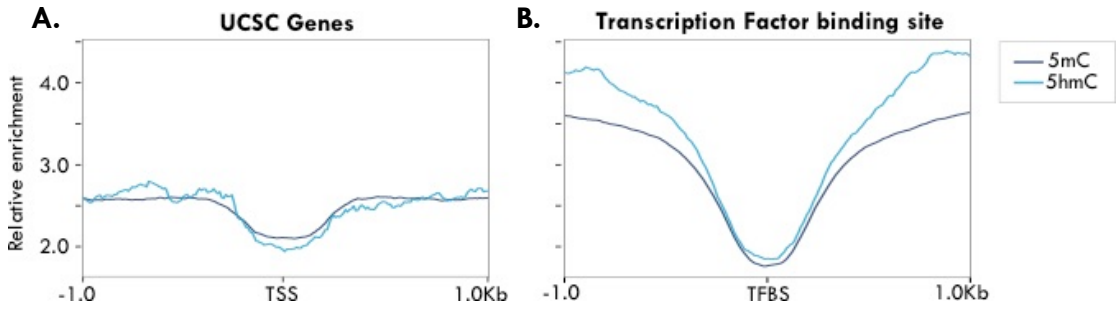



**Figure 5.6. Relationship between 5mC and 5hmC and transcription start sites and transcription factor binding sites**

The results of the MeDIP and hMeDIP sequencing experiments were aligned to all UCSC genes (**A.**) and transcription factor binding sites determined by the ENCODE consortium (**B.**) [247]. The computeMatrix and PlotProfile tools (Galaxy Versions 2.5.0.0) were used to create the graphs from the genomic annotations of all MeDIP and hMeDIP peaks using reference point scaling of 1kb either side [74].

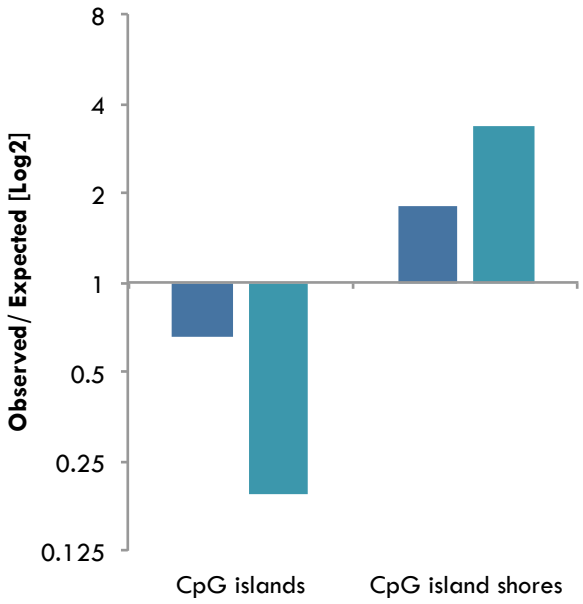



**Figure 5.7. 5mC and 5hmC are underrepresented at CpG islands but enriched in CpG island shores**

CpG islands (CGI) were downloaded from USCS table browser. Using BedTools package on galaxy CGI shores were generated by using the '*Slop*' function to extend CGIs by +/- 2kb followed by the '*Subtract*' tool to remove the CGI from the centre of these regions. The '*intersect intervals*' function was the used to determine the proportion of 5mC/ 5hmC-enriched regions (peaks) that overlapped (more than 80% of peak) with the CGI or CGI shores [175]. BEDTools '*RandomBed*' was used to create a random group of genomic intervals comparable with each 5(h)mC group. This was used to generate the expected number of overlaps. For each group the total number of overlaps (observed) is shown relative to corresponding expected group.

## 5.7 5hmC-enriched regions are associated with enhancers

5mC and 5hmC-enriched regions were investigated for overlap with functional chromatin domains of primary normal human skin epidermal keratinocytes (NHEK) obtained from the UCSC table browser [172]. The chromatin domains consist of 14 classifications determined from ChIP sequencing of CTCF and histone modifications, using a multivariate Hidden Markov Model (HMM) to model combinatorial chromatin patterns [172]. These can be combined into 6 main groups: promoters, enhancers, insulators, transcribed regions, repressed regions and inactive regions [172]. Both 5mC and 5hmC were enriched in transcribed regions and underrepresented in promoters (Figure 5.8). Furthermore, while both 5mC and 5hmC were most frequently found in inactive regions (Supplementary data 9), both were observed at a lower frequency than expected. This could be due to the lower CpG content of these regions [33-35]. In keeping with the literature, 5hmC-enriched regions were associated with enhancers, while 5mC-enriched regions were not (Figure 5.8)[52, 61].

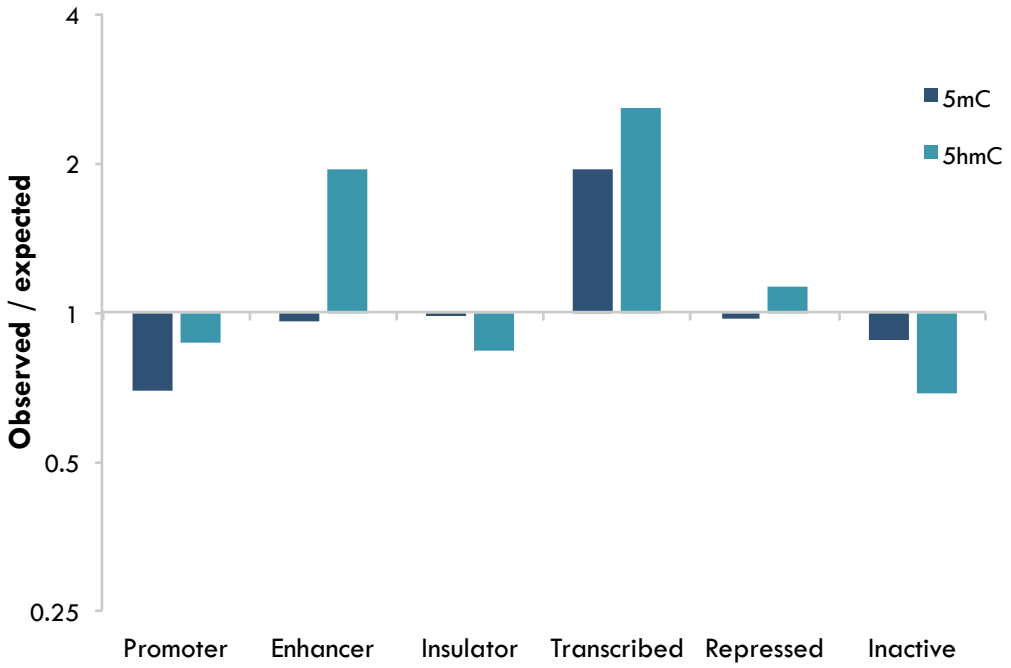



**Figure 5.8. Chromatin domains associated with 5mC and 5hmC-enriched regions**

Chromatin domains were downloaded from USCS table browser [172]. Using the BEDTools package on Galaxy the '*intersect intervals'* function was used to determine the proportion of each 5mC or 5hmC peaks that overlapped with each chromatin domain [175]. BEDTools '*RandomBed'* was used to create a random group of genomic intervals comparable with the 5mC and 5hmC data. This was used to generate the expected number of overlaps. For each group the total number of overlaps (observed) is shown relative to corresponding expected group, where anything <1 is underrepresented.

## 5.8 Regions enriched in DNA methylation and hydroxymethylation were primarily found at Alu elements

The results of the HOK MeDIP and hMeDIP experiments were separated based on detailed annotations to investigate for the presence of repetitive DNA. In HOK cells 94% of 5mC-enriched regions and 92% of 5hmC-enriched regions contain repetitive DNA and the vast majority of these are Alu elements (89% of 5mC and 77% of 5hmC peaks) (Figure 5.9). This value is considerably higher than the 10-11% of the human genome occupied by Alu (Figure 1.13). Alu elements can be split into three families: Alu J, the oldest; Alu S, middle aged but most frequent; and Alu Y, the youngest and most active [81]. When compared with the frequencies in the human genome, Alu J were underrepresented and Alu Y overrepresented in both the MeDIP and hMeDIP data (Figure 5.10). Alu Y was particularly enriched in 5hmC and Alu S was overrepresented in the MeDIP data, suggesting that while both modifications preferentially mark younger Alu elements, the control of the Alu S is primarily attributed to 5mC (Figure 5.10) [83].

DNA hydroxymethylation reads were aligned to Alu element subfamilies and clustered using the *hclust* method (R package) (Figure 5.11). A distinct band of 5hmC can be seen covering all Alu S elements and the majority of Alu Y elements. This is less distinct for Alu J. A small number of human-specific Alu Y elements lack this band, however these are underrepresented in the genome (Figure 5.11, Supplementary data 10).

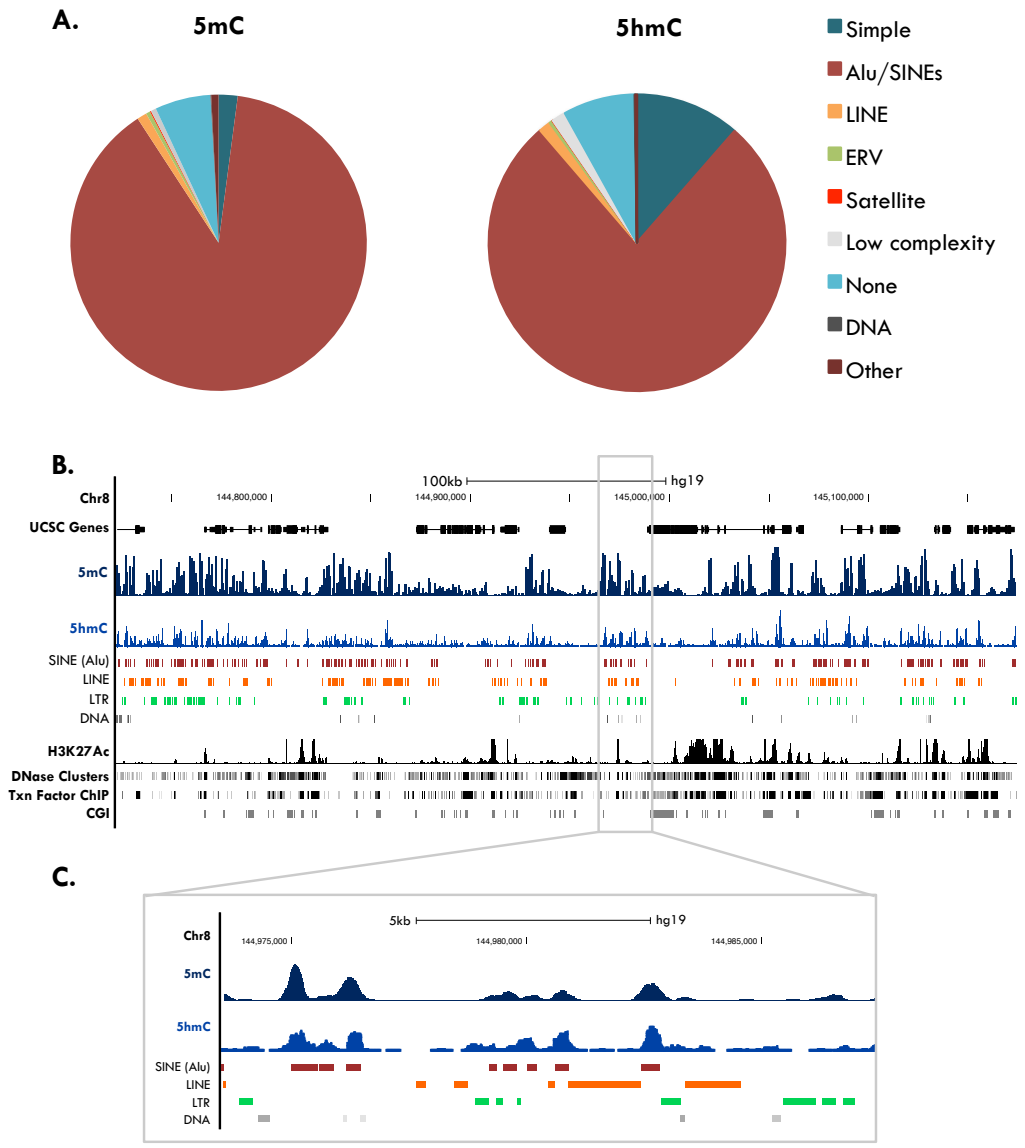



**Figure 5.9. MeDIP and hMeDIP peaks at repetitive DNA**

**A.** HOK MeDIP and hMeDIP peaks were separated based on detailed annotations and are shown as a pie chart with the number and percentage of total peaks shown alongside. MeDIP peaks were identified and detailed annotations assigned using MACS by Dr W. Croft while Homer was used by Dr S. Clokie for the hMeDIP data. **B.** UCSC browser shot representing an example of HOK MeDIP (dark blue) and hMeDIP (light blue) sequencing in relation to the location of repetitive elements and genes. **C.** Zoomed in image of 5hmC distribution at one cluster of Alu elements illustrates an overlap between 5hmC and Alu elements (red bars) but not other repetitive elements. UCSC browser shot of the section highlighted in **B.** The results are representative of two experiments combined prior to sequencing.

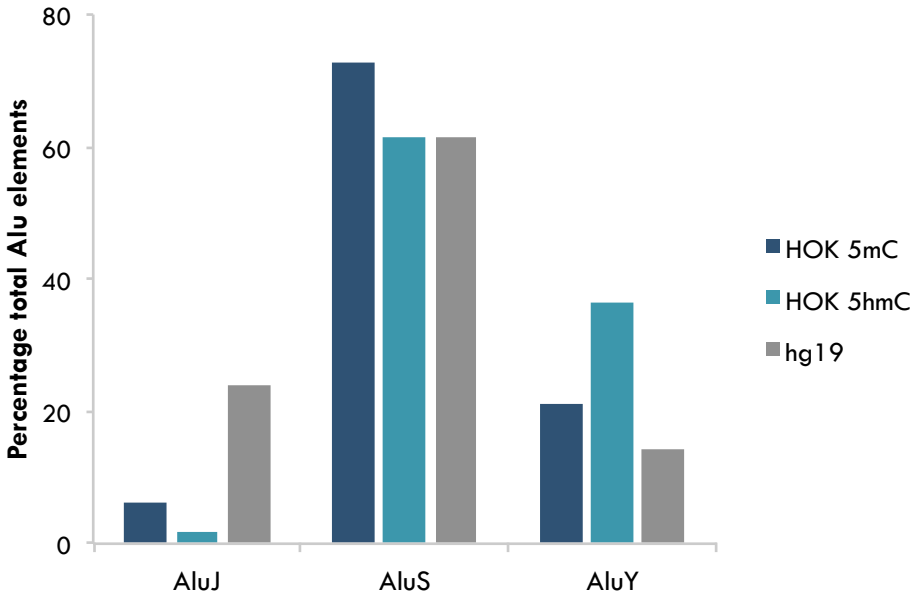



**Figure 5.10. Differential enrichment of 5mC and 5hmC in Alu element families**

The results of MeDIP and hMeDIP sequencing of HOK cells were separated based on Alu element families (J, S and Y) and are shown relative to the total number of Alu element containing 5hmC-enriched regions. The proportion of each group in the human genome as determined by Grover et al., 2004 [83] is shown alongside (hg19).

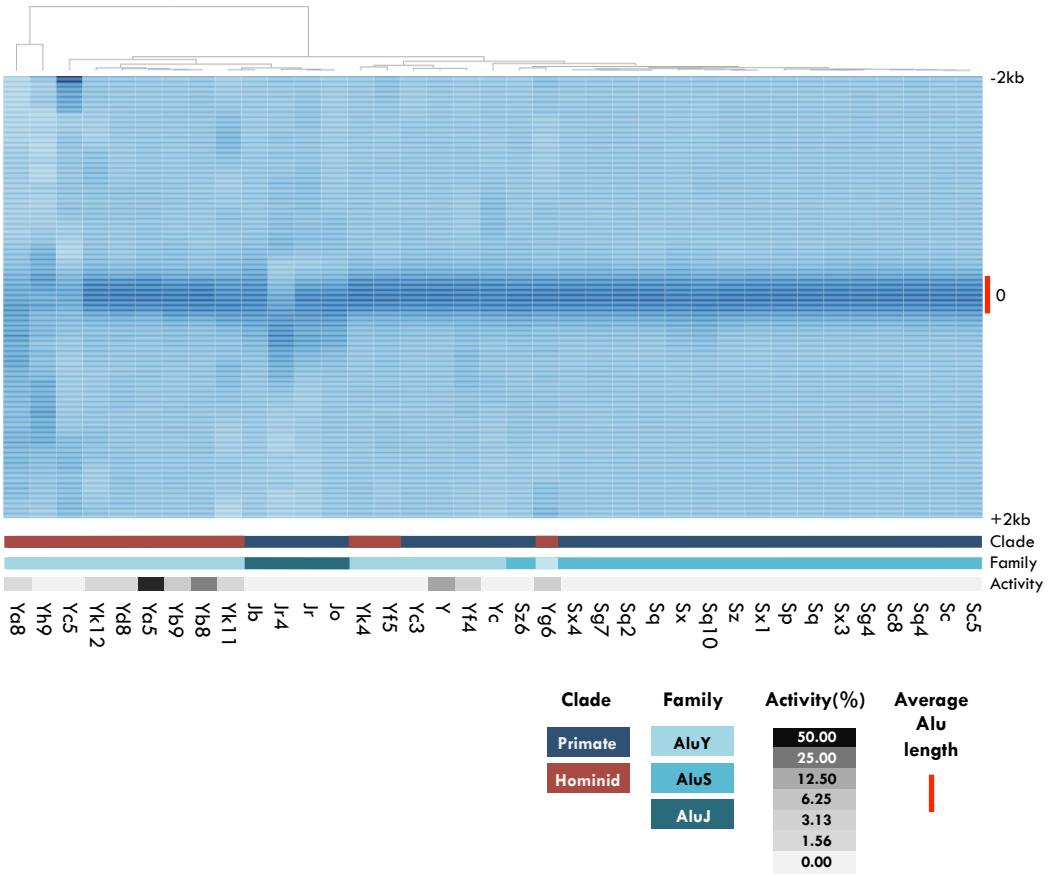



**Figure 5.11. Enrichment of 5hmC at Alu element subfamilies**

For each Alu element subfamily the 5hmC count from HOK hMeDIP sequencing is shown +/- 2kb of the centre of the repetitive element. hMeDIP reads were counted over each Alu element subfamily using Homer's histogram tool and are scaled to each 4kb region as discussed in section 2.14 [170]. The heatmap was generated using the d3heatmap R library and clustered by Alu subfamily. The average size of an Alu element is shown as a red bar at the side. The clade, family and activity of each element are shown below, where the clade and family were determined from Dfam [248] and activity from Konkel et al., 2015 [249]. Heatmap courtasy of Dr S. Clokie.

## 5.9 DNA methylation and hydroxymethylation at Alu elements occurs alongside but does not overlap regulatory elements.

Upon examining the genome browser it became apparent that in gene-rich regions Alu elements appear as clusters in areas adjacent to regulatory elements and are consistently marked with 5mC and 5hmC (Figure 5.12, upper panel). In gene-poor regions, Alu elements are more dispersed and lack 5hmC but partially retain 5mC (Figure 5.12, lower panel).

To further investigate the relationship between DNA modifications and Alu elements, four groups of regions were isolated for both 5mC and 5hmC: 5mC and 5hmC-enriched regions were analysed as i) all 5(h)mC (545,507 and 24,318 regions respectively); ii) 5(h)mC at Alu elements (483,059 and 18,749 regions) and; iii) 5(h)mC not at Alu elements (62,448 and 5,569 regions); (iv) the genomic locations of all Alu elements were downloaded from repeatmasker [250] and non-5mC and non-5hmC marked Alu elements were isolated (1,010,770 and 1,025,616 regions respectively). These were aligned to UCSC genes. Alu elements marked with 5mC and 5hmC were strongly excluded from the TSS (Figure 5.13). Interestingly, the drop in 5hmC at Alu elements did not directly align with the TSS; instead it was found slightly to the right (Figure 5.13). This was distinct for 5hmC over 5mC, suggesting that 5hmC at Alu elements is more commonly found upstream of the TSS, while 5mC is evenly distributed around it (Figure 5.13). The 5hmC in HOK cells was compared with ENCODE DNase hypersensitive sites (DHSs) from 125 cell lines; again 5hmC was found close to, but rarely overlapping with DHSs and 5hmC marked Alu elements

were excluded from the centre of the DHS (Figure 5.14, Figure 5.15). Of note, Alu Y elements are highly hydroxymethylated surrounding, but not overlapping the DHS sites (Figure 5.15). Regions enriched in 5mC or 5hmC, both within and outwith Alu elements were exempt from transcription factor binding sites (Figure 5.16). These results correspond to the distribution of 5mC and 5hmC observed in Figure 5.12; 5mC and 5hmC at Alu elements is found in close proximity, but not overlapping, regulatory elements. Therefore, the CpG modifications may function in regulating Alu elements in gene-rich regions.

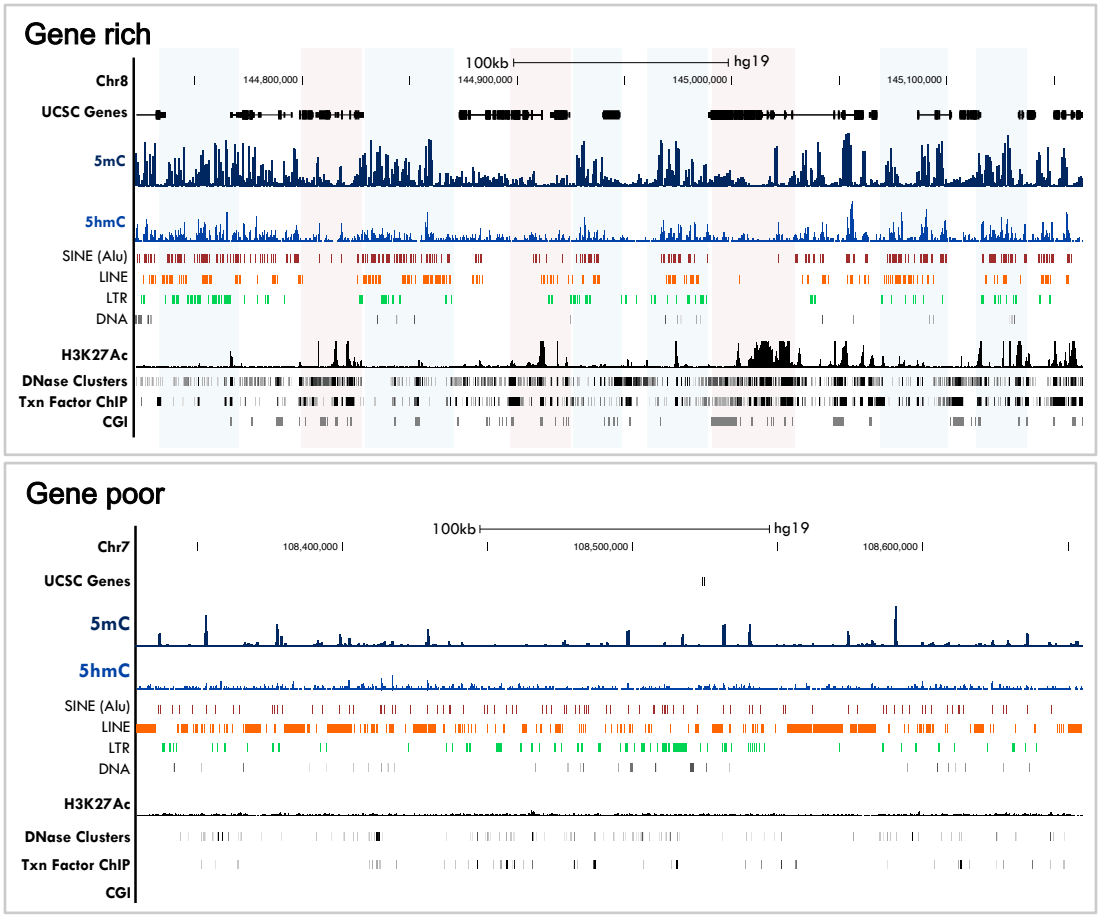



**Figure 5.12. 5mC and 5hmC mark Alu elements in gene-rich regions**

UCSC browser shots of MeDIP (dark blue track) and hMeDIP sequencing (light blue track) of HOK cells shown alongside repeatmasker tracks for the main classes of transposon (SINEs, LINEs, LTRs and DNA transposons), UCSC genes, CpG islands and tracks for transcription (Txn) factor ChIP, DNase clusters and layered H3K27Ac ChIP from all ENCODE cell lines. In the **top** panel pale blue bands highlight areas rich in Alu elements and 5mC and 5hmC, but poor in regulatory elements while pale pink bands represent areas rich in regulatory elements, but poor in 5hmC and Alu elements. In the **bottom** panel Alu elements in gene poor regions are more dispersed and marked only partially with 5mC.

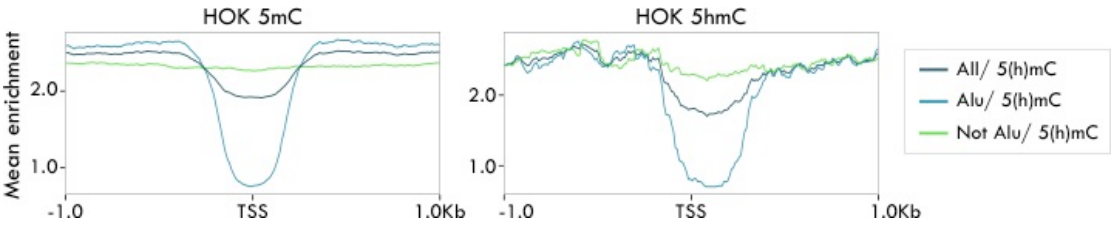



**Figure 5.13. 5mC and 5hmC at Alu elements were excluded from the TSS**

HOK 5mC and 5hmC-enriched regions (peaks) were separated based on the presence or absence of an Alu element. These were then aligned to the TSS of UCSC known genes using the computeMatrix and PlotProfile tools on Galaxy [176].

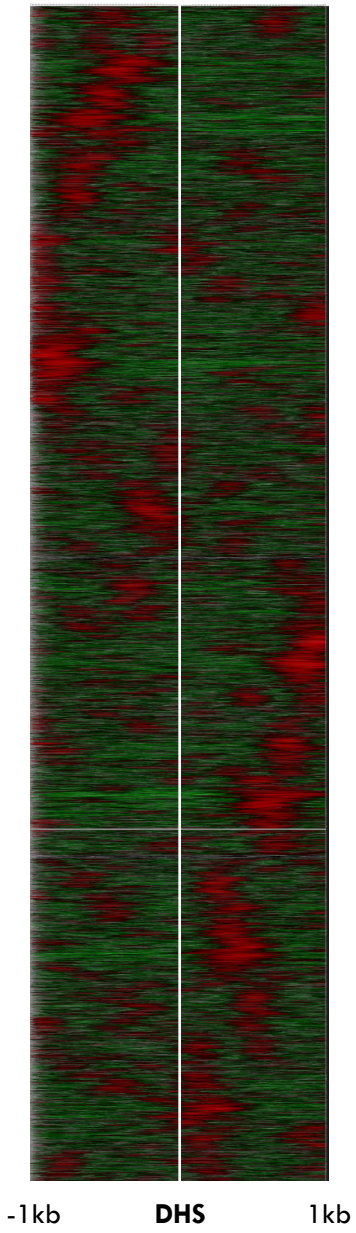



**Figure 5.14. 5hmC at Alu elements at DNase hypersensitive sites**

The hMeDIP reads that overlap with Alu elements downloaded from UCSC repeat masker were visualised using Homers' Cluster3, '*TreeView'* function. 5hmC reads were log transformed and centred on the subsampled (28,907) entries from WgEncodeAwgDnaseMasterSites list (DHS). Hierarchical clustering was performed based on correlation similarity and centroid linkage. Homers' '*-size given'* function was used to independently scale each 2kb region by separating the region into bins of 10bp. For each region, the colour represents the number of reads per bin relative to the total number of reads across the whole 2kb region, with red indicating an enrichment of 5hmC. Heatmap generated by Dr S. Clokie.

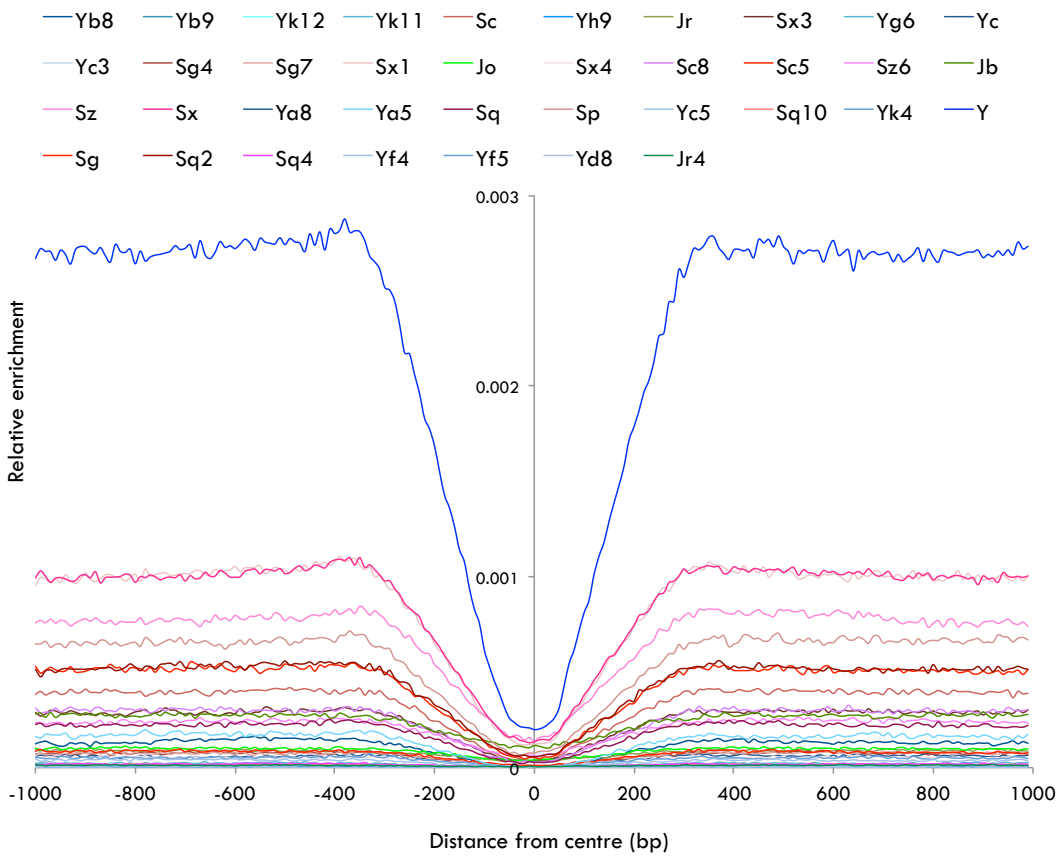



**Figure 5.15. 5hmC at Alu elements is excluded from DNase hypersensitive sites**

HOK hMeDIP reads that intersect with Alu elements (downloaded from UCSC repeat masker) were centred on Encode DNase I hypersensitive peaks (WgEncodeAwgDnaseMasterSites) to generate a histogram using the '*annotatepeaks*' program from Homer. For each Alu element subfamily the histogram is scaled within each 2kb region by separating the region into bins of 10bp and counting the total number of reads per bin relative to the total number of reads within the region. Analysis courtesy of Dr S. Clokie.

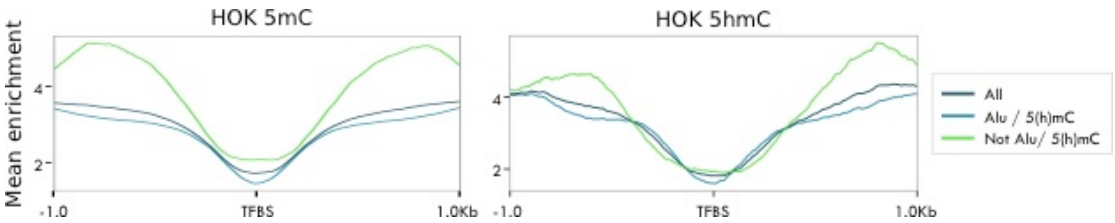



**Figure 5.16. 5mC and 5hmC enriched regions are excluded from transcription factor binding sites**

5mC and 5hmC enriched regions (peaks) were separated based on the presence or absence of an Alu element. The genomic coordinates for these groups were then used to align the peaks with the centre of transcription factor binding sites determined by the ENCODE consortium [247]. The computeMatrix and PlotProfile tools were used to create the graphs, using the 'reference point' mode to only count peaks 1kb either side of the TFBSs [176].

## 5.10 The function of 5mC and 5hmC may differ within and outside of Alu elements

DNA methylation/ hydroxymethylation/ Alu element groups from section 5.9 were investigated for overlap with functional chromatin domains of NHEK cells (Figure 5.17, Supplementary data 13) [172]. Heterochromatin is represented as expected at Alu elements without 5mC or 5hmC (5mC and 5hmC O/E=0.95) and underrepresented in regions marked by 5hmC (O/E: All=0.67; Alu=0.66; Not Alu=0.74) (Figure 5.17). Alu elements, 5mC and 5hmC are enriched in transcribed regions, with 5hmC marked Alu elements showing a higher association than the other groups (Figure 5.17). As discussed previously, 5hmC are enriched in enhancers while 5mC (and unmarked Alu elements) are not (Figure 5.17). Furthermore, distinct from regions containing Alu elements, non-Alu 5mC-enriched regions are associated with all promoter groups and non-Alu 5mC and 5hmC groups occupy insulators, polycomb-repressed regions and poised promoters (Figure 5.17). Therefore, the roles of 5mC and 5hmC may differ based on the presence or absence of an Alu element.

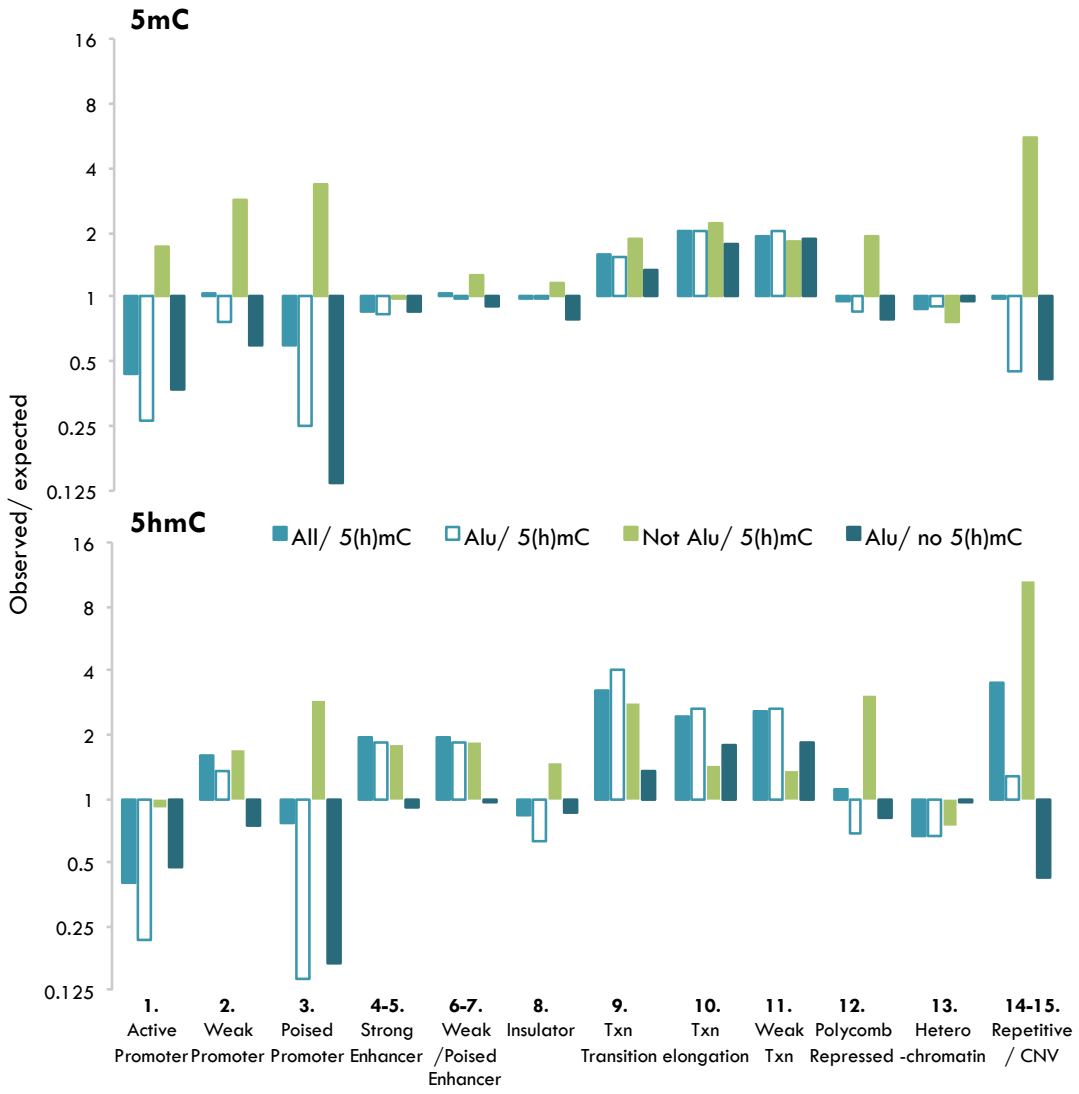



**Figure 5.17. Chromatin domains associated with 5mC and 5hmC at Alu elements**

HOK MeDIP and hMeDIP peaks were separated based on whether they overlap with an Alu element (Alu 5(h)mC) or not (not Alu 5(h)mC). Repeatmasker was used to download all genomic Alu elements and those not overlapping with 5(h)mC were retained for the 'Alu no 5(h)mC' group [251]. NHEK Chromatin domains were downloaded from USCS table browser [172]. Using the BEDTools package on Galaxy the '*intersect intervals*' function was used to determine the proportion of each Alu/5(h)mC group that overlapped with each chromatin domain [175]. BEDTools '*RandomBed*' was used to create a random group of genomic intervals comparable with each 5(h)mC/ Alu group. This was used to generate the 'expected' number of overlaps. For each group the total number of overlaps (observed) is shown relative to corresponding expected group.

## 5.11 5hmC marked Alu elements are found in regions associated with active histone modifications

Using the BedTools '*intersect intervals*' function [175] the overlap between 5mC, 5hmC and Alu groups with histone modifications from NHEK cells was determined (Figure 5.18, Supplementary data 14) [174]. Alu elements and 5hmC were found to be associated with active histone modifications (H3K4me3, H3K36me3, H3K4me1, H3K27me3) more than repressive ones (H3K27me3, H3K9me3) (Figure 5.18, Figure 5.19). Indeed, for all bar the non-Alu 5hmC group, repressive histone modifications were underrepresented in the 5hmC data (observed/expected<1) (Figure 5.18B). Total 5mC was also associated with H3K36me3 and underrepresented in H3K9me3. The inverse relationship between 5mC and H3K9me3 is distinct from the literature, and may simply represent the difference in CpG content between euchromatin and heterochromatin [33, 34]. H3K36me3 is linked to transcriptional elongation and was found to accompany 5mC, 5hmC and Alu elements (Figure 5.18). The observed/expected ratio was higher for Alu elements marked with 5hmC than either Alu or 5hmC alone (Figure 5.18). Similarly, when viewed across the whole chromosome (using chr8 as an example) 5hmC at Alu elements co-localises with active histone modifications but is exempt from regions of H3K9me3 (Figure 5.19).

DNA methylation and hydroxymethylation groups were aligned with NHEK histone modification peaks to determine how these regions overlap (Figure 5.20). In contrast to the results described above (Figure 5.18), 5hmC at Alu was exempt from the centre of most active histone modification peaks and 5hmC outside of Alu co-localised with the centre of H3K9me3 peaks (Figure 5.20). Similarly, 5mC was exempt

from the centre of active histone modifications, and co-localised with the heterochromatin mark H3K9me3 only outside of Alu elements (Figure 5.20).

In conclusion, upon examining the UCSC genome browser, clusters of 5mC and 5hmC-marked Alu elements tend to be located within or besides genes; and are associated with, but upon closer investigation not overlapping, active transcription start sites and regulatory elements (Figure 5.21). This suggests that Alu elements have a function in gene rich regions, and are controlled by DNA modifications.

