UNIVERSITY OF LIVERPOOL

# Genome-Wide DNA Methylation and Gene Expression Profiling of Cancer-Associated Myofibroblasts

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

**Hanna Najgebauer**

January 2016

# Acknowledgements

First and foremost I would like to thank my primary supervisor Professor Chris Sanderson for giving me the opportunity to embark on this adventure and learn about cancer biology, epigenetics and bioinformatics. I am thankful for all the help and scientific guidance throughout this PhD project. I would also like to thank my secondary supervisor Professor Andrea Varro for providing the cells for this study and for helping set up the myofibroblast cultures and proliferation and migration assays.

Special thanks to Dr Lakis Liloglou for his expert advice and time spent helping me set up pyrosequencing and qPCR assays. Also to Dr Jithesh Puthen for his initial help and advice regarding the bioinformatics analysis workflow, and to Dr Andy Jarnuczak for helping with proteomic data processing and analysis.

I would like to thank all who have helped me in different ways during the last 4 years: Dave, Emily, Jo, Amy, Jen, Bron, Olivier, Ben, Amelia, Israa, Bubaraye and Ahmed. Thank you for your help in the lab, it was much appreciated. You have all made this experience memorable and made day-to-day lab life enjoyable.

A huge thank you to the Cancer Research UK for funding this work.

Most importantly, I would like to thank my family and Andy for their continuous support.

# Contents

**Acknowledgements** ..... **i**

**Contents** ..... ii

**List of Figures** ..... ix

**List of Tables** ..... xv

**Appendices** ..... xviii

**Abbreviations** ..... xix

**Abstract** ..... xxii

**Introduction** ..... 1

**1.1 Overview** ..... 2

**1.2 Cancer** ..... 3

1.2.1 Molecular basis of cancer ..... 3

1.2.2 Hallmarks of cancer ..... 4

1.2.3 Gastric cancer ..... 4

1.2.3.1 Risk factors ..... 5

1.2.3.2 Classification ..... 5

1.2.3.3 Clinical outcomes ..... 7

1.2.3.4 Molecular biomarkers ..... 8

**1.3 Tumour microenvironment** ..... 9

1.3.1 Components of tumour microenvironment ..... 9

1.3.2 Stromal-epithelial interactions ..... 10

1.3.3 Myofibroblasts ..... 12

1.3.3.1 Myofibroblasts in wound healing ..... 12

1.3.3.2 Myofibroblasts in cancer ..... 13

1.3.4 Hypoxia ..... 17

1.3.4.1 Cellular responses to hypoxia ..... 17

1.3.4.2 Hypoxia-associated epigenetic changes ..... 18

1.3.4.3 Hypoxia in gastric cancer ..... 19

1.3.5 Targeting the tumour microenvironment ..... 20
1.3.5.1 Prognostication and diagnosis ..... 20
1.3.5.2 Stroma targeted therapy ..... 20
**1.4 Epigenetics** ..... 22
1.4.1 Mechanisms of gene regulation ..... 22
1.4.2 DNA methylation ..... 24
1.4.2.1 Genomic distribution of CpG methylation ..... 25
1.4.2.2 DNA methyltransferases ..... 27
1.4.2.3 Dynamics of DNA methylation ..... 27
1.4.3 DNA methylation in cancer ..... 28
1.4.4 DNA methylation in gastric cancer ..... 29
1.4.5 Epigenetic cancer therapy ..... 29
1.4.5.1 DNA methyltransferase inhibitors ..... 30
1.4.5.2 Histone deacetylase inhibitors ..... 30
**1.5 Personalized cancer therapy** ..... 31
1.5.1 Gene expression profiling ..... 31
1.5.2 DNA methylation profiling ..... 32
**1.6 Objectives and Aims** ..... 34

**Materials and Methods** ..... 35
**2.1 Materials** ..... 36
**2.2 Tissue Culture** ..... 37
2.2.1 Generation of human primary gastric myofibroblasts ..... 37
2.2.2 Gastric myofibroblast cell culture ..... 37
2.2.3 Gastric cancer cell culture ..... 39
**2.3 Integrated multi-omics experiments** ..... 39
2.3.1 Patient information ..... 39
2.3.2 Myofibroblast cell culture ..... 40
2.3.3 Conditioned media preparation for secretome profiling ..... 41
2.3.4 DNA extraction for Illumina Infinium HumanMethylation450k array ..... 42

2.3.5 RNA extraction for Illumina Human HT-12 v4 array ..... 43
**2.4 Myofibroblast conditioned media preparation** ..... 44
2.4.1 Standard conditioned media ..... 44
2.4.2 Hypoxic conditioned media ..... 44
**2.5 Cancer cell based assays** ..... 45
2.5.1 Cancer cell migration assay ..... 45
2.5.2 Cancer cell proliferation assay ..... 46
**2.6 Omics data processing and analysis** ..... 47
2.6.1 Illumina Infinium 450k data processing ..... 47
2.6.2 Differential methylation analysis ..... 48
2.6.3 Illumina HT-12 Expression data processing ..... 48
2.6.4 Differential gene expression analysis ..... 49
2.6.5 Integration of DNA methylation and gene expression data ..... 50
2.6.6 Secretome data processing ..... 50
2.6.7 Integration of gene expression and secretome data ..... 52
2.6.8 R/Bioconductor ..... 52
2.6.9 Ingenuity Pathway Analysis ..... 52
2.6.10 ConsensusPathDB ..... 54
**2.7 DNA methylation data analysis** ..... 54
2.7.1 Gene ontology enrichment analysis ..... 54
2.7.2 Associating differentially methylated CpG loci with genes ..... 55
2.7.3 Pathway Analysis for common differentially methylated genes in gastric and oesophageal CAMs ..... 55
**2.8 Gene expression data analysis** ..... 56
2.8.1 Gene ontology enrichment analysis ..... 56
2.8.2 Gene set enrichment analysis ..... 56
**2.9 Comparative transcriptome and secretome analysis** ..... 57
2.9.1 Gene ontology enrichment analysis ..... 57
2.9.2 Gene set enrichment analysis ..... 57
2.9.3 Ingenuity Pathway Analysis ..... 58

**2.10 Pyrosequencing DNA methylation analysis** ..... 59

2.10.1 DNA extraction ..... 59

2.10.2 Pyrosequencing analysis ..... 59

**2.11 TaqMan gene expression analysis** ..... 62

2.11.1 RNA extraction ..... 62

2.11.2 TaqMan qPCR analysis ..... 62

**Identification of Genome-Wide DNA Methylation Signatures in Gastric and Oesophageal Cancer-Associated Myofibroblasts** ..... 65

**3.1 Introduction** ..... 66

**3.2 Aims** ..... 67

**3.3 Results** ..... 68

3.3.1 Comparison of global DNA methylation between CAMs and ATMs from gastric and oesophageal cancers ..... 68

3.3.2 Identification of differentially methylated loci in gastric CAMs, ATMs and NTMs ..... 68

3.3.2.1 Characterization of differentially methylated loci in gastric CAMs compared to patient-matched ATMs ..... 71

3.3.2.2 Technical validation of identified differentially methylated CpG loci ..... 73

3.3.2.3 Gene ontology enrichment analysis of differentially methylated loci identified in gastric CAMs compared to ATMs ..... 75

3.3.2.4 Identification of CpG loci that might serve as proxies for gastric CAM identification ..... 77

3.3.3 Identification of differentially methylated loci in oesophageal CAMs and ATMs ..... 80

3.3.3.1 Characterization of identified differentially methylated loci in oesophageal CAMs compared to ATMs ..... 81

3.3.3.2 Gene ontology enrichment analysis for differentially methylated loci in oesophageal CAMs compared to ATMs ..... 83

3.3.4 Identification of conserved DNA methylation patterns in gastric and oesophageal CAMs ..... 84

3.3.4.1 Conserved methylation pattern at the individual CpG loci in gastric and oesophageal CAMs ..... 85

3.3.4.2 Identification of common differentially methylated genes in gastric and oesophageal CAMs ..... 89

**3.4 Discussion** ..... 95

**Identification of Genes that Exhibit Correlated Changes in Gene Expression and DNA Methylation in Gastric Myofibroblasts** ..... 99

**4.1 Introduction** ..... 100

**4.2 Aims** ..... 101

**4.3 Results** ..... 102

4.3.1 Identification of differential gene expression profiles in gastric CAMs, ATMs and NTMs ..... 102

4.3.2 Technical validation of identified differentially expressed genes ..... 104

4.3.3 Potential biological significance of the identified changes in gene expression profiles ..... 107

4.3.3.1 Gene ontology enrichment analysis ..... 107

4.3.3.2 Gene set enrichment analysis ..... 111

4.3.3.3 Pathway analysis ..... 114

4.3.4 Integration of differentially methylated loci with differential gene expression profiles in gastric CAMs compared to ATMs ..... 117

4.3.4.1 Analysis of the relationship between DNA methylation and gene expression in gastric CAMs ..... 119

4.3.4.2 Identification of genes exhibiting correlated changes in promoter methylation and gene expression ..... 121

4.3.4.3 Identification of genes exhibiting correlated changes in gene-body methylation and gene expression ..... 127

**4.4 Discussion** ..... 133

**Validation of DNA Methylation Patterns that Correlate with CAM-specific Gene Expression Profiles** ..... 137

**5.1 Introduction** ..... 138

**5.2 Aims** ..... 139

**5.3 Results** ..... 140

5.3.1 Validation of CAM differentially methylated CpG loci by targeted pyrosequencing analysis ..... 140

5.3.2 Pyrosequencing analysis identifies concordant DNA methylation changes within interrogated regions of gastric CAMs ..... 142

5.3.3 Locus specific changes in CAM DNA methylation also affect expression of neighbouring genes ..... 145

5.3.3.1 Selection of target regions ..... 145

5.3.3.2 Promoter hypermethylation may repress *SMAD3* expression in gastric CAMs ..... 145

5.3.3.3 Promoter hypomethylation may induce *SPON2* expression in gastric CAMs ..... 148

5.3.3.4 Promoter hypermethylation may repress *ZNF536* expression in gastric CAMs ..... 151

5.3.3.5 DNA hypermethylation may repress the expression of *FOXF1* and *FENDRR* in gastric CAMs ..... 154

**5.4 Discussion** ..... 157

**Effects of Hypoxia on Gene Expression and Secretion of Pro-tumorigenic Factors in Gastric Stromal Myofibroblasts** ..... 163

**6.1 Introduction** ..... 164

**6.2 Aims** ..... 166

**6.3 Results** ..... 167

6.3.1 Hypoxia enhances CAM-induced cancer cell migration and promotes CAM-like properties in NTMs ..... 167

6.3.1.1 Gastric cancer cell migration assays ..... 167

6.3.1.2 Gastric cancer cell proliferation assays ..... 167

6.3.2 Identification of hypoxia-induced gene expression signatures in gastric CAMs, ATMs and NTMs ..... 168

6.3.2.1 Differential hypoxia-induced myofibroblast phenotypes are not due to induced changes in DNA methylation profiles ..... 171

6.3.3 Potential biological significance of hypoxia-induced gene expression signatures in CAMs, ATMs and NTMs ..... 172

6.3.3.1 Universal hypoxia-induced gene expression signatures in CAMs, ATMs and NTMs ..... 173
6.3.3.2 Unique CAM, ATM and NTM hypoxia-induced gene expression signatures ..... 176
6.3.3.3 Predicted biological effects of hypoxia on CAMs, ATMs and NTMs ..... 183
6.3.3.4 Prediction of factors that may be secreted by hypoxic CAMs and NTMs ..... 183
6.3.4 Secretome analysis of CAM and NTM hypoxic conditioned media ..... 186
6.3.4.1 Identification and quantification of CAM and NTM secreted proteins ..... 186
6.3.4.2 Hypoxia–induced secretome signatures in gastric CAMs and NTMs ..... 188
6.3.4.3 Potential biological effects of the hypoxia-induced CAM and NTM secretomes ..... 191
6.3.4.4 Integration of hypoxia-induced CAM and NTM secretomes with CAM and NTM hypoxia-induced gene expression signatures ..... 194
**6.4 Discussion** ..... 196

**Final Discussion** ..... 203

7.1 Summary and Conclusions ..... 204
7.2 DNA methylation profiling of CAMs ..... 206
7.3 Global hypomethylation in CAMs ..... 207
7.4 DNA methylation in CAMs from different tissues ..... 209
7.5 Stromal DNA methylation signatures ..... 211
7.6 Limitations of the study ..... 212
7.7 Concluding remarks ..... 213

**Supplementary Data** ..... 214

*Chapter II* ..... 215
*Chapter III* ..... 219
*Chapter IV* ..... 223
*Chapter V* ..... 225
*Chapter VI* ..... 236
S6.1 Gene Set Enrichment Analysis (GSEA) ..... 237

S6.2 Leading Edge Analysis ..... 238

**References** ..... 242

# List of Figures

**Figure 1.1** Parts of the stomach. ..... **7**

**Figure 1.2** Components of the tumour microenvironment. ..... **10**

**Figure 1.3** Stromal-epithelial interactions in normal tissue and during malignant transformation ..... **11**

**Figure 1.4** Tumour promoting properties of cancer associated myofibroblasts ..... **15**

**Figure 1.5** Interplay between different aspects of epigenetic regulation of gene expression. ..... **24**

**Figure 1.6** Distribution of CpG sites in genomic regions. ..... **26**

**Figure 1.7** Schematic representations of DNA methylation patterns in normal and cancer cells ..... **28**

**Figure 2.1** Bioinformatics analysis workflow–integration of DNA methylation and gene expression data ..... **51**

**Figure 2.2** Bioinformatics analysis workflow–integration of gene expression and secretome data ..... **53**

**Figure 3.1** Global DNA methylation of gastric and oesophageal patient-matched CAMs and ATMs ..... **69**

**Figure 3.2** Differentially methylated CpG loci in gastric myofibroblasts purified from different tissue microenvironments ..... **70**

**Figure 3.3** Distribution of differentially methylated CpG loci identified in gastric CAMs vs ATMs. ..... **72**

**Figure 3.4** Genomic region and gene associations of differentially methylated CpG loci in gastric CAMs vs ATMs ..... **73**

**Figure 3.5** Correlations between Illumina 450k array data and pyrosequencing analysis in gastric myofibroblasts ..... **74**

**Figure 3.6** Heatmap representations of CpG loci that may serve as proxies for gastric CAM, ATM and NTM identification ..... **78**

**Figure 3.7** Genomic region and gene associations of CpG proxies that might be used to distinguish gastric CAMs from ATMs and NTMs ..... **80**

**Figure 3.8** Differentially methylated CpG loci identified in oesophageal CAM vs ATM comparison.....**82**
**Figure 3.9** Genomic region and gene associations of differentially methylated CpG loci in oesophageal CAMs vs ATMs. .....**82**
**Figure 3.10** Overlap of differentially methylated CpG sites identified in gastric and oesophageal CAMs vs ATMs .....**86**
**Figure 3.11** Representative conserved DNA methylation patterns in gastric and oesophageal patient-matched CAM and ATM samples .....**87**
**Figure 3.12** Genome-wide overview of DNA methylation changes in stromal myofibroblasts .....**88**
**Figure 3.13** Differentially methylated genes in gastric and oesophageal CAMs. .....**89**
**Figure 3.14** Comparison of gene ontology (GO) biological process (BP) terms identified for gastric and oesophageal differentially methylated CpG loci .....**92**
**Figure 4.1** Differential gene expression signatures in gastric myofibroblasts purified from different tissue microenvironments. .....**103**
**Figure 4.2** Quantitative PCR validations of genes identified as differentially expressed in CAM vs ATM and CAM vs NTM comparisons .....**106**
**Figure 4.3** Gene ontology (GO) cellular component (CC) enrichment for differentially expressed genes identified in gastric CAM vs ATM comparison. .....**109**
**Figure 4.4** GSEA enrichment plots for gene sets enriched in CAM and ATM phenotypes, which gene members are reported to be epigenetically regulated. ..**113**
**Figure 4.5** Unfolded protein response pathway.....**115**
**Figure 4.6** Reactome and KEGG pathways overrepresented in CAM vs ATM comparison.....**116**
**Figure 4.7** Identification of a subset of genes, which expression may be regulated by DNA methylation changes in gastric CAMs.....**118**
**Figure 4.8** Simultaneous negative correlation between gene expression and DNA methylation in promoter region and positive correlation in gene body of *TGFBR2* in patient-matched CAM and ATM samples. .....**120**
**Figure 4.9** Negative correlations identified between gene expression and promoter methylation in patient-matched CAM and ATM samples. .....**122**

**Figure 4.10** Positive correlations identified between gene expression and gene-body methylation in patient-matched CAM and ATM samples. .....**128**
**Figure 5.1** Representative pyrograms from comparative pyrosequencing DNA methylation assays. .....**141**
**Figure 5.2** Methylation changes at CpG loci identified in gastric CAMs and ATMs by Illumina 450k array are concordant within a broader local genomic region. .....**143**
**Figure 5.3** DNA methylation in the *SMAD3* promoter region regulates *SMAD3* expression in gastric CAMs and ATMs. .....**147**
**Figure 5.4** DNA methylation in the *SPON2* promoter region correlates with *SPON2* gene expression profiles in gastric CAMs and ATMs. .....**150**
**Figure 5.5** DNA methylation levels within the *ZNF536* promoter region correlate with *ZNF536* gene expression profiles in gastric CAMs and ATMs. .....**153**
**Figure 5.6** DNA methylation pattern in the genomic region associated with regulation of *FOXF1* and *FENDRR* expression in gastric CAMs and ATMs. .....**156**
**Figure 6.1** Differential effects of CAM/NTM hypoxic conditioned media on AGS gastric cancer cell migration or proliferation. .....**169**
**Figure 6.2** Hypoxia-induced gene expression signatures in gastric myofibroblasts purified from different tissue microenvironments.....**170**
**Figure 6.3** Comparative global DNA methylation profiles of primary gastric myofibroblasts following exposure to hypoxic or normoxic conditions. .....**172**
**Figure 6.4** Comparison of hypoxia-induced gene expression signatures identified in gastric CAMs, ATMs and NTMs .....**173**
**Figure 6.5** Universal changes induced by hypoxia in gastric CAMs, ATMs and NTMs .....**174**
**Figure 6.6** GSEA enrichment plots for most significantly enriched hallmark gene set and canonical pathway gene set.....**175**
**Figure 6.7** GSEA enrichment plots for the most significantly enriched hallmark gene sets .....**180**
**Figure 6.8** IPA canonical pathways significantly enriched in unique CAM, ATM and NTM hypoxia-induced gene expression profiles.....**182**
**Figure 6.9** Predicted biological effects of hypoxia on gastric CAMs, ATMs and NTMs .....**184**

**Figure 6.10** Predicted pro-migratory and pro-proliferative factors secreted by hypoxic CAMs and hypoxic NTMs. .....**185**
**Figure 6.11** Venn diagram representation of database searches used to classify proteins identified in CAM and NTM conditioned media obtained from normoxia and hypoxia. .....**187**
**Figure 6.12** Volcano plots of differentially secreted proteins .....**189**
**Figure 6.13** Comparison of differentially secreted proteins identified in CAM-hypoxic-CM and NTM-hypoxic-CM compared to their respective normoxic-control-CM.....**190**
**Figure 6.14** Predicted biological effects of the hypoxia-induced CAM and NTM secretomes.....**193**
**Figure 6.15** Predicted activation of HIF-1α regulates expression of proteins secreted by hypoxic NTMs. .....**194**
**Figure 6.16** Integration of secretome and gene expression data.....**195**
**Figure 7.1** Novel insights into molecular mechanisms underlying the tumour-promoting phenotype of gastric CAMs that have emerged from this study .....**206**

***Figure S2.1*** Quality of DNA samples used for Illumina Infinium HumanMethylation450k array.....**215**
***Figure S3.1*** Increased migration of gastric cancer cells in response to CAM conditioned media (CM) compared to ATM-CM and NTM-CM.....**219**
***Figure S3.2*** Increased proliferation of gastric cancer cells in response to CAM conditioned media (CM) compared to ATM-CM and NTM-CM.....**220**
***Figure S3.3*** Heatmap representation of differentially methylated regions (DMRs) identified in gastric CAMs compared to their patient-matched ATMs. .....**221**
***Figure S3.4*** Heatmap representation of differentially methylated regions (DMRs) identified in oesophageal CAMs compared to their patient-matched ATMs.....**222**
***Figure S4.1*** Representative examples of genes showing consistent changes in gene expression profiles in two independent studies of gastric CAM vs ATM gene expression profiles. .....**223**

***Figure S4.2*** PCP pathway .....**224**

***Figure S5.1*** *SMAD3* expression in gastric and oesophageal CAMs compared to their parent-matched ATMs .....**226**

***Figure S5.2*** Negative correlations between gene expression (qPCR data) and promoter methylation (pyrosequencing data) in patient-matched CAMs and ATMs .....**227**

***Figure S5.3*** Differentially methylated *SPON2* promoter-region identified by Illumina 450k array in oesophageal patient-matched CAM and ATM samples .....**228**

***Figure S5.4*** Simultaneous negative correlation between promoter methylation and *SPON2* gene expression and positive correlation between gene-body methylation and *SPON2* gene expression levels in patient-matched CAM and ATM samples....**229**

***Figure S5.5*** Negative correlation between *SPON2* gene expression and promoter methylation in patient-matched CAMs and ATMs and NTMs.....**230**

***Figure S5.6*** Simultaneous negative correlation between promoter methylation and *ZNF536* gene expression and positive correlation between gene-body methylation and *ZNF536* gene expression levels in patient-matched CAM and ATM samples. .**231**

***Figure S5.7*** Simultaneous negative correlation between promoter methylation and *ZNF536* gene expression and positive correlation between gene-body methylation and *ZNF536* gene expression levels in patient-matched CAMs and ATMs in comparison to NTMs.....**232**

***Figure S5.8*** Comparative DNA methylation profiles within a genomic region on chromosome 16 spanning 526,340bp identified by Illumina 450k array in gastric stromal myofibroblasts. .....**233**

***Figure S5.9*** Comparative DNA methylation profiles within a genomic region on chromosome 16 spanning 9,673bp identified by Illumina 450k array in oesophageal patient-matched CAM and ATM samples.....**234**

***Figure S5.10*** Pyrosequencing analysis of the *FOXF1* promoter region in gastric cancer cell lines .....**234**

***Figure S5.11*** Differential DNA methylation data for stomach adenocarcinoma from the Cancer Genome Atlas (TCGA) .....**235**

***Figure S6.1*** Universal gene expression changes induced under hypoxia in gastric CAMs, ATMs and NTMs.....**236**

***Figure S6.2*** Comparison of enriched MSigDB gene sets identified for unique CAM, ATM and NTM hypoxia-induced gene expression profiles .....**237**
***Figure S6.3*** Volcano plots of differentially secreted proteins .....**239**
***Figure S6.4*** Differential expression and secretion of gelsolin, vascular endothelial growth factor A and glucose-6-phosphate isomerase in hypoxic CAMs and hypoxic NTMs compared to respective control-normoxic myofibroblasts.....**240**
***Figure S6.5*** Differential expression and secretion of angiopoietin-related protein 4, insulin-like growth factor-binding protein 3, insulin-like growth factor-binding protein 6 and thrombospondin-2 in gastric CAMs and NTMs under hypoxia and normoxia .....**240**
***Figure S6.6*** Differential expression and secretion of enzyme lysyl oxidase in gastric CAMs and NTMs under hypoxia and normoxia. .....**241**
***Figure S6.7*** Global DNA methylation after long-term hypoxia treatment assessed by LINE-1 pyrosequencing assays .....**241**

# List of Tables

**Table 2.1** List of materials used in this study. ..... **36**

**Table 2.2** Patient information relating to age, gender, tumour location and tumour clinical assessment for gastric cancer patients who provided CAM and ATM cells used in this study. ..... **38**

**Table 2.3** Information relating to age, gender and cell origin for post-mortem organ donors who provided NTM cells used in this study. ..... **39**

**Table 2.4** List of samples used for integrated multi-omic experiments. ..... **41**

**Table 2.5** Nucleotide sequences of PCR primer sets and sequencing primers used for pyrosequencing DNA methylation assays ..... **61**

**Table 2.6** List of TaqMan assays and nucleotide sequences used in this study. ..... **64**

**Table 3.1** Gene ontology (GO) enrichment analysis for differentially methylated CpG loci identified in gastric CAMs compared to ATMs. ..... **76**

**Table 3.2** Gene ontology (GO) enrichment analysis for differentially methylated CpG loci identified in oesophageal CAMs compared to ATMs. ..... **84**

**Table 3.3** KEGG and Reactome pathways affected by DNA methylation changes in gastric and oesophageal CAMs. ..... **90**

**Table 3.4** Common gene ontology (GO) biological process (BP) terms associated with gastric and oesophageal differentially methylated loci in CAMs compared to patient–matched ATMs ..... **93**