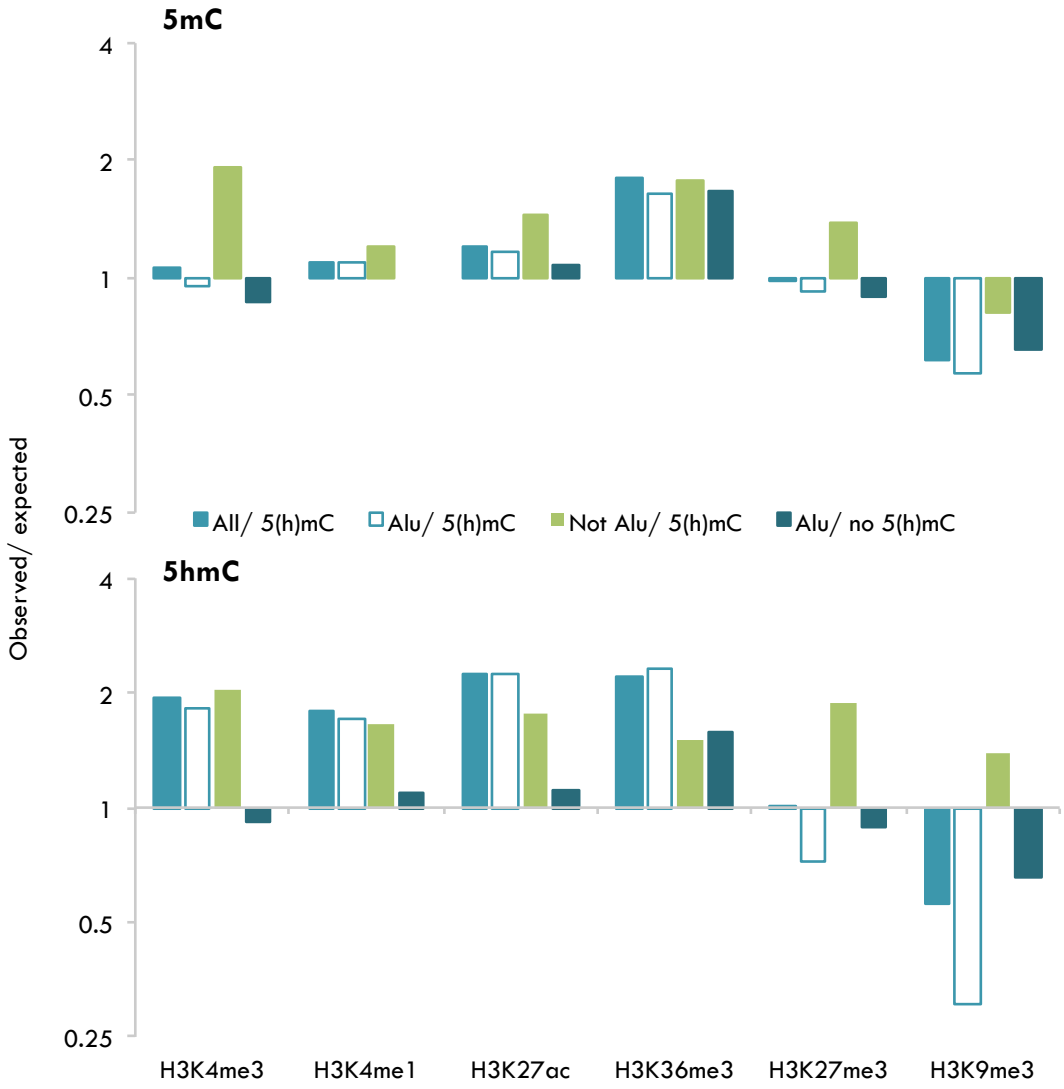



**Figure 5.18. Histone modifications associated with 5mC and 5hmC-enriched Alu elements**

MeDIP and hMeDIP peaks were separated depending on whether they overlap with an Alu element; likewise, Alu elements which did not overlap with a MeDIP or hMeDIP peak were determined. BEDTools '*intersect intervals*' function was used to determine the proportion of HOK 5mC, 5hmC and Alu regions that overlap with histone modifications from ENCODE ChIP data of NHEK cells [175]. The overlap of 5(h)mC /Alu groups is shown relative to that of a corresponding RandomBed file.

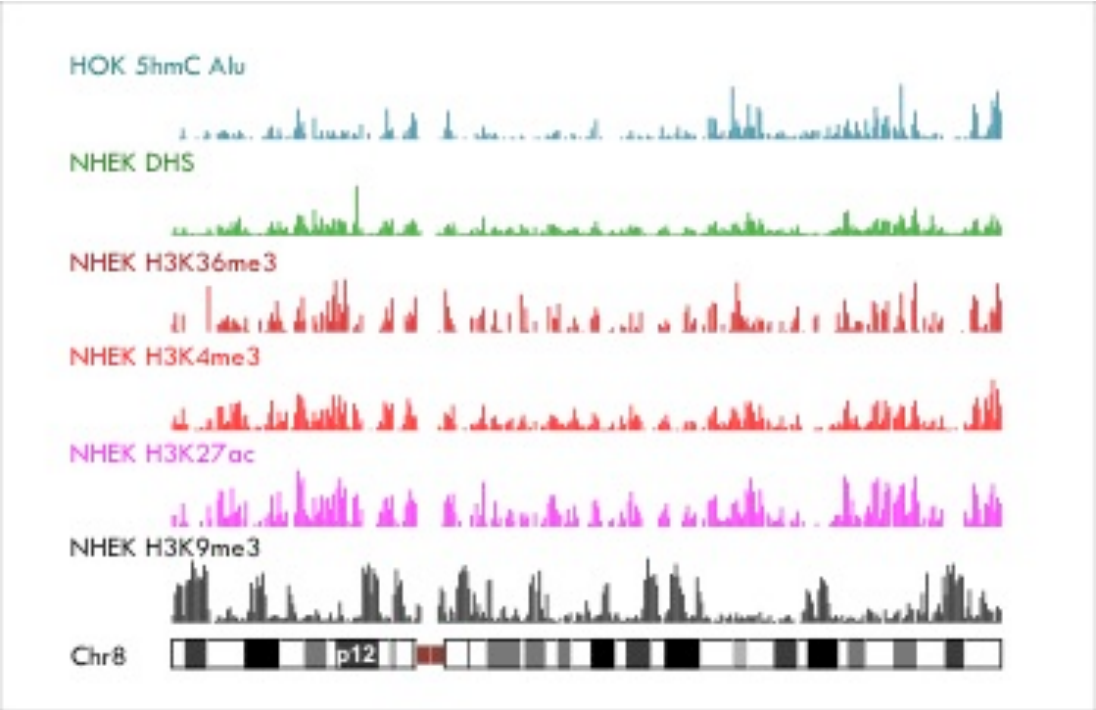



**Figure 5.19. Example of co-localisation of 5hmC-marked Alu elements and active histone modifications**

Chromosome 8 is used as a representative example to show the distribution of 5hmC peaks at Alu elements (blue) in relation to NHEK DNase hypersensitive sites (DHS green), NHEK H3K36me3 (dark red), NHEK H3K4me3 (red), NHEK H3K27ac (pink) and NHEK H3K9me3 (grey). Data was visualised using the Roadmaps epigenome project and WASHU Epigenome Browser.

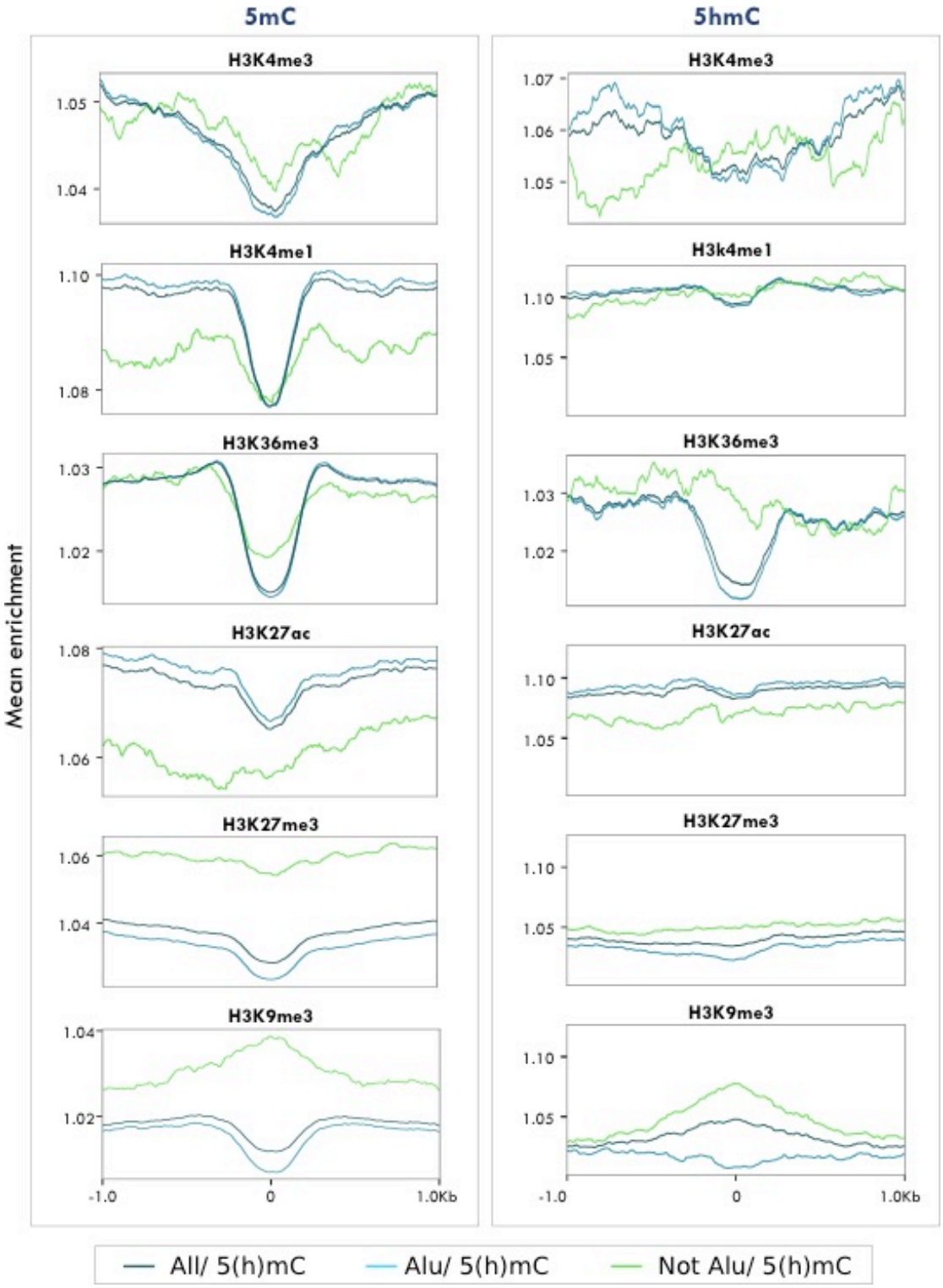



**Figure 5.20. The relationship between 5mC and 5hmC-enriched regions and histone modification peaks**

5mC and 5hmC-enriched regions were separated based on the presence or absence of an Alu element. These were aligned to NHEK histone modification peaks determined by ENCODE using the '*computeMatrix*' and '*PlotProfile*' tool on Galaxy [176]. Regions were separated into bins of 50bp and scaled relative to total peaks per region (mean enrichment).

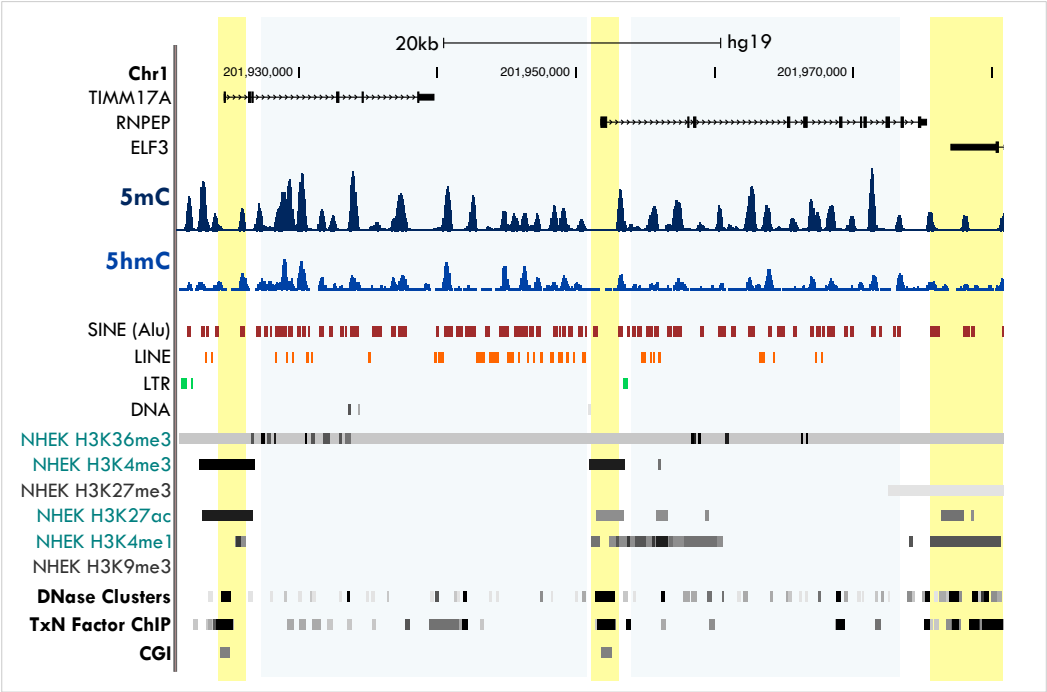



**Figure 5.21. 5mC and 5hmC at Alu elements occurs close to regulatory regions**

UCSC browser shot of HOK MeDIP (dark blue) hMeDIP sequencing (light blue) in relation to common repetitive elements and NHEK histone modification and regulatory regions. Yellow bars show active transcription start sites, enriched in H3K4me3, H3K27ac, TFBS (TxN Factor ChIP), DNase clusters and CpG islands (CGI). Regions enriched in 5mC, 5hmC and Alu elements (highlighted in blue), are found alongside, but rarely overlapping these active sites.

## 5.12 DNA methylation and hydroxymethylation outside of Alu elements are associated with bivalent chromatin

Enhancers and promoters that are 'poised' are occupied by both active (H3K4me3) and repressive (H3K27me3) histone modifications, referred to as bivalent chromatin [64]. In ESCs 5hmC has also been found at these sites [64]. Interestingly, in HOK cells, outside of Alu elements both 5mC and 5hmC associate with the active histone modification H3K4me3 in NHEK cells (Figure 5.18). This may represent differences in activation or repression of genes between HOK cells and NHEK cells. In HOK cells, both 5mC and 5hmC not associated with Alu elements are more commonly found within bivalent chromatin, than within fully active chromatin (containing both H3K4me3 and H3K27ac) (Figure 5.22).

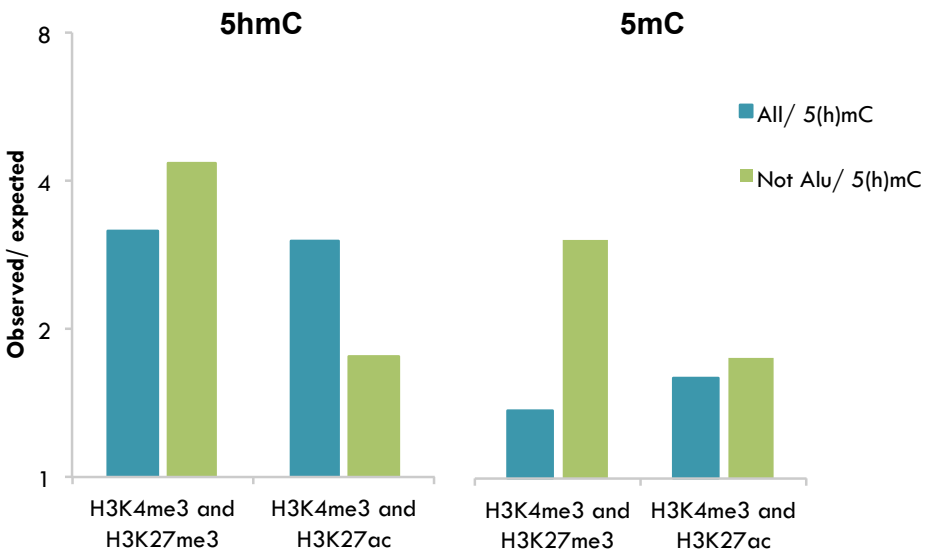



**Figure 5.22. 5mC and 5hmC and bivalent histone modifications**

BEDTools '*intersect intervals'* function [175] was used to determine whether 5mC, 5hmC not Alu groups that overlapped with the active histone modification H3K4me3 also overlapped with the repressive modification H3K27me3 or the active modification H3K27ac.

## 5.13 5hmC is associated with tissue specific enhancers

A subset of 5hmC (2890 regions) was associated with enhancers (Figure 5.17). The online software GREAT was used to assign genes to genomic co-ordinates of 5hmC-enriched regions that overlapped with the 'strong enhancer' chromatin domains and perform gene ontology analysis [177]. 5hmC-enriched enhancers were associated with disease ontology terms associated with epithelial cancers, including HNSCC (Figure 5.23). Furthermore, 5hmC-enriched enhancers were associated with molecular function and cellular component ontological groupings related to epithelial cell structure and movement (Figure 5.23).

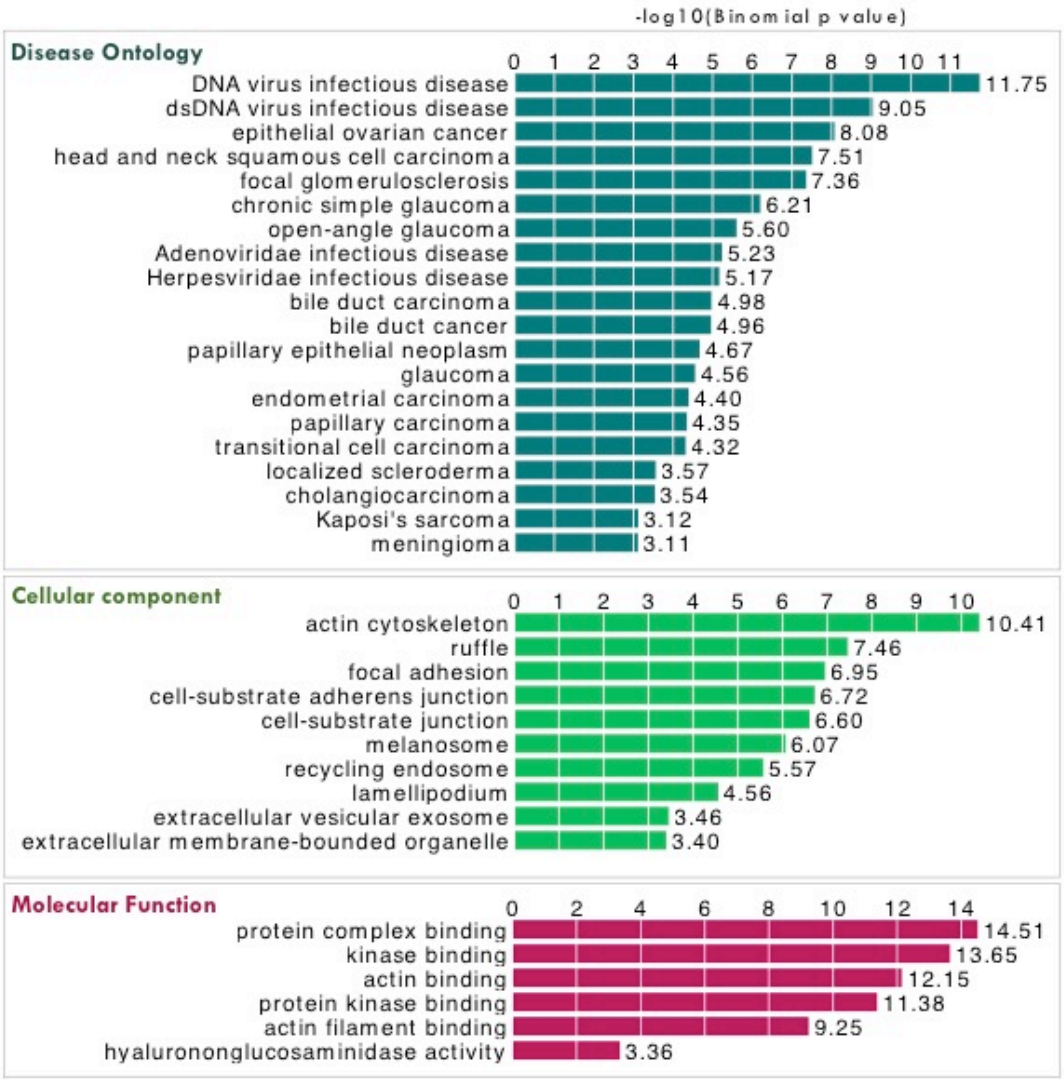


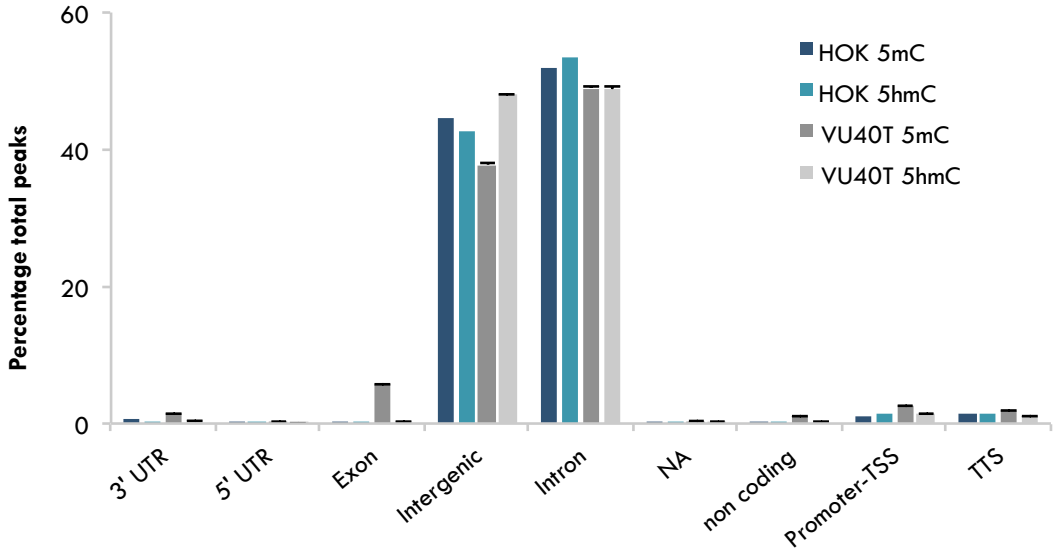

**Figure 5.23. Gene ontology analysis of 5hmC-enriched enhancers**

HOK 5hmC-enriched regions that overlap with the chromatin domain 'strong enhancers' were assigned genes and gene ontology performed using the online software GREAT human genome as a background sequence [177]. Significantly associated terms for disease ontology, cellular component and molecular function are shown.

## 5.14 DNA methylation is gained at promoters and CpG islands in an HNSCC cell line

DNA methylation and hydroxymethylation are disrupted in HNSCC [57, 99, 100, 102, 110]. To investigate the relationship between cytosine modifications and HNSCC, MeDIP and hMeDIP sequencing was performed on the HNSCC cell line, VU40T, and the results compared with the HOK (h)MeDIP results. In VU40T cells, 5mC-enriched regions were found at decreased frequency in intergenic regions, and more abundantly in promoters and exons than in HOK cells (Figure 5.24). Correspondingly, 5mC-enriched regions overlapped with CpG islands more in HNSCC than in HOK cells, with 5mC underrepresented in CGIs in HOK cells and enriched in CGIs in VU40T cells (Figure 5.25). 5hmC-enriched regions in VU40T were found less in introns and more in intergenic regions than in HOK cells (Figure 5.24). Furthermore, while 5mC-enriched regions were found closer to the TSS, associated with a higher CpG content in VU40T cells compared with HOK cells, the opposite was observed for 5hmC (Figure 5.26).

A.




B.

| | 5mC | | | | 5hmC | | | |
|---|---|---|---|---|---|---|---|---|
| | VU40T | | HOK | | VU40T | | HOK | |
| | No. Peaks | % | No. Peaks | % | No. Peaks | % | No. Peaks | % |
| **Promoter (TSS)** | 12492 | 2.45 | 6071 | 1.11 | 417 | 1.35 | 382 | 1.57 |
| **Exon** | 29321 | 5.74 | 1543 | 0.28 | 69 | 0.22 | 12 | 0.05 |
| **TTS** | 9685 | 1.90 | 7643 | 1.40 | 335 | 1.08 | 338 | 1.39 |
| **Intron** | 250776 | 49.11 | 282498 | 51.79 | 15135 | 48.73 | 13011 | 53.51 |
| **Intergenic** | 192928 | 37.78 | 242606 | 44.47 | 14884 | 47.87 | 10381 | 42.69 |
| **Other** | 15478 | 3.03 | 5144 | 0.94 | 232 | 0.75 | 193 | 0.79 |

**Figure 5.24. The distribution of 5mC and 5hmC in HNSCC cells**

MeDIP and hMeDIP was performed on HNSCC cells (VU40T) and compared with normal human oral keratinocytes (HOK). The results were separated based on the 'annotation' assigned to each peak and are shown as **A.** pie charts and **B.** a table. VU40T results are the mean of three independently sequenced and analysed hMeDIP experiments and two MeDIP experiments, while HOK cells represent two (h)MeDIP experiments that were combined prior to sequencing. MeDIP peaks were identified and annotations assigned using MACS by Dr Wayne Croft and hMeDIP peaks were identified and annotations assigned using Homer by Dr Sam Clokie. Error bars represent STD.

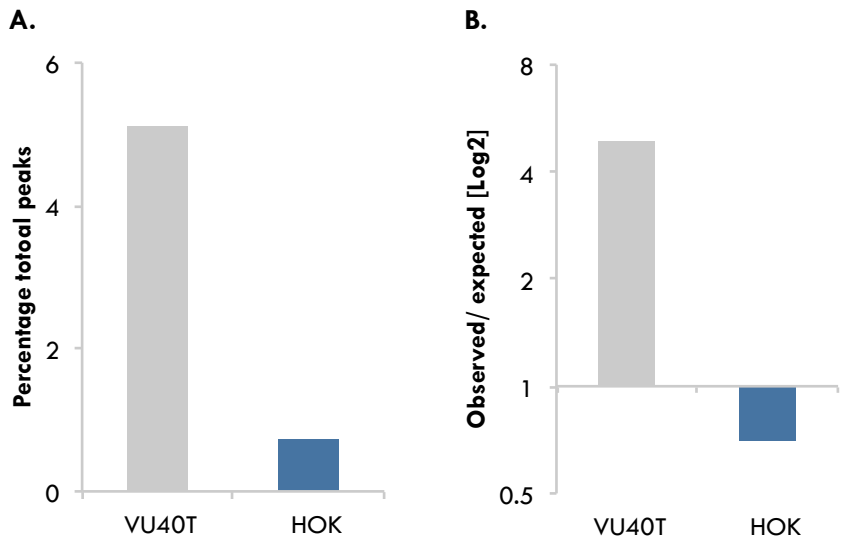



**Figure 5.25. 5mC is gained in CpG islands in HNSCC**

The UCSC table browser was used to download CpG islands onto Galaxy. Using the Bedtools '*intersect intervals*' function, the amount of overlap between VU40T MeDIP and HOK MeDIP peaks with these regions was determined. The results are shown as **A.** the percentage of total MeDIP peaks and **B.** normalised to a corresponding RandomBed file.

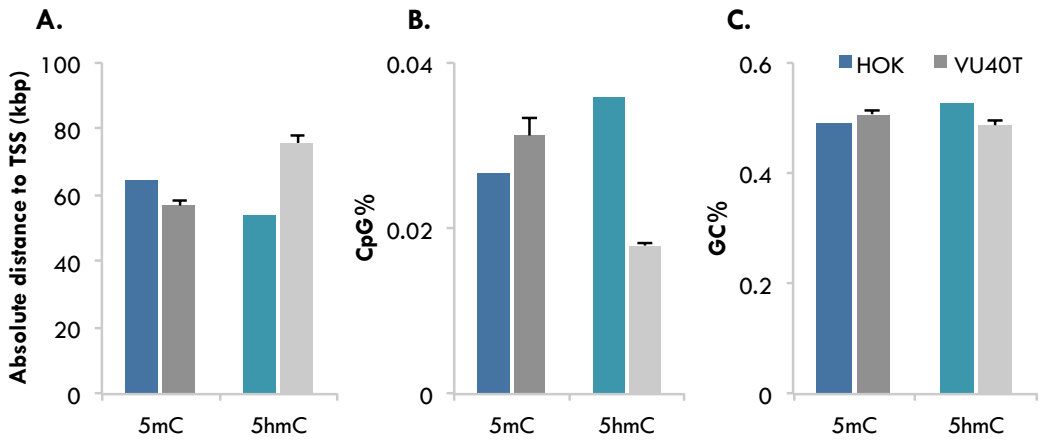



**Figure 5.26. Characteristics of 5mC and 5hmC-enriched regions in VU40T cells**

For peaks obtained from the HOK and VU40T MeDIP and hMeDIP sequencing the mean absolute distance to the transcription start site (TSS) (**A.**), CpG content (**B.**) and GC% (**C.**) were calculated. VU40T results were averaged from two MeDIP and three hMeDIP sequencing experiments. Error bars show STD.

## 5.15 Promoters that gain 5mC in HNSCC are associated with decreased expression in VU40T cells

Gene ontology and KEGG pathway analysis was performed on promoters that gained methylation in VU40T cells (Figure 5.27). The KEGG pathway term 'nicotine addiction' should be noted, as smoking is one of the main risk factors for HNSCC (Figure 5.27). However, in general, most ontological groupings did not correspond with tumour suppressive functions (Figure 5.27). To further investigate the function of promoter methylation in VU40T cells, MeDIP peaks containing promoters were separated into three groups: those which gained methylation in HNSCC (VU40T specific: 4999 genes); those which lost methylation in HNSCC (HOK specific: 3036 genes); and those where the methylation was unchanged (constant: 2483 genes). Using the VU40T transcriptome, the expression level (FPKM) of genes with 5mC marked promoters was determined for each group (Figure 5.28). VU40T specific promoters were associated with expressed genes (FPKM>1) at a lower frequency than the other two groups, and in general corresponded with lower FPKM values (Figure 5.28). Therefore, promoters that gain 5mC in HNSCC are less likely to be associated with genes expressed in VU40T cells (Figure 5.28). However, as no comparison with the HOK transcriptome was performed we cannot determine whether these genes would have been expressed prior to 5mC gain. (Figure 5.28). However, the majority of genes (77%) that lost promoter methylation in HNSCC (HOK specific) were not expressed in VU40T cells, suggesting that methylation loss alone is insufficient to activate gene expression (Figure 5.28).

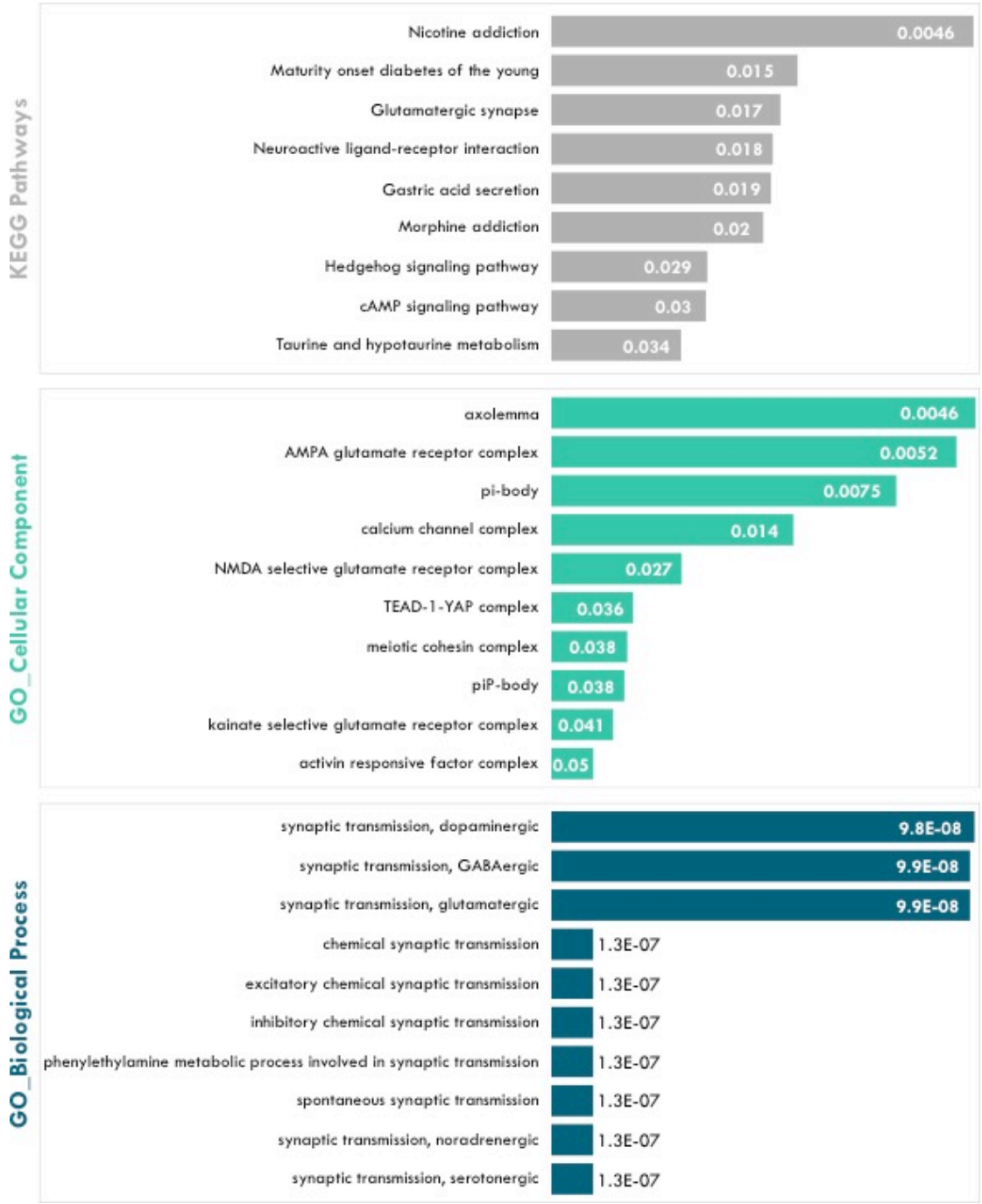
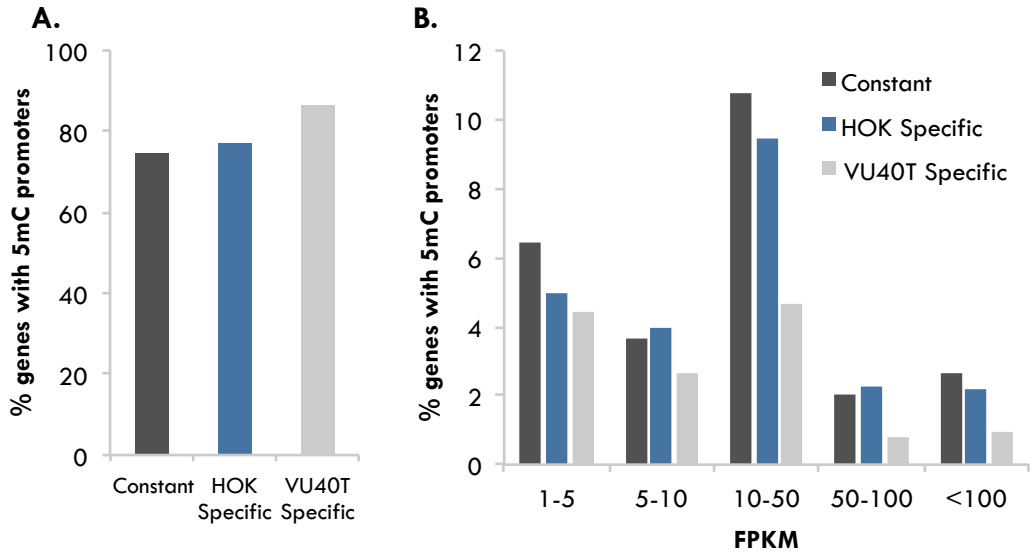



**Figure 5.27. Gene ontology analysis of VU40T specific promoters**

Peaks obtained from the VU40T cell and HOK MeDIP studies were separated based on the presence of a 5mC peak in either the HOK sample only (HOK specific), the VU40T cell sample only (VU40T specific) or both samples (constant). Promoters marked with 5mC in the VU40T specific group were subject to gene ontology analysis for cellular component and biological processes and KEGG pathway analysis using the online software enrichr [252, 253]. The top 10 significant terms are shown for each analysis type and the corresponding p-values are shown in white in the bars.




**C.**

| FPKM | Constant | | HOK Specific | | VU40T Specific | |
|---|---|---|---|---|---|---|
| | No. genes | % | No. genes | % | No. genes | % |
| <1 | 1849 | 74.47 | 2341 | 77.11 | 4325 | 86.53 |
| 1-5 | 160 | 6.44 | 150 | 4.94 | 221 | 4.42 |
| 5-10 | 91 | 3.66 | 121 | 3.99 | 132 | 2.64 |
| 10-50 | 267 | 10.75 | 288 | 9.49 | 232 | 4.64 |
| 50-100 | 51 | 2.05 | 69 | 2.27 | 41 | 0.82 |
| >100 | 65 | 2.62 | 67 | 2.21 | 47 | 0.94 |

**Figure 5.28. Expression of genes with methylated promoters**

The results of the VU40T and HOK MeDIP experiments were separated based on the presence of a 5mC peak in only the HOK sample (HOK specific), only the VU40T sample (VU40T specific) or in both samples (constant). For each group the expression level (FPKM) of promoters marked with 5mC was determined from the VU40T transcriptome. The results were separated based on FPKM values, where an FPKM <1 is not considered to be expressed. Results are shown as **A.** Bar graph showing percentage of the total 5mC marked promoters that are not expressed (FPKM<1) and **B.** Bar graph showing the FPKM values of expressed genes. **C.** Table showing both the total numbers of genes in each group and the percentage.

## 5.16 In HNSCC cells Alu elements are depleted of 5mC and 5hmC

As previously discussed (Section 5.8), 5mC and 5hmC are primarily found at Alu elements in normal oral keratinocytes (HOK cells). In HNSCC, only 32% of 5mC and 5hmC-enriched regions contained Alu elements (Figure 5.29). Instead, 5mC and 5hmC are more enriched in LINEs, simple repeats and regions not containing repetitive DNA, and 5mC is increased in LTRs (Figure 5.29). Indeed, almost 50% of 5mC-enriched regions in VU40T cells do not contain repetitive DNA, suggesting that DNA methylation is lost from these repetitive elements (Figure 5.29). Both 5mC and 5hmC are lost from Alu S and Alu Y elements, with a substantial drop in 5mC at Alu S. Interestingly, the most evolutionarily old family, Alu J, show decreased 5mC but increased 5hmC in VU40T cells compared to HOK cells (Figure 5.30). These data suggest that the regulation of Alu elements by 5mC and 5hmC is impaired in HNSCC.

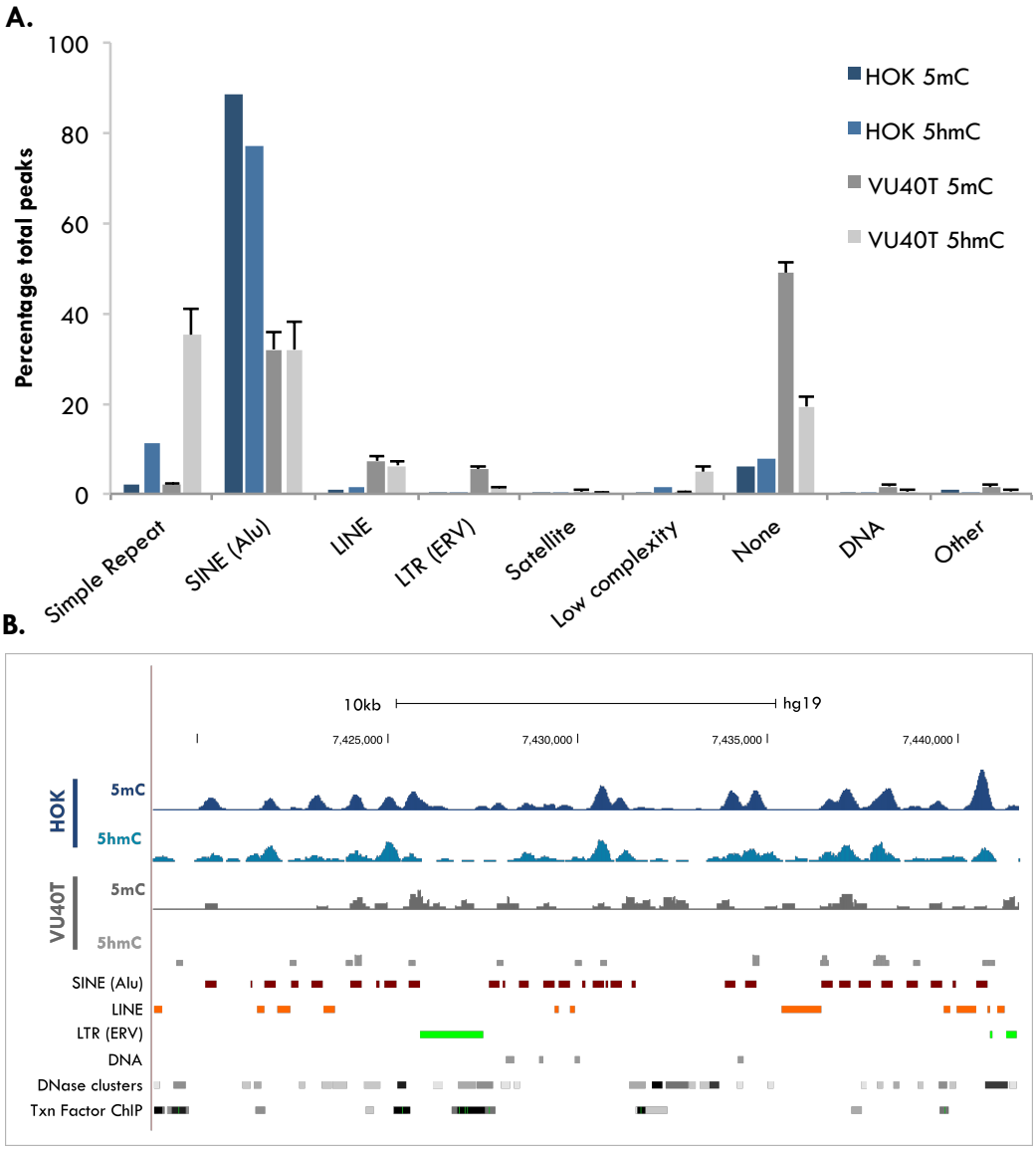



**Figure 5.29. DNA methylation and hydroxymethylation of repetitive DNA in HNSCC cells**

**A.** The MeDIP and hMeDIP sequencing results of VU40T and HOK cells were separated based on detailed annotations and are shown as a graph. VU40T MeDIP (n=2) and hMeDIP (n=3) peaks were analysed separately and the percentage total peaks averaged, error bars show STD. **B**. UCSC browser shot showing an example of the HOK (blue) and VU40T (grey) MeDIP (dark shade) and hMeDIP (lighter shade) sequencing results alongside repeatmasker repetitive DNA and ENCODE regulatory elements. MeDIP peaks were identified and detailed annotations assigned using MACS by Dr W.Croft and hMeDIP peaks were identified and detailed annotations assigned using Homer by Dr S. Clokie.

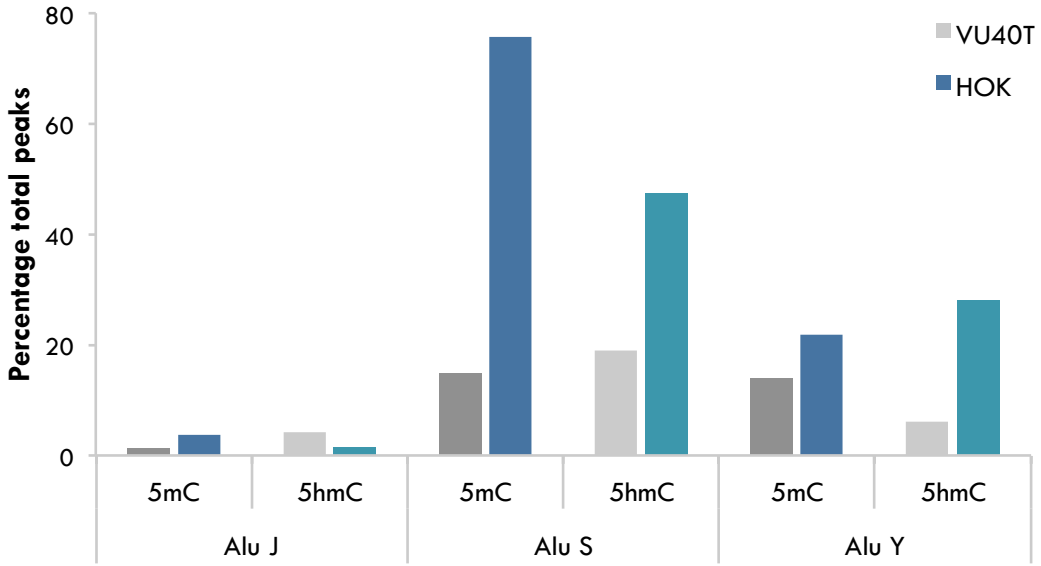



**Figure 5.30. Enrichment of 5mC and 5hmC across Alu elements families in normal oral keratinocytes and HNSCC cells**

Alu elements from 5mC and 5hmC-enriched regions were separated based on families (Alu J, S, Y) and are shown here as a percentage of the total 5(h)mC-enriched regions (peaks).

## 5.17 Discussion and Conclusion

Transposable elements occupy a vast proportion of the human genome, considerably higher than that occupied by genes [79]. In humans, the primate specific Alu elements, have expanded to cover ~10-11% of the genome and are enriched in CG-rich DNA, often found neighbouring or within genes [81, 83, 238]. In addition to the mutagenic potential of free Alu elements, embedded Alu elements have enhancer-like characteristics, transcription factor binding sites and can form new exons [81, 82, 238, 245, 254]. Therefore, epigenetic mechanisms must tightly regulate access to Alu elements to maintain normal cellular function.

This chapter has described the enrichment of Alu elements in MeDIP and hMeDIP sequencing data of normal oral keratinocytes. In gene-poor regions, Alu elements are evenly distributed and associated with 5mC, but not 5hmC. However, in gene-rich regions, Alu elements tend to be found in clusters, marked with both 5mC and 5hmC. These clusters of Alu elements were found neighbouring, but rarely overlapping, regulatory regions such as TSSs, DNase hypersensitive sites or transcription factor binding sites. Alu elements marked with 5mC and 5hmC are associated with areas of activating and not repressive histone marks but do not overlap with the centre of the peaks. Both 5mC and 5hmC are underrepresented in Alu J elements. This is the oldest family of Alu elements and has been inactive in the human genome for ~50 million years; therefore it is likely that the epigenetic control of these transposons is less important [79]. Interestingly, the middle aged Alu S family is enriched in the MeDIP data but not the hMeDIP data, suggesting that a more stable form of silencing may have evolved for these elements, while younger

elements rely on a more transient form of repression. Indeed, the most recently evolved Alu family, Alu Y, are enriched in the less repressive DNA modification, 5hmC.

In a study on the distribution of DNA modifications in mouse embryonic fibroblasts (MEFs) and mouse embryonic stem cells (mESCs), Papin *et al.,* (2017) determined that mouse-specific SINEs were enriched in DNA modifications and clustered around CpG islands [255]. These SINEs were methylated in MEFs and hydroxymethylated in mESCs and the researchers proposed that they might act as boundary elements that protect CGIs against pervasive DNA methylation, controlled by alternating the DNA modifications [255]. In human evolution, the primate-specific Alu elements have been preferentially selected for in regions rich in genes and regulatory elements [79]. Furthermore, while Alu elements are largely excluded from unmethylated domains they tend to occupy the boundaries [256, 257]. This chapter has described an enrichment of 5mC and 5hmC at Alu elements found alongside, but not overlapping, active chromatin marks and regulatory elements. Therefore, it is hypothesised that, like mouse-specific SINEs, Alu elements act as boundaries controlled by DNA methylation and hydroxymethylation that help maintain access to regulatory elements.

In HNSCC, this control of Alu elements is disrupted; the proportion of MeDIP and hMeDIP peaks associated with Alu elements is reduced to ~30% in VU40T cells compared to ~80% in HOK cells. Hypomethylation of Alu elements is a characteristic common to many cancers [87, 111, 114, 256]. Loss of 5mC at transposable elements has considerable mutagenic potential as it can lead to chromosomal instability and

increased transposition (summarised in Figure 1.15) [111]. Furthermore, increased expression of Alu elements in cancer has been associated with the production of small RNAs that silence genes containing intragenic Alu elements [258]. As the majority of genes contain at least one intragenic Alu, this could cause considerable changes to the transcriptome [84, 240]. The relationship between 5hmC, Alu elements and cancer has not previously been described and further work is required to determine the functional consequences of Alu 5hmC loss.

In a comparison of HOK and VU40T cells, DNA methylation appears to be preferentially lost from Alu S elements in HNSCC cells. Recent Alu retrotransposition events come from members of the Alu Y family; however, other Alu elements (particularly Alu S) are expressed as RNA that is unable to be inserted into the genome [245, 249, 259]. This Alu RNA can silence genes containing intragenic Alu elements [258]. Therefore, demethylation of Alu S family members may have been selected for during tumour progression as a means of gene silencing [258]. Conversely, Alu J elements gain 5hmC in VU40T cells. Many Alu J elements have evolved into enhancers or proto-enhancers [238]. As discussed here, and elsewhere, 5hmC is associated with active enhancers [61, 106]. Therefore, 5hmC gain at Alu J elements may create novel enhancers in HNSCC.

Non-repeat regions gain cytosine modifications in HNSCC. Promoters and CpG islands contain more 5mC in VU40T than HOK cells. Promoter 5mC gain in HNSCC cells is associated with decreased expression of these genes in VU40T cells; however, promoter 5mC loss is not accompanied by an equivalent increase in expression. This probably reflects the combinatorial nature of gene silencing and suggests that while

methylation gain can be sufficient to prevent transcription, the loss of 5mC alone is not always sufficient to reactivate gene expression [27]. Instead, corresponding changes in histone modifications, chromatin structure and transcription factor availability are needed to reactivate genes. To further these results, a comparison with the HOK transcriptome is required to determine whether promoters that gain 5mC in HNSCC were expressed in normal oral keratinocytes.

The MeDIP and hMeDIP results described here overlapped considerably. This is to be expected, as 5hmC is the product of 5mC. However, with the exception of DNA modification loss at Alu elements, the differences between HOK and VU40T cells were distinct for each modification (Figure 5.31). In HNSCC, 5mC was gained in regions associated with a higher CpG content, found closer to, and often within, the transcription start site, while 5hmC was gained in intergenic regions, associated with a lower CpG content and a greater distance from the TSS (Figure 5.31). These data are in keeping with the literature; in lung, liver and kidney cancer 5hmC changes are inversely correlated with 5mC changes [260, 261].

When considering these results, it is important to recognise that DIP-sequencing experiments are subject to bias towards short, unmodified DNA repeats (generally repeating dinucleotides such as CA repeats) [262]. This can be attributed to an intrinsic bias in the IgG antibodies [262]. In the analysis performed here, peaks lacking CpG dinucleotides were not considered, partially resolving this issue. However, this does not remove the possibility of nonspecific binding to repeats neighbouring CpG sites. Therefore, comparing the results to DIP-sequencing of a matched isotype control would be recommended. Furthermore, as different library

preparations and peak calling approaches were used for the VU40T samples compared to the HOK samples, some differences between the two data sets could be attributed to this.

Alu elements are repeats of around 280-300bp long and therefore are smaller than the average fragment generated by the MeDIP and hMeDIP experiments. During sequencing, short reads are produced from either end of a DNA fragment and aligned by taking into account the original size of the fragment. This has allowed Alu elements to be included in the analysed dataset as differences in the neighbouring regions of DNA allow the Alu elements to be mapped to unique regions in the genome.

In conclusion, this chapter has described how the distribution of 5mC and 5hmC is altered in HNSCC (Figure 5.31, Figure 5.32). In normal cells 5mC and 5hmC are excluded from the TSS and transcription factor binding sites and enriched in clusters of Alu elements in gene-rich regions (Figure 5.32). The function of DNA modifications appear to differ based on the presence or absence of an Alu element, with 5hmC outside of Alu elements associated with tissue specific enhancers and bivalent chromatin. However, in HNSCC both 5mC and 5hmC are lost from Alu elements, and 5mC is gained in LTRs and non-repeat regions such as promoters and CpG islands (Figure 5.32).

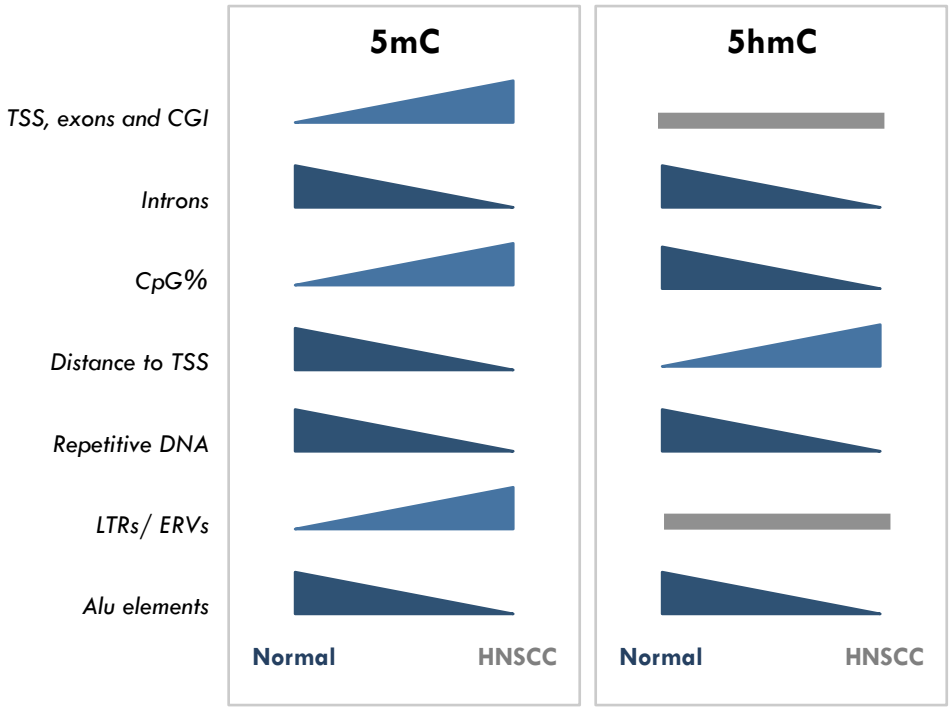



**Figure 5.31. 5mC and 5hmC changes in HNSCC**

The changes in 5mC and 5hmC between normal (HOK) and HNSCC (VU40T) cells are represented as a schematic. Increasing (light blue) or decreasing (dark blue) features are represented by blue triangles while aspects that are unaffected by HNSCC are represented by a grey bar.

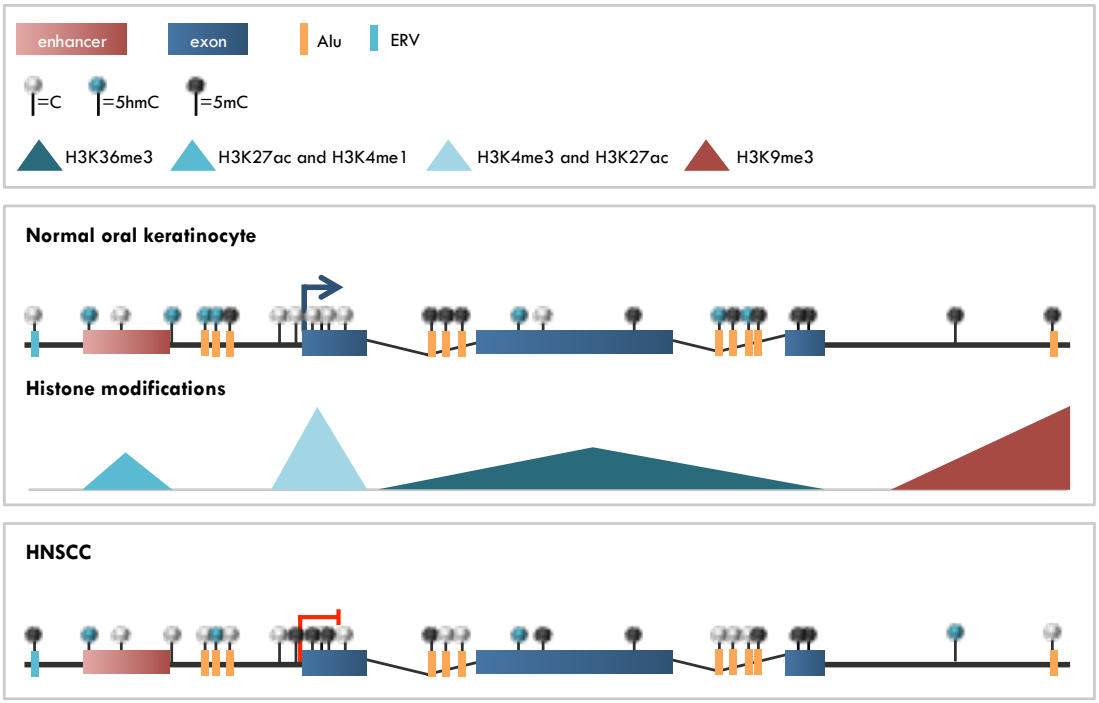



**Figure 5.32. DNA modifications in normal oral keratinocytes and HNSCC cells**

**Top panel:** The distribution of 5mC and 5hmC in a gene rich region in HOK cells is represented as a schematic alongside the associated histone modifications. Active genes are characterised by unmethylated CGIs and active histone modifications (H3K4me3 and H3K27ac) at the TSS (blue arrow), and gene bodies enriched in DNA methylation and H3K36 methylation. Enhancers are enriched in H3K27ac and H3K4me1, neighbouring 5hmC. In HOK cells, Alu elements are found to cluster in and around genes and regulatory regions and are consistently marked with 5mC and 5hmC. Outside of genes, Alu elements are dispersed and only marked with 5mC. **Bottom panel:** The distribution of 5mC and 5hmC are changed in HNSCC. 5mC is gained at LTRs and Alu elements lose both 5mC and 5hmC. At promoters 5mC gain is associated with reduced gene expression.

# Chapter 6. DAC partially restores the distribution of 5mC and 5hmC in an HNSCC cell line to one characteristic of normal oral keratinocytes

DAC is a DNA demethylating agent and DNMT inhibitor, which, as shown in Chapter 3, can demethylate a subset of HNSCC cell lines. In cancers, the efficacy of the drug has been attributed to the reactivation of silenced tumour suppressor genes. Indeed, genes inactivated in cancer appear to be the most responsive to reactivation by DAC [96, 263, 264]. However, the response to DNMTi therapy is frequently delayed and a number of genes not controlled by a methylated promoter are upregulated by DNMTi treatment, suggesting an indirect mechanism may also be at work [158, 265]. This led to the proposition of a new mechanism to explain the efficacy of DNMTi treatment in cancer, termed 'viral mimicry' [158, 180].

## 6.1. Viral mimicry

In an analysis of the transcriptome and methylome of 63 solid cancer cell lines after treatment with the DNMTi 5-azacytidine (AZA), Li and colleagues found an upregulation of immune related genes largely not controlled by a methylated CGI promoter [265]. They defined this set of genes as Aza IMmune genes (AIM) and

found that AIM genes were increased in DNMT double knockout cells; *in vitro,* after treatment with HDACi; and in breast and colorectal cancer patients subject to combined treatment with AZA and an HDACi [265].

Two papers expanded upon this discovery in 2015 to propose a new mechanism to explain the anticancer effects of low dose DNMTi [158, 180]. In solid tumour cancer cells *in vitro,* several days after the removal of low dose DAC or AZA, the expression of immune related genes including interferon-β1 (*IFN-β1*) and interferon-stimulated genes (ISGs) was increased [158, 180]. This was not accompanied by promoter 5mC changes; instead methylation was lost from endogenous retroviruses (ERVs), leading to expression of the ERVs as dsRNA. This ERV dsRNA triggers the dsRNA-sensing proteins TLR3 and MDA5 to activate IFN-β1 and bring about an immune response that kills the cells [158, 180]. A 'viral mimicry' term was introduced to explain this mechanism of delayed response to DNMTis (Figure 6.1)[158, 180]. A similar result was observed after AZA treatment in a mouse model of epithelial ovarian cancer, leading to increased CD45+ immune cells, active CD8+ T cells and natural killer cells and a reduced tumour burden [266]. Similarly, the anticancer agent, RRx-001, can induce a viral mimicry response in colon cancer cells via decreased expression of DNMT1 and -3A [267]. Therefore, the anti-tumourigenic action of DNMTi could be due to both the direct effects of DNA damage and demethylation of tumour suppressor genes and the indirect effects of interferon activation via ERV demethylation [158, 180, 266].

While immune therapy is an attractive cancer therapeutic option, in most cancer types fewer than half of patients respond [266]. When comparing the expression of

AIM genes in solid tumour patient samples, the tumours clustered into groups with high or low AIM gene expression suggesting an immune evasion phenotype that could be reversed by epigenetic therapy [265]. Indeed, in a mouse model of ovarian cancer, treatment with AZA and HDACi reduced the immunosuppressive tumour microenvironment and improved the response to the immune checkpoint inhibitor α-PD-1 [266].

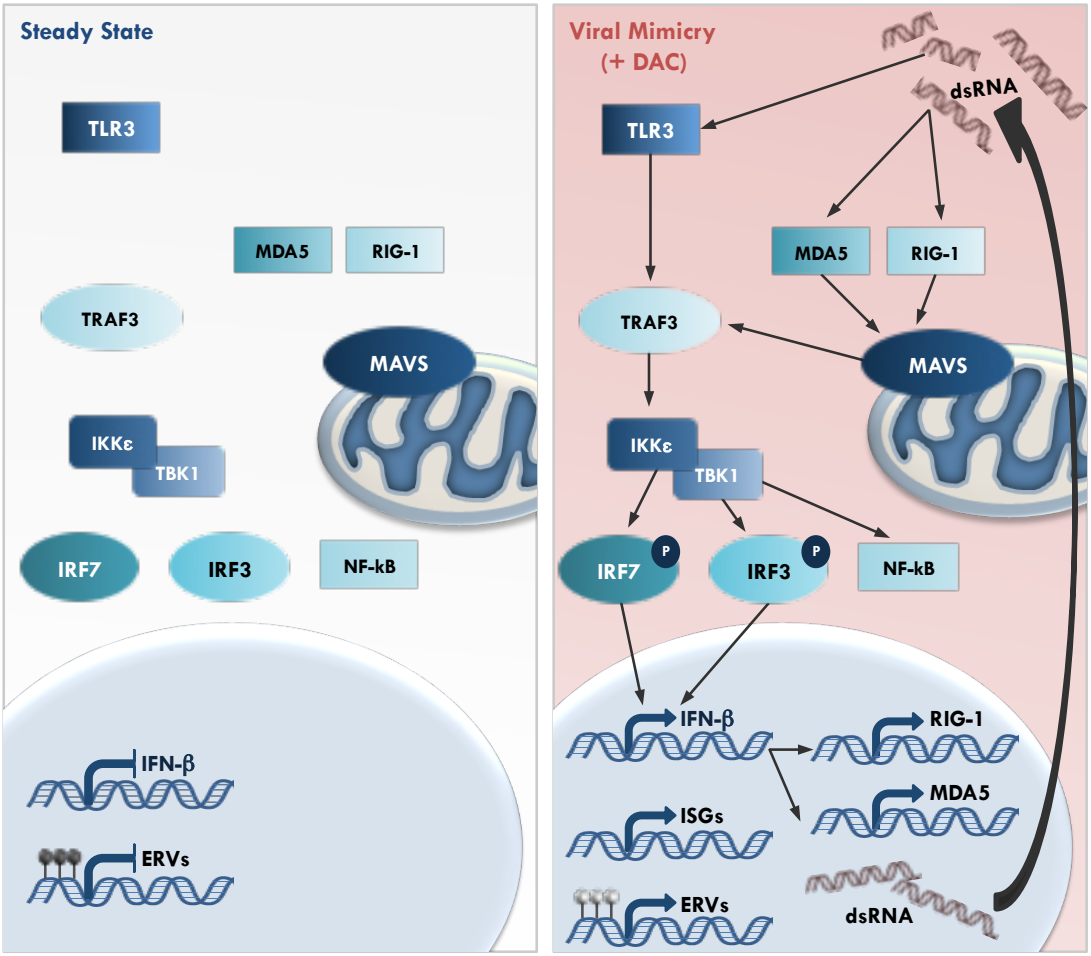



**Figure 6.1. Viral mimicry**

DAC treatment decreases methylation of ERVs leading to their expression as dsRNA. These dsRNAs can activate TLR3, MDA5 and RIG-1 in the cytoplasm, which then activate TRAF3 signalling which triggers IKKε and TBK1 and promotes the phosphorylation of IRF7 and IRF3. IRF7 and IRF3 enter the nucleus and stimulate the expression of IFN-β. IFN-β then activates the expression of interferon-stimulated genes (ISG) including RIG-1 and MDA5, which feed back into the dsRNA response, and activates genes that trigger cell death [158, 180].

## 6.2. 5hmC and DAC treatment

As described in Chapter 3, treatment with DAC leads to global 5mC loss. DNA hydroxymethylation is the oxidation product of 5mC [53, 54]. Therefore, it could be predicted that treatment with DAC would also decrease 5hmC. Interestingly, *in vitro* evidence indicates the opposite; DAC treatment has been show to increase the levels of 5hmC [268, 269]. Indeed, Chowdhury *et al., (2015)* determined a paradoxical increase in 5hmC after DAC treatment [269]. This was not accompanied by a change in *TET* gene expression. Instead, the researchers hypothesised that TET proteins have a preference for hemi-methylated DNA and confirmed that the fraction of TET bound to DNA increased after DAC [269]. Therefore, DAC causes an increase in hemi-methylated DNA and TET proteins preferentially bind this DNA to convert 5mC into 5hmC [269]. In hepatocellular carcinoma, the DNMTi, AZA was also found to increase 5hmC [270]. Similarly, DAC and the TET cofactor, ascorbic acid (vitamin C), have been described to work in synergy to increase apoptosis in cancer cells [268]. Combined treatment was accompanied by increased ERV dsRNA and immune gene expression, greater than that seen by DAC alone; and an increase in global and ERV-specific 5hmC, greater than that seen by vitamin C treatment alone [268]. Therefore, the synergistic relationship can be attributed to increased TET activity by vitamin C that promotes viral mimicry and increased conversion of 5mC to 5hmC by DAC [268].

## 6.3. Aims and objectives

In previous chapters the potential use of DAC in the treatment of HNSCC were discussed, and it was noted that DAC is effective at low doses in some HNSCC lines. This chapter aims to expand upon this result by investigating the effect of relatively high (1μM) and low (100nM) dose DAC on the distribution of DNA methylation and hydroxymethylation in the DAC-responsive HNSCC cell line, VU40T. Studies focus on specific changes in 5mC and 5hmC between normal oral keratinocytes (HOK) and VU40T cells discussed in Chapter 5, such as the loss of modifications at Alu elements and the gain in 5mC at promoters. Furthermore, it was explored as to whether a viral mimicry response is observed in VU40T cells after treatment with DAC.

## 6.4. MeDIP and hMeDIP of DAC treated cells

The HNSCC cell line VU40T was treated with relatively low (100nM) and high (1μM) dose DAC for 96h and MeDIP and hMeDIP sequencing was performed. Experiments was performed in duplicate for MeDIP and triplicate for hMeDIP and qPCR used to check the efficacy of each reaction prior to library preparation and sequencing (Supplementary data 8). MeDIP experiments were also undertaken in normal human oral keratinocytes (HOK cells) treated with relatively high and low dose DAC, while hMeDIP was performed only on untreated HOK cells. HOK MeDIP and hMeDIP experiments were performed in triplicate and samples pooled prior to sequencing. The total number of peaks obtained for each treatment, the number of peaks discarded due to lack of CpGs and the number of peaks retained for analysis are presented in Table 6.1. Non-uniquely mapped reads were discarded from the analysis and MeDIP peaks were called and annotated by Dr W. Croft using MACS, while hMeDIP data was analysed by Dr S. Clokie using Homer.

**Table 6.1. MeDIP and hMeDIP output**

The total number of 5mC or 5hmC enriched regions ('peaks') obtained for each MeDIP and hMeDIP experiment. **First column**: before filtering; **second**: the number of peaks removed due to lack of CpGs; **third**: the total number of peaks retained for further analysis; **last column:** the percentage of total peaks retained for analysis. MeDIP peaks were called using MACS by Dr W. Croft and hMeDIP peaks were identified using Homer by Dr S. Clokie

| | | Total Peaks | CpG =0 | Retained Peaks | % Retained Peaks |
|---|---|---|---|---|---|
| MeDIP | **HOK NT** | 547344 | 1839 | 545505 | 99.66 |
| | **HOK 100nM DAC** | 606061 | 2588 | 603473 | 99.57 |
| | **HOK 1μM DAC** | 627469 | 1399 | 626070 | 99.78 |
| | **VU40T NT** | 512502 | 1822 | 510680 | 99.64 |
| | **VU40T 100nM DAC** | 648254 | 11216 | 637038 | 98.27 |
| | **VU40T 1μM DAC** | 700376 | 5090 | 695286 | 99.27 |
| hMeDIP | **HOK NT** | 26678 | 2358 | 24320 | 91.16 |
| | **VU40T NT** | 41562 | 10481 | 31081 | 74.78 |
| | **VU40T 100nM DAC** | 43978 | 10745 | 33233 | 75.57 |
| | **VU40T 1μM DAC** | 38353 | 7825 | 30528 | 79.60 |

## 6.5. DAC treatment alters the genomic distribution of 5mC and 5hmC in HNSCC cells

DNA dot blot analysis demonstrated a dose dependent decrease in global 5mC in VU40T cells after treatment with DAC (Figure 6.2). In leukemic cells a paradoxical increase in 5hmC has been described [269]. However, in VU40T cells, DAC treatment did not significantly alter the global levels of 5hmC (Figure 6.2).

The distribution of 5mC and 5hmC after DAC treatment in VU40T cells was determined using MeDIP and hMeDIP sequencing (Figure 6.3). Interestingly, DAC caused considerable alterations to the distribution of both 5mC and 5hmC (Figure 6.3). In VU40T cells *TET2* was expressed (FPKM= 7.7) while *TET1* was not (FPKM <1) and DAC caused an increase in *TET2* gene expression (Figure 6.4). This could partially explain the disruption to 5hmC detected in hMeDIP-seq experiment and the lack of 5hmC decrease observed in the DNA dot blot after treatment with DAC (Figure 6.2-Figure 6.4).

To further determine the difference in 5mC and 5hmC after DAC, MeDIP and hMeDIP peaks were separated based on their annotations (Figure 6.5). After both high and low dose DAC, 5mC was found less frequently in promoters and exons, and more frequently in introns and intergenic regions (Figure 6.5). Furthermore, 5mC was found less frequently at CpG islands (Figure 6.6). Compared with the untreated control, 5hmC was found more often in introns and less frequently in intergenic regions after DAC (Figure 6.5). While 5mC-enriched regions were further from the TSS, associated with a lower CpG content after DAC (particularly low dose DAC), the

opposite was observed for 5hmC (Figure 6.7). In conclusion, the results of the MeDIP and hMeDIP sequencing of VU40T cells suggest that DAC treatment is shifting the distribution of 5mC and 5hmC towards a distribution characteristic of normal oral keratinocytes (HOK cells) (Figure 6.5-6.7).

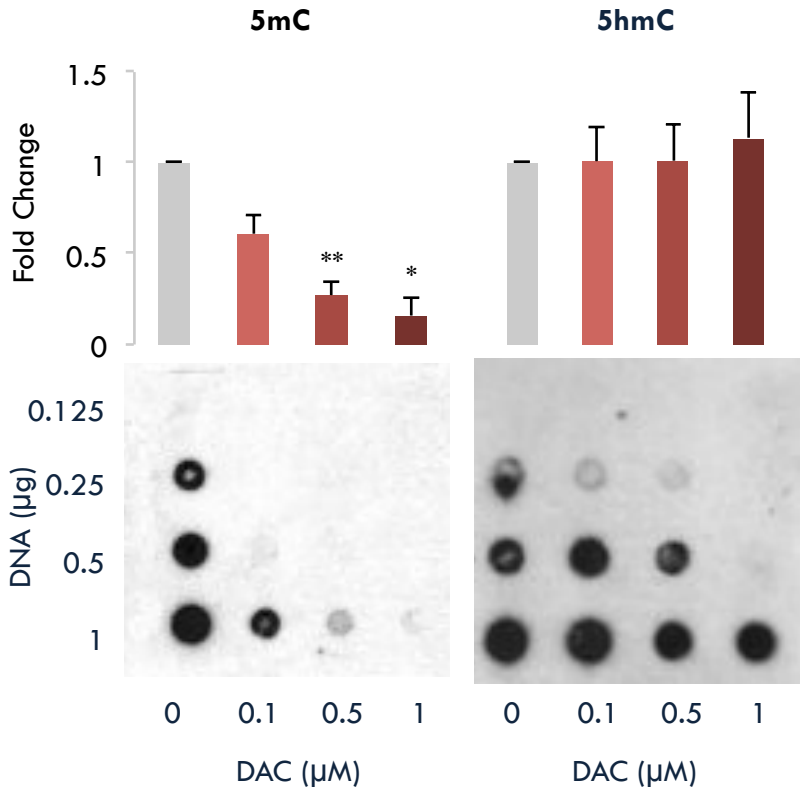



**Figure 6.2. Global levels of 5mC and 5hmC in DAC treated VU40T cells**

DNA was extracted from VU40T cells treated with a range of concentrations of DAC for 96h. Following this, a DNA dot blot was performed using an antibody against 5mC or 5hmC. Methylene blue was used as a loading control. The relative intensity of the 5mC or 5hmC dots was normalized to the corresponding methylene blue dot. The blot shown is a representative image, while the bar graph shows the mean ± SEM (n=3).

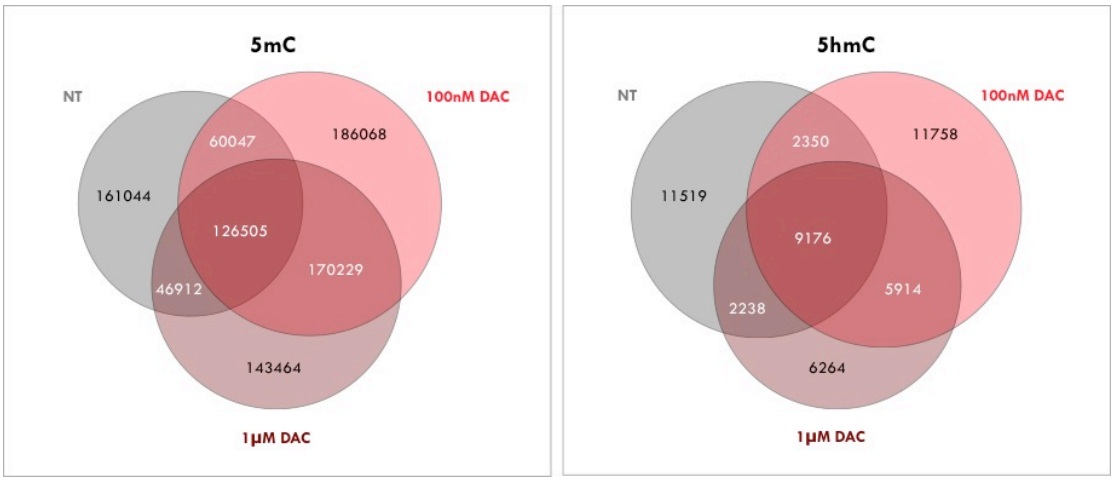



**Figure 6.3. DNA methylation and hydroxymethylation changes in response to DAC**

VU40T cells were treated with 100nM DAC, 1µM DAC or an equivalent volume of the vehicle, acetic acid (NT) for 96h. DNA was extracted and MeDIP and hMeDIP sequencing performed. The overlap between peaks from the different treatment groups are shown as a Venn diagram for 5mc and 5hmC. Numbers show the total number of peaks for each section. MeDIP peaks were identified using MACS by Dr W. Croft and hMeDIP peaks were called using Homer by Dr S. Clokie.

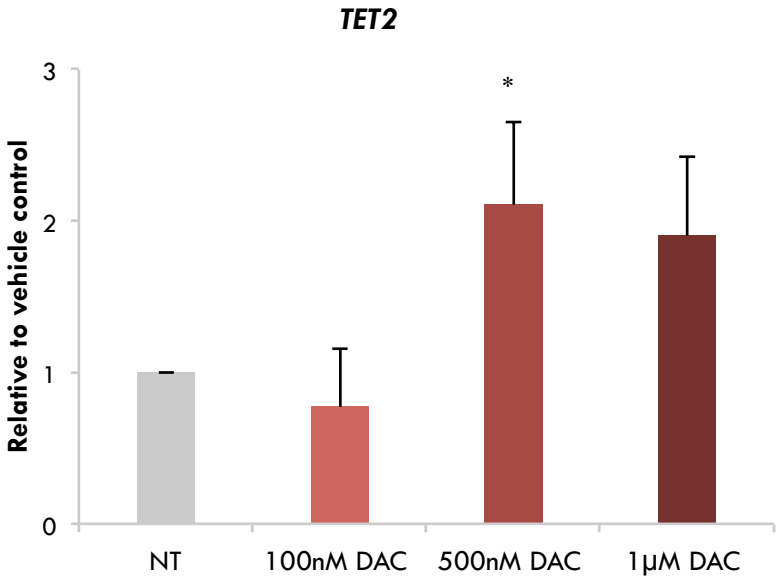
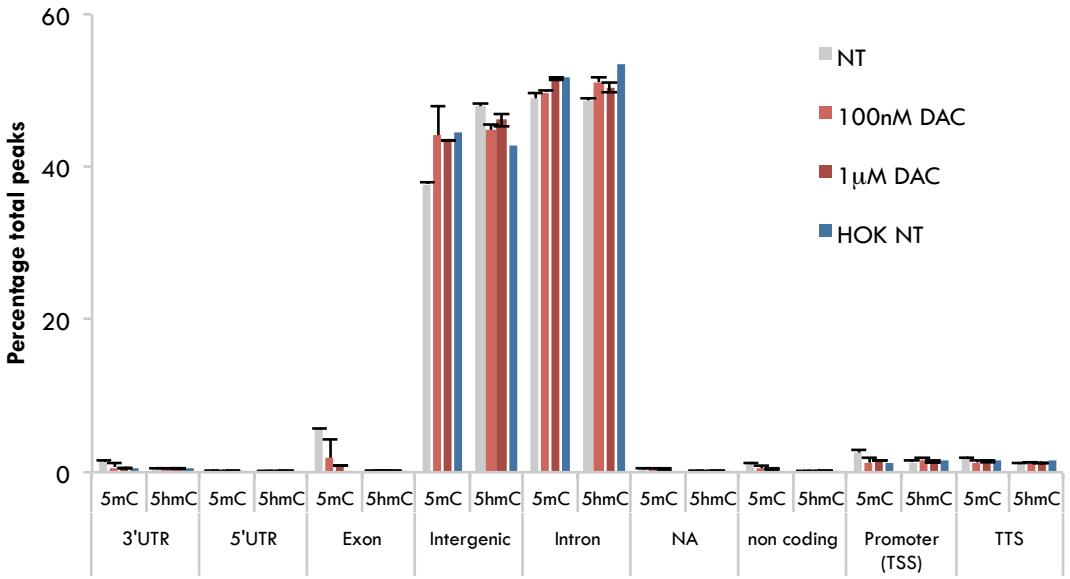



**Figure 6.4. DAC treatment increases *TET2* RNA**

VU40T cells were treated with 100nM, 500nM or 1µM DAC, or an equivalent volume of the vehicle for 96h. Following this, RNA was extracted and qRT-PCR performed for *TET2.* Data is shown as a fold change normalise against the vehicle only control (NT). Error bars represent SEM (n=3) and significance was determined using a paired t-test with Bonferroni correction where *p<0.05.

A.




B.

| | | 5mC | | | | 5hmC | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | HOK | NT | 100nM DAC | 1 μM DAC | HOK | NT | 100nM DAC | 1 μM DAC |
| No. Peaks | 3UTR | 2732 | 7120 | 3825 | 3920 | 100 | 122 | 181 | 148 |
| | 5UTR | 163 | 1119 | 414 | 506 | 7 | 12 | 21 | 14 |
| | Exon | 1543 | 29321 | 7903 | 5724 | 12 | 69 | 73 | 58 |
| | Intergenic | 242606 | 192928 | 286848 | 302331 | 10381 | 14884 | 14741 | 14053 |
| | Intron | 282498 | 250776 | 317967 | 358097 | 13011 | 15135 | 17099 | 15389 |
| | NA | 921 | 1758 | 2337 | 2083 | 55 | 65 | 37 | 34 |
| | non coding | 1328 | 5481 | 2614 | 2394 | 31 | 33 | 56 | 34 |
| | Promoter (TSS) | 6071 | 12492 | 7281 | 10406 | 382 | 417 | 586 | 431 |
| | TTS | 7643 | 9685 | 7849 | 9825 | 338 | 335 | 433 | 359 |
| Percentage Total Peaks | 3UTR | 0.50 | 1.39 | 0.69 | 0.56 | 0.41 | 0.40 | 0.52 | 0.48 |
| | 5UTR | 0.03 | 0.22 | 0.08 | 0.07 | 0.03 | 0.04 | 0.06 | 0.05 |
| | Exon | 0.28 | 5.73 | 1.79 | 0.83 | 0.05 | 0.22 | 0.25 | 0.20 |
| | Intergenic | 44.47 | 37.81 | 44.16 | 43.50 | 42.69 | 47.87 | 44.76 | 46.08 |
| | Intron | 51.79 | 49.00 | 49.84 | 51.48 | 53.51 | 48.73 | 51.12 | 50.42 |
| | NA | 0.17 | 0.37 | 0.40 | 0.31 | 0.23 | 0.21 | 0.12 | 0.10 |
| | non coding | 0.24 | 1.08 | 0.51 | 0.35 | 0.13 | 0.11 | 0.15 | 0.12 |
| | Promoter (TSS) | 1.11 | 2.50 | 1.25 | 1.50 | 1.57 | 1.35 | 1.71 | 1.38 |
| | TTS | 1.40 | 1.90 | 1.28 | 1.40 | 1.39 | 1.08 | 1.31 | 1.17 |

**Figure 6.5. In HNSCC DAC treatment shifts the distribution of 5mC and 5hmC back towards that of normal oral keratinocytes**

The genomic location of 5mC and 5hmC across different genomic regions is shown as a bar graph (**A**) and a table (**B**). The MeDIP and hMeDIP results of DAC treated VU40T cells are shown alongside untreated HOK cells for comparison. NT= vehicle only treated VU40T cells. VU40T MeDIP (n=2) and hMeDIP (n=3) results were analysed separately and the mean is shown. Error bars represent STD. MeDIP peaks were identified and annotations assigned using MACS by Dr W. Croft and hMeDIP analysis was carried out using Homer by Dr S. Clokie.