**Table 4.1** List of Illumina HT-12 probes identified as differentially expressed in CAM vs ATM and CAM vs NTM comparisons that were validated by TaqMan qPCR assays ..... **105**

**Table 4.2** Gene ontology (GO) biological process (BP) enrichment for differentially expressed genes identified in gastric CAM vs ATM comparison ..... **108**

**Table 4.3** GSEA results showing a list of significantly enriched hallmark gene sets in CAM and ATM phenotypes. ..... **112**

**Table 4.4** Genes that are hypomethylated and transcriptionally induced in CAMs compared to ATMs. ..... **124**

**Table 4.5** Genes that are hypermethylated and transcriptionally repressed in CAMs compared to ATMs. ..... **126**

**Table 4.6** Top 10 genes that are hypomethylated in gene-body and transcriptionally repressed in CAMs compared to ATMs. ..... **130**

**Table 4.7** Top 10 genes that are hypermethylated in gene-body and transcriptionally induced in CAMs compared to ATMs. ..... **132**

**Table 6.1** GSEA result summary. Significantly enriched hallmark gene sets and canonical pathways for universal CAM, ATM and NTM gene expression profiles ..... **175**

**Table 6.2** Gene ontology (GO) biological process (BP) enrichment for unique hypoxia-induced gene expression signatures identified in CAMs, ATMs or NTMs. ..... **177**

**Table 6.3** GSEA result summary. Significantly enriched hallmark gene sets for CAM, ATM and NTM unique hypoxia-induced gene expression profiles. ..... **181**

**Table 6.4** Gene ontology (GO) biological process (BP) enrichment for differentially secreted proteins by hypoxic NTMs ..... **192**

***Table S2.1*** Quality control information for RNA and DNA samples used for Ilumina HumanHT-12 v4 Expression and Illumina Infinium HumanMethylation450k arrays ..... **216**

***Table S2.2*** Patient information relating to the scoring details. ..... **217**

***Table S2.3*** Patient information relating to age, gender, tumour location and tumour clinical assessment for oesophageal cancer patients who provided CAM and ATM cells used in this study. ..... **218**

***Table S5.1*** Reproducibility of pyrosequencing assays. Correlation $R^2$ between methylation levels for biological replicates. ..... **225**

***Table S6.1*** Summary of leading edge analysis (LEA) of enriched MSigDB gene sets identified for unique CAM, ATM and NTM hypoxia-induced gene expression profiles. ..... **238**

# Appendices

List of supplementary files provided on CD:

| | |
|---|---|
| ***Appendix I*** | Quality control reports for the 'gastric' and 'oesophageal' Illumina Infinium HumanMethylation450 BeadChip arrays |
| ***Appendix II*** | Lists of differentially methylated CpG sites and genomic regions, including promoter, gene, island and tilling identified in gastric (CAM vs ATM, CAM vs NTM, ATM vs NTM) and oesophageal (CAM vs ATM) comparisons |
| ***Appendix III*** | Quality control report for the Illumina HumanHT-12v4 Expression BeadChip array |
| ***Appendix IV*** | Lists of differentially expressed genes identified in the following comparisons: CAM vs ATM, CAM vs NTM, ATM vs NTM, CAM hypoxia vs CAM normoxia, ATM hypoxia vs ATM normoxia and NTM hypoxia vs NTM normoxia |
| ***Appendix V*** | Table of identified and quantified proteins |
| ***Appendix VI*** | *File 1 – Superpathway of Cholesterol Biosynthesis*<br>*File 2 – Integrin Signalling* |

# Abbreviations

| | |
|---|---|
| **5-mC** | 5-methylcytosine |
| **α-SMA** | alpha-smooth muscle actin |
| **ASPH** | aspartate beta-hydroxylase |
| **ATM** | Adjacent Tissue Myofibroblasts |
| **AURKA** | aurora kinase A |
| **B4GALT6** | beta-1,4-galactosyltransferase |
| **bp** | base pair |
| **CAF** | Cancer Associated Fibroblast |
| **CAM** | Cancer Associated Myofibroblast |
| **CM** | Conditioned Media |
| **CPDB** | ConsensusPathDB |
| **CpG** | Cytosine-phosphate-Guanine |
| **DACOR1** | DNMT1-associated colon cancer repressed lncRNA 1 |
| **DAG** | directed acyclic graph |
| **DEPTOR** | DEP domain containing MTOR-interacting protein |
| **DMEM** | Dulbecco's Modified Eagle's Medium |
| **DMR** | Differentially Methylated Region |
| **DNA** | deoxyribonucleic acid |
| **DNMT** | DNA methyltransferase |
| **dNTP** | deoxynucleoside triphosphate |
| **ECM** | extracellular matrix |
| **EdU** | 5-ethynyl-2'-deoxyuridine |
| **ER** | endoplasmic reticulum |
| **FBS** | fetal bovine serum |
| **FDR** | false discovery rate |
| **FENDRR** | FOXF1 adjacent non-coding developmental regulatory RNA |
| **FGF** | fibroblast growth factor |
| **FOXC1** | forkhead box C1 |
| **FOXF1** | forkhead box F1 |

| | |
|---|---|
| **GO** | Gene Ontology |
| **GO_BP** | Gene Ontology Biological Process |
| **GO_CC** | Gene Ontology Cellular Component |
| **GO_MF** | Gene Ontology Molecular Function |
| **GOrilla** | Gene Ontology enRIchment anaLysis and visuaLizAtion tool |
| **GREAT** | Genomic Regions Enrichment of Annotations Tool |
| **GSEA** | Gene Set Enrichment Analysis |
| ***H. pylori*** | *Helicobacter pylori* |
| **HGF** | hepatocyte growth factor |
| **HIF** | hypoxia-inducible factor |
| **HOXA5** | homeobox A5 |
| **HSPA5** | heat shock protein family A (Hsp70) member 5 |
| **IKB** | Ingenuity Knowledge Base |
| **IPA** | Ingenuity Pathway Analysis |
| **kb** | kilobase |
| **KEGG** | Kyoto Encyclopedia of Genes and Genomes |
| **LC-MS/MS** | liquid chromatography–mass spectrometry |
| **LEA** | Leading Edge Analyses |
| **LFQ** | label free quantification |
| **lncRNA** | long non-coding RNA |
| **MetazSecKB** | Metazoa Secretome and Subcellular Proteome Knowledge Base |
| **miR** | microRNA |
| **MSigDB** | Molecular Signatures Database |
| **MUC2** | mucin 2 |
| **NTM** | Normal Tissue Myofibroblasts |
| **PBS** | phosphate buffered saline |
| **PCR** | polymerase chain reaction |
| **PDGF** | platelet-derived growth factor |
| **PDGF-R** | PDGF receptor |
| **PFA** | paraformaldehyde |
| **RNA** | ribonucleic acid |

| | |
|---|---|
| **ROS** | reactive oxygen species |
| **SAM** | S-adenosylmethionine |
| **SEM** | standard error of the mean |
| **SMAD3** | SMAD family member 3 |
| **SNP** | single nucleotide polymorphisms |
| **SPON2** | spondin 2 |
| **SULF** | sulfatase |
| **TCGA** | The Cancer Genome Atlas |
| **TGF-β** | transforming growth factor β |
| **TGFBR2** | TGF-β receptor type II |
| **TSS** | transcription start site |
| **UPR** | unfolded protein response |
| **VEGF** | vascular endothelial growth factor |
| **VPS28** | vacuolar protein sorting 28 homolog |
| **ZNF536** | zinc finger protein 536 |

# Abstract

In recent years it has become increasingly apparent that tumour development and metastasis are not simply driven by mutations within cancer cells. Factors produced by stromal myofibroblasts play a key role in the development and metastasis of many forms of cancer. However, our knowledge of the range of molecular mechanisms that drive paracrine communication between cancer and stromal cells remains incomplete. Evidence from previous studies show that myofibroblasts derived from gastric tumours (CAMs) not only retain their ability to enhance the proliferation and migration of cancer cells *in vitro* but also have inherently different gene expression profiles compared to patient-matched adjacent tissue myofibroblasts (ATMs), or normal tissue myofibroblasts (NTMs). In addition, it was shown that CAMs are characterized by global loss of DNA methylation when compared to patient-matched ATMs. Taken together, these observations indicate that epigenetic programming of myofibroblasts within the tumour microenvironment may contribute to their tumour-promoting properties.

In this study, genome-wide DNA methylation and gene expression profiling was performed on a collection of primary patient-matched gastric CAM and ATM samples and normal NTM samples in order to identify CAM-specific DNA methylation signatures and to assess the extent to which these genome-wide DNA methylation changes may regulate CAM-specific gene expression profile. The differential methylation analysis identified widespread alterations of DNA methylation in gastric CAMs compared to ATMs and NTMs. In addition, the integrated analysis of DNA methylation and gene expression showed that DNA methylation is involved in epigenetic regulation of biological pathways and processes involved in the tumour–promoting function of gastric CAMs. In particular, promoter DNA hypomethylation emerged as regulatory mechanism for transcriptional activation of genes involved in secretion and transport of molecules while promoter DNA hypermethylation emerged as regulatory mechanism for

transcriptional repression of genes involved in pathology of gastrointestinal cancers and regulation of developmental processes.

Additionally, given that hypoxic conditions may be common within the tumour microenvironment, we were interested to know what role hypoxia may play in regulating paracrine communication between CAMs and gastric cancer cells. Using cell based assays and a combination of genomic, transcriptomic and proteomic approaches we provided a novel insight into the profile of changes in gene expression and protein secretion that result from exposure of normal (NTMs) or cancer-associated myofibroblasts (CAMs) to hypoxic conditions and identified hypoxia-induced factors that may drive CAM-, ATM- and NTM-specific hypoxic responses, which have differential effects on cancer cell migration and proliferation. Importantly, data from this study show that gastric myofibroblasts purified from different tissue microenvironments (CAMs, ATMs, NTMs) have distinct responses to hypoxia, which most probably stem from differential epigenetic profiles providing, once again, strong evidence that epigenetic programming of CAMs is in part mediated by interactions with cancer cells. However, hypoxia is unlikely to be a driving factor in conferring cancer related changes in DNA methylation profiles of gastric myofibroblasts.

Taken together, this study is the first to perform a comparative genome-wide analysis of DNA methylation and gene expression patterns in gastric cancer-associated myofibroblasts. Significantly, widespread epigenomic and transcriptomic alterations were observed in CAMs compared to their patient-matched ATMs, providing potential clues as to the molecular mechanism of cancer mediated stromal cell programming. Finally, CAM-specific DNA methylation patterns identified in this study may prove useful in future clinical practice as biomarkers for improved diagnosis and prognosis of gastric cancer.

# Chapter I

# Introduction

## 1.1 Overview

Recent technological advancements in high-throughput 'omics' techniques have transformed the field of cancer research. Scientists can now routinely use these techniques to profile the genome, epigenome, transcriptome, proteome and metabolome of cancer cells. As such, data generated using these technologies have provided a valuable insight into the broader mechanisms of cancer biology. Multi-institutional collaborative projects, such as The Cancer Genome Atlas (TCGA), the Cancer Genome Project, and the International Cancer Genome Consortium, were established to generate and integrate those data and comprehensively characterise common types of human cancers with the aim to provide a resource to guide and improve current clinical practice.

Gastric cancer is one of the most common cancers worldwide. Despite its declining incidence the mortality rate for gastric cancer remains high, mainly due to late diagnosis and lack of precise prognostic biomarkers. In recent years, high-throughput large-scale gene expression profiling technology has significantly improved our understanding of gastric cancer biology. The analysis of gastric cancer gene expression data has identified gene expression signature that can be used to determine, differentiate and categorize gastric cancer subtypes, thereby complementing traditional morphological and clinical assessment and providing real potential for improving patient prognosis and enhancing our ability to rationally design more appropriate tailored therapies (Brettingham-Moore, Duong et al. 2011, Lin, Zhao et al. 2015)

Significantly, it is now evident that tumour growth and metastasis are not simply driven by cancer cells alone. The tumour microenvironment plays a profound role in these processes. In particular, cancer associated myofibroblasts have been shown to promote the migration and proliferation of gastric cancer cells, through a complex mechanism of reciprocal communication with the developing tumour.

As such, the tumour stroma offers a better potential for the identification of consistent stage specific markers and novel approaches to therapeutic intervention (Zhi, Shen et al. 2010, Holmberg, Quante et al. 2012). Although a number of studies have demonstrated the importance of CAMs in gastric cancer progression (Yan, Wang et al. 2015) the detailed mechanisms underlying tumour-promoting phenotype of gastric CAMs remain elusive.

## 1.2 Cancer

With an ever-aging population cancer remains a major human health problem and a leading cause of death worldwide. We now understand that cancer is a multistep and multifactorial disease that results from a combination of genetic and epigenetic changes that affect key regulatory pathways in the cell. Although cancer can develop in almost any tissue and each type of cancer has unique features, there are common processes involved in the progression of many forms of cancer. For all forms of cancer there is a direct link between stage of detection and patient prognosis. Therefore, there is a pressing need to develop a better understanding of the sequential processes that contribute to each stage of tumour development, in order to facilitate the development of improved diagnostic and therapeutic strategies.

### 1.2.1 Molecular basis of cancer

Cancer arises from a single cell which accumulates multiple genetic and epigenetic changes in the expression or regulation of oncogenes and/or tumour suppressor genes. These alterations provide the transformed cell with important growth advantages that allow cell clones to form a malignant mass, which ultimately dominates and outcompetes neighbouring cells (Jones and Baylin 2002, Vogelstein and Kinzler 2004). The progression of cancer involves a pre-invasive phase, followed

by invasion of the surrounding stroma, entry to blood and lymphatic vessels and metastasis to secondary sites (Chaffer and Weinberg 2011). The intratumour heterogeneity that arises during carcinogenesis introduces significant challenges contributing to resistance to radio- and chemotherapy (Bedard, Hansen et al. 2013).

### 1.2.2 Hallmarks of cancer

The hallmarks of cancer, as defined by Hanahan and Weinberg, include sustaining proliferative signalling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, activating invasion and metastasis, genome instability and mutation, tumour promoting inflammation, reprogramming energy metabolism and evading immune destruction (Hanahan and Weinberg 2011).

### 1.2.3 Gastric cancer

Gastric cancer is one of the most common cancers worldwide. Despite its declining incidence (Bertuccio, Chatenoud et al. 2009), it remains the third leading cause of cancer death worldwide (Ferlay, Soerjomataram et al. 2015) and the tenth most common cause of cancer death in the United Kingdom (Cancer Research UK 2012). Also, the prevalence of gastric cancer continues to be much higher in Asia and South America (Brenner, Rothenbacher et al. 2009) and treatment options for gastric cancer remain limited. The typical treatment strategy for the non-metastatic disease includes endoscopic treatment, surgical resection and chemotherapy. However, the disease is usually diagnosed at advanced stages, which explains its elevated mortality rate (American Cancer Society 2015), with a median survival for patients with advanced stage gastric cancer of 8–10 months (Lorenzen and Lordick 2011). Cancer Research UK reported in 2010-2011 that only 15% of gastric cancer

patients in England and Wales are expected to survive their disease for more than ten years (Cancer Research UK).

### 1.2.3.1 Risk factors

Gastric cancer is a very complex and a highly heterogeneous disease, which tends to develop slowly over many years, therefore emergency presentation are the most common mechanism of diagnosis. The risk of gastric cancer is strongly linked with age and sex. Around 95% of gastric cancer cases are diagnosed in people aged 55 or older. It is more common in men than women. In United Kingdom, it is the seventh most common cause of cancer death in men and the twelfth most common cause of cancer death in women (Cancer Research UK 2012). An estimated 75% of gastric cancer cases in the United Kingdom are linked to lifestyle factors, including *Helicobacter pylori* (*H. pylori*) infection, high salt intake, and tobacco/smoking. Other risk factors include ionising radiation, certain occupational exposures, low fruit and vegetables intake, alcohol, overweight and obesity as well as certain medical conditions (Cancer Research UK 2012). Also, people with first-degree relatives who have had gastric cancer are more likely to develop the disease (American Cancer Society 2015).

### 1.2.3.2 Classification

Adenocarcinoma, which originating from the glandular epithelium of the stomach lining, is the most common histological type of gastric cancer, accounting for 95% of all gastric malignancies (Cancer Research UK). Gastric adenocarcinomas can be further divided into two forms based on Lauren classification, diffuse and intestinal, which have different clinicopathological and prognostic features (Lauren 1965).

The intestinal type of gastric cancer is often found in the corpus and antrum of the stomach (Figure 1.1) (Wagner and Moehler 2009) and is associated with *H. pylori* infection, obesity, and certain dietary factors, such as high salt intake, preserved food and smoked meats. It is believed to arise through a long-term multistep progression from chronic gastritis to chronic atrophy to intestinal metaplasia to dysplasia and finally to gastric tumour (Correa, Haenszel et al. 1975). This form of gastric cancer is well differentiated and occurs more frequently in older patients (Crew and Neugut 2006).

In contrast, the diffuse type of gastric cancer is usually located in the proximal stomach (cardia, fundus, corpus) and gastroesophageal junction (Figure 1.1) (Wagner and Moehler 2009). This form of gastric cancer is poorly differentiated and confers a comparatively worse prognosis. It is more prevalent in the obese, young male population suffering from chronic reflux disease (Crew and Neugut 2006). However, studies showed that *H. pylori* infection also plays a role in the development of diffuse gastric cancer as a consequence of chronic inflammation, but is not associated with the occurrence of intermediate steps, such as gastric atrophy and intestinal metaplasia (Milne, Carneiro et al. 2009).

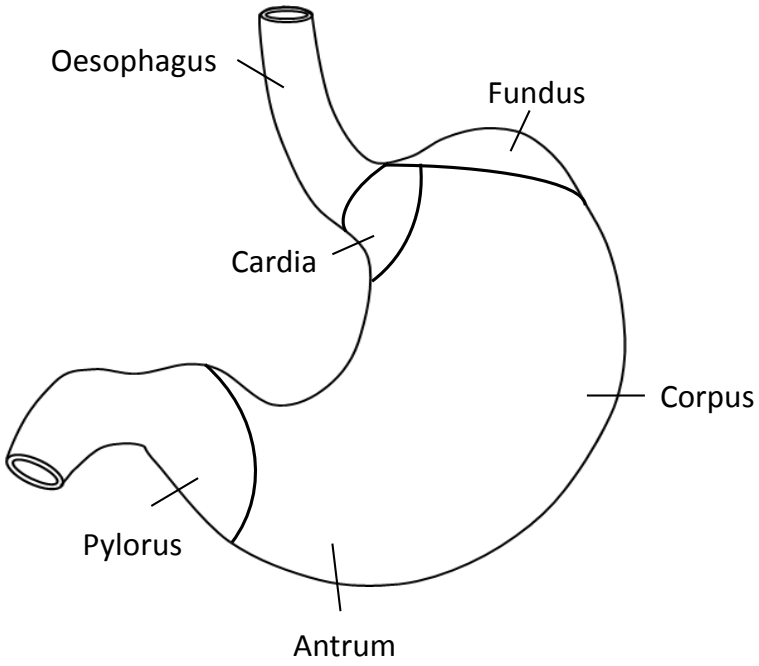


**Figure 1.1 Parts of the stomach.**

### 1.2.3.3 Clinical outcomes

Gastric cancer patients from different parts of the world show different clinical outcomes. Traditionally, these differences have been attributed to differences in clinical management and disease stage. However, recent meta-analysis of >1600 gastric cancers revealed that Asian and non-Asian gastric cancers exhibit distinct molecular signatures related to T-cell function (Lin, Gagnon-Bartsch et al. 2015). These findings might impact the design of future gastric cancer trials as they may well have an effect on treatment responses and clinical outcomes. Also, cancers starting in different parts of the stomach can cause different symptoms and tend to have different clinical outcomes. The location of the tumour can also affect the treatment options (American Cancer Society 2015). Gastric cancers can grow through the stomach wall and invade adjacent organs, or metastasize to lymph vessels and nearby lymph nodes, which form a very rich network in the stomach.

In advanced stages cancer cells can invade the bloodstream and spread to the liver, lungs and bone (American Cancer Society 2015).

### 1.2.3.4 Molecular biomarkers

The high mortality rate of gastric cancer is mostly attributed to the lack of early detection and effective medical treatment of advanced stages of the disease. At present serum carcinoembryonic antigen, carbohydrate antigen (CA) 19-9 and CA72-4 are the most widely used gastric cancer biomarkers (Duraes, Almeida et al. 2014), however they have a relatively low sensitivity and specificity in gastric cancer diagnosis and prognosis (Emoto, Ishigami et al. 2012, Kanda, Fujii et al. 2014). Therefore, a lot of efforts have been put into identification of more effective biomarkers which could be used to detect early gastric cancer and predict both recurrence and chemosensitivity. Recent progress in molecular techniques have provided new insight into the aberrant expression of gastric cancer-related molecules, including oncogenes and tumour suppressor genes, microRNAs, long non-coding RNAs, and aberrant DNA methylation, as potential novel molecular biomarkers that may have clinical value and therefore may be used for gastric cancer screening, diagnosis, tumour classification, prognosis and the prediction of therapeutic responses (Pinheiro, Silva Ferreira et al. 2014, Kanda and Kodera 2015). However, despite these extensive efforts, there are only a few defined gastric cancer biomarkers that are currently used in clinical practice. For that reason, it was proposed that several different biomarkers (genetic, epigenetic and proteomic) should be used in combination for each patient, in order to offer a much more reliable option for improved patient stratification (Pinheiro, Silva Ferreira et al. 2014).

## 1.3 Tumour microenvironment

It is now widely recognised that solid tumours are highly complex tissues composed of multiple distinct cell types that participate in dynamic and conditional heterotypic interactions. These changing interactions contribute to most aspects of tumour development, including initiation, progression, metastasis and chemoresistance (Hanahan and Weinberg 2011).

### 1.3.1 Components of tumour microenvironment

The tumour microenvironment consists of a dynamic system of various non-transformed cells, including fibroblasts, myofibroblasts, endothelial cells, immune cells and bone marrow-derived cells, as well as a specialised extracellular matrix (ECM) and various soluble paracrine factors, such as growth factors, cytokines, chemokines, enzymes and metabolites (Figure 1.2) (Bhowmick, Neilson et al. 2004, Allen and Jones 2011). Tumour progression is directly linked to special and temporal changes in hypoxia and acidosis, which have a significant impact on the bidirectional interactions between cancer cells and stromal components (Casazza, Di Conza et al. 2014, Kharaishvili, Simkova et al. 2014). Therefore, better understanding of the specific interactions between cancer cells and the surrounding microenvironment is needed to identify the mechanisms underlying tumour development and progression as well as driving the development of novel prognostic and therapeutic targets.

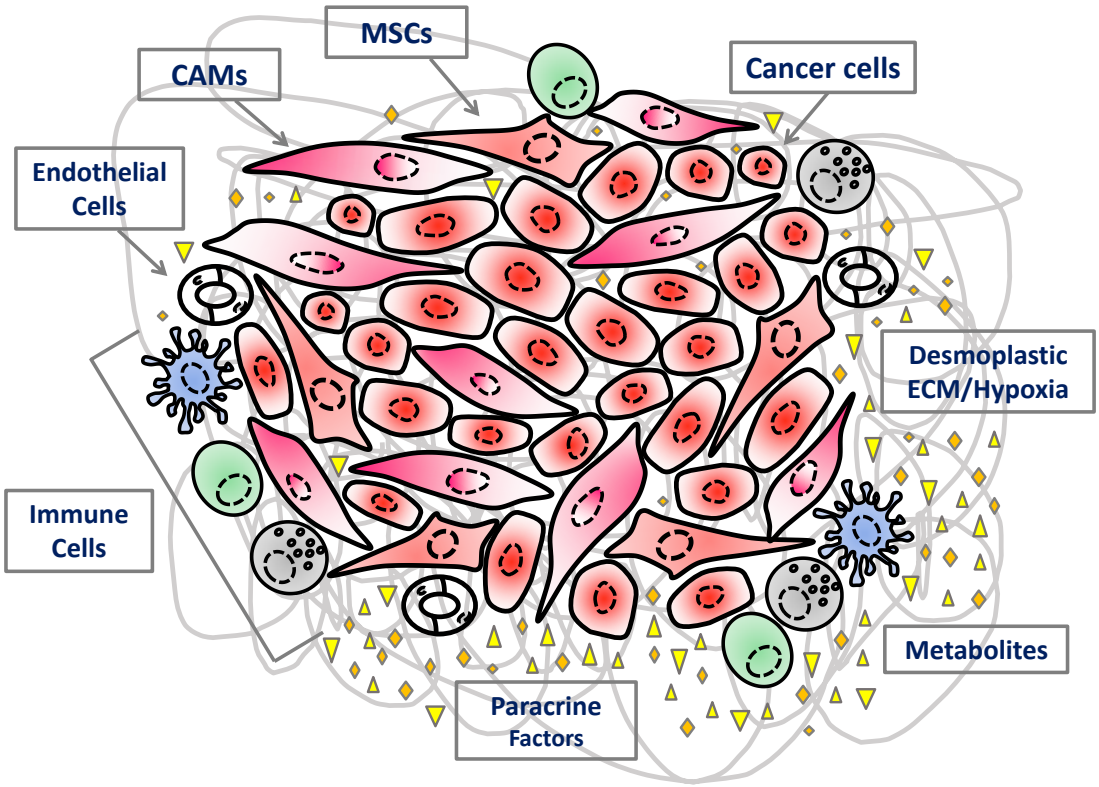


**Figure 1.2 Components of the tumour microenvironment.**
Elevated levels of cancer cell-derived growth factors and cytokines recruit diverse tumour-associated cells, which secrete additional growth factors, forming a positive feedback loop that promotes tumour cell progression and metastasis.