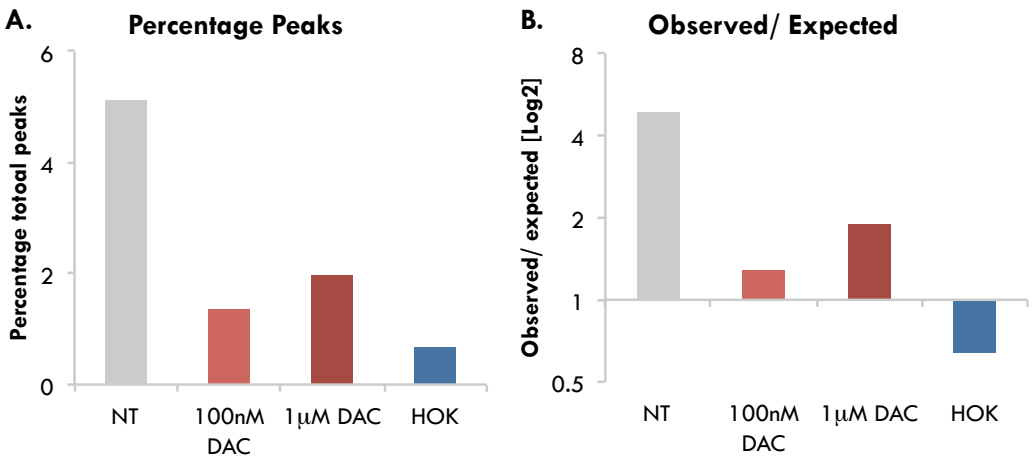



**Figure 6.6. Differential enrichment of 5mC at CpG islands in DAC treated VU40T cells**

The UCSC table browser was used to download CpG islands onto Galaxy. Using the Bedtools '*intersect intervals*' function, the amount of overlap between these regions and MeDIP peaks was determined. The results are shown as **A.** the percentage of total MeDIP peaks and **B.** normalised to a corresponding RandomBed file. NT= VU40T cells treated with vehicle only.

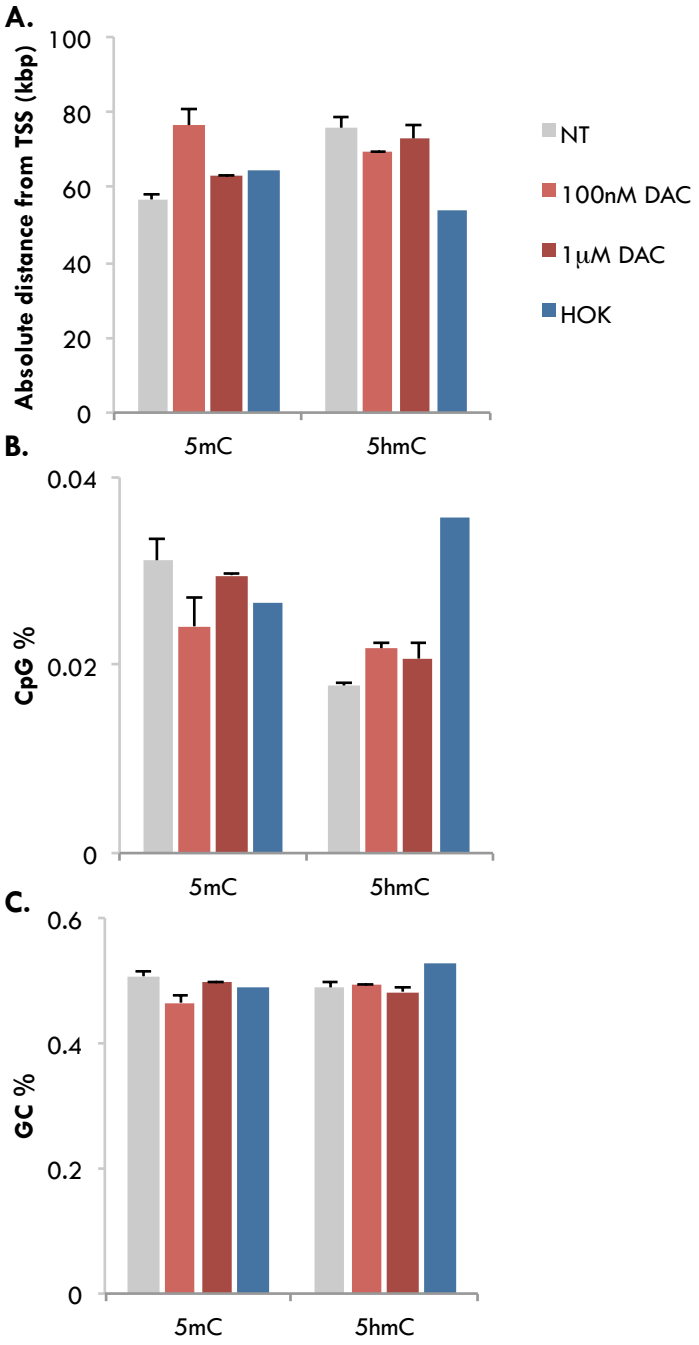



**Figure 6.7. Characteristics of DAC treated VU40T cells**

MeDIP and hMeDIP experiments were used to determine the location of 5mC and 5hmC in VU40T and HOK cells. Untreated (NT), 100nM DAC treated and 1μM DAC treated samples were compared in regards to: **A.** Mean distance from the TSS. **B.** Mean CpG content. **C.** Mean GC%. VU40T results were analysed separately and an average of the replicates is shown where error bars represent STD.

## 6.6. DAC treatment does not reactivate tumour suppressor genes commonly inactivated in HNSCC

A panel of tumour suppressor genes commonly inactivated in HNSCC were compiled from the literature [98-102]. Additionally, the TCGA data from HNSCC patients was downloaded from cBioPortal and genes with copy number variations (CNV) with more than 2% frequency in HNSCC and mutations with more than 8% frequency were added to the literature list [8, 222, 271]. This list was compared with the MeDIP data for untreated VU40T cells and those with a 5mC peak in the promoter were retained (Figure 6.8). The VU40T transcriptome, before and after 500nM DAC, was then used to compare changes in 5mC after DAC treatment to changes in gene expression (Figure 6.8). Interestingly, in this panel, decreased promoter methylation after DAC did not correlate with increased gene expression (Figure 6.8). This is indicative of the combinatorial nature of gene silencing, but also suggests that the efficacy of DAC in VU40T cells is not attributed to the reactivation of tumour suppressor genes.

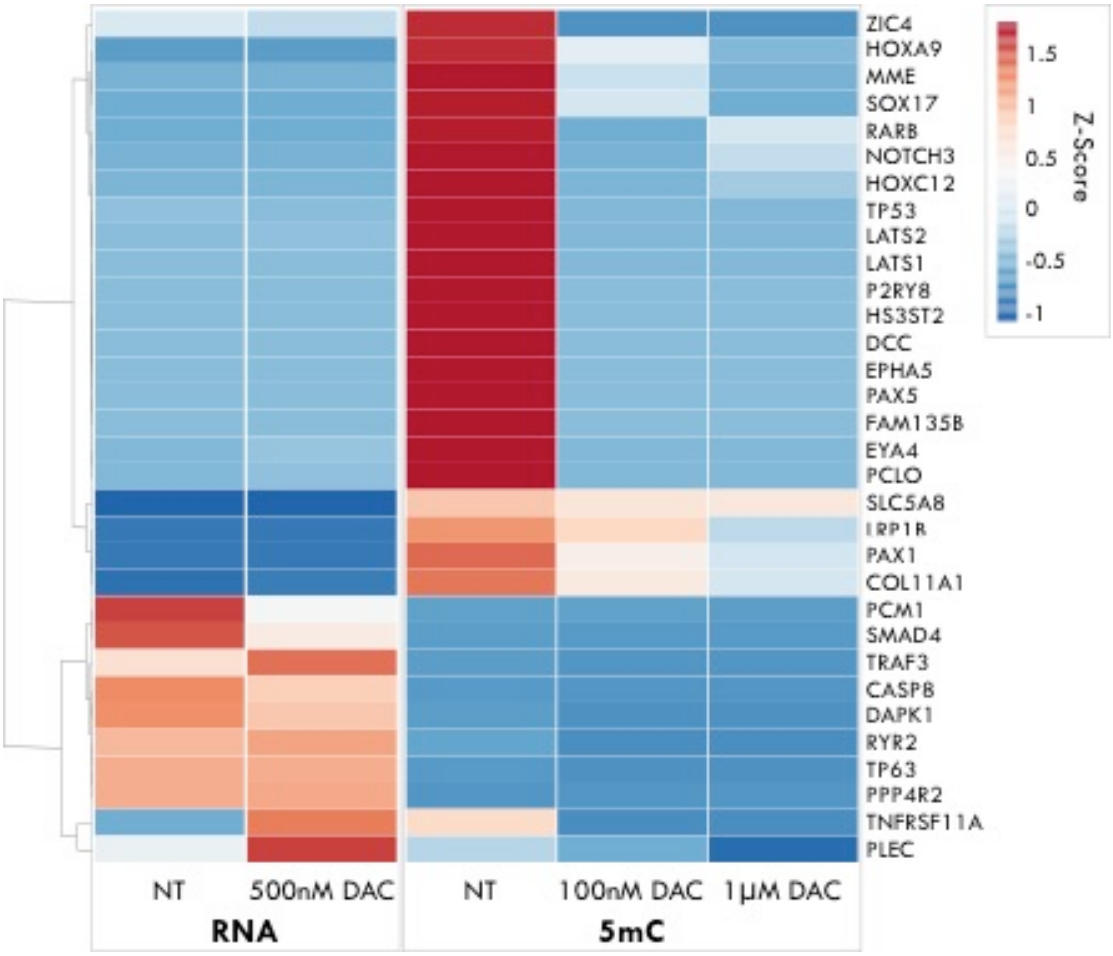



**Figure 6.8. A comparison of the expression and methylation changes in VU40T cells in response to DAC**

A panel of tumour suppressor genes from the literature, and genes with common CNVs (>2% frequency) or mutations (>8% frequency) in HNSCC was compiled. Those with a 5mC-enrichced region (peak) in the promoter of vehicle only VU40T cells (NT) are shown here. Changes in promoter methylation after 100nM and 1μM DAC treatment were compared with the VU40T cell transcriptome before and after 500nM DAC. The ClustVis software was used to cluster genes using correlation distance and average linkage [167]. Unit variance scaling was applied across both 5mC and RNA groups. RNA changes do not correspond with 5mC changes.

## 6.7. DAC treatment induces a viral mimicry response in VU40T cells

In solid tumours, the efficacy of low dose DNMTi has been attributed to the demethylation of endogenous retroviruses (ERVs) [2, 3]. These are expressed as double stranded RNA (dsRNA) that triggers an immune response that results in cell death [2, 3]. Therefore, this study sought to determine whether a similar viral mimicry response occurs in VU40T cells after treatment with DAC. DAC demethylates ERVs, particularly after 1µM treatment (Figure 6.9). The MeDIP data was compared with the transposable element transcriptome of VU40T cells before and after 500nM DAC and for a subset of ERV subfamilies, decreased 5mC corresponded with increased ERV RNA (Figure 6.10). This corresponds with a global increase in dsRNA (Figure 6.11). DeSeq2 analysis was used to compare the VU40T transcriptome before and after 500nM DAC. The results were subject to gene ontology analysis and were enriched in terms of interferon signalling and the viral life cycle (Table 6.2). Using qRT-PCR the RNA levels of several genes understood to be involved in the viral mimicry response were examined before and after treatment with DAC (Figure 6.12). Increased expression of interferon β (*IFN β*); its transcriptional activator, interferon regulatory factor 7 (*IRF7*); and the dsRNA sensing protein, Melanoma Differentiation-Associated protein 5 (MDA5, encoded for by *IFIH1* gene) were observed in VU40T cells after DAC treatment (Figure 6.12). Therefore, the efficacy of DAC treatment in VU40T cells could be partially attributed to a viral mimicry response (Figure 6.13).

The expression level of interferon signalling genes upregulated by DAC treatment was determined in TCGA data from 523 HNSCC patient samples (Figure 6.14). More

than half (57%) of the samples had alterations in at least one gene, and the samples formed two groups: those with relatively high IFN pathway expression and those with low expression (Figure 6.14). Therefore, these genes could be used as a biomarker for DAC response in HNSCC.

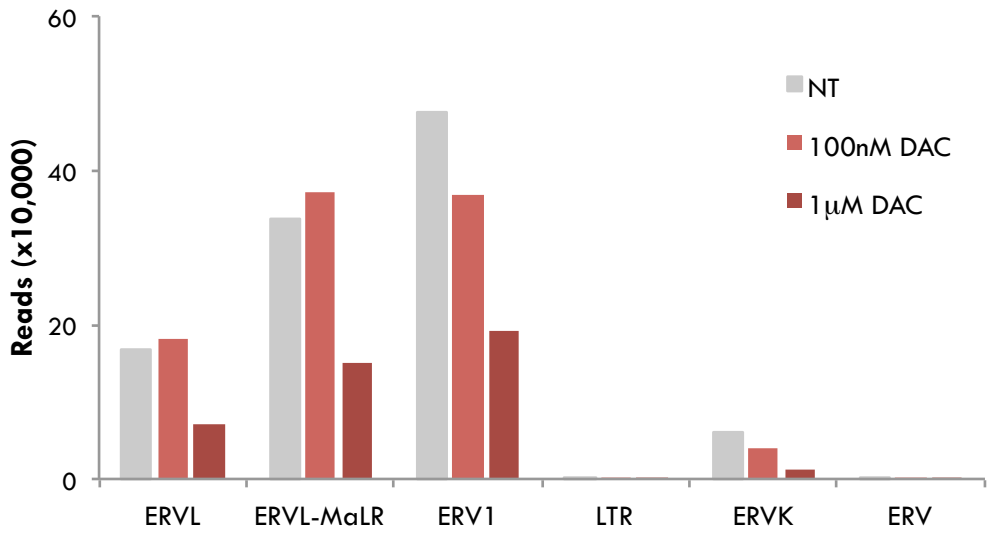



**Figure 6.9. 5mC is lost from endogenous retroviruses (ERVs) after treatment with DAC**

The number of MeDIP reads was calculated for indicated ERV families was determined by counting the number of reads intersecting with the repeatmasker data from the UCSC browser, using '*multicov*' function from the bedTools suite. The results are shown as a bar graph. MeDIP sequencing experiments are representative of two experiments, combined for analysis, while hMeDIP experiments are representative of three experiments. This analysis was performed by Dr S. Clokie.

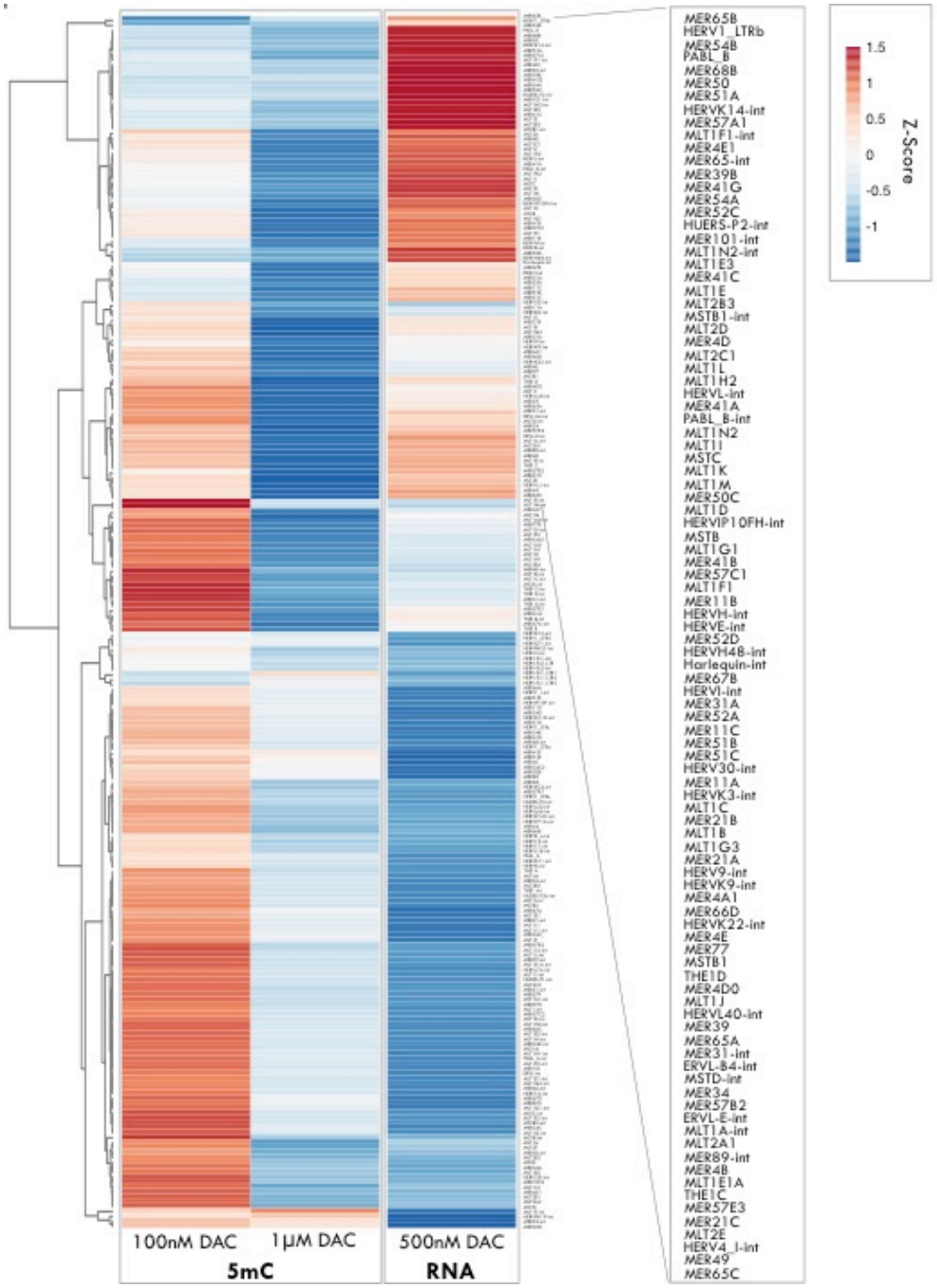



**Figure 6.10. Decreased methylation correlates with increased RNA at some ERVs**

The number of MeDIP reads for different subfamilies of ERV was calculated. The treated samples were normalised against the untreated sample. RNA sequencing was performed on VU40T cells treated with 500nM DAC or an equivalent volume of the vehicle. The aligned RNA was compared to a transposable element reference sequence using CuffCompare. For ERV elements with an FPKM>1, the 500nM DAC sample was normalized against the control. The ClustVis software was used to apply unit variance scaling across the 5mC and RNA data to visualise how 5mC changes correlated with RNA changes of ERV subfamilies [167]. Clustering was applied using orrelation distance and average linkage.

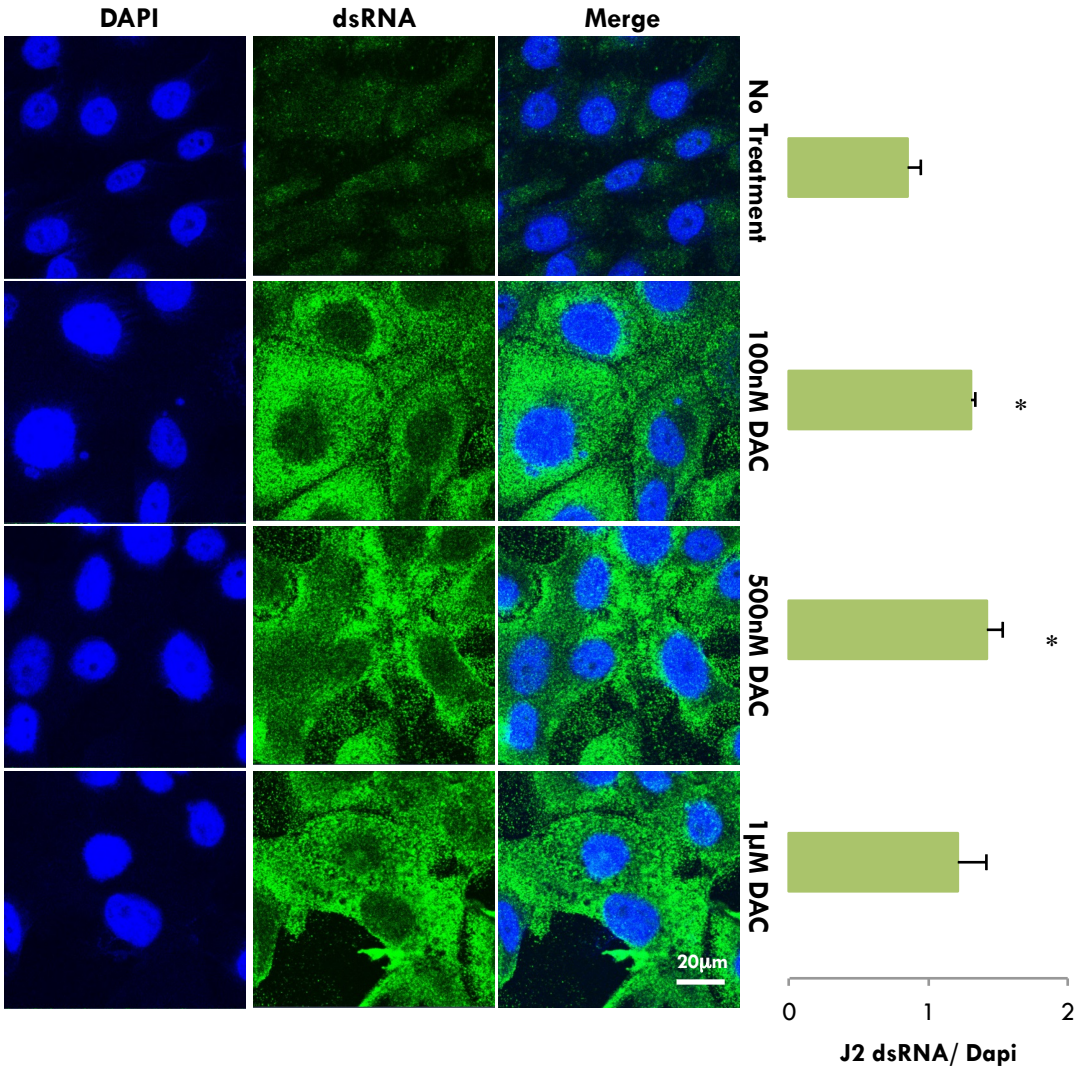



**Figure 6.11. DAC treatment increases the levels of double stranded RNA**

VU40T cells treated with 100nM, 500nM or 1μM DAC or an equivalent volume of vehicle control, were cultured on sterilised coverslips for 96h. These were then subject to immunofluorescent analysis using the J2 dsRNA antibody. Representative images are shown. Relative intensity was determined using the ImageJ software and the J2 dsRNA result was normalised against the corresponding DAPI result [129]. The graph shows the mean of three experiments, with error bars representing SEM. Significance was determined for each treated sample versus the vehicle treated control (NT) using a paired t-test with Bonferroni correction where *P<0.05.

**Table 6.2. DAC treatment is associated with an increased expression of genes involved in the immune system activation**

DeSeq2 analysis [163] was used to determine differences in the RNA sequencing results of VU40T cells treated with 500nM DAC and vehicle control. Those with a fold change <-1 were considered upregulated by DAC treatment, and those with a fold change >1 were considered downregulated by DAC treatment. These genes were subject to ontology analysis for biological processes and the top enriched terms are shown here. Terms of interest are boxed in red.

| Increased expression with DAC | | Decreased expression with DAC | |
|---|---|---|---|
| **GO Biological Process** | **Fold Enrichment** | **GO Biological Process** | **Fold Enrichment** |
| response to interferon-alpha (GO:0035455) | 8.83 | DNA replication-independent nucleosome assembly (GO:0006336) | 6.38 |
| type I interferon signalling pathway (GO:0060337) | 6.64 | viral transcription (GO:0019083) | 4.15 |
| cellular response to type I interferon (GO:0071357) | 6.64 | viral gene expression (GO:0019080) | 4.08 |
| response to type I interferon (GO:0034340) | 6.26 | translational initiation (GO:0006413) | 3.65 |
| endodermal cell differentiation (GO:0035987) | 5.89 | mitotic cell cycle checkpoint (GO:0007093) | 3.44 |
| endoderm formation (GO:0001706) | 5.1 | sister chromatid segregation (GO:0000819) | 3.16 |
| extracellular matrix disassembly (GO:0022617) | 4.97 | DNA replication (GO:0006260) | 3.04 |
| endoderm development (GO:0007492) | 4.24 | nuclear chromosome segregation (GO:0098813) | 2.97 |
| cornification (GO:0070268) | 4.17 | chromosome segregation (GO:0007059) | 2.72 |
| formation of primary germ layer (GO:0001704) | 3.96 | protein targeting (GO:0006605) | 2.67 |
| cell junction organization (GO:0034330) | 3.8 | ribosome biogenesis (GO:0042254) | 2.62 |
| cell junction assembly (GO:0034329) | 3.69 | translation (GO:0006412) | 2.61 |
| regulation of viral life cycle (GO:1903900) | 3.6 | peptide biosynthetic process (GO:0043043) | 2.54 |
| gastrulation (GO:0007369) | 3.48 | establishment of protein localization to organelle (GO:0072594) | 2.52 |
| extracellular matrix organization (GO:0030198) | 3.44 | DNA repair (GO:0006281) | 2.48 |
| extracellular structure organization (GO:0043062) | 3.43 | amide biosynthetic process (GO:0043604) | 2.48 |
| regulation of viral process (GO:0050792) | 3.37 | cell division (GO:0051301) | 2.45 |
| cell-cell junction organization (GO:0045216) | 3.21 | ribonucleoprotein complex biogenesis (GO:0022613) | 2.35 |
| regulation of symbiosis, encompassing mutualism through parasitism (GO:0043903) | 2.94 | cellular component disassembly (GO:0022411) | 2.34 |
| response to endoplasmic reticulum stress (GO:0034976) | 2.49 | peptide metabolic process (GO:0006518) | 2.27 |

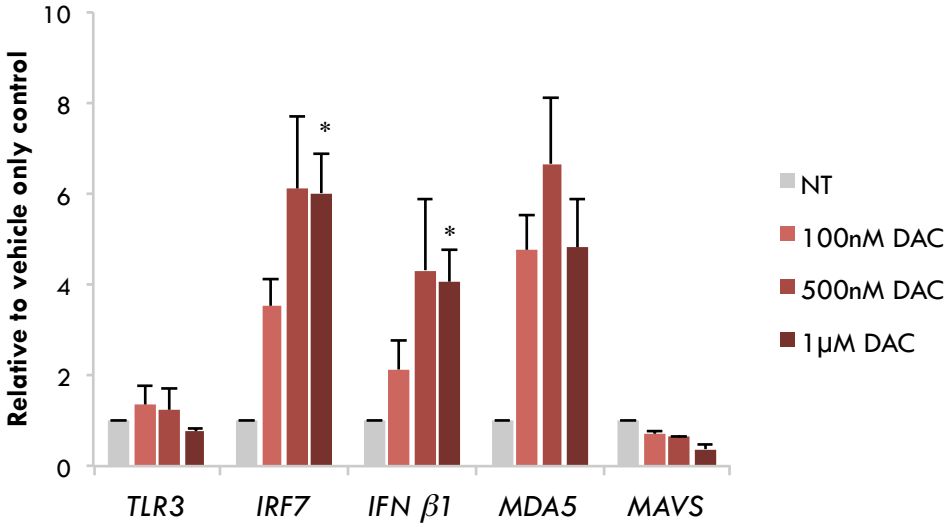



**Figure 6.12. DAC treatment increases the expression of dsRNA and viral response genes**

RNA was extracted and qRT-PCR performed on VU40T cells treated with 100nM DAC, 500nM DAC or an equivalent volume of vehicle. Results are shown relative to the vehicle control with error bars representing SEM (n=3). SPSS software was used to calculate paired t-tests with Bonferroni correction of each treated samples against the untreated control (NT) where * $p<0.05$.

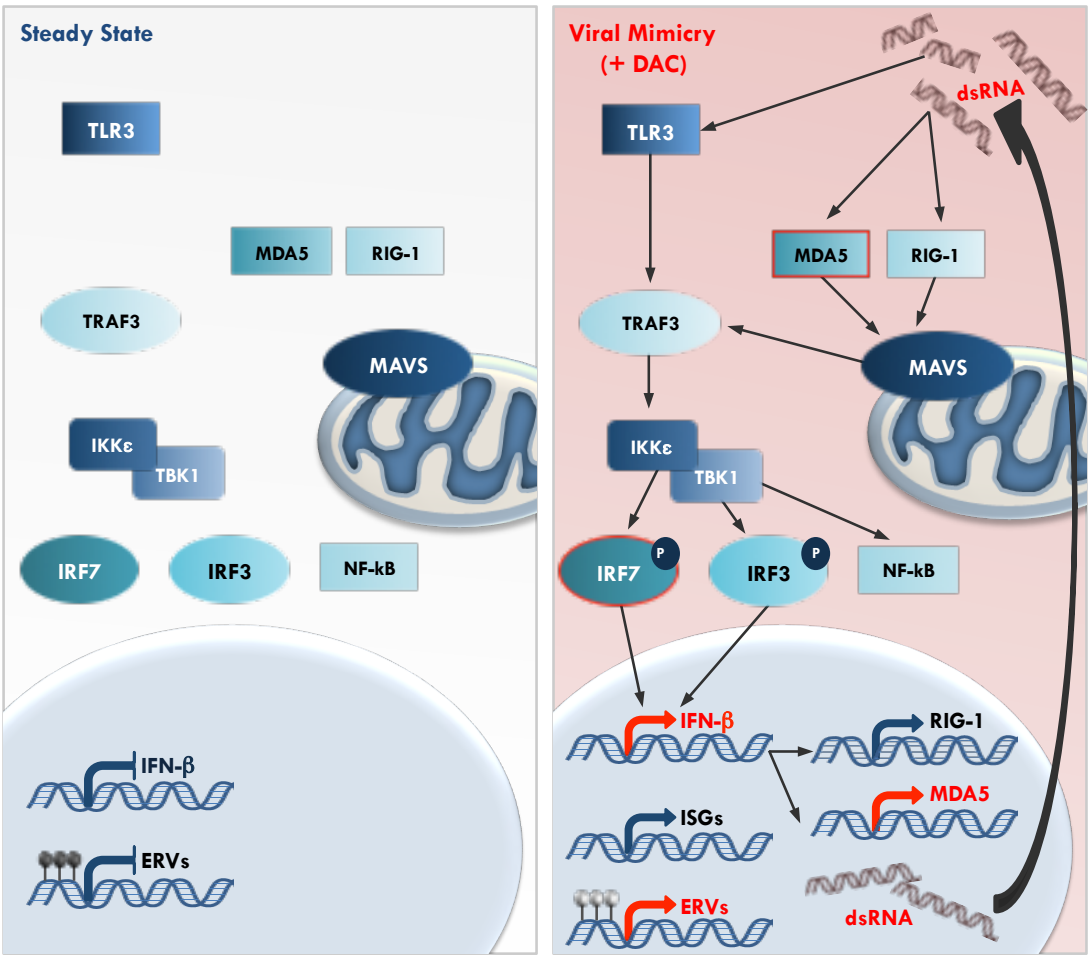



**Figure 6.13. In VU40T cells DAC induces a viral mimicry response**

DNMTi treatment is understood to cause a viral mimicry response. This is represented as a schematic. The effect of DAC on this pathway in VU40T cells, as discussed in the results above, is shown in red. In VU40T cells DAC demethylated ERVs, leading to an increase in dsRNA. A corresponding increase in expression of *IFIH1* (MDA5), *IRF7*, *IFN-β* and interferon related genes were also observed in VU40T cells after DAC treatment.

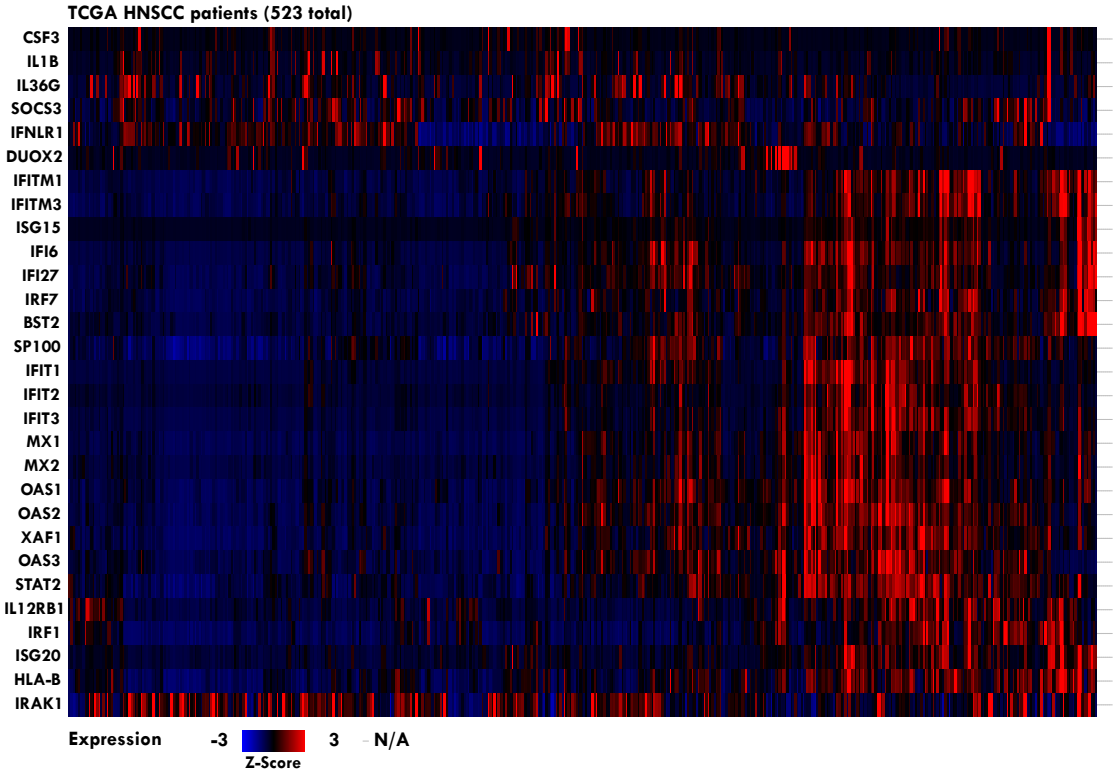



**Figure 6.14. Immune genes upregulated by DAC treatment cluster together in HNSCC patient samples**

Interferon pathway genes upregulated by DAC treatment were entered into the cBioPortal website [222, 271]. The heatmap shows the expression of these genes (rows) in 523 HNSCC patient samples (columns) from TCGA data, which are sorted based on RNA levels. Scaling shows mRNA expression Z-scores.

## 6.8. In HNSCC cells 5mC and 5hmC are partially restored at Alu elements after DAC treatment

Chapter 5 has discussed how 5mC and 5hmC are lost from repetitive DNA, particularly Alu elements, in HNSCC. Therefore the MeDIP and hMeDIP results for DAC treated VU40T cells were separated based on detailed annotations. Interestingly, treatment with DAC increased the proportion of both 5mC and 5hmC at Alu elements (Figure 6.15). As discussed above, the proportion of 5mC at non-repeat regions (e.g. exons and promoters) and LTRs (e.g. ERVs) was decreased by DAC (Figure 6.15). Conversely, 5hmC was lost from simple repeats (Figure 6.15). The relationship between 5mC and 5hmC after DAC treatment is distinct for SINEs compared with other forms of repetitive DNA (Figure 6.16). For the majority of repetitive DNA classes 5mC is decreased, and 5hmC unchanged by DAC (Figure 6.16). However, SINEs (and in particular Alu elements) show increased methylation and hydroxymethylation after DAC treatment (Figure 6.16). Interestingly, L1 elements also show an increase in 5mC and 5hmC (Figure 6.16).

MeDIP and hMeDIP peaks were separated into three groups based on the response to DAC treatment: peaks unchanged by DAC (constant); peaks gained by DAC treatment (DAC-specific); and peaks lost by DAC treatment (NT-specific) and the detailed annotations of these groups was determined (Figure 6.17). The results were similar for both low and high dose DAC treatment (Figure 6.17). Encouragingly, regions that lose 5mC after DAC are primarily associated with non-repeat regions, suggesting that DAC treatment does not severely alter genome stability (Figure 6.17). However, regions that gain 5mC after DAC are enriched in Alu elements

(Figure 6.17). Interestingly, for 5hmC both NT-specific and DAC-specific groups were associated with Alu elements, while the constant group was not, suggesting that Alu elements are more prone to changes in 5hmC than other regions of the genome (Figure 6.17).

The proportion of Alu element subfamilies within the MeDIP and hMeDIP data was determined (Figure 6.18). DNA methylation was lost from the oldest Alu family, Alu J (Figure 6.18). Both 5mC and 5hmC were gained at both Alu S and Alu Y elements (Figure 6.18). These results suggest that in VU40T cells DAC treatment is partially restoring the distribution of 5mC and 5hmC at Alu elements to that of normal oral keratinocytes (HOK cells) (Figure 6.18).

The VU40T transcriptome, before and after treatment with 500nM DAC was compared with a transposable element reference sequence using CuffCompare to determine the expression of Alu element subfamilies (Figure 6.19). In general, increased 5mC and 5hmC after 1μM DAC treatment correlated with decreased RNA after 500nM DAC (Figure 6.19). This suggests that in HNSCC, treatment with DAC can re-establish the epigenetic control of Alu element expression.