## 1.3.2 Stromal-epithelial interactions

The development of the tumour and progression towards advanced stages of the disease involves the co-evolution of both cancer cells and stromal cells (Shimoda, Mellody et al. 2010). In normal tissues paracrine communication between epithelial and stromal cells is required for the maintenance of normal tissue function (Figure 1.3A). During malignant transformation epithelial cells acquire mutations that dysregulate normal cell-cell communication. These mutations not only alter intrinsic cellular signalling pathways but also change the way cells respond to signals emerging from the neighbouring stromal cells. Ultimately, the altered paracrine signalling leads to tissue remodelling and changes within the surrounding

stroma, which then drive tumour progression (Figure 1.3B) (Wiseman and Werb 2002).

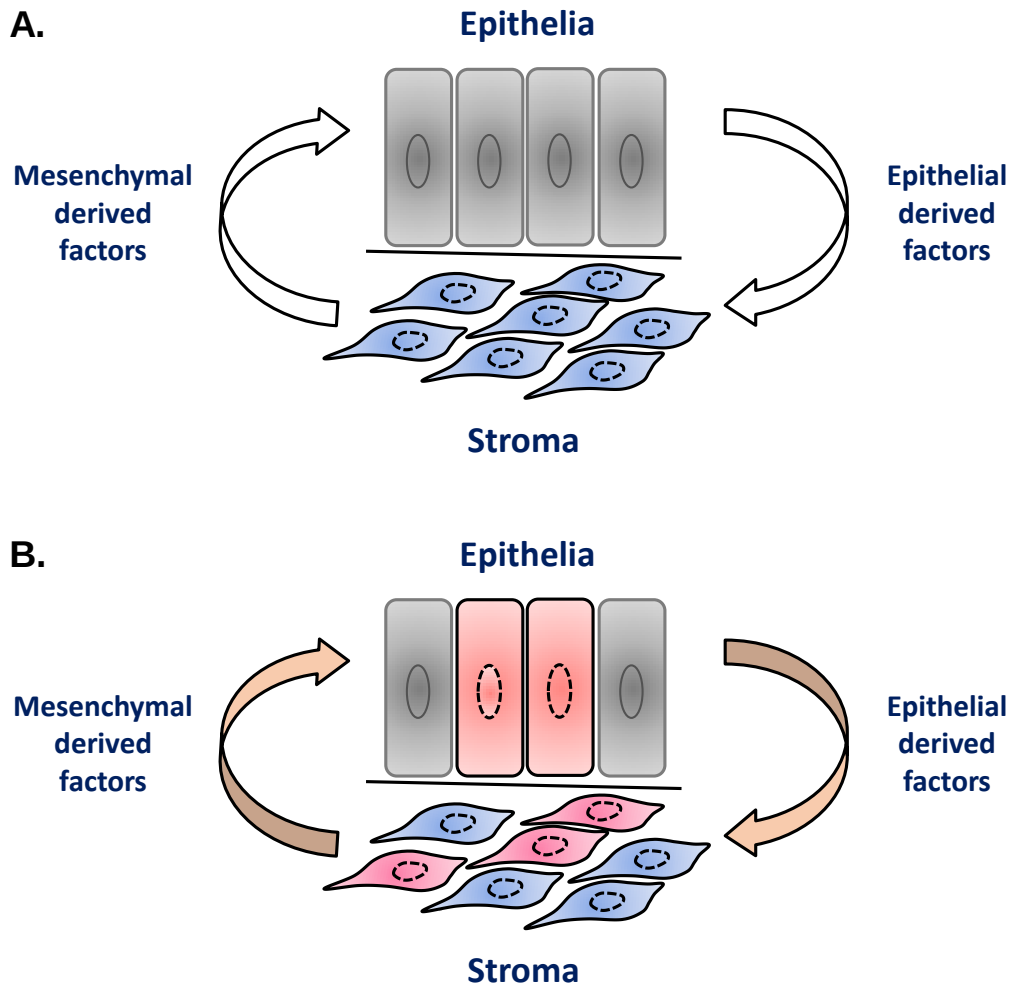


**Figure 1.3 Stromal-epithelial interactions** in **A.** normal tissue and **B.** during malignant transformation.

## 1.3.3 Myofibroblasts

Myofibroblasts, also known as activated fibroblasts, are stromal spindle-shaped cells that are relatively rare in most normal healthy tissues (Eyden 2008). They are characterised by the expression of α-smooth muscle actin (α-SMA) that distinguish them from 'normal' tissue fibroblasts (Skalli, Schurch et al. 1989). Myofibroblasts secrete ECM components, cytokines, chemokines, growth factors, proteases and their inhibitors (Tomasek, Gabbiani et al. 2002, Hinz 2007). They were first discovered in wound granulation tissue of healing skin wounds, where it was hypothesized that they were responsible for the wound contraction (Gabbiani, Ryan et al. 1971). Since then, myofibroblasts have been described in a number of different pathologies, including cancer. Although myofibroblasts have similar morphology, function and biochemical repertoire regardless of their tissue of residence, they may express some phenotypic and functional heterogeneity depending on the initial cause or process of their activation, such as tissue inflammation or proximity to a developing tumour. Notably, in contrast to other tissues, the gastrointestinal tract was reported to have a consistent but potentially dynamic population of activated myofibroblasts, which help to maintain normal epithelial function (Powell, Mifflin et al. 1999).

### 1.3.3.1 Myofibroblasts in wound healing

Myofibroblasts are activated as part of a normal wound healing response (Hinz 2007). When tissue is repaired myofibroblasts are either cleared by apoptosis (Desmouliere, Redard et al. 1995), or may become deactivated (Hecker, Jagirdar et al. 2011). Dysregulation of this normal process of repair can lead to persistent myofibroblast activation, for instance by chronic inflammation or mechanical stress in the tissue, which is associated with many fibrotic and scarring disorders (Noble, Barkauskas et al. 2012, Weber, Sun et al. 2013), and cancer (Cox and Erler 2011).

### 1.3.3.2 Myofibroblasts in cancer

Myofibroblasts are the most abundant cell type in tumour stroma and can be isolated from various cancer types, including breast cancer, prostate cancer, pancreatic cancer, cholangiocarcinoma, lung cancer, gastric cancer, and colorectal cancer. However, myofibroblasts are relatively rare in brain, renal and ovarian cancers (Shiga, Hara et al. 2015). The presence of myofibroblasts in large numbers is often correlated with poor prognosis (Yamashita, Ogawa et al. 2012, Ha, Yeo et al. 2014) and may be involved in cancer cell resistance to therapy (Li, Hu et al. 2015).

Cancer associated myofibroblasts (CAMs), also sometimes referred to as cancer associated fibroblasts (CAFs), are functionally distinct and exhibit an altered phenotype when compared to non-cancerous tissue-derived myofibroblasts at least in part due to epigenetic changes (Jiang, Gonda et al. 2008). This includes changes in CAM transcriptome (Peng, Zhao et al. 2013) and microRNA expression profile (Zhao, Sun et al. 2012) as well as a characteristic changes in CAM metabolome (Martinez-Outschoorn, Lisanti et al. 2014) and secretome (Holmberg, Quante et al. 2012). Human tumour xenograft models have shown that co-injection of CAMs with cancer cells enhances tumour growth compared to injection of non-cancer associated or non-activated fibroblasts (Orimo, Gupta et al. 2005).

It is now well established that tumour-promoting attributes of CAMs are not a consequence of genetic alterations analogous to these observed in cancer cells (Campbell, Polyak et al. 2009). As CAMs are more genetically stable then cancer cells, they represent high priority targets for biomarker discovery and development of novel therapeutic strategies (De Vlieghere, Verset et al. 2015) as discussed in section 1.3.5.

#### 1.3.3.2.1 Origin of CAMs

Cancer associated myofibroblasts form a heterogeneous population of cells, which can be derived from several cell types, including resident tissue fibroblasts (Ronnovjessen and Petersen 1993, Mitra, Zillhardt et al. 2012), adipocytes (Kidd, Spaeth et al. 2012), epithelial cells (Petersen, Nielsen et al. 2003), endothelial cells (Zeisberg, Potenta et al. 2007), and bone marrow derived mesenchymal stem cells (Mishra, Mishra et al. 2008). The recruitment and activation of CAMs is mainly driven by cancer-derived growth factors and cytokines, such as transforming growth factor β (TGF-β), platelet-derived growth factor (PDGF) and interleukin 6 (IL-6) (Ronnovjessen and Petersen 1993, Shao, Nguyen et al. 2000, Giannoni, Bianchini et al. 2010). However, recently other mechanisms have also been reported such as hypoxia and reactive oxygen species (ROS). ROS have been demonstrated to promote conversion of fibroblast into myofibroblasts through the accumulation of the hypoxia-inducible factor HIF1-α, while treatment with antioxidants have been found to reduce HIF1-α levels and supress numerous myofibroblasts features (Toullec, Gerald et al. 2010). In addition, cancer-derived exosomes have also been shown to contribute to the recruitment and differentiation of fibroblasts and mesenchymal stem cells to CAMs (Webber, Steadman et al. 2010, Gu, Qian et al. 2012).

#### 1.3.3.2.2 Tumour-promoting properties of CAMs

Numerous studies have reported that primary stromal myofibroblasts derived from various types of tumours, exhibit tumour-promoting properties and actively contribute to tumour progression and metastasis compared to non-cancerous tissue-derived myofibroblasts. As such, CAMs have been shown to influence several emerging hallmarks of cancer (Figure 1.4). It was demonstrated that multiple CAM-derived factors sustain pro-proliferative signalling in cancer cells, support the

resistance of cancer cells to cell death and growth suppressors and promote cancer cell invasion and metastasis. In addition, CAMs have been strongly implicated in tumour angiogenesis, immunosuppression, inflammation and ECM remodelling (Pietras and Ostman 2010, Polanska and Orimo 2013, Ohlund, Elyada et al. 2014). Significantly, CAMs have also been reported to play a role in deregulating cellular energetics and to contribute to genome instability and the accumulation of mutations in cancer cells (Martinez-Outschoorn, Balliet et al. 2010, Martinez-Outschoorn, Lisanti et al. 2014). In this study the effects of CAM-derived secretory factors on gastric cancer cell migration and proliferation are being investigated as highlighted in Figure 1.4.

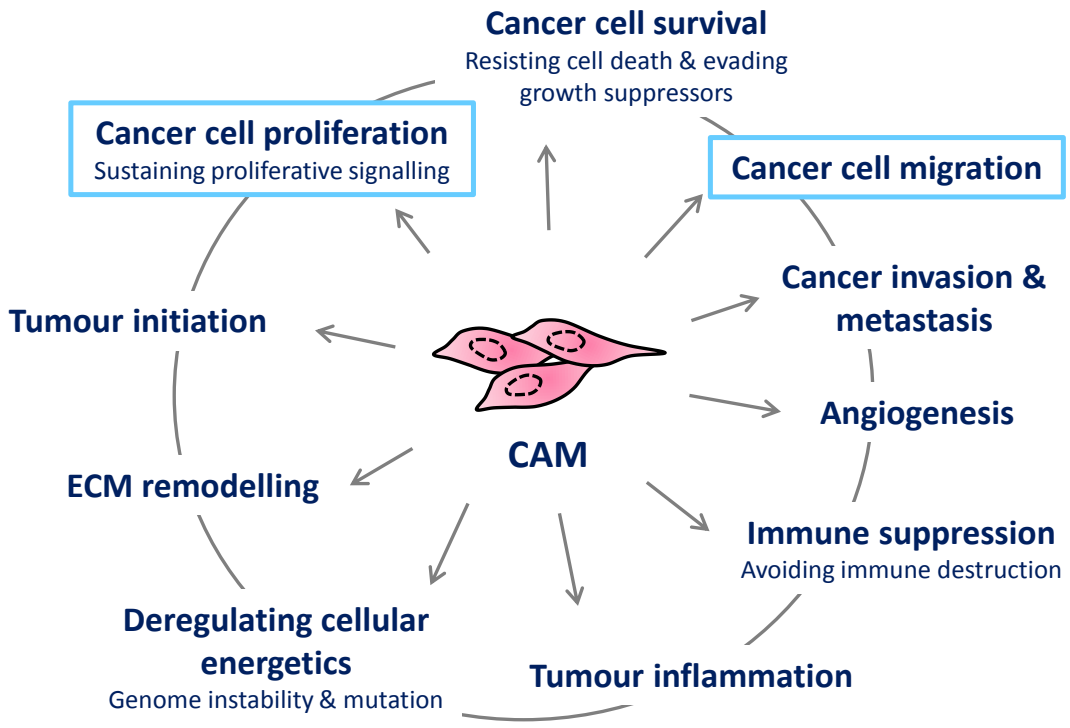


**Figure 1.4 Tumour promoting properties of cancer associated myofibroblasts (CAMs) reported in numerous studies.** Properties highlighted in boxes are being investigated in this study.

### 1.3.3.2.3 Myofibroblasts in gastric cancer

Myofibroblasts are well represented in normal gastric tissue, where they are closely allied to the epithelium. However in gastric cancers, myofibroblasts exhibit disordered architecture, morphology and function (Balabanova, Holmberg et al. 2014). Studies conducted in Professor Varro's group (University of Liverpool) showed that gastric cancer associated myofibroblasts (CAMs) secrete factors that increase the migration and proliferation of gastric cancer cells when compared to adjacent tissue myofibroblasts (ATMs), or normal tissue myofibroblasts (NTMs) (Holmberg, Quante et al. 2012). In addition, a comparative secretome analysis of myofibroblast conditioned media shows that gastric CAMs have differential protein secretion compared to ATMs. This study identified a decreased level of the extracellular matrix ECM adaptor protein like transforming growth factor β-induced gene-h3 (TGFβig-h3) in the CAM secretome, which also correlated with lymph node metastasis, worse prognosis and shorter patient survival (Holmberg, Quante et al. 2012).

Significantly, a number of other factors that distinguish gastric CAMs from their normal non-cancerous counterparts have also been reported. One such factor is galectin-1 which shows strong expression in CAMs and was shown to enhance gastric cancer cell migration and invasion by upregulating integrin-β1 expression (Xu-Jun, Hou-Quan et al. 2014). In addition, CAM-derived galectin-1 was shown to promote angiogenesis in gastric cancer (Tang, Gao et al. 2015). Other examples include miR-106b and fibroblast growth factor 9 (FGF-9) which are upregulated in gastric CAMs (Ting-Song, Xiao-Hu et al. 2014, Sun, Fukui et al. 2015). FGF-9 was shown to promote the anti-apoptosis and invasive capability of gastric cancer cells (Sun, Fukui et al. 2015) whereas upregulated miR-106b was associated with poor patient prognosis as it was shown to enhance gastric cancer cell migration and invasion (Ting-Song, Xiao-Hu et al. 2014). Immunohistochemistry and real-time PCR experiments showed that CAMs are frequently accumulated in gastric cancer tissue,

and the prevalence of CAMs is positively correlated with tumour size, depth and metastasis (Zhi, Shen et al. 2010). Also, studies in scirrhous gastric cancer, which has the worst prognosis among types of gastric cancer, revealed that CAMs might regulate the stemness of cancer stem cells (CSCs) (Tsuyoshi, Masakazu et al. 2014) and identified Asporin as an unique CAM-derived secretory protein that promotes coordinated invasion of CAMs and cancer cells (Satoyoshi, Kuriyama et al. 2014).

Taken together, these findings demonstrate that CAMs play a vital role in various aspects of gastric cancer pathogenesis. It is therefore critical to further elucidate the role of gastric CAMs in tumour progression and patient prognosis.

### 1.3.4 Hypoxia

Hypoxia is a condition in which a region of the body is deprived of oxygen and is rare in normal tissue. In solid tumours, however regions of hypoxia are common due to high rate of tumour growth that cannot be sustained by a limited oxygen supply and abnormal tumour vasculature (Harris 2002). Hypoxic tumours exhibit increased aggressiveness and metastasis, and are often associated with poor patient prognosis and resistance to therapy (Subarsky and Hill 2003, Vaupel and Mayer 2007).

#### 1.3.4.1 Cellular responses to hypoxia

Hypoxia activates the hypoxia-inducible factor (HIF) family of transcription factors, which plays a central role in the adaptation of tumour cells to hypoxic condition by regulating diverse cellular processes, including metabolism, angiogenesis, cell proliferation, apoptosis, tissue remodelling and metastasis (Carmeliet, Dor et al. 1998, Bertout, Patel et al. 2008, Gilkes, Semenza et al. 2014). Under hypoxic

conditions, HIF-1α is stabilized and forms a heterodimer with the oxygen-independent subunit HIF-β, translocates to the nucleus, and directly activates the expression of hundreds of target genes (Semenza 2003). Significantly, hypoxia was reported to induce the expression of a number of microRNAs in cancer cells (Kulshreshtha, Ferracin et al. 2007) and promote the release of cancer-derived exosomes (King, Michael et al. 2012). In addition, exosomes secreted under hypoxic condition were found to enhance cancer cell invasiveness, stemness, and induce microenvironment changes in contrast to exosomes secreted under normal oxygen condition (Ramteke, Ting et al. 2015).

### 1.3.4.2 Hypoxia-associated epigenetic changes

It is increasingly evident that hypoxic tumour cells display a distinct epigenetic profile. A number of genes have been reported to be transcriptionally repressed by promoter hypermethylation and various histone modifications in cancer cells under hypoxic conditions (Ramachandran, Ient et al. 2015). Interestingly, chromatin immunoprecipitation (ChIP) analyses identified a hypoxia-induced signature of chromatin modifications which may play a role in gene regulation in proliferating tumour cells undergoing cyclic periods of hypoxia and reoxygenation (Johnson, Denko et al. 2008). The authors found that low oxygen level causes widespread repression of transcription and globally induces a mixture of histone modifications commonly associated with either transcriptional activation or repression. In general, the performed ChIP analyses of core promoters of hypoxia-repressed and hypoxia-activated genes showed that gene-specific profiles of repressed or activated chromatin structure correlate with hypoxia-regulated, decreased or increased gene expression. However, at each hypoxic-responsive promoters tested, whether activated or repressed, the authors found increased trimethylation of histone 3 lysine 4 (H3K4me3) and decreased trimethylation of histone3 lysine 27 (H3K27me3) induced by hypoxia. These specific exceptions to commonly accepted principles of 'histone code', which associate H3K4me3 with transcriptional

activation and H3K27me3 with transcriptional repression, reveal a hypoxia-induced histone signatures that may be characteristic for the often transient state of tumour hypoxia (Johnson, Denko et al. 2008). In addition, low oxygen concentration has also been linked to an increase in global DNA hypomethylation in hypoxic cancer cells (Shahrzad, Bertrand et al. 2007, Liu, Liu et al. 2011).

#### 1.3.4.3 Hypoxia in gastric cancer

Hypoxia is an important factor in gastric cancer pathogenesis. Immunohistochemical study of gastric biopsies show that HIF-1α expression is an early event in gastric carcinogenesis as its expression increases with progression from normal mucosa to gastric cancer (Griffiths, Pritchard et al. 2007). Inhibition of HIF-1α activity was shown to impair gastric tumour growth, angiogenesis and vessel maturation (Stoeltzing, McCarty et al. 2004). It was also suggested that HIF-1α expression and *H. pylori* infection may play a synergistic role as both are closely related with gastric cancer occurrence, the depth of cancer cell invasion and lymph node metastasis (Wu, Zheng et al. 2013). In addition, hypoxic microenvironment was reported to induce epithelial–mesenchymal transition (EMT), enhance stem-like properties of gastric cancer cells and promote gastric cancer cell invasion and metastasis (Matsuoka, Yashiro et al. 2013, Guo, Wang et al. 2015).

## 1.3.5 Targeting the tumour microenvironment

Targeting of the tumour microenvironment is an area of the considerable interest for both patient stratification and the identification of new targets for pharmacological inhibition.

### 1.3.5.1 Prognostication and diagnosis

It is now well documented that higher amounts of stroma components within the tumour bulk have negative prognostic impact on patient survival (Labiche, Heutte et al. 2010, Wang, Ma et al. 2012). In addition, numerous single biomarkers, either CAM-specific of CAM-derived, which demonstrated independent association with patient survival have been identified in immunohistochemical studies (Hale, Hayden et al. 2013, Paulsson and Micke 2014, De Vlieghere, Verset et al. 2015). Significantly, gene expression profiling of CAMs derived from breast cancer tumours have been reported to identify stroma signatures that are correlated with clinical outcome (Finak, Bertos et al. 2008) and can predict therapy response (Farmer, Bonnefoi et al. 2009). Also, it was shown recently that a population of circulating CAMs (cCAM) could be identified in the peripheral blood of patients with metastatic breast cancer. Therefore, cCAMs together with circulating tumour cells (CTC) might be used as a clinically relevant biomarker for the detection of cancer metastasis (Ao, Shah et al. 2015).

### 1.3.5.2 Stroma targeted therapy

The reactive tumour stroma is increasingly being viewed as a potential therapeutic target, and strategies are being developed to disrupt tumour-stroma interactions (Gonda, Varro et al. 2010). Stroma targeted therapy might be a valid complement to conventional cancer cell–directed treatments and might have broad clinical

implications as a therapeutic strategy. For example, in gastric cancer tissue, it was shown that cancer cells express platelet-derived growth factor (PDGF), whereas stromal cells, including CAMs, pericytes and lymphatic endothelial cells express PDGF receptor (PDGF-R) (Sumida, Kitadai et al. 2011). Using an orthotopic nude mouse model of human gastric cancer, it was demonstrated that oral administration of PDGF-R tyrosine kinase inhibitor in combination with intraperitoneal injection of cytotoxic chemotherapeutic agent not only significantly inhibited the growth of tumours but also the incidences of lymph node and peritoneal metastasis (Sumida, Kitadai et al. 2011). In addition to disrupting the crosstalk between tumour cells and stroma components, other treatment strategies include: i) inhibiting or sequestering stroma-derived factors that stimulate neoplastic cell behaviour and drug resistance, such as vascular endothelial growth factor (VEGF), which is one of the most potent factors contributing to angiogenesis (Hurwitz, Fehrenbacher et al. 2004, Crawford, Kasman et al. 2009); ii) targeting the hypoxic milieu by targeting HIF-1α (Semenza 2003), and iii) targeting the stromal barrier to increase drug delivery (Provenzano, Cuevas et al. 2012).

## 1.4 Epigenetics

In the past few decades, epigenetics has emerged as an exciting new field in development and disease. Epigenetics is usually defined as the study of heritable changes in gene expression that are not due to direct changes in the primary DNA sequence (Bird 2007).

It is now well recognized that the genetic information itself is not sufficient to explain many complex phenotypes and causes of human diseases. Epigenetic mechanisms, including DNA and histone modifications, must also be considered as they play an important role in regulating distinct gene expression profiles and therefore control diverse normal cellular phenotypes and functions. As such, alterations of epigenetic code are central to many common human diseases, including cancer (Robertson 2005, Esteller 2007, Feinberg 2007). Better understanding of the interplay between epigenetic changes, gene expression and cancer hold promise for the development of new approaches to improve molecular diagnosis and targeted therapies (Egger, Liang et al. 2004, Feinberg 2007, Gal-Yam, Saito et al. 2008).

### 1.4.1 Mechanisms of gene regulation

Epigenetic mechanisms that control gene expression include DNA methylation, histone modifications, chromatin remodelling and regulation by small and long non-coding RNAs (Han, Witmer et al. 2007, Jones and Baylin 2007).