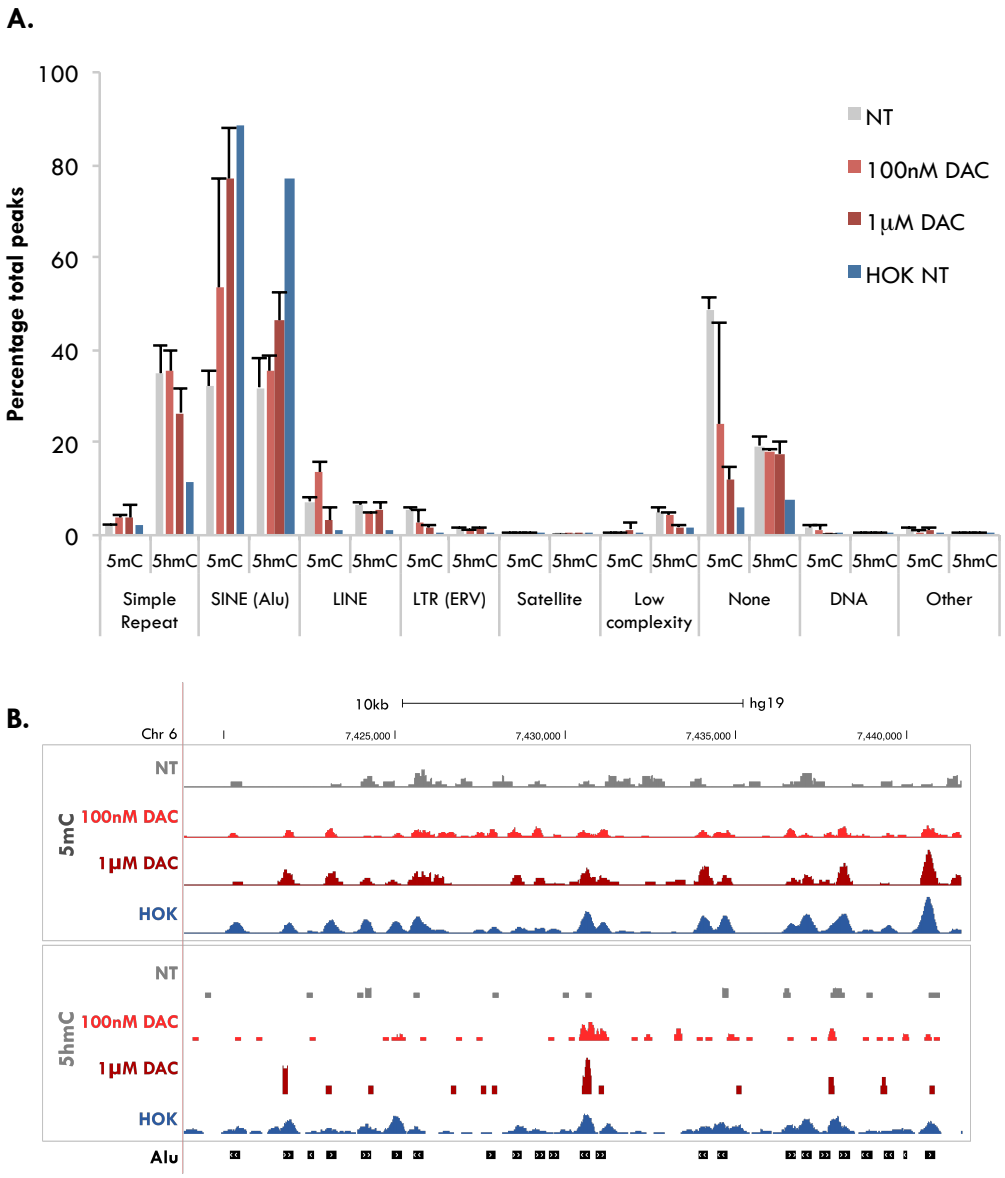



**Figure 6.15. DNA methylation and hydroxymethylation at repetitive DNA after treatment with DAC**

MeDIP and hMeDIP peaks were separated based on detailed annotation. For MeDIP detailed annotations were assigned using MACS by Dr W. Croft and hMeDIP peaks were assigned using Homer by Dr S. Clokie **A.** The bar graph shows the relative proportion of repetitive elements in HOK cells and VU40T cells treated with varying concentrations of DAC. MeDIP sequencing experiments for VU40T cells were performed in duplicate and compiled for analysis, and in triplicate for hMeDIP sequencing experiments. Error bars show STD. For HOK cells, MeDIP and hMeDIP experiments were performed in triplicate and pooled prior to sequencing. **B.** An example UCSC browser shot showing 5mC and 5hmC lost between HOK (blue tracks) and VU40T cells (grey tracks) and regained after treatment with DAC (red tracks).

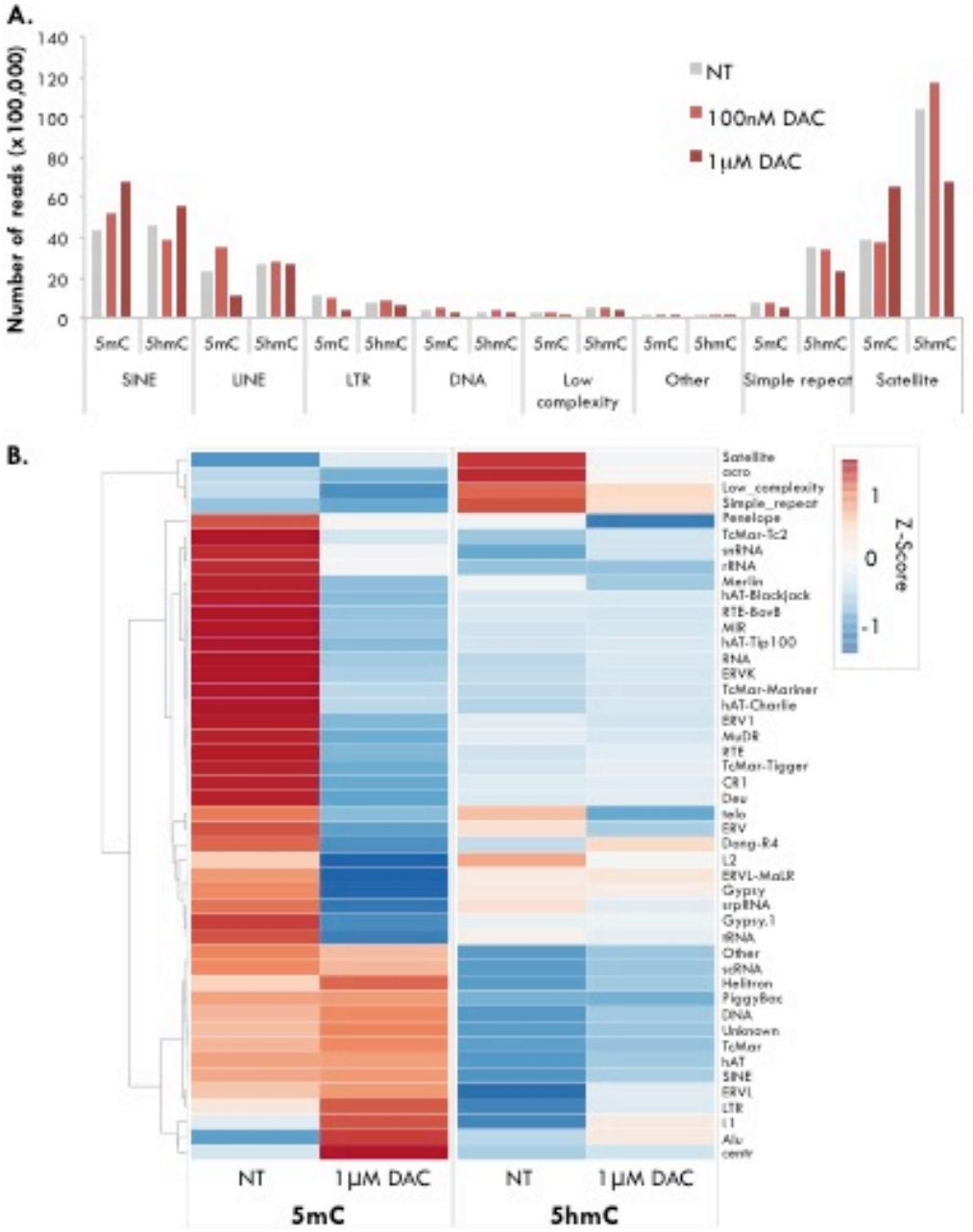



**Figure 6.16. Changes in methylation and hydroxymethylation at repetitive DNA in VU40T cells after DAC treatment**

The total number of both uniquely and non-uniquely mapped 5hmC and 5mC reads within repetitive element subfamilies was determined by counting the number of reads intersecting with the 'repeatmasker' data from the UCSC browser, using '*multicov*' function from the bedTools suite (Dr S. Clokie). **A.** Bar graph showing the number of 5hmC (n=3) and 5mC (n=2) reads for each repetitive element family. **B.** The ClustVis software was used to generate a heatmap where rows were clustered using correlation distance and average linkage for untreated (NT) and 1µM DAC treated samples [167]. Unit variance scaling was applied across 5mC and 5hmC rows.

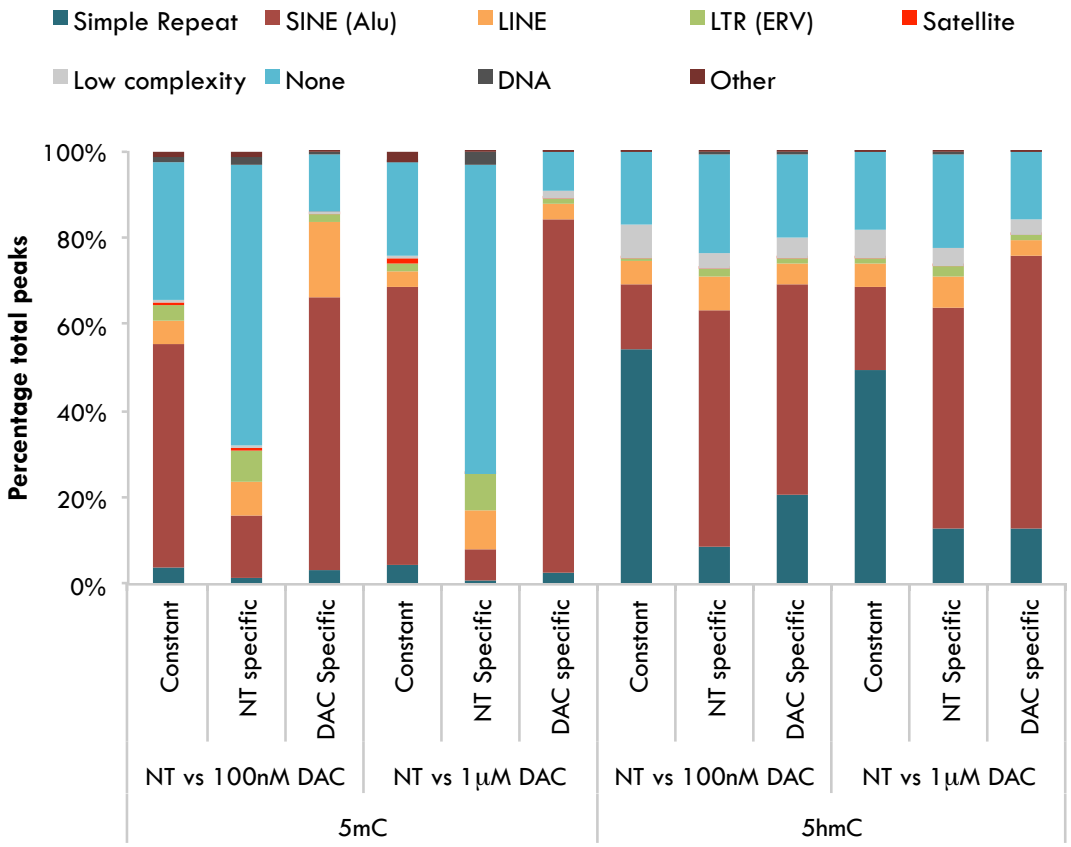



**Figure 6.17. Changes in the distribution of 5mC and 5hmC over repetitive DNA in VU40T cells in response to DAC**

VU40T cells MeDIP and hMeDIP peaks were separated based on the response to high (1μM) and low (100nM) dose DAC. The detailed annotations for each group were then determined and are shown as a stacked column bar graph to represent the relative levels of repetitive elements classes within each group.

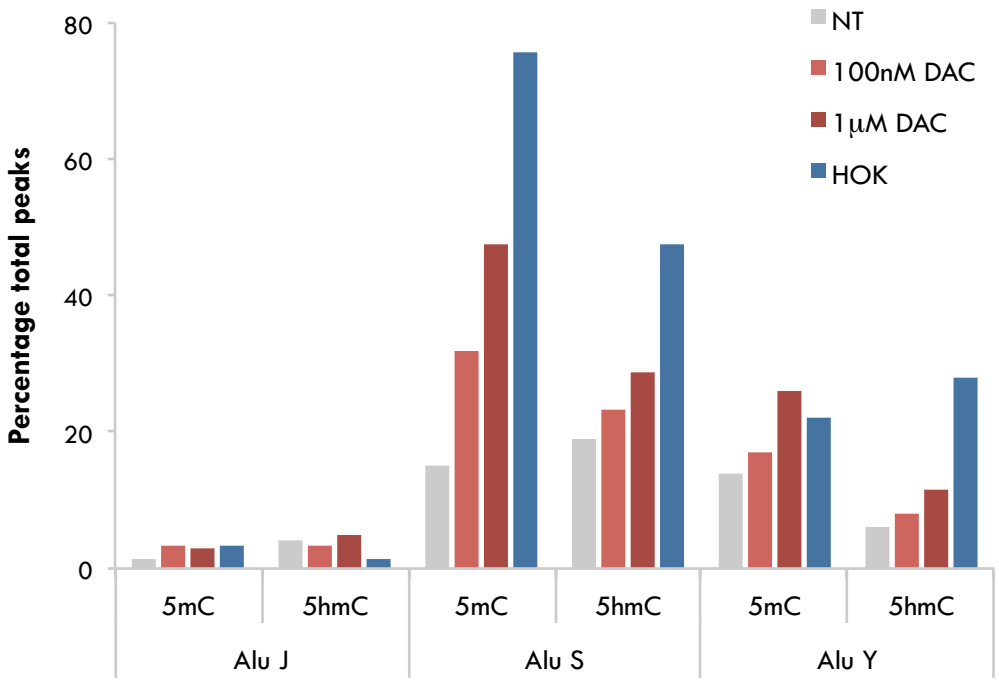



**Figure 6.18. Differential enrichment of 5mC and 5hmC at Alu element families in VU40T cells treated with DAC**

The results of MeDIP (n=2) and hMeDIP (n=3) sequencing of HOK and untreated (NT) and 100nM and 1μM DAC treated VU40T cells were separated based on the presence of an Alu element. The proportions of each Alu element family (J, S and Y) relative to total 5mC or 5hmC peaks are shown as a bar graph.

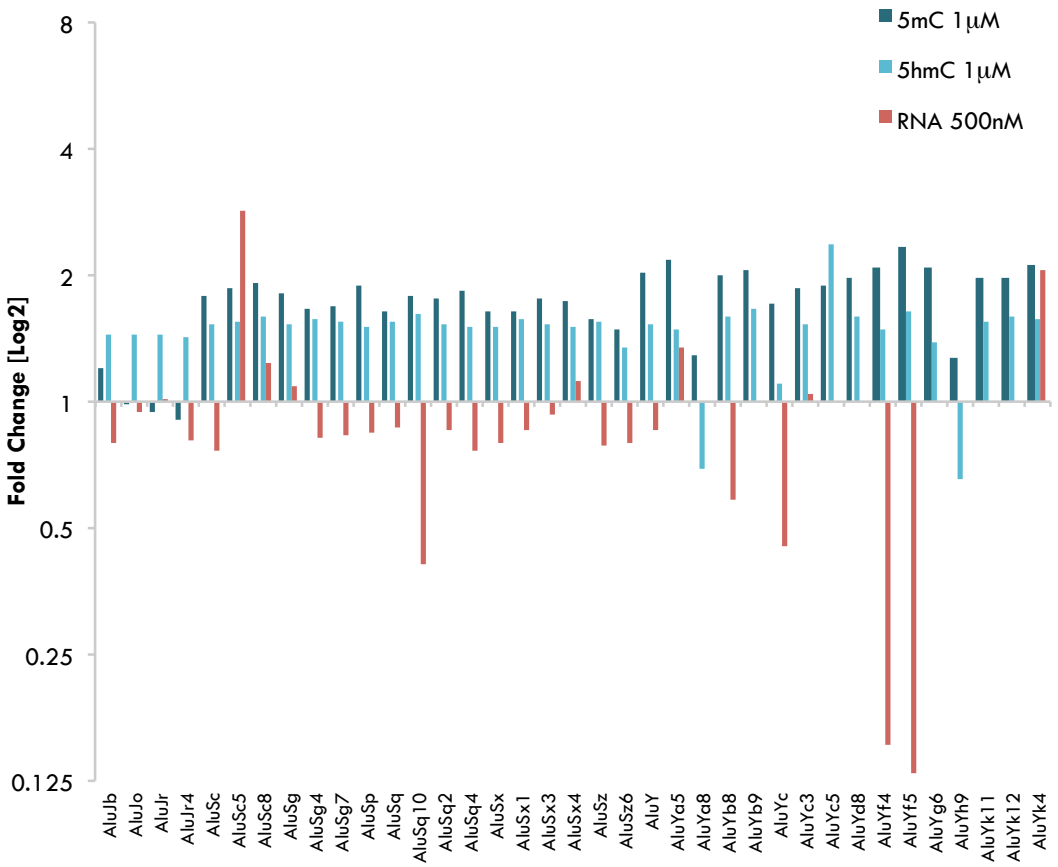



**Figure 6.19. In HNSCC, increased 5mC and 5hmC at Alu elements correlates with decreased Alu expression**

The number of MeDIP and hMeDIP reads for different subfamilies of Alu was calculated. The 1μM DAC treated samples were normalised against the untreated sample. RNA sequencing was performed on VU40T cells treated with 500nM DAC or an equivalent volume of the vehicle. The aligned RNA was compared with a transposable element reference sequence using CuffCompare. For Alu elements with an FPKM>1, the 500nM DAC sample was normalized against the control. A bar graph shows the relationship between 5mC and 5hmC changes and Alu element expression as a Log2 fold change after DAC treatment.

## 6.9. DAC treatment does not alter the distribution of DNA methylation in HOK cells

Treatment of primary human oral keratinocytes (HOK cells) with high (1µM) or low (100nM) dose DAC caused only minimal changes in the distribution of 5mC (Figure 6.20). Similarly, HOK cells did not show the 5mC distribution changes characteristic of DAC treatment in VU40T cells: 5mC was not gained in promoters and exons; the CpG content and distance to the TSS was unaltered; and 5mC was not gained at Alu elements (Figure 6.21-Figure 6.23). As discussed in Chapter 3, HOK cells are less responsive to DAC treatment than VU40T cells. This could be attributed to the lack of 5mC changes described here (Figure 6.21-Figure 6.23).

The 5mC results discussed in this chapter can be summarised by the UCSC browser shot shown in Figure 6.24. In HNSCC, 5mC is gained in promoters, CGI and exons and lost from Alu elements (Chapter 5, Figure 6.24). In VU40T cells treatment with DAC reduces the 5mC at the CGI and increases 5mC over Alu elements (Figure 6.6, Figure 6.15, Figure 6.24). However, in HOK cells 5mC peaks are enriched in Alu elements, and scarcely altered by treatment with DAC (Chapter 5, Figure 6.24).

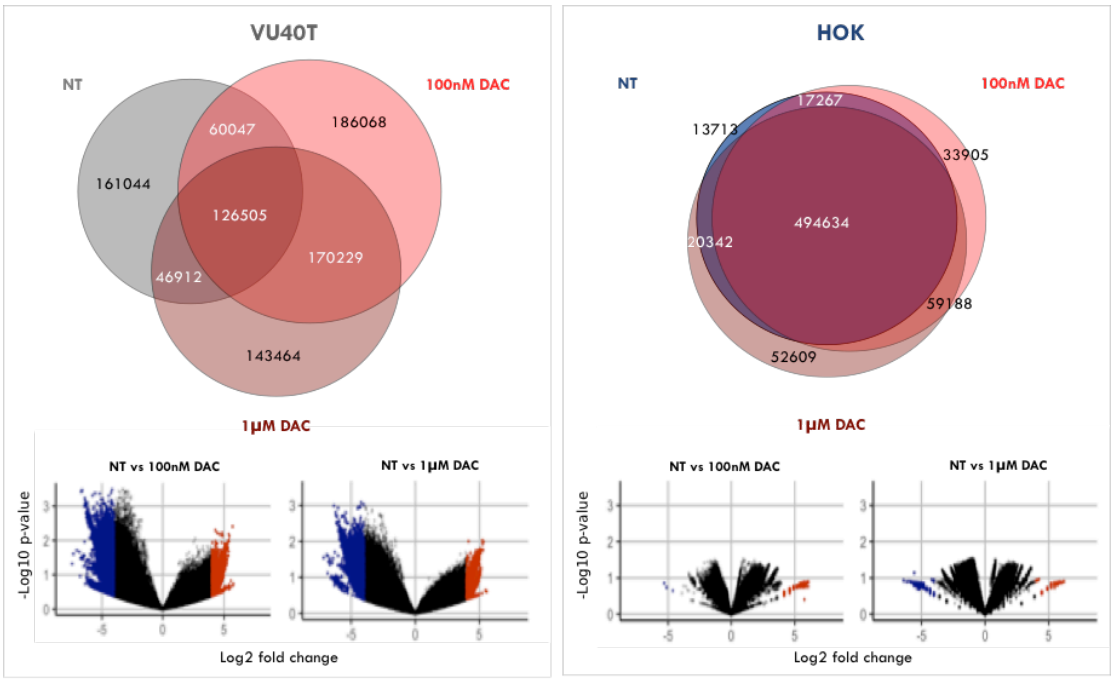



**Figure 6.20. DAC treatment leads to minimal changes in 5mC in HOK cells**

MeDIP was performed on VU40T and HOK cells after 96h treatment with 1µM DAC, 100nM DAC or an equivalent volume of vehicle. **Top**: The overlap of 5mC-enriched regions between treatment groups is shown as a Venn diagram. Numbers show total numbers of peaks present in each category. **Bottom:** Volcano plots show 5mC changes between untreated and DAC treated cells. For VU40T cells the MeDIP sequencing experiment was performed in duplicate and the results combined for analysis. For HOK cells, the MeDIP experiment was performed in triplicate, but combined prior to sequencing. Peaks were identified by Dr Wayne Croft.

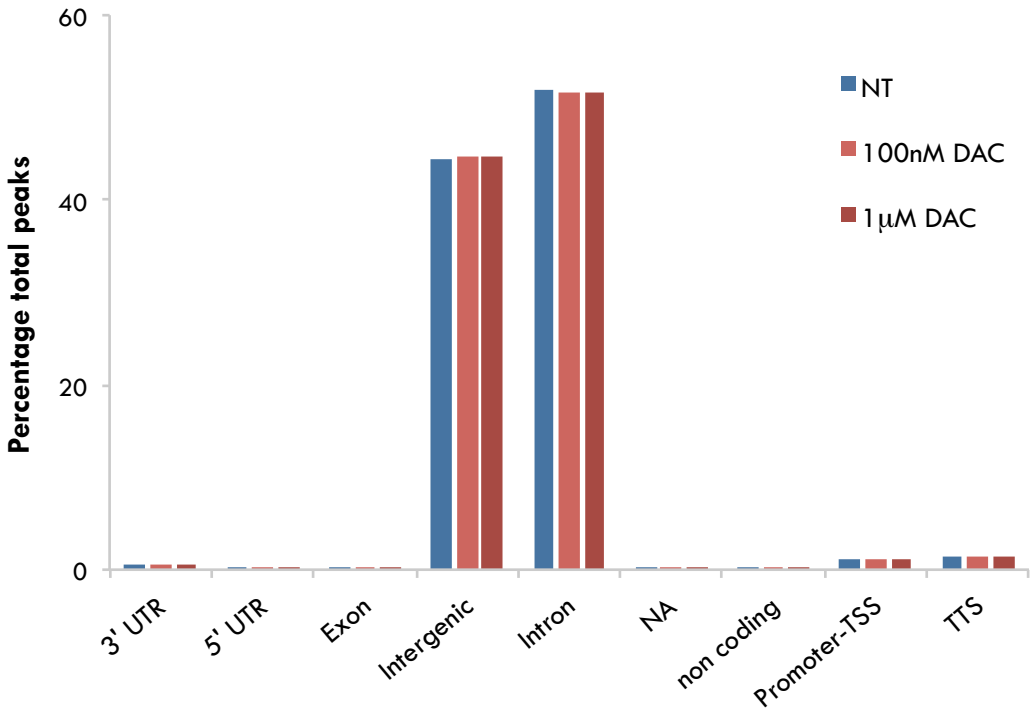



**Figure 6.21. The distribution of 5mC in HOK cells after DAC treatment**

HOK MeDIP results were separated by genomic annotations and are shown as a bar graph relative to the total number of peaks for each treatment group. NT=vehicle only treated HOK cells.

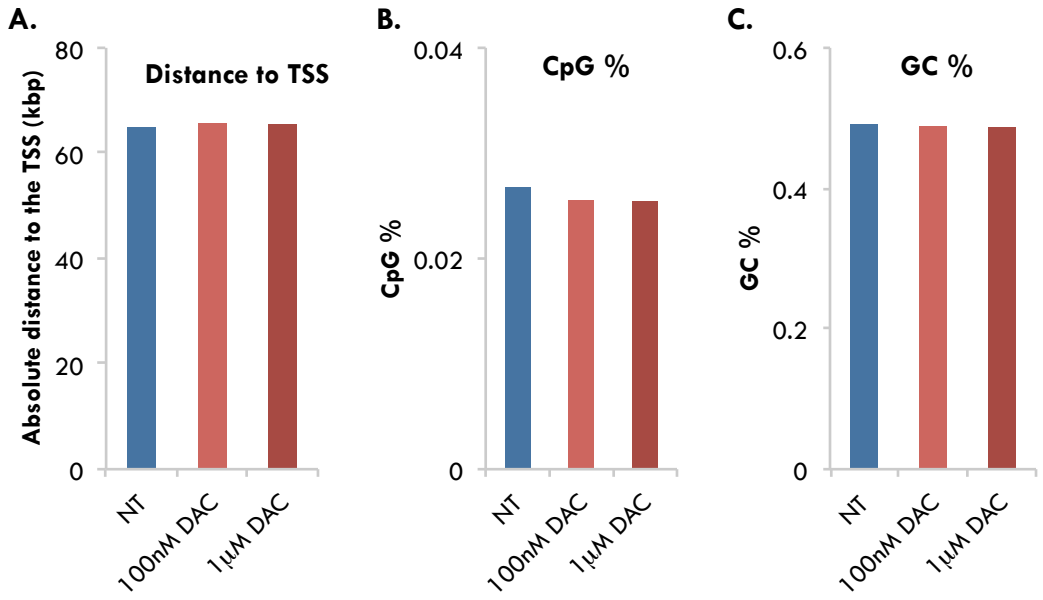



**Figure 6.22. Characteristics of HOK MeDIP peaks after DAC treatment**

For HOK 5mC-enriched regions the average distance to the TSS (**A**); CpG% (**B**); and GC% (**C**) were calculated. HOK MeDIP experiments were performed in triplicate and pooled prior to sequencing, while VU40T MeDIP experiments were sequenced separately and pooled for analysis.

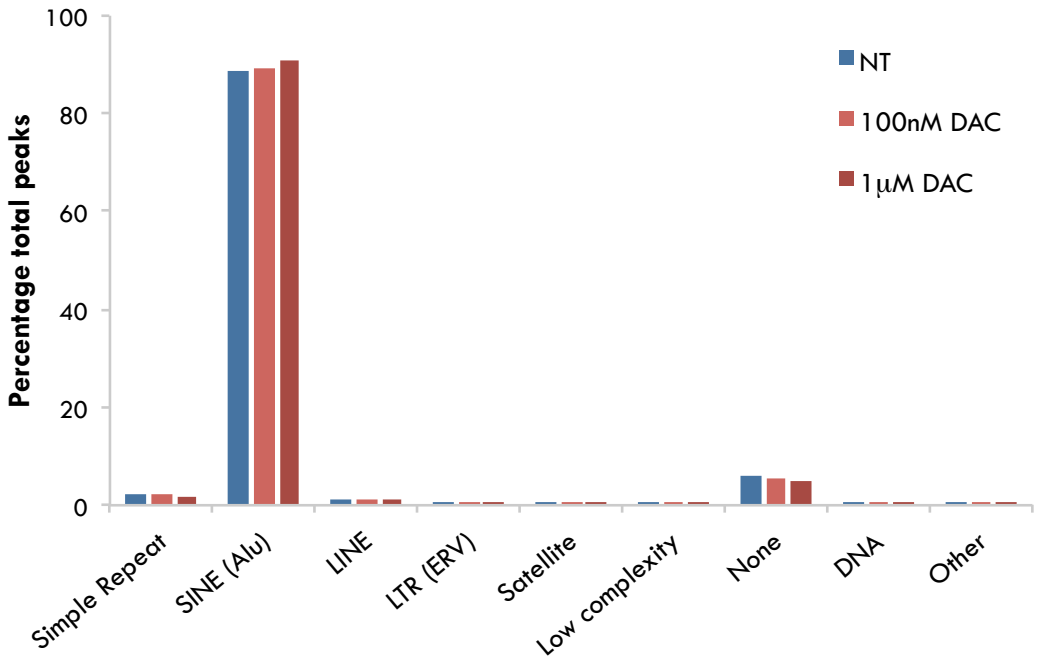



**Figure 6.23. DNA methylation of repetitive DNA in HOK cells after treatment with DAC**

HOK MeDIP peaks were separated based on detailed annotations and are shown as a bar graph representing the proportion of each repeat class in the total number of peaks. HOK MeDIP experiments (n=3) were pooled prior to sequencing, while VU40T MeDIP experiments (n=2) were sequenced individually and pooled for analysis.

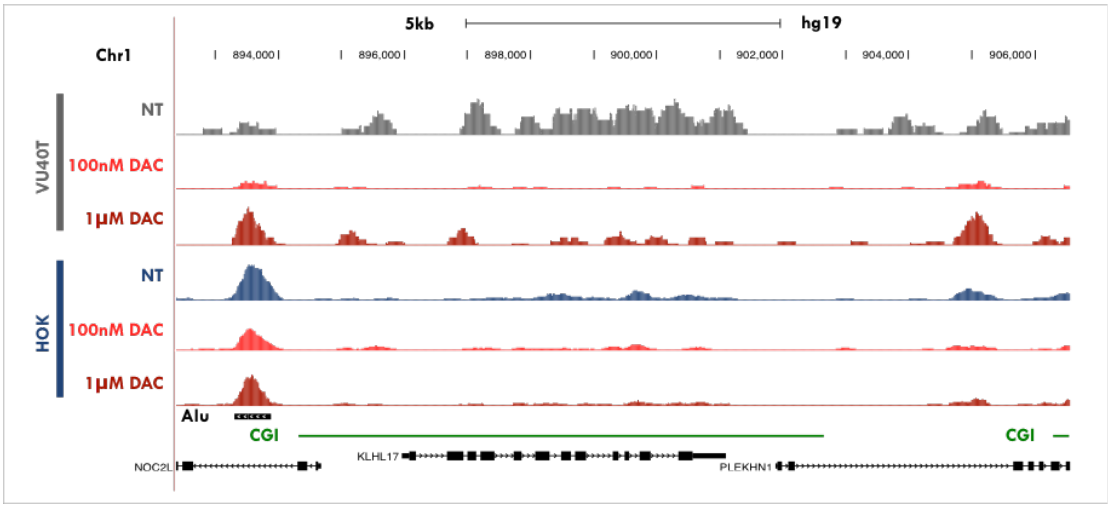



**Figure 6.24. DAC treatment alters the distribution of 5mC in VU40T cells, but not HOK cells.**

An example UCSC browser shot of 5mC in VU40T and HOK cells before and after DAC treatment in relation to genes, Alu elements (black bars) and CpG islands (green bars). While VU40T specific 5mC peaks are lost in DAC treated cells, HOK specific peaks are not.

## 6.10. Discussion and conclusion

The efficacy of DAC treatment has been attributed to its ability to demethylate DNA. Indeed, in HNSCC cell lines, DAC causes a global reduction in 5mC. The work presented here demonstrates that in an HNSCC cell line DAC does not indiscriminately demethylate the genome. Instead, DAC partially restores 5mC and 5hmC patterns to those seen in normal oral keratinocytes (Figure 6.25, Figure 6.26). Regions that gain 5mC in cancer are understood to be more susceptible to demethylation via DAC [96, 263, 264]. In VU40T cells, promoters, CGIs and ERVs gain methylation. This was lost after DAC treatment. Additionally, an unexpected relative increase of 5mC and 5hmC at Alu elements was observed. This led to changes in Alu expression and could affect functioning of adjacent regulatory elements.

In a comparison of the methylome of HNSCC cells with that of normal oral keratinocytes, 5mC was detected more frequently in promoters, exons and CpG islands (CGI) (Chapter 5). This is in keeping with the literature and is understood to contribute to HNSCC development and progression [11, 99-102]. In VU40T cells, losses in 5mC after DAC primarily occur at non-repeat regions. DAC demethylates promoters, exons and CGIs to levels similar to those of HOK cells. From these results, it can be hypothesised that in this system DAC does not cause global 5mC loss and genomic instability and instead targets aberrantly methylated promoters and CGIs. However, when the methylation and RNA levels of tumour suppressor genes previously implicated in HNSCC were examined, promoter demethylation by DAC was not sufficient to reactivate gene expression. The epigenetic control of gene expression is a multi-layered process, therefore, in addition to 5mC loss, changes in

histone modifications, chromatin structure and transcription factor binding could be needed to activate repressed genes [27]. Further work is required to determine whether other genes are reactivated by promoter demethylation after DAC and what pathways these are involved in. However, the results presented here suggest that promoter demethylation is not the primary mechanism of DAC efficacy in VU40T cells.

Interestingly, the response to low dose DNMTi treatment is frequently delayed and is accompanied by the expression of numerous genes not controlled by a methylated promoter [158, 265]. This led to the proposition that DNMTi can induce viral mimicry, where ERVs are expressed as dsRNA that triggers an immune response [158, 180, 266-268]. Previous work in this area has focused on the delayed response, where methylation and RNA are measured several days after removal of the drug [158, 180]. However, in VU40T cells, 96h treatment with both high and low dose DAC is sufficient to directly cause some aspects of viral mimicry: ERVs are demethylated and expressed; dsRNA is increased and; immune and anti-viral genes are upregulated. This suggests that the efficacy of DAC in VU40T is at least partially attributed to viral mimicry, however, further work involving knockdowns of interferon or viral pathway genes are needed to confirm this.

In normal oral keratinocytes (HOK cells), Alu elements are enriched in 5mC and 5hmC (Chapter 5). Methylated and hydroxymethylated Alu elements are found in gene rich regions where they form clusters around regulatory elements, potentially acting as boundaries which protect the regulatory regions from invasive methylation (Chapter 5, Figure 6.26). In HNSCC, 5mC and 5hmC are dramatically reduced at Alu;

however, this is partially repaired by DAC treatment (Figure 6.26). Increased 5mC and 5hmC was observed in VU40T cells after both high and low dose DAC. Therefore, DAC treatment helps to re-establish the control of Alu elements by DNA modifications. This corresponds with a decrease in Alu RNA, suggesting that the modifications function in repressing Alu transcription. Alu RNA can insert into the genome causing a variety of mutations and Alu RNA that cannot retrotranspose can form small RNAs that silence Alu containing genes [84, 258]. In VU40T cells, DAC treatment corresponds with decreased expression of both Alu S elements that lack retrotransposition potential and Alu Y elements that remain active. Therefore, DAC could help protect HNSCC cells from the mutagenic and silencing effects of Alu RNA while restoring the control of Alu boundary elements by DNA modifications.

The majority of 5mC changes in HNSCC are distinct from 5hmC changes. These findings are in keeping with the literature where cancer and tissue specific differences in 5mC inversely correlate with 5hmC [260, 261]. The same can be observed in VU40T cells after DAC treatment; while 5mC is decreased in promoters and CpG rich regions close to the TSS, 5hmC is decreased in intergenic regions and CpG poor regions further from the TSS (Figure 6.25). This could be attributed to decreased *TET2* RNA after DAC. The exception is at Alu elements, where both 5mC and 5hmC are decreased in HNSCC and increased by DAC treatment (Figure 6.25). This suggested that Alu elements have a distinct form of epigenetic control.

When comparing the results of MeDIP and hMeDIP sequencing experiments, it should be noted that the resulting differences only represent differences in the proportion of reads at each region and not absolute differences. For example, for the

Alu element result discussed here, it cannot be determined whether 5mC and 5hmC are gained at Alu elements after DAC, or if less DNA modifications are lost here than in the rest of the genome. Therefore, instead of being the sites of active modification gain after DAC; Alu elements may be protected from global modification loss. Mechanistically, this could be attributed to preferential degradation of DNMT1 by DAC over DNMT3A and -B [272]. DNMT3B is specifically enriched at the gene bodies where it prevents spurious transcription from alternate start sites and transposable elements [46, 273]. Interestingly, this is frequently where Alu elements are located, and therefore could partially explain the protection of Alu elements from global DNA modification loss described here. However, further experiments involving oxidative bisulfite conversion of specific Alu elements are required in order to determine whether DAC treatment causes DNA modification gain at Alu or if Alu elements are protected from global DNA modification loss.

In previous chapters the potential efficacy of DAC treatment in HNSCC has been described (Chapter 3). While some HNSCC cell lines respond to nanomolar concentrations of the drug, normal oral keratinocytes (HOK cells) are resistant to treatment. In this chapter these results have been expanded upon to describe DNA modification changes in response to DAC. Encouragingly, while HNSCC cells show considerable changes in the distribution of 5mC after DAC treatment, normal oral keratinocytes do not. DAC partially restores the distribution of 5mC and 5hmC in VU40T cells to those of normal oral keratinocytes and re-establishes the positioning of both modifications at Alu elements. Therefore, the efficacy of DAC treatment in VU40T cells could be attributed both to the induction of a viral mimicry response

through demethylation of ERVs and to protection from the mutagenic and silencing effects of Alu RNA.

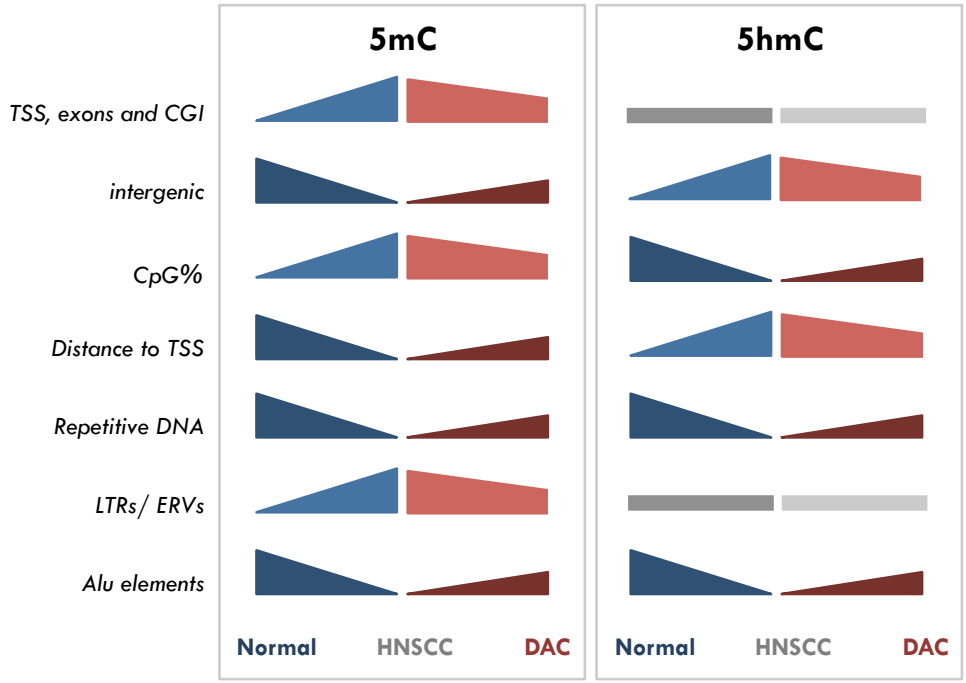



**Figure 6.25. The distribution of 5mC and 5hmC is altered in HNSCC, and partially repaired by treatment with DAC**

Diagram representing changes in the distribution of 5mC 5hmC in normal, HNSCC and DAC treated HNSCC cells. Blue bars show how the percentage of 5(h)mC peaks associated with different characteristics change between normal and HNSCC cells, while red bars represent the effect of DAC treatment on these characteristics in HNSCC. Aspects of 5(h)mC distribution lost between normal and HNSCC cells are partially regained by treatment with DAC.