DNA methylation was the first recognized and one of the most studied epigenetic modifications. It regulates the expression of both protein-coding and non-protein coding RNAs, such as microRNAs ((Han, Witmer et al. 2007). Conversely, microRNAs

can also regulate histone modifications and the expression of DNA methyltransferases, which regulate DNA methylation (Benetti, Gonzalo et al. 2008, Sinkkonen, Hugenschmidt et al. 2008). Long noncoding RNAs may modulate the activity of DNA methyltransferases and interact with the histone machinery (Rinn and Chang 2012, Merry, Forrest et al. 2015). Furthermore, specific histone modifications such as acetylation, methylation and phosphorylation contribute to the remodelling of chromatin structure, thus regulate its accessibility and compactness (Kouzarides 2007). In turn, the state of chromatin can control the expression of non-protein coding RNAs (Chen, Fu et al. 2014). Histone modifications can also attract DNA methyltransferases to initiate cytosine methylation, which in turn can reinforce histone modification patterns conducive to silencing (Cedar and Bergman 2009). This complex interplay between the different types of epigenetic mechanisms and modifications creates an 'epigenetic landscape', which ultimately results in the regulation of phenotype-specific gene expression profiles (Figure 1.5), which are important for a given phenotype, or play an important role in response to internal or external stimuli.

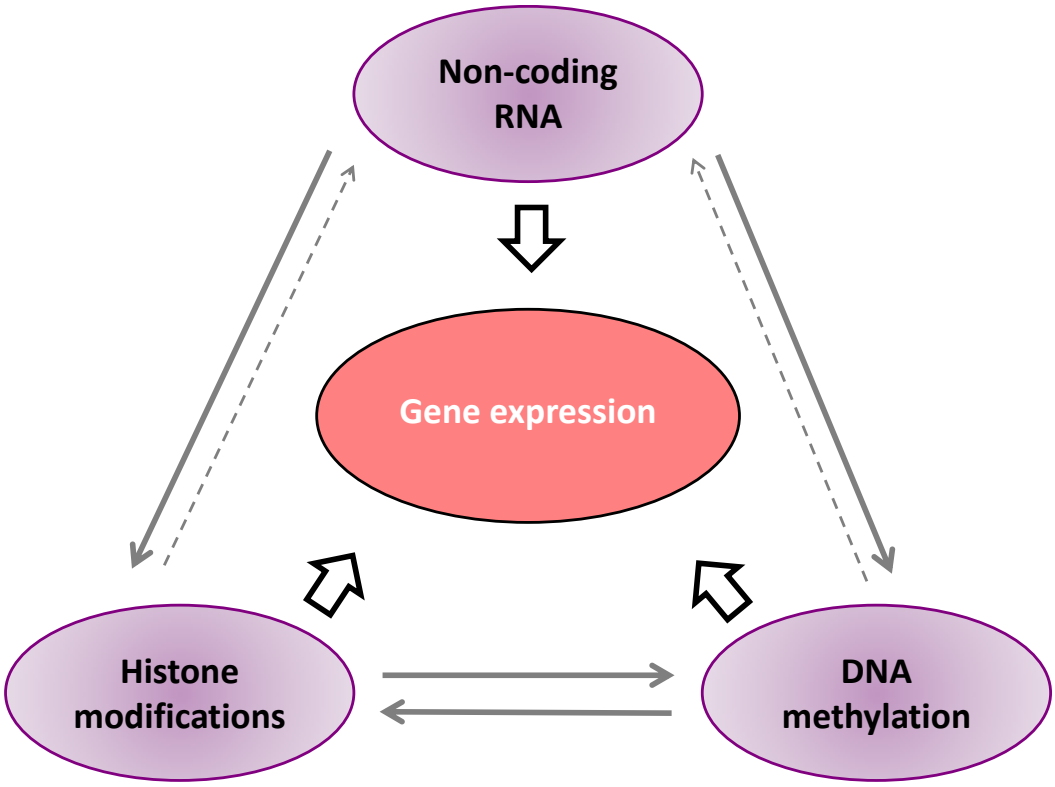


**Figure 1.5 Interplay between different aspects of epigenetic regulation of gene expression.**

## 1.4.2 DNA methylation

DNA methylation is a stable epigenetic mark which results from addition of the methyl group (-$CH_3$) at the 5-carbon ring of the cytosine nucleotides (Ehrlich and Wang 1981). This conversion of cytosine to 5-methylcytosine (5-mC) is catalysed by DNA methyltransferases (DNMTs) and is associated with gene silencing (Herman and Baylin 2003). DNA methylation is essential for normal development and plays important role in a number of key cellular processes, including genomic imprinting, X-chromosome inactivation, suppression of repetitive elements and preservation of chromosome stability (Robertson 2005, Smith and Meissner 2013). It regulates gene expression either by recruiting repressive methyl-binding proteins or by impairing the binding of transcription factors to their regulatory sequences (Klose

and Bird 2006). More recently, it has also been found to play a role in alternative splicing (Maor, Yearim et al. 2015).

#### 1.4.2.1 Genomic distribution of CpG methylation

In somatic cells, 5-mC occurs primarily in CpG contexts, where a cytosine residue is followed by guanine base. The genomic distribution of CpG methylation is highly non-random as evidence suggests that the function of DNA methylation may vary with genomic context.

Early studies focused on CpG islands, stretches of DNA that have high CpG density in an otherwise CpG-sparse genome (Figure 1.6). CpG islands often co-localize with gene promoters and are typically unmethylated allowing gene transcription. Aberrant hypermethylation of these regions results in transcriptional gene repression (Figure 1.7) (Bird 1986, Jones and Baylin 2002). The same methylation pattern is observed for CpG island 'shores', which are defined as the 2kb regions flanking a CpG island (Figure 1.6). These regions have been reported to be more dynamic then the CpG islands. In addition, most of the tissue-specific DNA methylation seems to occur in CpG island shores (Doi, Park et al. 2009, Irizarry, Ladd-Acosta et al. 2009). Beyond shores are 'shelves', which are defined as the 2 kb regions flanking CpG island 'shores' (Figure 1.6) (Bibikova, Barnes et al. 2011).

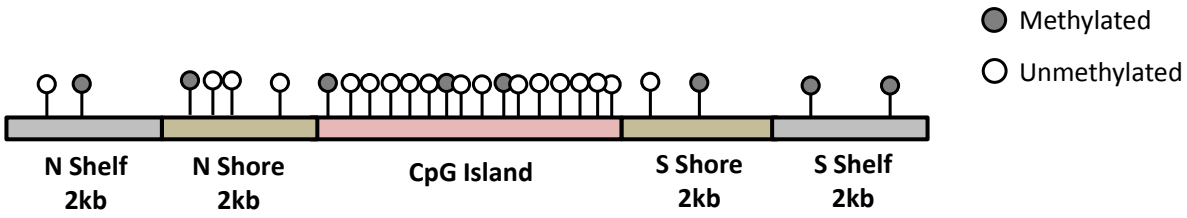


**Figure 1.6 Distribution of CpG sites in genomic regions.**

Contrary to promoter region, methylation at the gene body facilitates transcription and is common in ubiquitously expressed genes (Hellman and Chess 2007). It has been proposed that gene body methylation might be related to elongation efficiency and prevention spurious initiations of transcription (Zilberman, Gehring et al. 2007). In disease, the gene body tends to lose methylation (Figure 1.7), allowing transcription to be initiated at several incorrect sites.

Finally, DNA methylation is not only associated with gene regulation. Heavily methylated CpGs fraction is found in repetitive sequences. The hypermethylation at these regions is needed to protect chromosomal integrity which is achieved by preventing reactivation of endoparasitic sequences that cause chromosomal instability, translocations and gene disruptions (Esteller 2007). This methylation pattern is commonly altered in cancer (Figure 1.7)

### 1.4.2.2 DNA methyltransferases

Different members of the DNA methyltransferase (DNMT) family of enzymes are required either for maintenance or *de novo* establishment of DNA methylation patterns. DNMT1 is responsible for the maintenance of established DNA methylation patterns. It functions during DNA replication to copy the DNA methylation pattern from the existing DNA strand onto the newly synthesized daughter strand (Szyf, Bozovic et al. 1991). DNMT3a and DNMT3b are highly expressed in developing embryos as they mediate establishment of *de novo* DNA methylation patterns, putting the initial pattern of methyl group in place on a unmodified DNA sequence (Okano, Bell et al. 1999).

### 1.4.2.3 Dynamics of DNA methylation

In spite of being relatively stable, DNA methylation can also be quite dynamic, changing in response to cell environmental stimuli. One such example might be regular endurance training. A study in humans showed that regular exercise over a period of 3-months induces genome-wide DNA methylation changes in human skeletal muscles which are correlated with gene expression changes (Lindholm, Marabita et al. 2014). A similar study has been done in mice (Kanzleiter, Jaehnert et al. 2015). In fact, it was suggested that an analysis of specific genomic regions, which were identified as 'highly dynamic differentially methylated regions' across 30 diverse human cell and tissue types may prove useful for human health diagnostics and disease prognosis (Ziller, Gu et al. 2013).

## 1.4.3 DNA methylation in cancer

Aberrant patterns of DNA methylation are one of the characteristic features of cancer cells (Feinberg and Tycko 2004, Esteller 2007). Specifically, it was shown that cancer genomes are hypomethylated while promotors of tumour suppressor genes are often hypermethylated when compared to non-cancerous cells (Figure 1.7) (Feinberg and Vogelstein 1983, Jones and Baylin 2002). Global DNA hypomethylation alters chromatin architecture leading to activation of oncogenes and transposable elements whereas promoter hypermethylation represses the expression of tumour suppressor genes. Consequently, these epigenetic events provide cancer cells with a growth advantage, increasing their genetic instability and allowing them to metastasize (Herman and Baylin 2003, Esteller 2008). In fact, promoter hypermethylation and the resulting silencing of tumour suppressor genes as well as global loss of genomic methylation are frequent and early events in cancer and in some instances correlate with disease severity and metastatic potential (Chan, Broaddus et al. 2002, Widschwendter, Jiang et al. 2004).

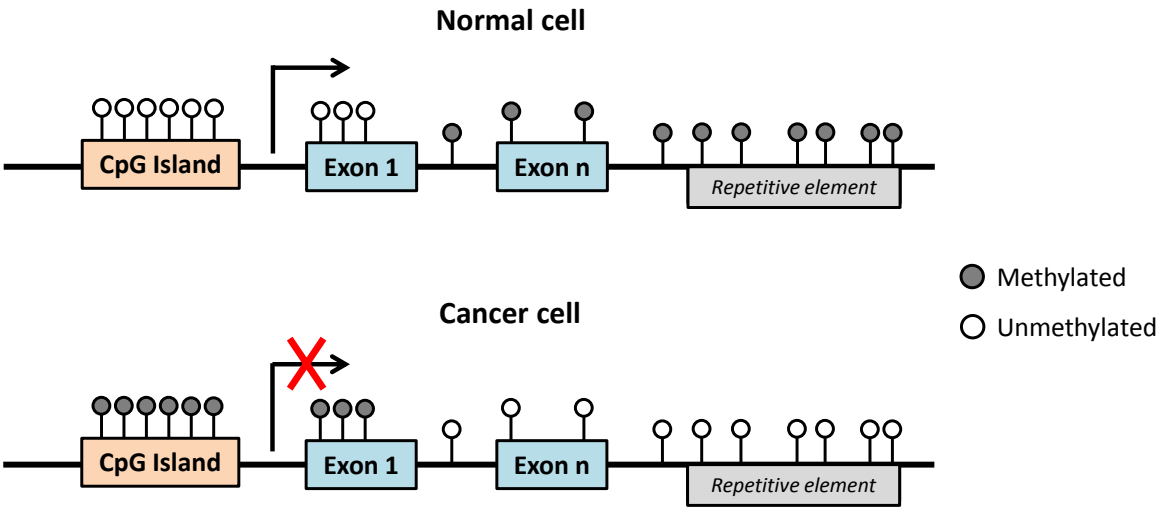


**Figure 1.7 Schematic representations of DNA methylation patterns in normal and cancer cells.** In normal cells (top), CpG islands associated with promoter regions usually remain unmethylated allowing gene transcription and methylation of repetitive sequences prevents genomic instability. In cancer cells (bottom), promoter regions become methylated resulting in transcriptional repression of many genes whereas global loss of methylation triggers genomic instability and aberrant transcriptional initiations.