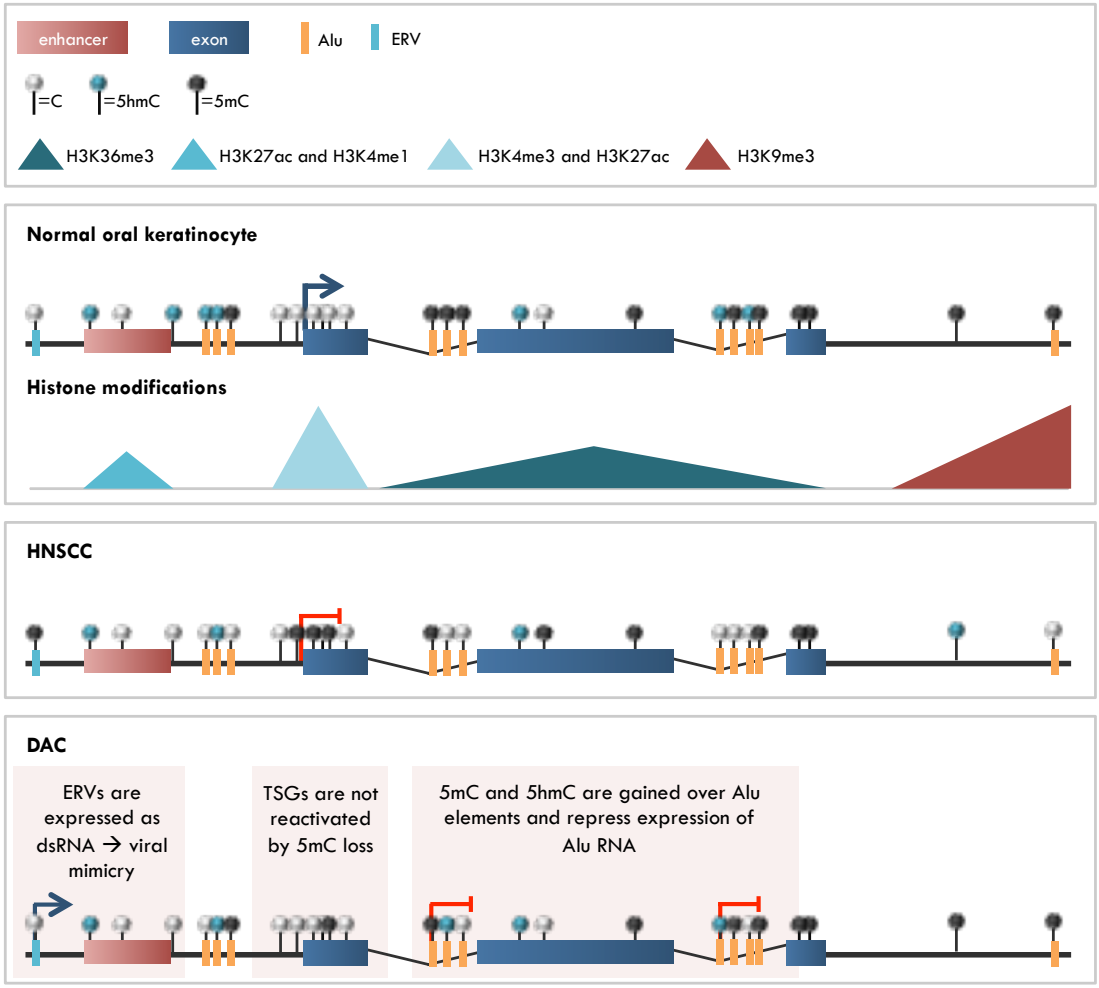



**Figure 6.26. The distribution of 5mC and 5hmC in normal oral keratinocytes, HNSCC cells and HNSCC cells treated with DAC**

**Top panel:** The distribution of 5mC and 5hmC in a gene rich region in HOK cells is represented as a schematic alongside the associated histone modifications. **Middle panel:** The distribution of 5mC and 5hmC are changed in HNSCC. **Bottom panel:** In HNSCC DAC treatment partially restores the distribution of 5mC and 5hmC to that of normal oral keratinocytes. This has three main outcomes represented by pink boxes: 1) DAC demethylates ERVs, allowing them to be expressed as dsRNA and activate the immune system; 2) DAC demethylates promoters and CGIs. However, common HNSCC tumour suppressor genes are not reactivated by DAC treatment. 3) DAC increases the proportion of 5mC and 5hmC at Alu elements. A corresponding decrease in gene expression is also seen at Alu element subfamilies that gain 5mC and 5hmC after DAC treatment.

# Chapter 7. Discussion and Future Work

## 7.1. In normal oral keratinocytes Alu elements are enriched in DNA methylation and hydroxymethylation

An important finding of this thesis was the enrichment of both 5mC and 5hmC at Alu elements in normal oral keratinocytes (HOK cells). Methylated and hydroxymethylated Alu elements were found in clusters in gene rich regions, neighbouring regulatory elements and active histone modifications. In the human genome Alu elements are most prevalent in GC-rich DNA, neighbouring and overlapping genes [79]. While Alu elements are excluded from unmethylated domains and CpG islands, they are found at the boundaries of these domains [256]. Therefore, in HOK cells, clusters of Alu elements may act as boundary elements controlled by DNA modifications. Moreover, Alu elements can facilitate nucleosome positioning and 5hmC increases the binding affinity of DNA for the histone octamer [239, 274]. Therefore, in HOK cells, methylated and hydroxymethylated Alu elements may anchor nucleosomes to regions neighbouring genes and regulatory elements; preventing transcription of the transposable element while protecting the neighbouring regulatory region from methylation and nucleosome invasion.

Interestingly, VU40T cells are characterised by less 5mC and 5hmC at Alu elements, and more 5mC at CGI. Similarly, in VU40T cells, 5mC is lost from CGI after DAC treatment and 5mC and 5hmC are gained at Alu elements. Therefore, it would be

interesting to determine whether CGI methylation changes correspond with changes in DNA modifications at flanking Alu elements. Furthermore, correlating the MeDIP and hMeDIP sequencing results with the results of DNase hypersensitive site sequencing, before and after DAC, would help determine how changes in Alu methylation or hydroxymethylation correspond with changes in chromatin structure. Additionally, a variation of the CRISPR-Cas9 system could be used to target Alu elements for demethylation and investigate the effect of this on nucleosome positioning and methylation of neighbouring regulatory regions. In this system, the catalytic domain of TET would be fused to dCas9 and guide RNAs designed against specific Alu elements [275].

Only a very small number of Alu elements are expressed as RNA, and the majority lack retrotransposition potential [245, 259]. This is understood to be due to epigenetic silencing. Compared with HOK cells, the enrichment of Alu elements in the VU40T MeDIP and hMeDIP data is considerably reduced. Lack of DNA modifications at Alu elements could lead to their expression, causing increased mutations and silencing of Alu containing genes [84, 258]. To confirm this hypothesis, the HOK transcriptome could be compared with the VU40T transcriptome to determine whether increased Alu RNA is observed in the HNSCC cell line. Encouragingly, DAC treatment partially restored 5mC and 5hmC at Alu elements. This corresponded with a decrease in Alu RNA expression. Therefore, DAC may partially restore the control of Alu element expression in HNSCC cells. To expand upon this result and determine the potential implications, both the

retrotransposition potential of expressed Alu elements, and the expression of Alu element containing genes would need to be investigated.

In conclusion, I propose that in gene rich regions, clusters of Alu elements act as boundaries that protect neighbouring regulatory regions from methylation and nucleosome occupancy. DNA modifications control both the expression of the underlying retrotransposons and the function of the boundary element. Indeed, after DAC treatment the gain in DNA modifications at Alu elements corresponds with both decreased Alu RNA levels and decreased CpG island methylation.

## 7.2. DAC as a potential therapeutic in HNSCC

### 7.2.1. DAC is effective in a subset of HNSCC cell lines

The overall aim of this thesis was to investigate the potential efficacy of DAC in the treatment of HNSCC. Encouragingly, half of the HNSCC cell lines tested responded to nanomolar concentrations of the drug. However, 2 cell lines were unresponsive. This corresponded with the ability of DAC to demethylate DNA, suggesting that efficacy is dependent on the uptake, activation or retention of the drug. The drug transport genes *SLC15A2, SLC22A5* and *SLC29A1* are candidate biomarkers for response to DAC in HNSCC, as their expression is higher in DAC responsive cell lines compared with unresponsive cells. However, the number of HNSCC cell lines examined here is relatively small and follow up experiments with more DAC treated samples is needed to determine which, if any, of these genes could predict response. This could be followed up with knock-down or knock-in experiments to determine if removing the

transporter makes DAC-responsive cell lines unresponsive to DAC treatment or adding the transporter makes unresponsive cell lines responsive.

### 7.2.2. DAC shifts the distribution of 5mC and 5hmC back towards that of normal cells

MeDIP and hMeDIP sequencing was used to determine the distribution of 5mC and 5hmC in normal oral keratinocytes (HOK) and an HNSCC cell line, VU40T, and to investigate the effect of DAC on this distribution. Interestingly, DNA methylation and hydroxymethylation changes that occurred in HNSCC were partially restored by treatment with DAC.

Promoters, exons and CpG islands gain methylation in HNSCC [11, 99-101]. Encouragingly, DAC preferentially demethylates non-repeat regions such as promoters and exons, and the enrichment of 5mC at CGI is reduced by DAC. However, in HNSCC cells, decreased promoter 5mC did not correlate with an increase in tumour suppressor gene (TSG) expression. Therefore, combined treatment with other epigenetic drugs may be required to activate these TSG. However, in VU40T cells, DAC alone is sufficient to decrease viability, therefore, it can be hypothesised that the reactivation of these TSGs is not required for DAC efficacy.

In VU40T cells, DAC results in a decrease in 5mC at ERVs, increased dsRNA and increased expression of immune responsive genes. These results suggest that viral mimicry is occurring [158, 180]. This process appears to centre on the interferon pathway, with increased expression of *IFN-β* and interferon pathway-associated

genes. Interestingly, in HNSCC patient samples, expression levels of these genes cluster together; patients either show relatively high or low expression of DAC-induced interferon genes. This suggests that a subset of HNSCC patients may be more responsive to DAC. To expand upon this work, it would be interesting to determine whether the expression of these genes correlates with response to DAC in HNSCC cell lines. Furthermore, this could suggest that DAC may sensitise some HNSCC patients to immune therapies.

As discussed in Section 7.1, the proportion of 5mC and 5hmC at Alu elements was dramatically reduced in VU40T cells and partially restored by DAC. This suggests that either Alu elements are somehow protected from the DNA demethylating action of DAC or that DAC leads to an active gain in DNA modifications at the transposable element sites. Oxidative bisulfite conversion and pyrosequencing of Alu elements would help to determine whether 5mC and 5hmC are increased at Alu elements after DAC. Furthermore, this would provide base pair resolution of 5mC and 5hmC distribution over the repeats. Interestingly, Alu elements show the most substantial changes in 5mC and 5hmC between HOK and VU40T cells, and in VU40T cells in response to DAC. Therefore, this could have considerable implications on the mechanism of DAC efficacy and suggests that at Alu elements DNA modifications are more dynamic than at other sites in the DNA.

### 7.2.3. Primary oral keratinocytes are not responsive to DAC treatment

Encouragingly, primary oral keratinocytes (HOK cells) do not respond to clinically relevant concentrations of DAC. Instead, HOK cells group with the HNSCC cell lines that were unresponsive to treatment. HOK MeDIP showed minimal 5mC distribution

changes after DAC treatment and 5mC changes characteristic of DAC treated VU40T cells were not observed in HOK cells; 5mC was not lost at promoters and ERVs or gained at Alu elements. As discussed in Section 7.2.2, in HNSCC cells, DAC shifts the distribution of 5mC back towards one characteristic of HOK cells. DNA methylation gained in cancer may have not fully developed the complex chromatin structure and epigenetic feedback loops that maintain 5mC silencing in normal cells, making these regions more susceptible to epigenetic therapy [27, 38, 45, 96]. Accordingly, DAC treatment is understood to preferentially reactivate genes silenced in cancer [96, 263, 264]. Therefore, DAC may target regions of 5mC specific to HNSCC, and thus HOK cells would be protected from the treatment. The results presented here discuss a single HNSCC cell line and relevant control, therefore to confirm this result the experiment would need to be repeated on a wider panel of HNSCC cell lines and controls.

### 7.2.4. The relationship between 5mC and 5hmC

The results of MeDIP and hMeDIP sequencing showed considerable overlap; this can be attributed to the fact that 5hmC is the oxidised product of 5mC. At Alu elements, changes in 5mC and 5hmC that occurred between HOK and VU40T cells, or between untreated and DAC treated cells corresponded with each other; both 5mC and 5hmC were decreased at Alu elements in HNSCC, and increased after DAC. However, elsewhere, changes in 5mC and 5hmC were specific for each modification. This is in keeping with results from the literature; regions hypermethylated in cancer are hypo-hydroxymethylated, often correlating with mutations in *TET* or *IDH* genes [260, 261]. Indeed, in VU40T cells, 5mC and 5hmC changes after DAC were accompanied

by an increase in *TET2* gene expression. Therefore, correlating the 5mC and 5hmC changes with the results of a TET2 ChIP would confirm the involvement of TET2 in converting 5mC→5hmC, and perhaps also 5hmC→5fC/ 5caC/C after DAC treatment. To further investigate the relationship between the DNA modifications, oxidative bisulfite sequencing could be performed to determine 5mC and 5hmC at base pair resolution. The results of the MeDIP and hMeDIP sequencing could be used to design a capture array to reduce the sequencing costs while still investigating the regions of interest.

## 7.3. DAC and Paracetamol work in synergy to reduce viability in HNSCC cell lines

The results of a DAC sensitising assay determined that the common analgesic, paracetamol functions in synergy with DAC to reduce viability in DAC responsive HNSCC cell lines. Combined treatment allowed toxic concentrations of either drug to be replaced by clinically relevant ones. The mechanism behind this synergistic relationship can be attributed to three pathways: DAC altered the cyclooxygenase pathway; DAC and paracetamol mimicked the effects of paracetamol overdose; and paracetamol decreased DNA methylation.

Paracetamol is understood to act on the cyclooxygenase pathway, primarily through COX-2 inhibition [192, 276]. In DAC responsive cell lines, DAC increased COX-2 expression causing a corresponding increase in the COX-2 product, $PGE_2$. Adding paracetamol or the COX-2 specific inhibitor, valdecoxib, restored $PGE_2$ levels to that of untreated cells. The COX-2/$PGE_2$ pathway is associated with inflammation, growth and survival. Therefore, in HNSCC cells, paracetamol may prevent COX-2/$PGE_2$

pathway promotion by DAC. This cannot fully account for the reduction in viability described since other COX inhibitors do not sensitise HNSCC cells to DAC. However, to confirm this, further experiments using more specific COX inhibitors and a wider range of concentrations are required.

Paracetamol overdose occurs when the levels of its toxic metabolite, NAPQI, overwhelm glutathione (GSH) stores [224]. A small amount of all paracetamol consumed is converted into NAPQI by CPY450 enzymes, primarily CYP2E1 [223]. In DAC responsive cells DAC demethylates and increases the expression of CYP2E1, leading to a reduction in GSH. Therefore, DAC may increase the conversion of paracetamol to NAPQI and mimic the effects of paracetamol overdose at therapeutic concentrations. The CYP450 inhibitor, disulfiram, did not rescue VU40T cells from combined treatment. However, GSH levels were not investigated, therefore, it cannot be confirmed that disulfiram can restore the rate of conversion of paracetamol→NAPQI. Further work in this area could include investigating whether GSH or the antidote to paracetamol overdose, NAC, are able to prevent the synergistic relationship between DAC and paracetamol.

Interestingly, in DAC responsive cell lines, paracetamol treatment led to a reduction in 5mC via reduced *DNMT* expression. Therefore, the synergistic relationship between the two drugs could be due to increased DNA demethylation. MeDIP and DNMT ChIP could be performed on DAC and paracetamol treated cells to determine how the addition of paracetamol to DAC treatment alters the distribution of 5mC. Several different mechanisms have been proposed to regulate *DNMT* gene expression, including activation by SP-1/-3 and inhibition by miRNAs [277-282]. The

effect of paracetamol on these pathways needs to be investigated in order to determine how paracetamol causes reduced *DNMT* expression in HNSCC cells.

In conclusion, the mechanistic studies presented here show that several pathways are altered by the combined treatment, and suggests that these cooperate to produce the synergistic relationship described.

Further work involving mouse models and clinical trials are needed to confirm the clinical efficacy of DAC and paracetamol in HNSCC. However the implications of this finding are considerable. Firstly, clinical trials on the efficacy of DAC in solid tumours are already underway, and as paracetamol is a widely used analgesic, it is probable that they have already been used in combination [121]. Therefore, the synergistic effect of DAC and paracetamol may confound the outcomes of past or ongoing clinical trials. Furthermore, it is imperative to determine whether DAC and paracetamol act in synergy in haematological cancers, where DAC treatment is already in use. Finally, paracetamol is an appealing drug for repurposing as it is widely available and inexpensive. For HNSCC, combined treatment with DAC and paracetamol may prove to be the additional therapeutic option that is urgently required.

## 7.4. Conclusion

The primary aim of this thesis was to investigate the relationship between DNA modifications and HNSCC and to examine the efficacy of the DNA demethylating agent, DAC, in HNSCC treatment. A subset of HNSCC cell lines responded to nanomolar concentrations of DAC. This corresponded with a shift in the distribution of 5mC and 5hmC towards one characteristic of normal oral keratinocytes. In HNSCC

cells DAC increased ERV expression and induced a viral mimicry response. Furthermore, DAC re-established 5mC and 5hmC at Alu elements, leading to their repression. Finally, DAC and paracetamol were found to work in synergy to reduce viability in HNSCC cell lines, providing an appealing new therapeutic option. Therefore, DAC has potential for use in the treatment of HNSCC, either as a single agent, or combined with paracetamol (Figure 7.1).

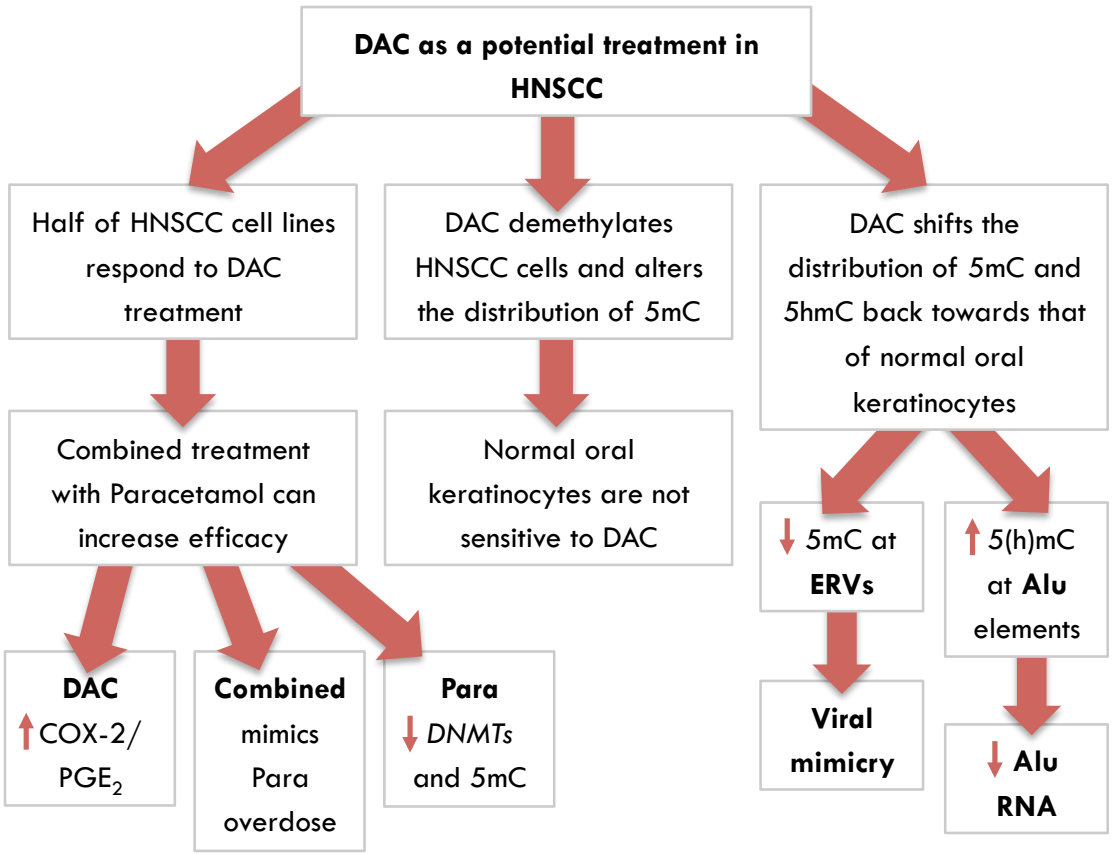



**Figure 7.1. DAC as a potential therapeutic option in the treatment of HNSCC**

The main conclusions of this thesis are represented as a schematic. DAC has potential as a therapeutic in HNSCC. **1**. Half of HNSCC cell lines tested respond to DAC treatment and in DAC responsive cell lines efficacy can be increased through combined treatment with paracetamol. **2**. Primary oral keratinocytes are not responsive to DAC treatment. **3**. DAC shifts the distribution of 5mC and 5hmC in HNSCC cells.

# Supplementary Information

## Supplementary data 1. FCM1 drug library

Drugs used in the sensitising assay alongside the numbers they are represented as in figures. Controls are highlighted in blue.

| | | | |
|---|---|---|---|
| prednisolone | 1 | metoclopromide HCL | 51 |
| amantidine hydrochloride | 2 | domperidone | 52 |
| folic acid | 3 | thalidomide | 53 |
| thiamine HCL, | 4 | chloroquine, | 54 |
| ranitidine/zantac | 5 | metformin hydrochloride, | 55 |
| fluoxetine | 6 | niclosamide | 56 |
| dexamethasone | 7 | pravastatin sodium | 57 |
| DMSO | 8 | nortryptyline | 58 |
| phytomenadione/ vitamin K1 | 9 | dantrolene sodium | 59 |
| carbamazepine | 10 | omeprazole | 60 |
| propanolol hydrochloride, | 11 | diclofenac sodium, | 61 |
| erythromycin | 12 | ritodrine hydrochloride, | 62 |
| retinol | 13 | selegiline hydrochloride, | 63 |
| bendroflumethiazide, | 14 | mebendazole | 64 |
| propylthiouracil | 15 | flupentixol | 65 |
| nicotinic acid, | 16 | acipimox | 66 |
| theophylline, | 17 | ETHOH | 67 |
| nicotinamide, | 18 | desferrioxamine mesilate | 68 |
| ascorbic acid | 19 | imipramine, | 69 |
| Acyclovir | 20 | artemisinin | 70 |
| allopurinol | 21 | propantheline bromide | 71 |
| chlorpheniramine disodium salt | 22 | diltiazem hydrochloride | 72 |
| neostigmine | 23 | ampicillin sodium salt | 73 |
| alpha tocopheryl acetate | 24 | methanol | 74 |
| 2-bromo-a-ergocryptine methanesulfonate salt/ bromocriptine | 25 | amphotericin b | 75 |
| cyclophosphamide, | 26 | danazol | 76 |
| imatinib | 27 | phenoxymethyl penicillin, | 77 |
| zinc acetate | 28 | 5'aminosalicyclic acid/ mesalazine | 78 |
| valproic acid, | 29 | finasteride | 79 |
| rifampicin | 30 | colchicine, | 80 |
| metronidazole | 31 | levothyroxine sodium | 81 |
| praziquantel | 32 | methotrexate, | 82 |
| flutamide, | 33 | mifepristone | 83 |
| clomipramine | 34 | nicotine | 84 |
| testosterone, | 35 | cefaclor | 85 |
| calciferol/ergocalciferol | 36 | cyanocobalamin (vitamin B12) | 86 |
| doxycycline, | 37 | medroxyprogesterone acetate, | 87 |
| tetracycline | 38 | bezafibrate | 88 |
| ibuprofen, | 39 | paroxetine, | 89 |
| norethisterone | 40 | trimethoprim | 90 |
| itraconazole | 41 | water | 91 |
| methyldopa | 42 | naloxone hydrochloride | 92 |
| fenofibrate, | 43 | simvastatin, | 93 |
| clofibric acid | 44 | flecainide acetate | 94 |
| clobetasol propionate | 45 | pilocarpine HCl | 95 |
| paracetamol (acetaminophen), | 46 | fluconazole | 96 |
| alverine citrate | 47 | acitretin | 97 |
| mefenamic acid | 48 | Memantine | 98 |
| prochlorperazine | 49 | Bortezamib | 99 |
| chlorambucil | 50 | phenelzine | 100 |

## Supplementary data 2. RNA sequencing reads

Total number of sequenced reads and reads aligned using HiSAT.

| | Total number reads: | Not Aligned | Aligned n=1 | Aligned n>1 |
|---|---|---|---|---|
| **VU40T** | 34941803 | 1675948 | 28457784 | 4808071 |
| **HN12** | 16587553 | 2461265 | 12229921 | 1896367 |
| **UDSCC2** | 18726353 | 2478684 | 14146982 | 2100687 |
| **SCC040** | 16556164 | 2348947 | 12288730 | 1918487 |
| **VU40T 500nM DAC** | 22397597 | 2241080 | 17458914 | 2697603 |
| **VU40T 132.3uM Para** | 18505914 | 1853625 | 14322656 | 2329633 |
| **VU40T Para + DAC** | 13463532 | 1975441 | 9932820 | 1555271 |

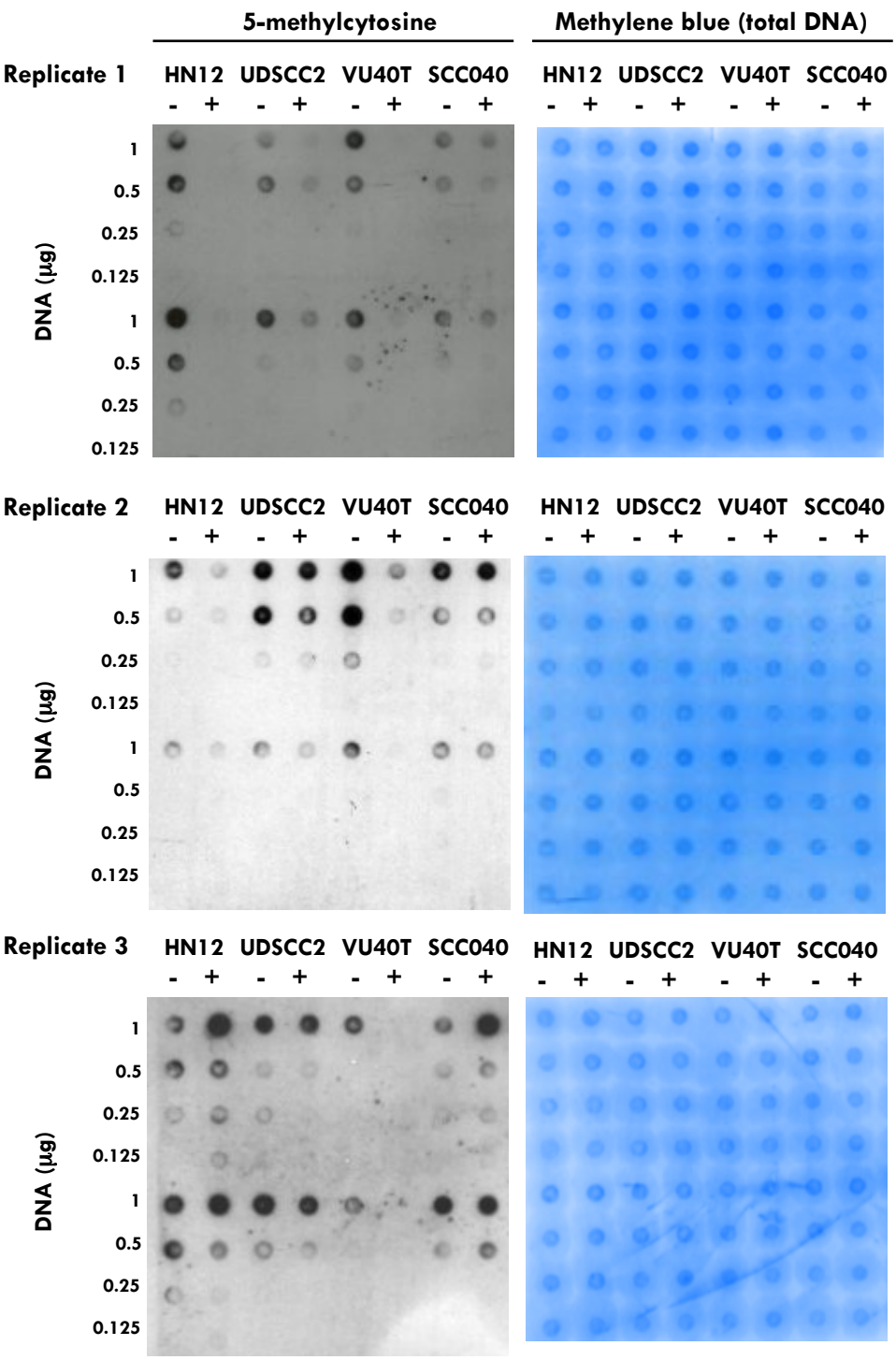
### Supplementary data 3. MeDIP and hMeDIP sequenced reads

The total number of reads sequenced for MeDIP, hMeDIP and input DNA are outlined for all samples analysed. Using Bowtie2, Dr Sam Clokie aligned reads to the human genome (hg19). The table also shows the number and percentage of reads aligned.

| hMeDIP | Replicate | Total number of reads sequenced | Aligned reads | |
|---|---|---|---|---|
| | | | Number | Percentage |
| VU40T NT | 1 | 24,476,992 | 23,244,909 | 94.97 |
| | 2 | 25,460,459 | 24,521,995 | 96.31 |
| | 3 | 8,336,836 | 7,870,773 | 94.41 |
| VU40T 100nM DAC | 1 | 25,987,848 | 25,169,967 | 96.85 |
| | 2 | 26,683,840 | 25,957,794 | 97.28 |
| VU40T 1uM DAC | 1 | 15,676,959 | 13,778,876 | 87.89 |
| | 2 | 27,395,777 | 25,988,864 | 94.86 |
| | 3 | 10,319,152 | 9,527,629 | 92.33 |
| HOK NT | 1 | 36,493,086 | 36,010,648 | 98.68 |

| MeDIP | Replicate | Total number of reads sequenced | Aligned reads | |
|---|---|---|---|---|
| | | | Number | Percentage |
| VU40T NT | 1 | 23,799,568 | 23,799,568 | 98.37 |
| | 2 | 13,835,182 | 13,533,894 | 97.82 |
| VU40T 100nM DAC | 1 | 21,670,412 | 21,252,088 | 98.07 |
| | 2 | 16,048,337 | 15,689,139 | 97.76 |
| VU40T 1uM DAC | 1 | 45,402,331 | 44,511,849 | 98.04 |
| | 2 | 16,737,593 | 16,470,694 | 98.41 |
| HOK NT | 1 | 29,986,781 | 29,755,476 | 99.23 |
| HOK 100nM DAC | 1 | 39,121,678 | 38,808,718 | 99.20 |
| HOK 1uM DAC | 1 | 24,379,915 | 24,199,440 | 99.26 |

| Input | Replicate | Total number of reads sequenced | Aligned reads | |
|---|---|---|---|---|
| | | | Number | Percentage |
| HOK NT | 1 | 31,846,500 | 31,694,473 | 99.52 |
| VU40T NT | 1 | 17,567,454 | 17,166,100 | 97.72 |




**Supplementary data 4. DNA dot blot analysis of 5mC levels in HNSCC cell lines**

HNSCC cell lines were treated with 500nM DAC (+) or an equivalent volume of vehicle (-) for 96h following which DNA was extracted and DNA dot blotting performed using an antibody against 5mC. After 5mC analysis blots were incubated in methylene blue to stain for total DNA.

## Supplementary data 5. Gene ontology terms associated with DAC responsive cell lines

Genes expressed (FPKM>1) in only in DAC responsive (HN12 and VU40T) cell lines were subject to gene onotology biological processes (GO BP) analysis using DAVID.

| Term | P-Value | Benjamini |
|---|---|---|
| positive regulation of catagen | 0.00021 | 0.17 |
| negative regulation of protein kinase activity | 0.0097 | 0.99 |
| positive regulation of mesenchymal cell apoptotic process | 0.017 | 0.99 |
| AMP catabolic process | 0.017 | 0.99 |
| activation of meiosis | 0.017 | 0.99 |
| embryonic hindlimb morphogenesis | 0.023 | 0.99 |
| cardioblast differentiation | 0.025 | 0.99 |
| embryonic nail plate morphogenesis | 0.025 | 0.99 |
| palate development | 0.026 | 0.98 |
| cellular response to estradiol stimulus | 0.028 | 0.97 |
| embryonic forelimb morphogenesis | 0.029 | 0.96 |
| drug transport | 0.041 | 0.98 |
| negative regulation of fat cell differentiation | 0.048 | 0.99 |
| cell differentiation involved in embryonic placenta development | 0.049 | 0.98 |
| BMP signaling pathway involved in heart development | 0.049 | 0.98 |
| pore complex assembly | 0.049 | 0.98 |
| translation | 0.062 | 0.99 |
| positive regulation of protein localization to Cajal body | 0.065 | 0.99 |
| protein folding | 0.065 | 0.99 |
| positive regulation of translation | 0.073 | 0.99 |
| polyamine biosynthetic process | 0.073 | 0.99 |
| positive regulation of osteoblast differentiation | 0.09 | 0.99 |
| cell morphogenesis | 0.096 | 0.99 |
| negative regulation of bone resorption | 0.096 | 0.99 |
| positive regulation of DNA damage response, signal transduction by p53 class mediator | 0.096 | 0.99 |
| response to caffeine | 0.096 | 0.99 |

**Supplementary data 6. Gene ontology terms associated with genes differentially expressed in DAC responsive versus DAC unresponsive HNSCC cell lines**

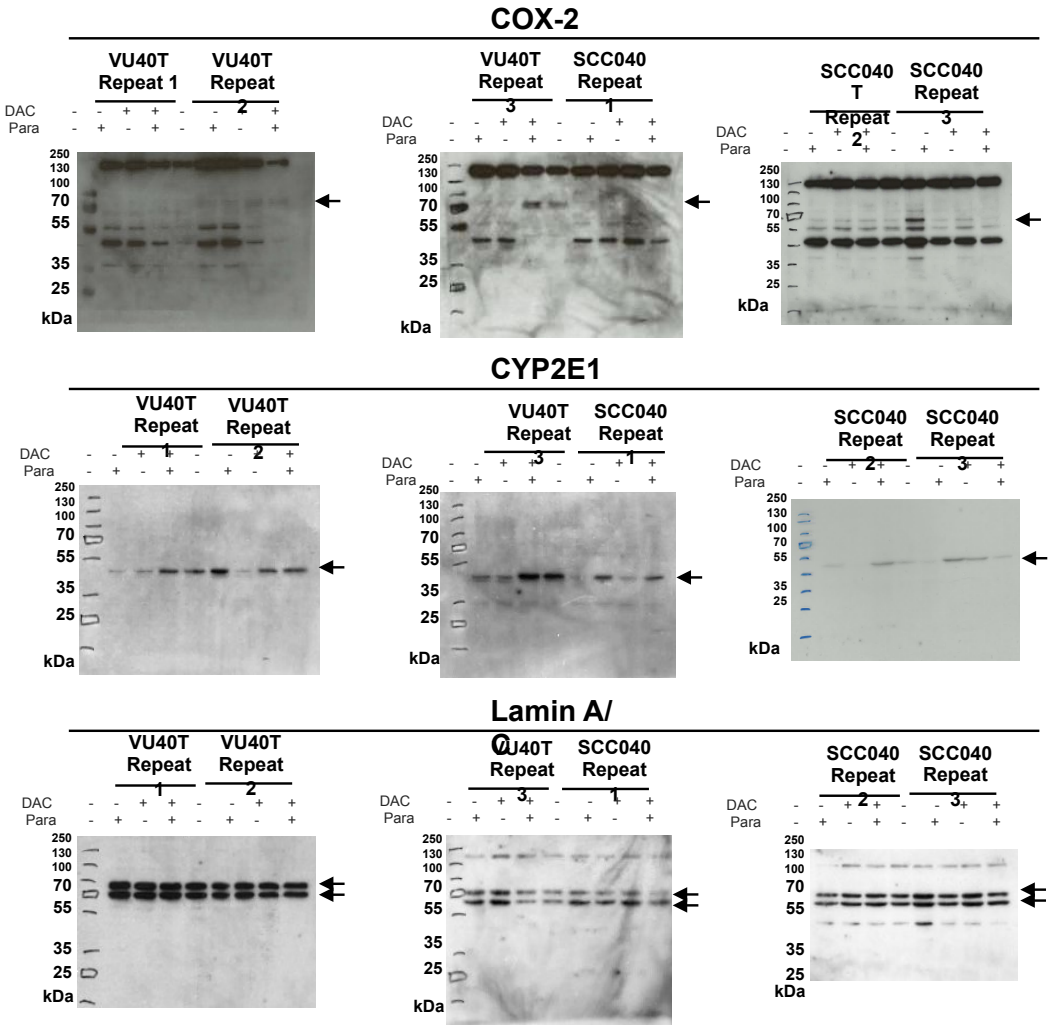
DeSeq2 was performed to determine differentially expressed genes in the DAC responsive cell lines versus the unresponsive cell lines. Each group was subject to gene ontology biological processes anaysis using Enrichr.

| DAC-unresponsive | | |
|---|---|---|
| **term ID** | **description** | **log10 p-value** |
| GO:0008150 | biological_process | −14.5243 |
| GO:0008152 | metabolic process | −3.065 |
| GO:0009987 | cellular process | −8.8761 |
| GO:0032502 | developmental process | -3.2636 |
| GO:0045892 | negative regulation of transcription, DNA-templated | −5.1518 |
| GO:0048856 | anatomical structure development | −2.4921 |
| GO:0050911 | detection of chemical stimulus involved in sensory perception of smell | −11.1851 |
| GO:0065007 | biological regulation | −6.5544 |
| GO:0071840 | cellular component organization or biogenesis | −4.762 |
| GO:0030030 | cell projection organization | −4.4547 |
| GO:0044237 | cellular metabolic process | −4.4895 |
| GO:1901360 | organic cyclic compound metabolic process | −3.1308 |
| GO:0009058 | biosynthetic process | −2.4908 |
| GO:0071704 | organic substance metabolic process | −1.7122 |
| GO:0006725 | cellular aromatic compound metabolic process | −2.3233 |
| GO:1901362 | organic cyclic compound biosynthetic process | −5.9586 |
| GO:0046483 | heterocycle metabolic process | −2.0783 |
| GO:0048519 | negative regulation of biological process | −3 |
| GO:0051606 | detection of stimulus | −8.1701 |
| GO:0007186 | G-protein coupled receptor signaling pathway | −4.8041 |
| GO:0006357 | regulation of transcription from RNA polymerase II promoter | −1.4711 |
| GO:0050789 | regulation of biological process | −5.1831 |
| GO:0018130 | heterocycle biosynthetic process | −5.3197 |
| GO:0016070 | RNA metabolic process | −2.0711 |
| GO:0019438 | aromatic compound biosynthetic process | −5.251 |
| GO:0070268 | cornification | −1.9872 |
| GO:0044271 | cellular nitrogen compound biosynthetic process | −2.9469 |
| GO:0006139 | nucleobase-containing compound metabolic process | −2.1858 |
| GO:0006996 | organelle organization | −2.8996 |
| GO:0044249 | cellular biosynthetic process | −2.5045 |
| GO:0006351 | transcription, DNA-templated | −6.007 |
| GO:2000112 | regulation of cellular macromolecule biosynthetic process | −4.2147 |
| GO:1901576 | organic substance biosynthetic process | −2.1524 |
| GO:0019222 | regulation of metabolic process | −2.983 |
| GO:0048858 | cell projection morphogenesis | −1.5901 |
| GO:0022607 | cellular component assembly | −1.3768 |
| **DAC- responsive** | | |
| **term ID** | **description** | **log10 p-value** |
| GO:0007606 | sensory perception of chemical stimulus | −5.767 |
| GO:0008150 | biological_process | −3.1593 |
| GO:0009987 | cellular process | −3.9469 |
| GO:0034660 | ncRNA metabolic process | −4.5452 |
| GO:0044085 | cellular component biogenesis | −4.1349 |
| GO:0050790 | regulation of catalytic activity | −1.7852 |
| GO:0051179 | localization | −2.3747 |
| GO:0051234 | establishment of localization | −1.6345 |
| GO:0071840 | cellular component organization or biogenesis | −5.1518 |
| GO:0007186 | G-protein coupled receptor signaling pathway | −1.6364 |
| GO:0016072 | rRNA metabolic process | −3.567 |

**Supplementary data 7. Gene ontology terms associated with genes differentially expressed after DAC, paracetamol or combined treatment**

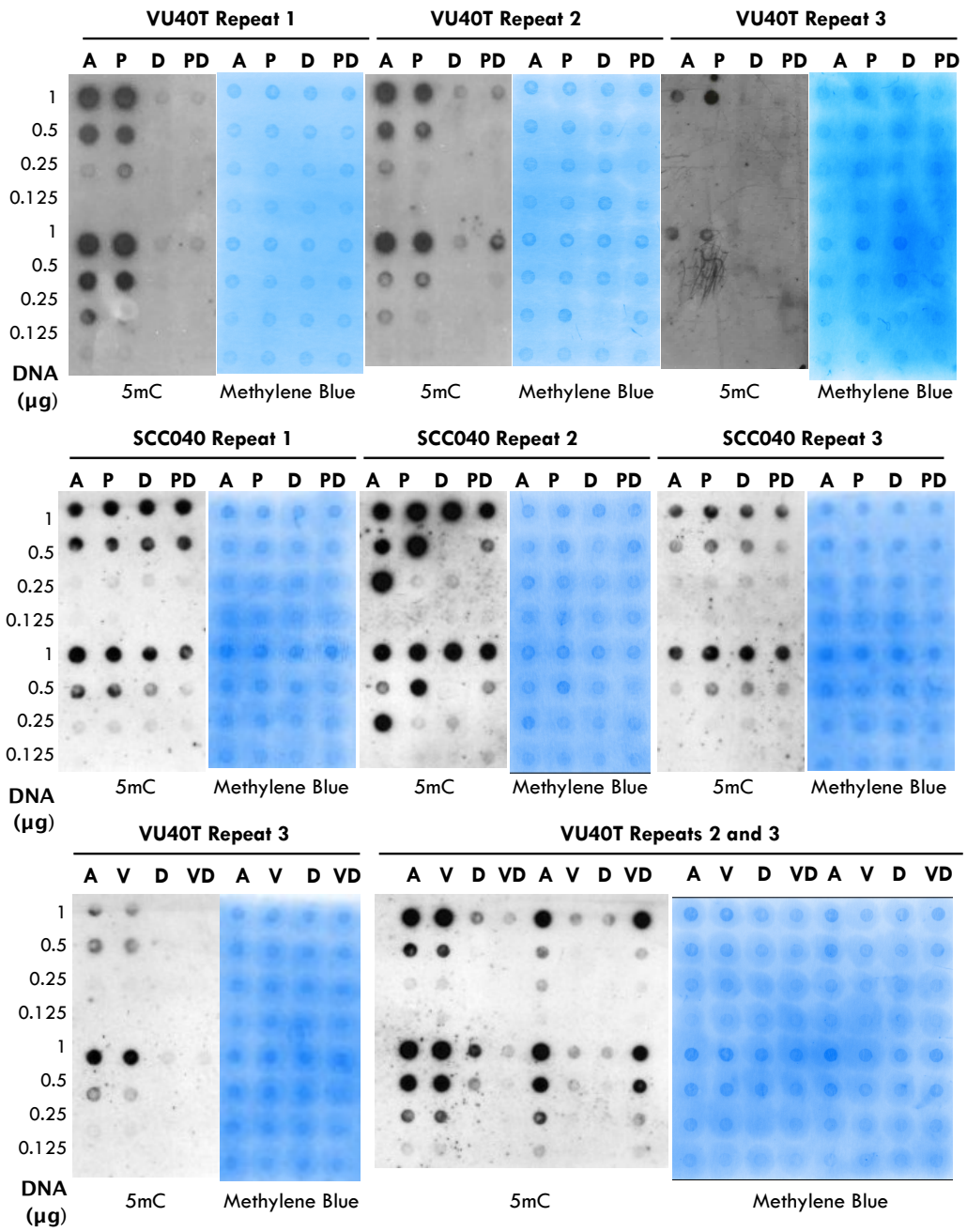
DeSeq2 was performed to determine differentially expressed genes between treatment groups. Each group was subject to gene ontology biological processes anaysis using Enrichr.

| | Increased by Treatment | | | Decreased by Treatment | | |
|---|---|---|---|---|---|---|
| | term ID | description | log10 p-value | term ID | description | log10 p-value |
| Paracetamol | GO:0046034 | ATP metabolic process | -7.48 | | | |
| | GO:0032981 | mitochondrial respiratory chain complex I assembly | -2.38 | | | |
| | GO:0006091 | generation of precursor metabolites and energy | -3.21 | | | |
| | GO:0022900 | electron transport chain | -6.29 | | | |
| | GO:0072521 | purine-containing compound metabolic process | -3.09 | | | |
| | GO:0055086 | nucleobase-containing small molecule metabolic process | -1.38 | | | |
| | GO:0019637 | organophosphate metabolic process | -1.63 | | | |
| DAC | GO:0002376 | immune system process | -9.21 | GO:0008152 | metabolic process | -6.63 |
| | GO:0008150 | biological_process | -2.80 | GO:0009987 | cellular process | -3.12 |
| | GO:0009987 | cellular process | -3.35 | GO:0023052 | signaling | -2.02 |
| | GO:0032501 | multicellular organismal process | -1.38 | GO:0034660 | ncRNA metabolic process | -10.90 |
| | GO:0032502 | developmental process | -2.61 | GO:0046907 | intracellular transport | -2.71 |
| | GO:0034097 | response to cytokine | -11.60 | GO:0071840 | cellular component organization or biogenesis | -10.95 |
| | GO:0042127 | regulation of cell proliferation | -6.56 | GO:0071353 | cellular response to interleukin-4 | -1.33 |
| | GO:0044699 | single-organism process | -6.75 | GO:0071704 | organic substance metabolic process | -6.78 |
| | GO:0050896 | response to stimulus | -2.63 | GO:0000278 | mitotic cell cycle | -6.07 |
| | GO:0051704 | multi-organism process | -2.18 | GO:0006996 | organelle organization | -7.70 |
| | GO:0065007 | biological regulation | -3.87 | GO:0008283 | cell proliferation | -1.95 |
| | GO:0006396 | RNA processing | -2.21 | GO:0009058 | biosynthetic process | -4.41 |
| | GO:0044763 | single-organism cellular process | -5.23 | GO:0019080 | viral gene expression | -3.14 |
| | GO:0043900 | regulation of multi-organism process | -4.12 | GO:0006807 | nitrogen compound metabolic process | -6.11 |
| | GO:0030155 | regulation of cell adhesion | -1.79 | GO:1901360 | organic cyclic compound metabolic process | -9.56 |
| | GO:0040012 | regulation of locomotion | -4.18 | GO:0044700 | single organism signaling | -2.01 |
| | GO:0051270 | regulation of cellular component movement | -4.91 | GO:0044238 | primary metabolic process | -6.53 |
| | GO:1901564 | organonitrogen compound metabolic process | -1.57 | GO:0006725 | cellular aromatic compound metabolic process | -9.16 |
| | GO:0051239 | regulation of multicellular organismal process | -6.15 | GO:0044237 | cellular metabolic process | -7.21 |
| | GO:0046903 | secretion | -2.63 | GO:0007059 | chromosome segregation | -1.69 |
| | GO:0050793 | regulation of developmental process | -5.35 | GO:0006413 | translational initiation | -2.60 |
| | GO:0032879 | regulation of localization | -2.04 | GO:0007049 | cell cycle | -5.37 |
| | GO:0048518 | positive regulation of biological process | -4.97 | GO:0043170 | macromolecule metabolic process | -2.64 |
| | GO:0006952 | defense response | -8.70 | GO:0046483 | heterocycle metabolic process | -9.03 |
| | GO:0048522 | positive regulation of cellular process | -3.72 | GO:0009124 | nucleoside monophosphate biosynthetic process | -2.12 |
| | GO:0048519 | negative regulation of biological process | -4.13 | GO:0034641 | cellular nitrogen compound metabolic process | -11.50 |
| | GO:0048523 | negative regulation of cellular process | -2.93 | GO:0000184 | nuclear-transcribed mRNA catabolic process, nonsense-mediated decay | -1.75 |
| | GO:0044710 | single-organism metabolic process | -2.84 | GO:0044260 | cellular macromolecule metabolic process | -3.64 |
| | GO:0051707 | response to other organism | -6.95 | GO:0044249 | cellular biosynthetic process | -4.63 |
| | GO:0006955 | immune response | -4.49 | GO:0043603 | cellular amide metabolic process | -5.01 |
| | GO:0009607 | response to biotic stimulus | -6.56 | GO:0006260 | DNA replication | -6.94 |
| | GO:1902578 | single-organism localization | -2.01 | GO:0090305 | nucleic acid phosphodiester bond hydrolysis | -1.83 |
| | GO:0009605 | response to external stimulus | -6.63 | GO:0006259 | DNA metabolic process | -6.99 |
| | GO:0007166 | cell surface receptor signaling pathway | -3.11 | GO:0010467 | gene expression | -1.78 |
| | GO:0048583 | regulation of response to stimulus | -1.87 | GO:0034470 | ncRNA processing | -8.31 |
| | GO:0050794 | regulation of cellular process | -3.84 | GO:0006312 | mitotic recombination | -2.63 |
| | GO:0006954 | inflammatory response | -4.19 | GO:0006396 | RNA processing | -7.26 |
| | | | | GO:0022616 | DNA strand elongation | -3.23 |
| | | | | GO:0032543 | mitochondrial translation | -2.16 |
| | | | | GO:0000722 | telomere maintenance via recombination | -2.02 |
| | | | | GO:0006139 | nucleobase-containing compound metabolic process | -9.48 |
| | | | | GO:1901566 | organonitrogen compound biosynthetic process | -4.40 |
| DAC+ Paracetamol | GO:0006396 | RNA processing | -2.67 | GO:0003008 | system process | -4.13 |
| | GO:0008150 | biological_process | -3.21 | GO:0006412 | translation | -37.77 |
| | GO:0009605 | response to external stimulus | -3.22 | GO:0007155 | cell adhesion | -2.00 |
| | GO:0009888 | tissue development | -5.65 | GO:0008150 | biological_process | -15.30 |
| | GO:0009987 | cellular process | -3.59 | GO:0008152 | metabolic process | -40.33 |
| | GO:0032501 | multicellular organismal process | -2.06 | GO:0009987 | cellular process | -22.36 |
| | GO:0032502 | developmental process | -2.59 | GO:0022610 | biological adhesion | -2.07 |
| | GO:0044699 | single-organism process | -6.12 | GO:0022613 | ribonucleoprotein complex biogenesis | -35.69 |
| | GO:0051239 | regulation of multicellular organismal process | -5.41 | GO:0023052 | signaling | -8.40 |
| | GO:0065007 | biological regulation | -5.06 | GO:0032501 | multicellular organismal process | -1.46 |
| | GO:0044763 | single-organism cellular process | -7.20 | GO:0071840 | cellular component organization or biogenesis | -35.17 |
| | GO:0051270 | regulation of cellular component movement | -3.44 | GO:0007059 | chromosome segregation | -12.80 |
| | GO:0040012 | regulation of locomotion | -2.53 | GO:0006974 | cellular response to DNA damage stimulus | -17.44 |
| | GO:0042127 | regulation of cell proliferation | -3.18 | GO:0007154 | cell communication | -7.39 |
| | GO:0051336 | regulation of hydrolase activity | -3.28 | GO:0044763 | single-organism cellular process | -4.46 |
| | GO:0032879 | regulation of localization | -4.58 | GO:0032259 | methylation | -5.06 |
| | GO:0046903 | secretion | -1.86 | GO:0051301 | cell division | -9.99 |
| | GO:0048523 | negative regulation of cellular process | -4.51 | GO:0030029 | actin filament-based process | -2.29 |
| | GO:0023051 | regulation of signaling | -2.53 | GO:0006091 | generation of precursor metabolites and energy | -2.17 |
| | GO:0016192 | vesicle-mediated transport | -1.46 | GO:0007049 | cell cycle | -23.95 |
| | GO:0010646 | regulation of cell communication | -1.93 | GO:0044237 | cellular metabolic process | -44.37 |
| | GO:0048518 | positive regulation of biological process | -2.60 | GO:0019637 | organophosphate metabolic process | -4.46 |
| | GO:0048522 | positive regulation of cellular process | -2.44 | GO:1901135 | carbohydrate derivative metabolic process | -1.58 |
| | GO:0048519 | negative regulation of biological process | -4.65 | GO:0009056 | catabolic process | -5.24 |
| | GO:0065008 | regulation of biological quality | -1.76 | GO:0006793 | phosphorus metabolic process | -2.17 |
| | GO:0006955 | immune response | -2.10 | GO:1901796 | regulation of signal transduction by p53 class mediator | -2.42 |
| | GO:1902578 | single-organism localization | -2.87 | GO:0006614 | SRP-dependent cotranslational protein targeting to membrane | -19.93 |
| | GO:0006952 | defense response | -3.00 | GO:0043624 | cellular protein complex disassembly | -11.89 |
| | GO:0010033 | response to organic substance | -2.41 | GO:0044710 | single-organism metabolic process | -1.92 |
| | GO:0048583 | regulation of response to stimulus | -2.00 | GO:0034660 | ncRNA metabolic process | -28.93 |
| | GO:0034660 | ncRNA metabolic process | -2.01 | GO:0051276 | chromosome organization | -20.43 |
| | GO:0050794 | regulation of cellular process | -3.73 | GO:0006839 | mitochondrial transport | -1.74 |
| | | | | GO:0006725 | cellular aromatic compound metabolic process | -38.98 |
| | | | | GO:0022411 | cellular component disassembly | -5.39 |
| | | | | GO:1901360 | organic cyclic compound metabolic process | -37.31 |
| | | | | GO:0046483 | heterocycle metabolic process | -38.33 |
| | | | | GO:0009058 | biosynthetic process | -26.33 |
| | | | | GO:0006807 | nitrogen compound metabolic process | -38.36 |
| | | | | GO:0050877 | neurological system process | -3.76 |
| | | | | GO:0055086 | nucleobase-containing small molecule metabolic process | -7.02 |
| | | | | GO:0051641 | cellular localization | -8.88 |
| | | | | GO:0044700 | single organism signaling | -8.37 |
| | | | | GO:0090305 | nucleic acid phosphodiester bond hydrolysis | -4.95 |
| | | | | GO:0009451 | RNA modification | -1.85 |
| | | | | GO:0019083 | viral transcription | -19.10 |
| | | | | GO:0006260 | DNA replication | -18.34 |
| | | | | GO:0043414 | macromolecule methylation | -1.55 |
| | | | | GO:0000956 | nuclear-transcribed mRNA catabolic process | -16.15 |
| | | | | GO:0044260 | cellular macromolecule metabolic process | -30.93 |
| | | | | GO:0044085 | cellular component biogenesis | -26.93 |
| | | | | GO:0016072 | rRNA metabolic process | -31.72 |
| | | | | GO:0043170 | macromolecule metabolic process | -25.55 |
| | | | | GO:0034641 | cellular nitrogen compound metabolic process | -53.60 |
| | | | | GO:0043603 | cellular amide metabolic process | -29.53 |
| | | | | GO:0006259 | DNA metabolic process | -25.75 |
| | | | | GO:0010629 | negative regulation of gene expression | -2.27 |
| | | | | GO:0016071 | mRNA metabolic process | -12.37 |
| | | | | GO:0015980 | energy derivation by oxidation of organic compounds | -2.46 |
| | | | | GO:1901576 | organic substance biosynthetic process | -26.27 |
| | | | | GO:0044238 | primary metabolic process | -39.78 |
| | | | | GO:0061641 | CENP-A containing chromatin organization | -2.01 |
| | | | | GO:0010467 | gene expression | -19.19 |
| | | | | GO:0006396 | RNA processing | -26.20 |
| | | | | GO:0071704 | organic substance metabolic process | -37.87 |
| | | | | GO:0097031 | mitochondrial respiratory chain complex I biogenesis | -3.17 |
| | | | | GO:0019538 | protein metabolic process | -6.53 |
| | | | | GO:0015833 | peptide transport | -2.28 |
| | | | | GO:0044267 | cellular protein metabolic process | -10.38 |
| | | | | GO:0000819 | sister chromatid segregation | -8.72 |
| | | | | GO:0043933 | macromolecular complex subunit organization | -17.79 |
| | | | | GO:0007186 | G-protein coupled receptor signaling pathway | -3.58 |
| | | | | GO:1901564 | organonitrogen compound metabolic process | -13.04 |
| | | | | GO:0007267 | cell-cell signaling | -1.80 |
| | | | | GO:0006996 | organelle organization | -25.40 |
| | | | | GO:0071826 | ribonucleoprotein complex subunit organization | -7.37 |
| | | | | GO:0042254 | ribosome biogenesis | -38.08 |
| | | | | GO:0071824 | protein-DNA complex subunit organization | -1.74 |
| | | | | GO:0000280 | nuclear division | -6.59 |
| | | | | GO:0007005 | mitochondrion organization | -7.68 |



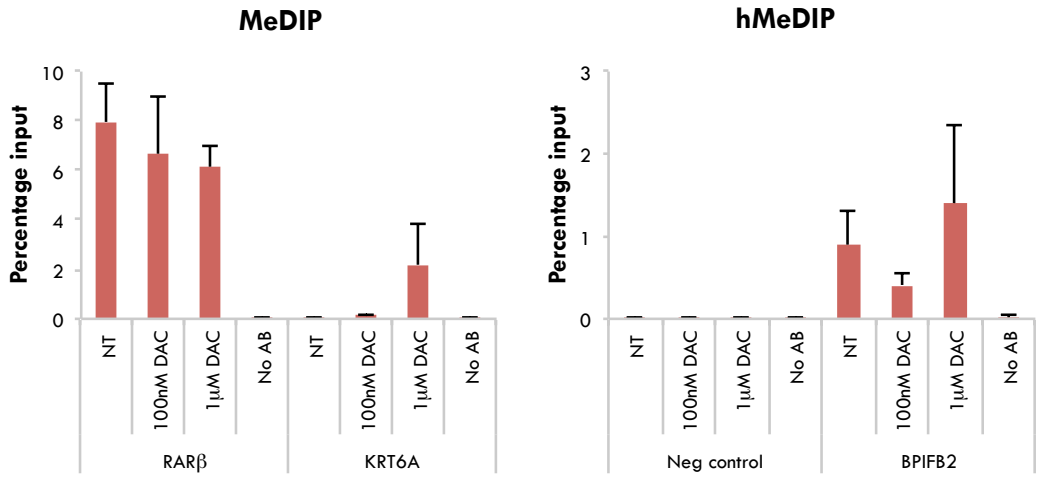

**Supplementary data 8. Western blot analysis**

VU40T and SCC040 cells were treated for 96h with 132.3µM paracetamol, 500nM DAC or both followed by protein extraction and western blotting for COX-2 (upper panels), CYP2E1 (middle panels) and Lamin A/C (lower panels) as loading control. Experiment was done in three biological replicates (shown in bold). Arrows depict the bands of interest.



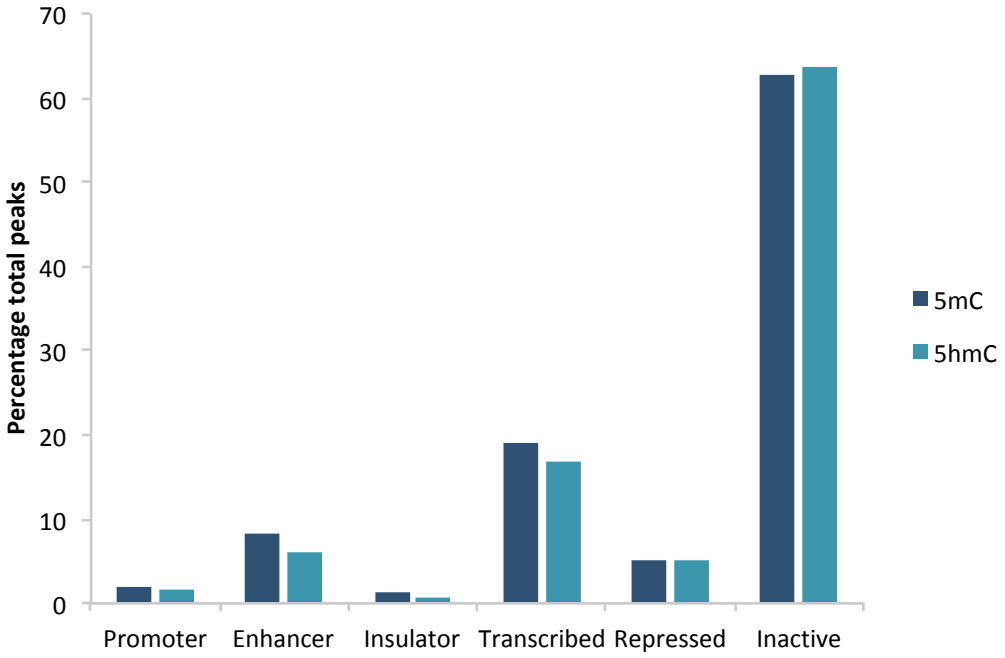

**Supplementary data 9. DNA dot blot analysis of DAC, paracetamol and valdecoxib treated HNSCC cell lines**

VU40T and SCC040 cells were treated for 96h with 132.3µM paracetamol (P), 500nM DAC (D), valdecoxib (V), combined treatment (PC/VD) or vehicle only (A). DNA dot blotting was performed for 5mC. Blots were incubated in methylene blue to stain for total DNA. Experiments were done in three biological replicates, with duplicate wells per experiment.



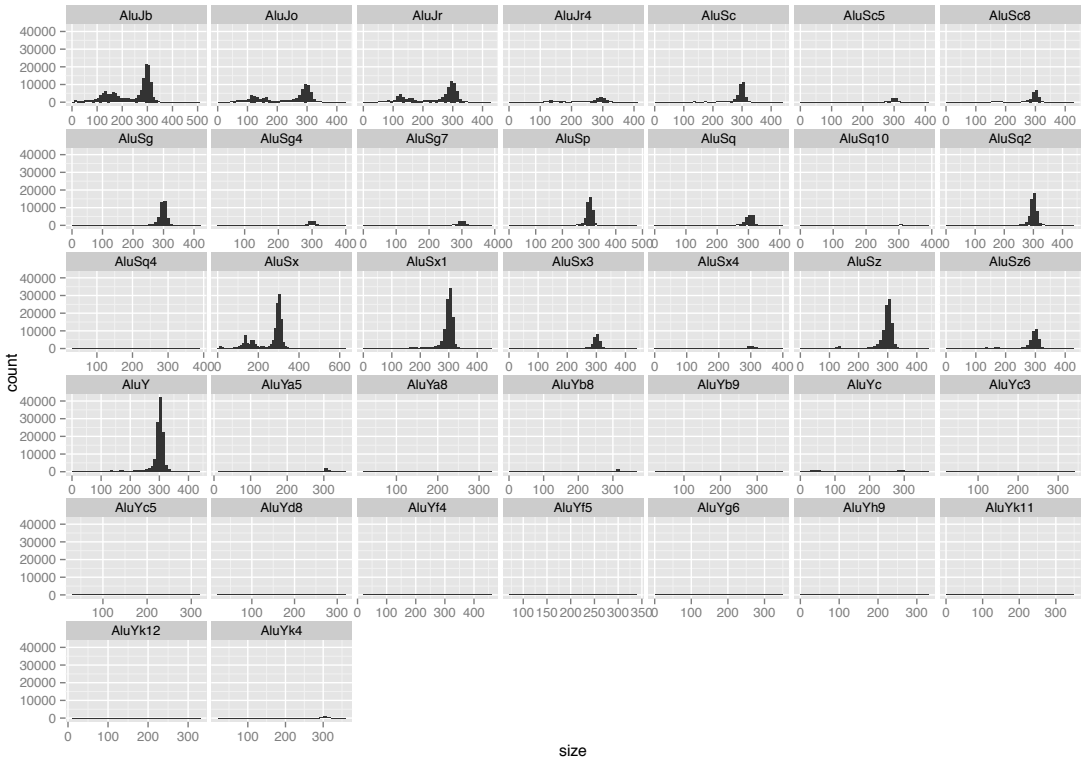

**Supplementary data 10. MeDIP and hMeDIP qRT-PCR check**

Prior to library preparation MeDIP and hMeDIP experiments were subject to qRT-PCR with primers corresponding to regions enriched in 5mC or 5hmC (RARβ and BFIPB2 respectively) or lacking 5mC or 5hmC (KRT6A and Neg control). MeDIP results are the mean of 2 experiments. Vehicle only (NT) and 1µM DAC treated hMeDIP results are the mean of 3 experiments, while 100nM DAC hMeDIP is the mean of 2 experiments. Results were normalised to total DNA results (input) and are shown alongside the (h)MeDIP experiment performed without an antibody (No AB). Error bars show SEM.



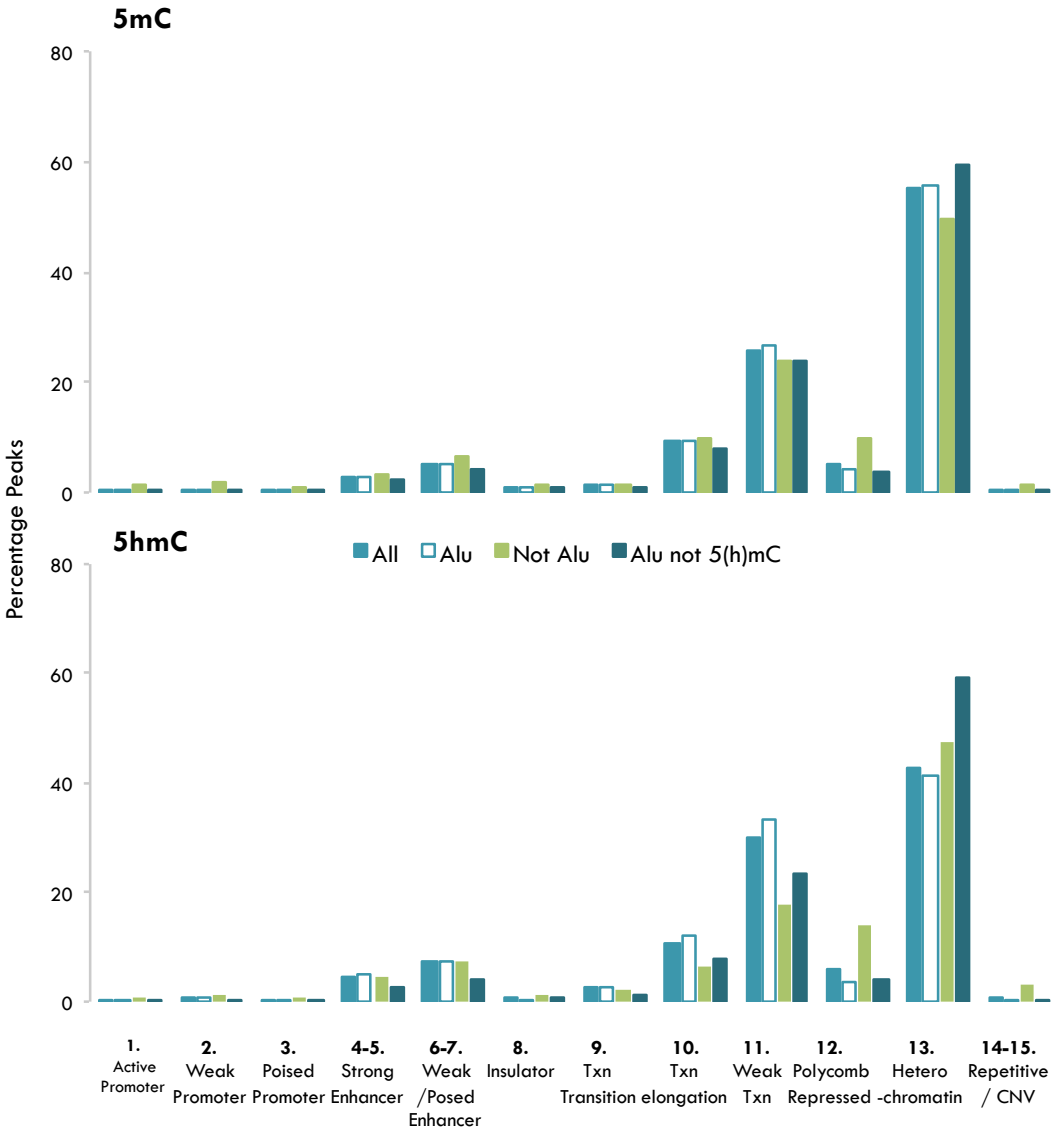

**Supplementary data 11. Overlap between 5mC and 5hmC peaks and NHEK chromatin domains**

NHEK chromatin domains were downloaded from the UCSC table browser. On Galaxy the BedTools '*intersect intervals*' function was used to determine the overlap between 5(h)mC peaks and NHEK chromatin domains. Results are shown as a percentage of the total 5(h)mC peaks.



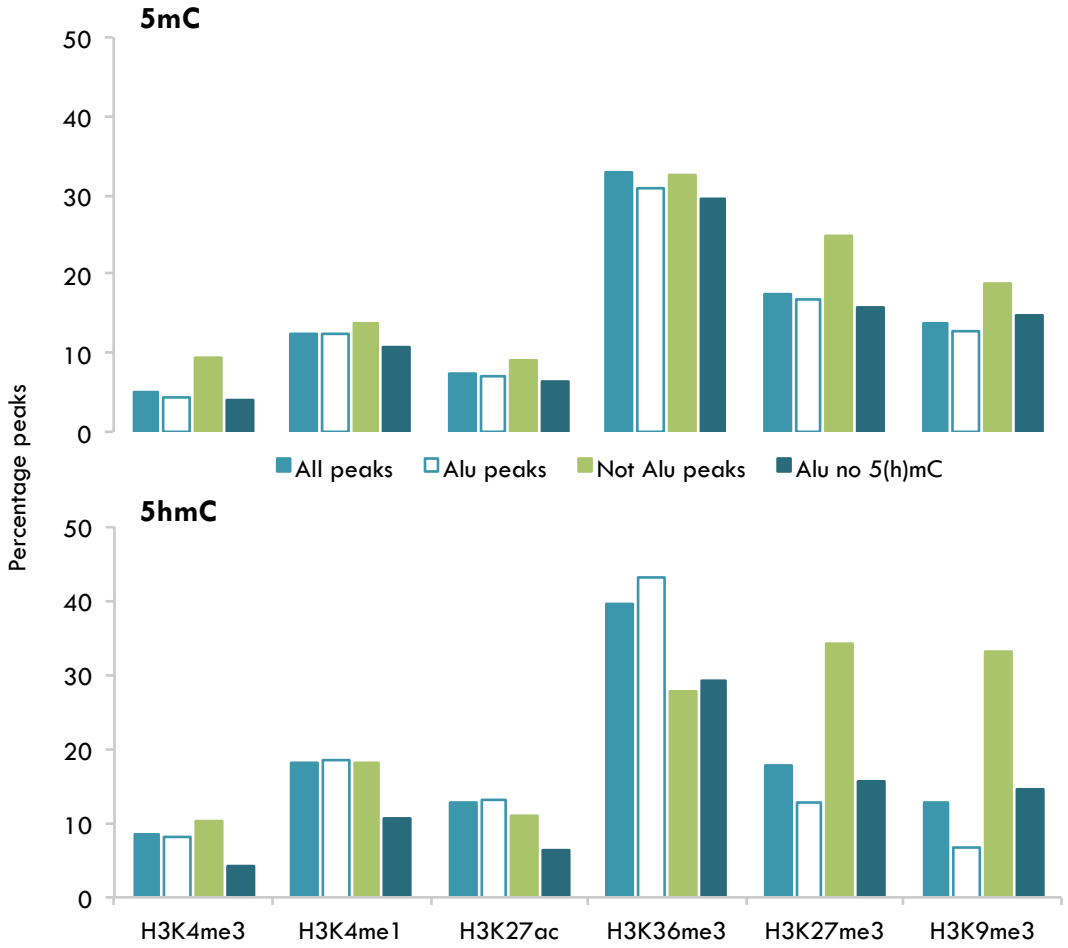

**Supplementary data 12. Alu element subfamilies**

All Alu elements were downloaded from repeatmasker and separated by subfamiles. The graphs show the number of Alu elements (count) of different sizes in each subfamily. This gives a visual representation of the amount of each subfamily present in the the human genome and shows that the majority of Alu elements are ~300bp.



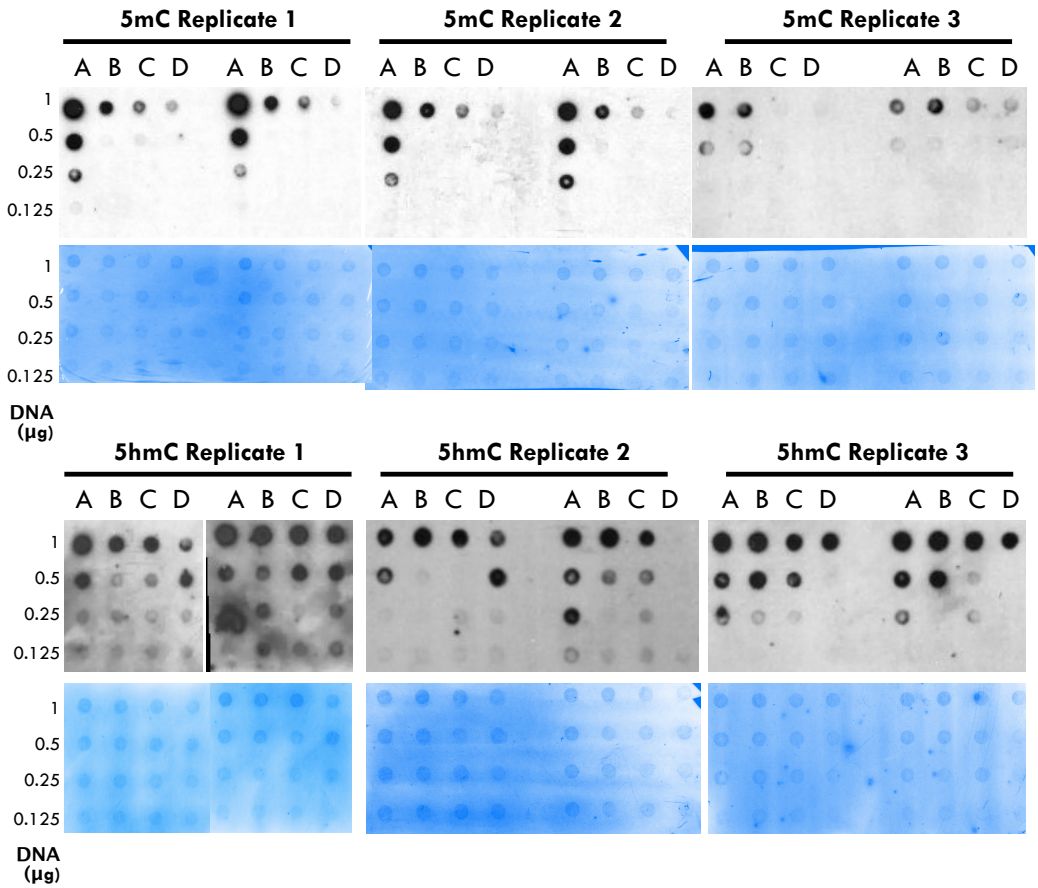

**Supplementary data 13. Overlap between HOK 5mC and 5hmC Alu groups and NHEK chromatin domains**

The results of HOK MeDIP and hMeDIP sequencing were separated based on the presence of an Alu element. All Alu elements were downloaded from repeat masker and those lacking 5mC or 5hmC in HOK cells were determined. NHEK chromatin domains were downloaded from the UCSC table browser. On Galaxy the BedTools 'intersect intervals' function was used to determine the overlap between 5(h)mC/Alu groups and NHEK chromatin domains. Results are shown as a percentage of total peaks.




**Supplementary data 14. Overlap between HOK 5mC and 5hmC Alu groups and NHEK histone modifications**

The results of HOK MeDIP and hMeDIP sequencing were separated based on the presence of an Alu element. All Alu elements were downloaded from repeat masker and those lacking 5mC or 5hmC in HOK cells were determined. NHEK histone modification peaks were downloaded from the UCSC table browser. On Galaxy the BedTools 'intersect intervals' function was used to determine the overlap between 5(h)mC/Alu groups and NHEK histone modifications. Results are shown as the percentage of total 5(h)mC/Alu peaks that overlap at all with the histone modification.




**Supplementary data 15. DNA dot blot analysis of 5mC and 5hmC of DAC treated VU40T cells**

VU40T cells were incubated with 100nM (B), 500nM(C), 1µM (D) DAC or an equivalent volume of vehicle (A) for 96h. DNA was extracted and DNA dot blotting performed using antibodies against 5mC (top) and 5hmC (bottom). Blots were incubated in methylene blue to stain for total DNA (bottom panels). Experiments were performed in triplicate, with duplicate wells per sample.

# References


1. Simard, E.P., L.A. Torre, and A. Jemal, *International trends in head and neck cancer incidence rates: differences by country, sex and anatomic site.* Oral Oncol, 2014. **50**(5): p. 387-403.
2. Castilho, R.M., C.H. Squarize, and L.O. Almeida, *Epigenetic Modifications and Head and Neck Cancer: Implications for Tumor Progression and Resistance to Therapy.* Int J Mol Sci, 2017. **18**(7).
3. Gong, W., et al., *Toward the use of precision medicine for the treatment of head and neck squamous cell carcinoma.* Oncotarget, 2017. **8**(2): p. 2141-2152.
4. Perez-Sayans, M., et al., *Genetic and molecular alterations associated with oral squamous cell cancer (Review).* Oncol Rep, 2009. **22**(6): p. 1277-82.
5. Lim, A., B. Solomon, and D. Rischin, *Molecular Insights Influencing the Management of Head and Neck Cancers.* Cancer Forum, 2015. **39**(1): p. 29-34.
6. Betiol, J., L.L. Villa, and L. Sichero, *Impact of HPV infection on the development of head and neck cancer.* Braz J Med Biol Res, 2013. **46**(3): p. 217-26.
7. Argiris, A., et al., *Head and neck cancer.* Lancet, 2008. **371**(9625): p. 1695-709.
8. Cancer Genome Atlas, N., *Comprehensive genomic characterization of head and neck squamous cell carcinomas.* Nature, 2015. **517**(7536): p. 576-82.
9. Yan, B., et al., *Signaling Networks of Activated Oncogenic and Altered Tumor Suppressor Genes in Head and Neck Cancer.* J Carcinog Mutagen, 2013. **Suppl 7**: p. 4.
10. Riaz, N., et al., *Unraveling the molecular genetics of head and neck cancer through genome-wide approaches.* Genes Dis, 2014. **1**(1): p. 75-86.
11. Bernstein, J.M., et al., *Molecular and cellular processes underlying the hallmarks of head and neck cancer.* Eur Arch Otorhinolaryngol, 2013. **270**(10): p. 2585-93.
12. Thomas, G.R., H. Nadiminti, and J. Regalado, *Molecular predictors of clinical outcome in patients with head and neck squamous cell carcinoma.* Int J Exp Pathol, 2005. **86**(6): p. 347-63.
13. Varelas, X. and M.A. Kukuruzinska, *Head and neck cancer: from research to therapy and cure.* Ann N Y Acad Sci, 2014. **1333**: p. 1-32.
14. Silva, P., et al., *Clinical and biological factors affecting response to radiotherapy in patients with head and neck cancer: a review.* Clin Otolaryngol, 2007. **32**(5): p. 337-45.

15. Gyawali, B., et al., *Chemotherapy in locally advanced head and neck squamous cell carcinoma.* Cancer Treat Rev, 2016. **44**: p. 10-6.
16. Du, Y., N.D. Peyser, and J.R. Grandis, *Integration of molecular targeted therapy with radiation in head and neck cancer.* Pharmacol Ther, 2014. **142**(1): p. 88-98.
17. Cohen, R.B., *Current challenges and clinical investigations of epidermal growth factor receptor (EGFR)- and ErbB family-targeted agents in the treatment of head and neck squamous cell carcinoma (HNSCC).* Cancer Treat Rev, 2014. **40**(4): p. 567-77.
18. Martinez-Useros, J. and J. Garcia-Foncillas, *The challenge of blocking a wider family members of EGFR against head and neck squamous cell carcinomas.* Oral Oncol, 2015. **51**(5): p. 423-30.
19. Ran, X. and K. Yang, *Inhibitors of the PD-1/PD-L1 axis for the treatment of head and neck cancer: current status and future perspectives.* Drug Des Devel Ther, 2017. **11**: p. 2007-2014.
20. Riggs, A.D., *X chromosome inactivation, differentiation, and DNA methylation revisited, with a tribute to Susumu Ohno.* Cytogenet Genome Res, 2002. **99**(1-4): p. 17-24.
21. Bird, A., *Perceptions of epigenetics.* Nature, 2007. **447**(7143): p. 396-8.
22. Goldberg, A.D., C.D. Allis, and E. Bernstein, *Epigenetics: a landscape takes shape.* Cell, 2007. **128**(4): p. 635-8.
23. Waddington, C.H., *The epigenotype. 1942.* Int J Epidemiol, 2012. **41**(1): p. 10-3.
24. Bernstein, B.E., A. Meissner, and E.S. Lander, *The mammalian epigenome.* Cell, 2007. **128**(4): p. 669-81.
25. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-3.
26. Moraes, F. and A. Goes, *A decade of human genome project conclusion: Scientific diffusion about our genome knowledge.* Biochem Mol Biol Educ, 2016. **44**(3): p. 215-23.
27. Jones, P.A., *Functions of DNA methylation: islands, start sites, gene bodies and beyond.* Nat Rev Genet, 2012. **13**(7): p. 484-92.
28. Felsenfeld, G. and M. Groudine, *Controlling the double helix.* Nature, 2003. **421**(6921): p. 448-53.
29. Gasche, J.A. and A. Goel, *Epigenetic mechanisms in oral carcinogenesis.* Future Oncol, 2012. **8**(11): p. 1407-25.
30. Li, E. and Y. Zhang, *DNA methylation in mammals.* Cold Spring Harb Perspect Biol, 2014. **6**(5): p. a019133.

31. Holliday, R. and J.E. Pugh, *DNA modification mechanisms and gene activity during development.* Science, 1975. **187**(4173): p. 226-32.
32. Bird, A.P., *Use of restriction enzymes to study eukaryotic DNA methylation: II. The symmetry of methylated sites supports semi-conservative copying of the methylation pattern.* J Mol Biol, 1978. **118**(1): p. 49-60.
33. Bird, A.P., *CpG-rich islands and the function of DNA methylation.* Nature, 1986. **321**(6067): p. 209-13.
34. Tsai, H.C. and S.B. Baylin, *Cancer epigenetics: linking basic biology to clinical medicine.* Cell Res, 2011. **21**(3): p. 502-17.
35. Coulondre, C., et al., *Molecular basis of base substitution hotspots in Escherichia coli.* Nature, 1978. **274**(5673): p. 775-80.
36. Bestor, T.H., *Cloning of a mammalian DNA methyltransferase.* Gene, 1988. **74**(1): p. 9-12.
37. Hermann, A., R. Goyal, and A. Jeltsch, *The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites.* J Biol Chem, 2004. **279**(46): p. 48350-9.
38. Probst, A.V., E. Dunleavy, and G. Almouzni, *Epigenetic inheritance during the cell cycle.* Nat Rev Mol Cell Biol, 2009. **10**(3): p. 192-206.
39. Subramaniam, D., et al., *DNA methyltransferases: a novel target for prevention and therapy.* Front Oncol, 2014. **4**: p. 80.
40. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.
41. Bird, A., *DNA methylation patterns and epigenetic memory.* Genes Dev, 2002. **16**(1): p. 6-21.
42. Auclair, G. and M. Weber, *Mechanisms of DNA methylation and demethylation in mammals.* Biochimie, 2012. **94**(11): p. 2202-11.
43. Jones, P.A. and G. Liang, *Rethinking how DNA methylation patterns are maintained.* Nat Rev Genet, 2009. **10**(11): p. 805-11.
44. Bird, A.P. and A.P. Wolffe, *Methylation-induced repression--belts, braces, and chromatin.* Cell, 1999. **99**(5): p. 451-4.
45. Ooi, S.K., et al., *DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA.* Nature, 2007. **448**(7154): p. 714-7.
46. Baubec, T., et al., *Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation.* Nature, 2015. **520**(7546): p. 243-7.
47. Baubec, T. and D. Schubeler, *Genomic patterns and context specific interpretation of DNA methylation.* Curr Opin Genet Dev, 2014. **25**: p. 85-92.

48. Padeken, J., P. Zeller, and S.M. Gasser, *Repeat DNA in genome organization and stability.* Curr Opin Genet Dev, 2015. **31**: p. 12-9.
49. de Koning, A.P., et al., *Repetitive elements may comprise over two-thirds of the human genome.* PLoS Genet, 2011. **7**(12): p. e1002384.
50. Wiench, M., et al., *DNA methylation status predicts cell type-specific enhancer activity.* EMBO J, 2011. **30**(15): p. 3028-39.
51. Lister, R., et al., *Human DNA methylomes at base resolution show widespread epigenomic differences.* Nature, 2009. **462**(7271): p. 315-22.
52. Shen, L. and Y. Zhang, *5-Hydroxymethylcytosine: generation, fate, and genomic distribution.* Curr Opin Cell Biol, 2013. **25**(3): p. 289-96.
53. Kriaucionis, S. and N. Heintz, *The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain.* Science, 2009. **324**(5929): p. 929-30.
54. Tahiliani, M., et al., *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.* Science, 2009. **324**(5929): p. 930-5.
55. Chen, Z., et al., *Decreased 5-hydroxymethylcytosine levels correlate with cancer progression and poor survival: a systematic review and meta-analysis.* Oncotarget, 2017. **8**(1): p. 1944-1952.
56. Haffner, M.C., et al., *Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers.* Oncotarget, 2011. **2**(8): p. 627-37.
57. Jawert, F., et al., *Loss of 5-hydroxymethylcytosine and TET2 in oral squamous cell carcinoma.* Anticancer Res, 2013. **33**(10): p. 4325-8.
58. Li, W. and M. Liu, *Distribution of 5-hydroxymethylcytosine in different human tissues.* J Nucleic Acids, 2011. **2011**: p. 870726.
59. Lian, C.G., et al., *Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma.* Cell, 2012. **150**(6): p. 1135-46.
60. Nestor, C.E., et al., *Rapid reprogramming of epigenetic and transcriptional profiles in mammalian culture systems.* Genome Biol, 2015. **16**: p. 11.
61. Putiri, E.L., et al., *Distinct and overlapping control of 5-methylcytosine and 5-hydroxymethylcytosine by the TET proteins in human cancer cells.* Genome Biol, 2014. **15**(6): p. R81.
62. Pfeifer, G.P., S. Kadam, and S.G. Jin, *5-hydroxymethylcytosine and its potential roles in development and cancer.* Epigenetics Chromatin, 2013. **6**(1): p. 10.
63. Ito, S., et al., *Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine.* Science, 2011. **333**(6047): p. 1300-3.

64. Branco, M.R., G. Ficz, and W. Reik, *Uncovering the role of 5-hydroxymethylcytosine in the epigenome.* Nat Rev Genet, 2011. **13**(1): p. 7-13.
65. Kohli, R.M. and Y. Zhang, *TET enzymes, TDG and the dynamics of DNA demethylation.* Nature, 2013. **502**(7472): p. 472-9.
66. Tan, L. and Y.G. Shi, *Tet family proteins and 5-hydroxymethylcytosine in development and disease.* Development, 2012. **139**(11): p. 1895-902.
67. Long, H.K., N.P. Blackledge, and R.J. Klose, *ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection.* Biochem Soc Trans, 2013. **41**(3): p. 727-40.
68. Zhang, H., et al., *TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine.* Cell Res, 2010. **20**(12): p. 1390-3.
69. Xu, Y., et al., *Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells.* Mol Cell, 2011. **42**(4): p. 451-64.
70. Hahn, M.A., P.E. Szabo, and G.P. Pfeifer, *5-Hydroxymethylcytosine: A stable or transient DNA modification?* Genomics, 2014. **104**(5): p. 314-23.
71. Fuks, F., et al., *The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation.* J Biol Chem, 2003. **278**(6): p. 4035-40.
72. Kitsera, N., et al., *Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemi-modified CpG dinucleotide in a gene promoter.* Nucleic Acids Res, 2017.
73. Spruijt, C.G., et al., *Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives.* Cell, 2013. **152**(5): p. 1146-59.
74. Iurlaro, M., et al., *A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation.* Genome Biol, 2013. **14**(10): p. R119.
75. Stresemann, C. and F. Lyko, *Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine.* Int J Cancer, 2008. **123**(1): p. 8-13.
76. Piccolo, F.M. and A.G. Fisher, *Getting rid of DNA methylation.* Trends Cell Biol, 2014. **24**(2): p. 136-43.
77. He, Y.F., et al., *Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA.* Science, 2011. **333**(6047): p. 1303-7.
78. Scarfo, I., et al., *Transposable elements: The enemies within.* Exp Hematol, 2016. **44**(10): p. 913-6.
79. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.

80. Burns, K.H., *Transposable elements in cancer.* Nat Rev Cancer, 2017. **17**(7): p. 415-424.
81. Deininger, P., *Alu elements: know the SINEs.* Genome Biol, 2011. **12**(12): p. 236.
82. Wang, C. and S. Huang, *Nuclear function of Alus.* Nucleus, 2014. **5**(2): p. 131-7.
83. Grover, D., et al., *Alu repeat analysis in the complete human genome: trends and variations with respect to genomic composition.* Bioinformatics, 2004. **20**(6): p. 813-7.
84. Luo, Y., X. Lu, and H. Xie, *Dynamic Alu methylation during normal development, aging, and tumorigenesis.* Biomed Res Int, 2014. **2014**: p. 784706.
85. Choi, S.H., et al., *Changes in DNA methylation of tandem DNA repeats are different from interspersed repeats in cancer.* Int J Cancer, 2009. **125**(3): p. 723-9.
86. Szpakowski, S., et al., *Loss of epigenetic silencing in tumors preferentially affects primate-specific retroelements.* Gene, 2009. **448**(2): p. 151-67.
87. Rodriguez, J., et al., *Genome-wide tracking of unmethylated DNA Alu repeats in normal and cancer cells.* Nucleic Acids Res, 2008. **36**(3): p. 770-84.
88. Baylin, S.B. and J.E. Ohm, *Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction?* Nat Rev Cancer, 2006. **6**(2): p. 107-16.
89. Arantes, L.M., et al., *Methylation as a biomarker for head and neck cancer.* Oral Oncol, 2014. **50**(6): p. 587-92.
90. Baylin, S.B. and P.A. Jones, *A decade of exploring the cancer epigenome - biological and translational implications.* Nat Rev Cancer, 2011. **11**(10): p. 726-34.
91. Yang, H., et al., *Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation.* Oncogene, 2013. **32**(5): p. 663-9.
92. Jones, P.A. and S.B. Baylin, *The fundamental role of epigenetic events in cancer.* Nat Rev Genet, 2002. **3**(6): p. 415-28.
93. Fraga, M.F., et al., *A mouse skin multistage carcinogenesis model reflects the aberrant DNA methylation patterns of human tumors.* Cancer Res, 2004. **64**(16): p. 5527-34.
94. Feinberg, A.P. and B. Vogelstein, *Hypomethylation distinguishes genes of some human cancers from their normal counterparts.* Nature, 1983. **301**(5895): p. 89-92.
95. Sproul, D. and R.R. Meehan, *Genomic insights into cancer-associated aberrant CpG island hypermethylation.* Brief Funct Genomics, 2013. **12**(3): p. 174-90.
96. Jones, P.A. and S.B. Baylin, *The epigenomics of cancer.* Cell, 2007. **128**(4): p. 683-92.

97. Ji, X., et al., *Diagnostic accuracy of DNA methylation for head and neck cancer varies by sample type and number of markers tested.* Oncotarget, 2016. **7**(48): p. 80019-80032.
98. Sanchez-Cespedes, M., et al., *Gene promoter hypermethylation in tumors and serum of head and neck cancer patients.* Cancer Res, 2000. **60**(4): p. 892-5.
99. Hasegawa, M., et al., *Patterns of gene promoter methylation in squamous cell cancer of the head and neck.* Oncogene, 2002. **21**(27): p. 4231-4236.
100. Steinmann, K., et al., *Frequent promoter hypermethylation of tumor-related genes in head and neck squamous cell carcinoma.* Oncology Reports, 2009. **22**(6): p. 1519-1526.
101. Guerrero-Preston, R., et al., *Key tumor suppressor genes inactivated by "greater promoter" methylation and somatic mutations in head and neck cancer.* Epigenetics, 2014. **9**(7): p. 1031-46.
102. Bennett, K.L., et al., *Frequently methylated tumor suppressor genes in head and neck squamous cell carcinoma.* Cancer Res, 2008. **68**(12): p. 4494-9.
103. Murata, A., et al., *TET family proteins and 5-hydroxymethylcytosine in esophageal squamous cell carcinoma.* Oncotarget, 2015. **6**(27): p. 23372-82.
104. Laird, A., et al., *5-hydroxymethylcytosine profiling as an indicator of cellular state.* Epigenomics, 2013. **5**(6): p. 655-69.
105. Kudo, Y., et al., *Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation.* Cancer Sci, 2012. **103**(4): p. 670-6.
106. Rasmussen, K.D., et al., *Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis.* Genes Dev, 2015. **29**(9): p. 910-22.
107. Ficz, G. and J.G. Gribben, *Loss of 5-hydroxymethylcytosine in cancer: Cause or consequence?* Genomics, 2014. **104**(5): p. 352-357.
108. Xu, W., et al., *Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases.* Cancer Cell, 2011. **19**(1): p. 17-30.
109. Figueroa, M.E., et al., *Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation.* Cancer Cell, 2010. **18**(6): p. 553-67.
110. Wang, Y., et al., *The level and clinical significance of 5-hydroxymethylcytosine in oral squamous cell carcinoma: An immunohistochemical study in 95 patients.* Pathol Res Pract, 2017. **213**(8): p. 969-974.
111. Ross, J.P., K.N. Rand, and P.L. Molloy, *Hypomethylation of repeated DNA sequences in cancer.* Epigenomics, 2010. **2**(2): p. 245-69.

112. Lee, E., et al., *Landscape of somatic retrotransposition in human cancers.* Science, 2012. **337**(6097): p. 967-71.
113. Daskalos, A., et al., *Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer.* Int J Cancer, 2009. **124**(1): p. 81-7.
114. Sirivanichsuntorn, P., et al., *LINE-1 and Alu hypomethylation in mucoepidermoid carcinoma.* BMC Clin Pathol, 2013. **13**: p. 10.
115. Cho, N.Y., et al., *Hypermethylation of CpG island loci and hypomethylation of LINE-1 and Alu repeats in prostate adenocarcinoma and their relationship to clinicopathological features.* J Pathol, 2007. **211**(3): p. 269-77.
116. Jones, P.A., J.P. Issa, and S. Baylin, *Targeting the cancer epigenome for therapy.* Nat Rev Genet, 2016. **17**(10): p. 630-41.
117. Tsai, H.C., et al., *Transient low doses of DNA-demethylating agents exert durable antitumor effects on hematological and epithelial tumor cells.* Cancer Cell, 2012. **21**(3): p. 430-46.
118. Kantarjian, H., et al., *Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study.* Cancer, 2006. **106**(8): p. 1794-803.
119. Juttermann, R., E. Li, and R. Jaenisch, *Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation.* Proc Natl Acad Sci U S A, 1994. **91**(25): p. 11797-801.
120. Juo, Y.Y., et al., *Epigenetic therapy for solid tumors: from bench science to clinical trials.* Epigenomics, 2015. **7**(2): p. 215-35.
121. Nervi, C., E. De Marinis, and G. Codacci-Pisanelli, *Epigenetic treatment of solid tumours: a review of clinical trials.* Clin Epigenetics, 2015. **7**: p. 127.
122. Issa, J.P., et al., *Phase 1 study of low-dose prolonged exposure schedules of the hypomethylating agent 5-aza-2'-deoxycytidine (decitabine) in hematopoietic malignancies.* Blood, 2004. **103**(5): p. 1635-40.
123. Leone, G., et al., *Inhibitors of DNA methylation in the treatment of hematological malignancies and MDS.* Clin Immunol, 2003. **109**(1): p. 89-102.
124. Yang, A.S., et al., *DNA methylation changes after 5-aza-2'-deoxycytidine therapy in patients with leukemia.* Cancer Res, 2006. **66**(10): p. 5495-503.
125. Cowan, L.A., S. Talwar, and A.S. Yang, *Will DNA methylation inhibitors work in solid tumors? A review of the clinical experience with azacitidine and decitabine in solid tumors.* Epigenomics, 2010. **2**(1): p. 71-86.

126. Cashen, A.F., et al., *Pharmacokinetics of decitabine administered as a 3-h infusion to patients with acute myeloid leukemia (AML) or myelodysplastic syndrome (MDS).* Cancer Chemother Pharmacol, 2008. **61**(5): p. 759-66.
127. Fenaux, P., et al., *Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study.* Lancet Oncol, 2009. **10**(3): p. 223-32.
128. Abele, R., et al., *The EORTC Early Clinical Trials Cooperative Group experience with 5-aza-2'-deoxycytidine (NSC 127716) in patients with colo-rectal, head and neck, renal carcinomas and malignant melanomas.* Eur J Cancer Clin Oncol, 1987. **23**(12): p. 1921-4.
129. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-82.
130. Appleton, K., et al., *Phase I and pharmacodynamic trial of the DNA methyltransferase inhibitor decitabine and carboplatin in solid tumors.* J Clin Oncol, 2007. **25**(29): p. 4603-9.
131. Viet, C.T., et al., *Decitabine rescues cisplatin resistance in head and neck squamous cell carcinoma.* PLoS One, 2014. **9**(11): p. e112880.
132. Clavel, M., et al., *5-Aza-2'-deoxycytidine (NSC 127716) in non-seminomatous testicular cancer. Phase II from the EORTC Early Clinical Trials Cooperative Group and Genito-Urinary Group.* Ann Oncol, 1992. **3**(5): p. 399-400.
133. Momparler, R.L., et al., *Pilot phase I-II study on 5-aza-2'-deoxycytidine (Decitabine) in patients with metastatic lung cancer.* Anticancer Drugs, 1997. **8**(4): p. 358-68.
134. Stathis, A., et al., *Phase I study of decitabine in combination with vorinostat in patients with advanced solid tumors and non-Hodgkin's lymphomas.* Clin Cancer Res, 2011. **17**(6): p. 1582-90.
135. Chu, B.F., et al., *Phase I study of 5-aza-2'-deoxycytidine in combination with valproic acid in non-small-cell lung cancer.* Cancer Chemother Pharmacol, 2013. **71**(1): p. 115-21.
136. Schwartsmann, G., et al., *A phase I trial of cisplatin plus decitabine, a new DNA-hypomethylating agent, in patients with advanced solid tumors and a follow-up early phase II evaluation in patients with inoperable non-small cell lung cancer.* Invest New Drugs, 2000. **18**(1): p. 83-91.
137. Pohlmann, P., et al., *Phase II trial of cisplatin plus decitabine, a new DNA hypomethylating agent, in patients with advanced squamous cell carcinoma of the cervix.* Am J Clin Oncol, 2002. **25**(5): p. 496-501.

138. Matei, D., et al., *Epigenetic resensitization to platinum in ovarian cancer.* Cancer Res, 2012. **72**(9): p. 2197-205.
139. Tawbi, H.A., et al., *Safety and efficacy of decitabine in combination with temozolomide in metastatic melanoma: a phase I/II study and pharmacokinetic analysis.* Ann Oncol, 2013. **24**(4): p. 1112-9.
140. Fang, F., et al., *A phase 1 and pharmacodynamic study of decitabine in combination with carboplatin in patients with recurrent, platinum-resistant, epithelial ovarian cancer.* Cancer, 2010. **116**(17): p. 4043-53.
141. Glasspool, R.M., et al., *A randomised, phase II trial of the DNA-hypomethylating agent 5-aza-2'-deoxycytidine (decitabine) in combination with carboplatin vs carboplatin alone in patients with recurrent, partially platinum-sensitive ovarian cancer.* Br J Cancer, 2014. **110**(8): p. 1923-9.
142. Gupta, S.C., et al., *Cancer drug discovery by repurposing: teaching new tricks to old dogs.* Trends Pharmacol Sci, 2013. **34**(9): p. 508-17.
143. Paul, S.M., et al., *How to improve R&D productivity: the pharmaceutical industry's grand challenge.* Nat Rev Drug Discov, 2010. **9**(3): p. 203-14.
144. DiMasi, J.A., R.W. Hansen, and H.G. Grabowski, *The price of innovation: new estimates of drug development costs.* J Health Econ, 2003. **22**(2): p. 151-85.
145. Sleire, L., et al., *Drug repurposing in cancer.* Pharmacol Res, 2017. **124**: p. 74-91.
146. Boguski, M.S., K.D. Mandl, and V.P. Sukhatme, *Drug discovery. Repurposing with a difference.* Science, 2009. **324**(5933): p. 1394-5.
147. Ashburn, T.T. and K.B. Thor, *Drug repositioning: identifying and developing new uses for existing drugs.* Nat Rev Drug Discov, 2004. **3**(8): p. 673-83.
148. Pantziarka, P., et al., *The Repurposing Drugs in Oncology (ReDO) Project.* Ecancermedicalscience, 2014. **8**: p. 442.
149. Oprea, T.I. and J. Mestres, *Drug repurposing: far beyond new targets for old drugs.* AAPS J, 2012. **14**(4): p. 759-63.
150. Gelijns, A.C., N. Rosenberg, and A.J. Moskowitz, *Capturing the unexpected benefits of medical research.* N Engl J Med, 1998. **339**(10): p. 693-8.
151. Hermsen, M.A., et al., *Centromeric breakage as a major cause of cytogenetic abnormalities in oral squamous cell carcinoma.* Genes Chromosomes Cancer, 1996. **15**(1): p. 1-9.
152. White, J.S., et al., *The influence of clinical and demographic risk factors on the establishment of head and neck squamous cell carcinoma cell lines.* Oral Oncol, 2007. **43**(7): p. 701-12.

153. Martin, D., et al., *The head and neck cancer cell oncogenome: a platform for the development of precision molecular therapies.* Oncotarget, 2014. **5**(19): p. 8906-23.

154. Wang, H., et al., *Role for EPS8 in squamous carcinogenesis.* Carcinogenesis, 2009. **30**(1): p. 165-74.

155. Khanim, F.L., et al., *Redeployment-based drug screening identifies the anti-helminthic niclosamide as anti-myeloma therapy that also reduces free light chain production.* Blood Cancer J, 2011. **1**(10): p. e39.

156. Chou, T.C., *Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies.* Pharmacol Rev, 2006. **58**(3): p. 621-81.

157. Brown, T., *Dot and slot blotting of DNA.* Curr Protoc Mol Biol, 2001. **Chapter 2**: p. Unit2 9B.

158. Roulois, D., et al., *DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts.* Cell, 2015. **162**(5): p. 961-73.

159. Untergasser, A., et al., *Primer3--new capabilities and interfaces.* Nucleic Acids Res, 2012. **40**(15): p. e115.

160. Afgan, E., et al., *The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update.* Nucleic Acids Res, 2016. **44**(W1): p. W3-W10.

161. Kim, D., B. Langmead, and S.L. Salzberg, *HISAT: a fast spliced aligner with low memory requirements.* Nat Methods, 2015. **12**(4): p. 357-60.

162. Pertea, M., et al., *Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown.* Nat Protoc, 2016. **11**(9): p. 1650-67.

163. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

164. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.

165. Supek, F., et al., *REVIGO summarizes and visualizes long lists of gene ontology terms.* PLoS One, 2011. **6**(7): p. e21800.

166. Oliveros, *Venny. An interactive tool for comparing lists with Venn's diagrams.* 2007-2015.

167. Metsalu, T. and J. Vilo, *ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap.* Nucleic Acids Res, 2015. **43**(W1): p. W566-70.

168. Mohn, F., et al., *Methylated DNA immunoprecipitation (MeDIP).* Methods Mol Biol, 2009. **507**: p. 55-64.

169. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
170. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Mol Cell, 2010. **38**(4): p. 576-89.
171. Karolchik, D., et al., *The UCSC Table Browser data retrieval tool.* Nucleic Acids Res, 2004. **32**(Database issue): p. D493-6.
172. Ernst, J. and M. Kellis, *Discovery and characterization of chromatin states for systematic annotation of the human genome.* Nat Biotechnol, 2010. **28**(8): p. 817-25.
173. Rosenbloom, K.R., et al., *ENCODE data in the UCSC Genome Browser: year 5 update.* Nucleic Acids Res, 2013. **41**(Database issue): p. D56-63.
174. Quinlan, A.R., *BEDTools: The Swiss-Army Tool for Genome Feature Analysis.* Curr Protoc Bioinformatics, 2014. **47**: p. 11 12 1-34.
175. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-2.
176. Ramirez, F., et al., *deepTools2: a next generation web server for deep-sequencing data analysis.* Nucleic Acids Res, 2016. **44**(W1): p. W160-5.
177. McLean, C.Y., et al., *GREAT improves functional interpretation of cis-regulatory regions.* Nat Biotechnol, 2010. **28**(5): p. 495-501.
178. Santi, D.V., C.E. Garrett, and P.J. Barr, *On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs.* Cell, 1983. **33**(1): p. 9-10.
179. Ferguson, A.T., et al., *Role of estrogen receptor gene demethylation and DNA methyltransferase.DNA adduct formation in 5-aza-2'deoxycytidine-induced cytotoxicity in human breast cancer cells.* J Biol Chem, 1997. **272**(51): p. 32260-6.
180. Chiappinelli, K.B., et al., *Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses.* Cell, 2015. **162**(5): p. 974-86.
181. Palii, S.S., et al., *DNA methylation inhibitor 5-Aza-2'-deoxycytidine induces reversible genome-wide DNA damage that is distinctly influenced by DNA methyltransferases 1 and 3B.* Mol Cell Biol, 2008. **28**(2): p. 752-71.
182. Zhu, W.G., et al., *5-aza-2'-deoxycytidine activates the p53/p21Waf1/Cip1 pathway to inhibit cell proliferation.* J Biol Chem, 2004. **279**(15): p. 15161-6.
183. Wu, L., et al., *High expression of the human equilibrative nucleoside transporter 1 gene predicts a good response to decitabine in patients with myelodysplastic syndrome.* J Transl Med, 2016. **14**: p. 66.

184. Qin, T., et al., *Mechanisms of resistance to decitabine in the myelodysplastic syndrome.* PLoS One, 2011. **6**(8): p. e23372.
185. Qin, T., et al., *Mechanisms of resistance to 5-aza-2'-deoxycytidine in human cancer cell lines.* Blood, 2009. **113**(3): p. 659-67.
186. Koczor, C.A., R.A. Torres, and W. Lewis, *The role of transporters in the toxicity of nucleoside and nucleotide analogs.* Expert Opin Drug Metab Toxicol, 2012. **8**(6): p. 665-76.
187. EU, E.M.A.-E., *Dacogen 50 mg powder for concentrate for solution for infusion.* 2012.
188. Huang, D.W., et al., *DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W169-75.
189. *[Venny Snellman, in memoriam].* Sykepleien, 1966. **53**(5): p. 121.
190. Gan, C.P., et al., *Valproic acid: growth inhibition of head and neck cancer by induction of terminal differentiation and senescence.* Head Neck, 2012. **34**(3): p. 344-53.
191. Graham, G.G., et al., *The modern pharmacology of paracetamol: therapeutic actions, mechanism of action, metabolism, toxicity and recent pharmacological findings.* Inflammopharmacology, 2013. **21**(3): p. 201-32.
192. Bertolini, A., et al., *Paracetamol: new vistas of an old drug.* CNS Drug Rev, 2006. **12**(3-4): p. 250-75.
193. Ghanem, C.I., et al., *Acetaminophen from liver to brain: New insights into drug pharmacological action and toxicity.* Pharmacol Res, 2016. **109**: p. 119-31.
194. Cebola, I. and M.A. Peinado, *Epigenetic deregulation of the COX pathway in cancer.* Prog Lipid Res, 2012. **51**(4): p. 301-13.
195. Schwartz, J.I., et al., *Inhibition of prostacyclin and thromboxane biosynthesis in healthy volunteers by single and multiple doses of acetaminophen and indomethacin.* Clin Pharmacol Drug Dev, 2015. **4**(5): p. 337-45.
196. Liu, B., L. Qu, and S. Yan, *Cyclooxygenase-2 promotes tumor growth and suppresses tumor immunity.* Cancer Cell Int, 2015. **15**: p. 106.
197. Boutaud, O., et al., *Determinants of the cellular specificity of acetaminophen as an inhibitor of prostaglandin H(2) synthases.* Proc Natl Acad Sci U S A, 2002. **99**(10): p. 7130-5.
198. Lucas, R., et al., *Cellular mechanisms of acetaminophen: role of cyclo-oxygenase.* FASEB J, 2005. **19**(6): p. 635-7.

199. Graham, G.G. and K.F. Scott, *Mechanisms of action of paracetamol and related analgesics.* Inflammopharmacology, 2003. **11**(4): p. 401-13.
200. Chandrasekharan, N.V., et al., *COX-3, a cyclooxygenase-1 variant inhibited by acetaminophen and other analgesic/antipyretic drugs: cloning, structure, and expression.* Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13926-31.
201. Kis, B., J.A. Snipes, and D.W. Busija, *Acetaminophen and the cyclooxygenase-3 puzzle: sorting out facts, fictions, and uncertainties.* J Pharmacol Exp Ther, 2005. **315**(1): p. 1-7.
202. Qin, N., et al., *Cloning, expression, and functional characterization of human cyclooxygenase-1 splicing variants: Evidence for intron 1 retention.* Journal of Pharmacology and Experimental Therapeutics, 2005. **315**(3): p. 1298-1305.
203. Hogestatt, E.D., et al., *Conversion of acetaminophen to the bioactive N-acylphenolamine AM404 via fatty acid amide hydrolase-dependent arachidonic acid conjugation in the nervous system.* J Biol Chem, 2005. **280**(36): p. 31405-12.
204. Mallet, C., et al., *Endocannabinoid and serotonergic systems are needed for acetaminophen-induced analgesia.* Pain, 2008. **139**(1): p. 190-200.
205. Mallet, C., et al., *TRPV1 in brain is involved in acetaminophen-induced antinociception.* PLoS One, 2010. **5**(9).
206. Gunawan, B.K., et al., *c-Jun N-terminal kinase plays a major role in murine acetaminophen hepatotoxicity.* Gastroenterology, 2006. **131**(1): p. 165-78.
207. Badmann, A., et al., *Role of TRAIL and the pro-apoptotic Bcl-2 homolog Bim in acetaminophen-induced liver damage.* Cell Death Dis, 2011. **2**: p. e171.
208. Neuwelt, A.J., et al., *Preclinical high-dose acetaminophen with N-acetylcysteine rescue enhances the efficacy of cisplatin chemotherapy in atypical teratoid rhabdoid tumors.* Pediatr Blood Cancer, 2014. **61**(1): p. 120-7.
209. Kobrinsky, N.L., et al., *Treatment of advanced malignancies with high-dose acetaminophen and N-acetylcysteine rescue.* Cancer Invest, 1996. **14**(3): p. 202-10.
210. Das, J., et al., *Acetaminophen induced acute liver failure via oxidative stress and JNK activation: protective role of taurine by the suppression of cytochrome P450 2E1.* Free Radic Res, 2010. **44**(3): p. 340-55.
211. Rothwell, P.M., et al., *Short-term effects of daily aspirin on cancer incidence, mortality, and non-vascular death: analysis of the time course of risks and benefits in 51 randomised controlled trials.* Lancet, 2012. **379**(9826): p. 1602-1612.

212. Algra, A.M. and P.M. Rothwell, *Effects of regular aspirin on long-term cancer incidence and metastasis: a systematic comparison of evidence from observational studies versus randomised trials.* Lancet Oncology, 2012. **13**(5): p. 518-527.

213. Rothwell, P.M., et al., *Effect of daily aspirin on risk of cancer metastasis: a study of incident cancers during randomised controlled trials.* Lancet, 2012. **379**(9826): p. 1591-1601.

214. Torti, D.C., et al., *Analgesic and nonsteroidal anti-inflammatory use in relation to nonmelanoma skin cancer: a population-based case-control study.* J Am Acad Dermatol, 2011. **65**(2): p. 304-12.

215. Jacobs, E.J., et al., *A large cohort study of long-term acetaminophen use and prostate cancer incidence.* Cancer Epidemiol Biomarkers Prev, 2011. **20**(7): p. 1322-8.

216. Saba, N.F., et al., *Role of cyclooxygenase-2 in tumor progression and survival of head and neck squamous cell carcinoma.* Cancer Prev Res (Phila), 2009. **2**(9): p. 823-9.

217. Posadas, I., et al., *Acetaminophen potentiates staurosporine-induced death in a human neuroblastoma cell line.* Br J Pharmacol, 2007. **150**(5): p. 577-85.

218. Wu, Y.J., et al., *Acetaminophen enhances cisplatin- and paclitaxel-mediated cytotoxicity to SKOV3 human ovarian carcinoma.* Anticancer Res, 2013. **33**(6): p. 2391-400.

219. Feng, L. and Z. Wang, *Chemopreventive effect of celecoxib in oral precancers and cancers.* Laryngoscope, 2006. **116**(10): p. 1842-5.

220. Mestre, J.R., et al., *Inhibition of cyclooxygenase-2 expression. An approach to preventing head and neck cancer.* Ann N Y Acad Sci, 1999. **889**: p. 62-71.

221. Kim, Y.Y., et al., *Anti-cancer effects of celecoxib in head and neck carcinoma.* Mol Cells, 2010. **29**(2): p. 185-94.

222. Gao, J., et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.* Sci Signal, 2013. **6**(269): p. pl1.

223. Manyike, P.T., et al., *Contribution of CYP2E1 and CYP3A to acetaminophen reactive metabolite formation.* Clin Pharmacol Ther, 2000. **67**(3): p. 275-82.

224. Dahlin, D.C., et al., *N-acetyl-p-benzoquinone imine: a cytochrome P-450-mediated oxidation product of acetaminophen.* Proc Natl Acad Sci U S A, 1984. **81**(5): p. 1327-31.

225. Fukuhara, K., et al., *A 1H NMR-based metabolomics approach for mechanistic insight into acetaminophen-induced hepatotoxicity.* Drug Metab Pharmacokinet, 2011. **26**(4): p. 399-406.

226. Win, S., et al., *c-Jun N-terminal kinase (JNK)-dependent acute liver injury from acetaminophen or tumor necrosis factor (TNF) requires mitochondrial Sab protein expression in mice.* J Biol Chem, 2011. **286**(40): p. 35071-8.

227. Nakamoto, N., et al., *Cyclooxygenase-2 inhibitor and interferon-beta synergistically induce apoptosis in human hepatoma cells in vitro and in vivo.* Int J Oncol, 2006. **29**(3): p. 625-35.

228. James, L.P., P.R. Mayeux, and J.A. Hinson, *Acetaminophen-induced hepatotoxicity.* Drug Metab Dispos, 2003. **31**(12): p. 1499-506.

229. Robertson, K.D., et al., *Differential mRNA expression of the human DNA methyltransferases (DNMTs) 1, 3a and 3b during the G(0)/G(1) to S phase transition in normal and tumor cells.* Nucleic Acids Res, 2000. **28**(10): p. 2108-13.

230. Plongthongkum, N., D.H. Diep, and K. Zhang, *Advances in the profiling of DNA modifications: cytosine methylation and beyond.* Nat Rev Genet, 2014. **15**(10): p. 647-61.

231. Skvortsova, K., et al., *Comprehensive evaluation of genome-wide 5-hydroxymethylcytosine profiling approaches in human DNA.* Epigenetics Chromatin, 2017. **10**: p. 16.

232. Meissner, A., et al., *Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis.* Nucleic Acids Res, 2005. **33**(18): p. 5868-77.

233. Huang, Y., et al., *The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing.* PLoS One, 2010. **5**(1): p. e8888.

234. Yu, M., et al., *Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome.* Cell, 2012. **149**(6): p. 1368-80.

235. Booth, M.J., et al., *Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution.* Science, 2012. **336**(6083): p. 934-7.

236. Song, C.X., C. Yi, and C. He, *Mapping recently identified nucleotide variants in the genome and transcriptome.* Nat Biotechnol, 2012. **30**(11): p. 1107-16.

237. Nestor, C.E. and R.R. Meehan, *Hydroxymethylated DNA immunoprecipitation (hmeDIP).* Methods Mol Biol, 2014. **1094**: p. 259-67.

238. Su, M., et al., *Evolution of Alu elements toward enhancers.* Cell Rep, 2014. **7**(2): p. 376-85.

239. Tanaka, Y., et al., *Effects of Alu elements on global nucleosome positioning in the human genome.* BMC Genomics, 2010. **11**: p. 309.

240. O'Connell, M.A., N.M. Mannion, and L.P. Keegan, *The Epitranscriptome and Innate Immunity.* PLoS Genet, 2015. **11**(12): p. e1005687.
241. Chen, L.L., J.N. DeCerbo, and G.G. Carmichael, *Alu element-mediated gene silencing.* EMBO J, 2008. **27**(12): p. 1694-705.
242. Zhang, L., J.G. Chen, and Q. Zhao, *Regulatory roles of Alu transcript on gene expression.* Exp Cell Res, 2015. **338**(1): p. 113-8.
243. Sela, N., et al., *Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome.* Genome Biol, 2007. **8**(6): p. R127.
244. Attig, J., et al., *Splicing repression allows the gradual emergence of new Alu-exons in primate evolution.* Elife, 2016. **5**.
245. Oler, A.J., et al., *Alu expression in human cell lines and their retrotranspositional potential.* Mob DNA, 2012. **3**(1): p. 11.
246. Ji, X., et al., *Diagnostic accuracy of DNA methylation for head and neck cancer varies by sample type and number of markers tested.* Oncotarget, 2016. **7**(48): p. 80019-80032.
247. Wang, J., et al., *Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium.* Nucleic Acids Res, 2013. **41**(Database issue): p. D171-6.
248. Hubley, R., et al., *The Dfam database of repetitive DNA families.* Nucleic Acids Res, 2016. **44**(D1): p. D81-9.
249. Konkel, M.K., et al., *Sequence Analysis and Characterization of Active Human Alu Subfamilies Based on the 1000 Genomes Pilot Project.* Genome Biol Evol, 2015. **7**(9): p. 2608-22.
250. Jurka, J., *Repbase update: a database and an electronic journal of repetitive elements.* Trends Genet, 2000. **16**(9): p. 418-20.
251. Tarailo-Graovac, M. and N. Chen, *Using RepeatMasker to identify repetitive elements in genomic sequences.* Curr Protoc Bioinformatics, 2009. **Chapter 4**: p. Unit 4 10.
252. Chen, E.Y., et al., *Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.* BMC Bioinformatics, 2013. **14**: p. 128.
253. Kuleshov, M.V., et al., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.* Nucleic Acids Res, 2016. **44**(W1): p. W90-7.
254. Hagan, C.R., R.F. Sheffield, and C.M. Rudin, *Human Alu element retrotransposition induced by genotoxic stress.* Nat Genet, 2003. **35**(3): p. 219-20.

255. Papin, C., et al., *Combinatorial DNA methylation codes at repetitive elements.* Genome Res, 2017. **27**(6): p. 934-946.
256. Jorda, M., et al., *The epigenetic landscape of Alu repeats delineates the structural and functional genomic architecture of colon cancer cells.* Genome Res, 2017. **27**(1): p. 118-132.
257. Rollins, R.A., et al., *Large-scale structure of genomic methylation patterns.* Genome Res, 2006. **16**(2): p. 157-63.
258. Moolhuijzen, P., et al., *The transcript repeat element: the human Alu sequence as a component of gene networks influencing cancer.* Funct Integr Genomics, 2010. **10**(3): p. 307-19.
259. Macia, A., et al., *Epigenetic control of retrotransposon expression in human embryonic stem cells.* Mol Cell Biol, 2011. **31**(2): p. 300-16.
260. Li, X., et al., *Whole-genome analysis of the methylome and hydroxymethylome in normal and malignant lung and liver.* Genome Res, 2016. **26**(12): p. 1730-1741.
261. Chen, K., et al., *Loss of 5-hydroxymethylcytosine is linked to gene body hypermethylation in kidney cancer.* Cell Res, 2016. **26**(1): p. 103-18.
262. Lentini, A., et al., *A reassessment of DNA-immunoprecipitation-based genomic profiling.* Nat Methods, 2018. **15**(7): p. 499-504.
263. Liang, G., et al., *Analysis of gene induction in human fibroblasts and bladder cancer cells exposed to the methylation inhibitor 5-aza-2'-deoxycytidine.* Cancer Res, 2002. **62**(4): p. 961-6.
264. Karpf, A.R., et al., *Inhibition of DNA methyltransferase stimulates the expression of signal transducer and activator of transcription 1, 2, and 3 genes in colon tumor cells.* Proc Natl Acad Sci U S A, 1999. **96**(24): p. 14007-12.
265. Li, H., et al., *Immune regulation by low doses of the DNA methyltransferase inhibitor 5-azacitidine in common human epithelial cancers.* Oncotarget, 2014. **5**(3): p. 587-98.
266. Stone, M.L., et al., *Epigenetic therapy activates type I interferon signaling in murine ovarian cancer to reduce immunosuppression and tumor burden.* Proc Natl Acad Sci U S A, 2017. **114**(51): p. E10981-E10990.
267. Zhao, H., et al., *The immunomodulatory anticancer agent, RRx-001, induces an interferon response through epigenetic induction of viral mimicry.* Clin Epigenetics, 2017. **9**: p. 4.
268. Liu, M., et al., *Vitamin C increases viral mimicry induced by 5-aza-2'-deoxycytidine.* Proc Natl Acad Sci U S A, 2016. **113**(37): p. 10238-44.

269. Chowdhury, B., et al., *The hypomethylating agent Decitabine causes a paradoxical increase in 5-hydroxymethylcytosine in human leukemia cells.* Sci Rep, 2015. **5**: p. 9281.

270. Sajadian, S.O., et al., *Induction of active demethylation and 5hmC formation by 5-azacytidine is TET2 dependent and suggests new treatment strategies against hepatocellular carcinoma.* Clin Epigenetics, 2015. **7**: p. 98.

271. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.* Cancer Discov, 2012. **2**(5): p. 401-4.

272. Ghoshal, K., et al., *5-Aza-deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the KEN box, bromo-adjacent homology domain, and nuclear localization signal.* Mol Cell Biol, 2005. **25**(11): p. 4727-41.

273. Teissandier, A. and D. Bourc'his, *Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription.* EMBO J, 2017. **36**(11): p. 1471-1473.

274. Mendonca, A., et al., *Hydroxymethylation of DNA influences nucleosomal conformation and stability in vitro.* Biochim Biophys Acta, 2014. **1839**(11): p. 1323-9.

275. Xu, X., et al., *A CRISPR-based approach for targeted DNA demethylation.* Cell Discov, 2016. **2**: p. 16009.

276. Anderson, B.J., *Paracetamol (Acetaminophen): mechanisms of action.* Paediatr Anaesth, 2008. **18**(10): p. 915-21.

277. Lin, R.K. and Y.C. Wang, *Dysregulated transcriptional and post-translational control of DNA methyltransferases in cancer.* Cell Biosci, 2014. **4**: p. 46.

278. Amodio, N., et al., *miR-29s: a family of epi-miRNAs with therapeutic implications in hematologic malignancies.* Oncotarget, 2015. **6**(15): p. 12837-61.

279. Ramassone, A., et al., *Epigenetics and MicroRNAs in Cancer.* Int J Mol Sci, 2018. **19**(2).

280. Garzon, R., et al., *MicroRNA-29b induces global DNA hypomethylation and tumor suppressor gene reexpression in acute myeloid leukemia by targeting directly DNMT3A and 3B and indirectly DNMT1.* Blood, 2009. **113**(25): p. 6411-8.

281. Fabbri, M., et al., *MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B.* Proc Natl Acad Sci U S A, 2007. **104**(40): p. 15805-10.

282. Wang, Y., T. Chen, and W. Tong, *miRNAs and their application in drug-induced liver injury.* Biomark Med, 2014. **8**(2): p. 161-72.