### 1.4.4 DNA methylation in gastric cancer

Epigenetic alterations contribute significantly to the development and progression of gastric cancer (Gigek, Chen et al. 2012, Kang, Song et al. 2014). Aberrant DNA methylation in gene promoters is the most well-defined epigenetic hallmark in gastric cancer. A large number of genes with different biological functions have been found to be methylated including genes involved in DNA repair, cell cycle, cell adhesion, invasion, migration, growth, differentiation, apoptosis, transcriptional regulation and other processes (Qu, Dang et al. 2013, Tahara and Arisawa 2015). In addition, aberrant methylation of a number of genes has been significantly associated with clinicopathological features and clinical outcomes (Qu, Dang et al. 2013, Tahara and Arisawa 2015). Significantly, recent reports have indicated that infection with *H. pylori* is associated with elevated levels of aberrant DNA methylation (Shin, Kim et al. 2011) and that DNA methylation transcriptionally regulate the number of microRNAs in gastric cancer (Bae, Kang et al. 2015). Given that aberrant DNA methylation is more frequent than mutations and can drive the pathogenesis of gastric cancer and contribute to the molecular heterogeneity of gastric cancers, a lot of effort has been put into defining promoter methylation signatures that can be used for early detection and diagnosis of gastric cancer in biopsy specimens and non-invasive body fluids such as serum and gastric washes as well as for prognostication and the prediction of treatment responses (Abbaszadegan, Moaven et al. 2008, Watanabe, Kim et al. 2009, Sapari, Loh et al. 2012, Schneider, Mera et al. 2015).

### 1.4.5 Epigenetic cancer therapy

Unlike genetic mutations, epigenetic alterations are potentially reversible. Several epigenetic drugs that target enzymes controlling epigenetic modifications have already shown promising anti-tumorigenic activity in some malignancies and are currently being tested in clinical trials (Egger, Liang et al. 2004, Mund and Lyko

2010). Epigenetic drugs that have already been approved by FDA for use in certain types of cancer are described below.

### 1.4.5.1 DNA methyltransferase inhibitors

DNA methyltransferase (DMNT) inhibitors, such as 5-azacytidine (azacytidine) and 5-aza-2'-deoxycytidine (decitabine) are cytidine analogues that function as demethylating agents. They are incorporated into DNA during replication and form a covalent bond with DNMT1, which triggers the degradation of the enzyme and leads to widespread reductions in DNA methylation (Santi, Norment et al. 1984, Egger, Liang et al. 2004). As these compounds integrate into DNA during replication, rapidly dividing cancer cells are more susceptible to their effects. Both 5-azacytidine (azacytidine) and 5-aza-2'-deoxycytidine (decitabine) have already been approved by the FDA for use in certain haematological malignancies (Kaminskas, Farrell et al. 2005, Malik and Cashen 2014).

### 1.4.5.2 Histone deacetylase inhibitors

Histone deacetylases (HDACs) catalyse the removal of acetyl groups from amino-terminal lysine residues of histones, which results in chromatin condensation and transcriptional repression. Histone deacetylase (HDAC) inhibitors form a diverse group of compounds. They inhibit not only HDACs but also a large number of non-histone transcription factors and transcriptional co-regulators modified by acetylation (Yoo and Jones 2006). HDAC inhibitors have been found to induce different phenotypes in various transformed cells, including growth arrest, activation of apoptosis, autophagic cell death, reactive oxygen species-induced cell death, mitotic cell death and senescence (Xu, Parmigiani et al. 2007). Two structurally unrelated HDAC inhibitors, vorinostat and istodax, have been approved by the FDA for the treatment of cutaneous T cell lymphoma (Grant, Easley et al.

2007, Bertino and Otterson 2011). Significantly, HDAC inhibitors have been found to sensitize cancer cells to a variety of cytotoxic drugs (Mataga, Rosenthal et al. 2012) thus they can be used in combinational therapy with other anti-cancer drugs which is expected to increase cancer cell drug specificity and reduce side effects (Sarkar, Goldgar et al. 2013).

## 1.5 Personalized cancer therapy

The latest advances in 'omics' technologies have led to some major breakthroughs in cancer research, which will have the potential to influence clinical practice and reshape how the disease is characterized and treated. This includes gastric cancer.

Personalized cancer therapy offers a tailored treatment strategy that aims to either increase the efficacy of the individual patient treatment or lower the potential side effects by selecting the appropriate treatment based on the patient's unique genetic characteristics and tumour related biomarkers.

### 1.5.1 Gene expression profiling

Numerous studies have recently used high-throughput large-scale gene expression profiling technology to identify unique 'gene expression signatures' that can be used to classify gastric tumours, determine response to treatments and predict patient survival (Brettingham-Moore, Duong et al. 2011, Lin, Zhao et al. 2015). Shah et al. demonstrated that distinct subtypes of gastric cancer, classified by histopathologic and anatomic criteria as proximal, diffuse and distal, can also be classified based on their unique gene expression profiles (Shah, Khanin et al. 2011). In this study, gene expression analysis of 36 primary gastric adenocarcinoma samples identified gene signatures that successfully distinguish the three gastric

cancer subtypes with more than 85% accuracy (Shah, Khanin et al. 2011). In a different study, gene expression profiling was used to identify gastric cancer intrinsic subtypes that might be associated with differences in patient survival and response to chemotherapy. The analysis of 37 gastric cancer cell lines identified gene signature that can predict two major intrinsic genomic subtypes: genomic intestinal (G-INT) and genomic diffuse (G-DIF), which were found to be partially associated with Lauren' histopathologic classification. The G-INT cell lines were significantly more sensitive to 5-fluorouracil and oxaliplatin, but more resistant to cisplatin, than the G-DIF cell lines (Tan, Ivanova et al. 2011). 5-Fluorouracil, oxaliplatin, and cisplatin are drugs presently used in the adjuvant and first-line palliative gastric cancer treatment. Notably, gene expression profiling was also used to predict the likelihood of relapse after gastric cancer surgery. Cho et al. analysed 65 gastric adenocarcinoma samples and developed a risk score based on the gene signature associated with relapse, which was successfully tested in an independent cohort (Cho, Lim et al. 2011).

### 1.5.2 DNA methylation profiling

The genome-wide DNA methylation profiling technologies allow interrogation of DNA methylation status over large genomic regions and are now commonly used to characterise epigenomes of human cancers and identify clinically useful diagnostic biomarkers. Notably, Yamanoi et al. demonstrated that distinct DNA methylation profiles, which may determine tumour aggressiveness and patient outcome, have already become established in non-cancerous tissue at the precancerous stage and may be inherited by gastric carcinomas themselves (Yamanoi, Arai et al. 2015). The genome-wide differential DNA methylation analysis of 109 non-cancerous gastric mucosa and 105 gastric tumour tissues identified widespread alterations of DNA methylation between the two groups. Using the identified DNA methylation signature of 3861 differentially methylated loci they clustered the normal mucosa samples into three subgroups: A, B1 and B2. Gastric carcinomas belonging to

subgroup B1 showed tumour aggressiveness more frequently than those belonging to subgroups A and B2. The recurrence-free and overall survival rates of patients in subgroup B1 were lower than those of patients in subgroups A and B2. Multivariate analyses revealed that DNA methylation level in tumour samples was a significant prognostic factor, being independent of clinicopathological parameters (Yamanoi, Arai et al. 2015).

## 1.6 Objectives and Aims

In this study, an integrated genomics, transcriptomics and proteomics approach was used to understand regulation and crosstalk in the tumour microenvironment. The primary objective of this work was to identify and characterise genome-wide DNA methylation patterns in primary cancer-associated myofibroblasts (CAMs) derived from gastric tumours and establish the extent to which DNA methylation may regulate gene expression in these cells. The secondary objective of this work was to identify and characterise the consequences of tumour hypoxia on the ability of stromal cells to induce gastric cancer cell migration and proliferation.

Specific aims were:

- To identify DNA methylation and gene expression changes in primary gastric myofibroblasts purified from different tissue microenvironments: CAM (cancer), ATM (adjacent tissue) and NTM (normal tissue) and to investigate how these differential DNA methylation and gene expression signatures contribute to the tumour-promoting properties of gastric CAMs.
- To determine the effects of CAM and NTM hypoxia-induced secretory factors on gastric cancer cell migration and proliferation.
- To identify hypoxia-induced gene expression signatures in gastric CAMs, ATMs and NTMs and investigate their potential biological consequences.
- To identify and quantify hypoxia-induced CAM and NTM secretomes and integrate them with hypoxia-induced CAM and NTM gene expression signatures.

# Chapter II

# Materials and Methods

# 2.1 Materials

**Table 2.1 List of materials used in this study.**

| Material | Supplier |
|---|---|
| AGS Gastric Adenocarcinoma Cell Line | American Type Culture Collection (VA, USA) |
| MKN45 Gastric Adenocarcinoma Cell Line | Riken BioResource (Wako, Japan) |
| Dulbecco's Modified Eagle's Medium (DMEM), Fetal Bovine Serum (FBS), Penicillin–Streptomycin (10 000U/ml), Antibiotic–Antimycotic (100x), MEM Non-Essential Amino Acid Solution (100x), Dulbecco's Phosphate Buffered Saline (PBS) (10x), Trypsin–EDTA (0.25%), Click-iT EdU Alexa Fluor 488 Imaging Kit, ProLong Gold Antifade Reagent, Quant-iT PicoGreen dsDNA Assay Kit, TaqMan Universal Master Mix II, TaqMan assays | Life Technologies (Paisley, UK) |
| BD BioCoat Control Inserts with 8.0 um PET Membrane in Two 24 Well Plates. | SLS (Hessle, UK) |
| Reastain Quick-Diff Kit | Reagena (Takojantie, Finland) |
| miRNeasy Mini Kit, DNase I, Proteinase K, RNase A, HotStarTaq Master Mix Kit, dNTPs, QuantiTect Reverse Transcription Kit | Qiagen (Manchester, UK) |
| Wizard SV Genomic DNA Purification System | Promega (Southampton, UK) |
| EZ DNA Methylation-Gold Kit | Zymo Research (CA, USA) |
| Custom Primers | Eurofins MWG (Ebersberg, Germany) |
| Agilent 2100 Bioanalyzer and RNA 6000 Nano kit | Agilent Technologies (Stockport, UK) |
| Sepharose Beads | GE Healthcare (Amersham, UK) |
| Agarose | Bioline (London, UK) |
| Lambda DNA/HindIII Marker | New England Biolabs (Hitchin, UK) |
| Formaldehyde 16% Solution | Agar Scientific (Essex, UK) |
| Chloroform, Absolute Ethanol, Triton X-100, Bovine Serum Albumin | Sigma-Aldrich (Dorset, UK) |
| Hypoxystation H35 | Don Whitley Scientific Limited (West Yorkshire, UK) |

## 2.2 Tissue Culture

### 2.2.1 Generation of human primary gastric myofibroblasts

Human primary myofibroblasts were derived from resected gastric tumours (CAMs) and matched adjacent tissue (ATMs) obtained from patients undergoing gastric cancer surgery at Department of Surgery, University of Szeged (Hungary). Myofibroblasts were prepared in the Department of Medicine, University of Szeged, (Hungary) as previously reported (McCaig, Duval et al. 2006). Tumour and adjacent tissues were characterized using the TNM classification. Normal tissue myofibroblasts (NTMs) were generated from deceased transplant donors with normal gastric morphology. The list of patient-matched CAM and ATM myofibroblast cell lines and normal tissue myofibroblast (NTM) cell lines used in this study is shown in Table 2.2 and Table 2.3, respectively. For histopathological assessments, myofibroblasts were defined as stellate/spindle-shaped cells with consistent α-SMA and vimentin co-expression (Holmberg, Quante et al. 2012). Smooth muscle fibres were excluded based on their characteristic morphology. The study was approved by the Ethics Committee of the University of Szeged (Hungary).

### 2.2.2 Gastric myofibroblast cell culture

Primary human gastric myofibroblasts (CAMs, ATMs and NTMs) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS), 1% penicillin-streptomycin, 1% antibiotic-antimycotic and 1% non-essential amino acid solution. This is referred to as growth media unless otherwise stated. Cells were incubated at 37°C in a humidified atmosphere with 5% $CO_2$ and were split when they reached 80-90% confluence by washing in 1x Dulbecco's Phosphate Buffered Saline (PBS) twice before adding 0.25% trypsin then the cell passage number was noted. The growth media was changed approximately every 60 hours. In all experiments cells were not passaged beyond passage 12.

**Table 2.2 Patient information relating to age, gender, tumour location and tumour clinical assessment for gastric cancer patients who provided CAM and ATM cells used in this study.** Tumour staging is defined in terms of pathology of tumour (pT) where 0 defines no sign of tumour and 3 maximum size and/or extensions. Similarly, the involvement of local and metastasis to proximal lymph nodes is stated as N0-N3 and metastasis severity by M0, no metastasis and M, distal metastasis.

| Patient ID | Age | Gender | Location of Tumour | Lauren Classification | *H.pylori* status | Prognostic Score | Tumour Staging | Survival (months) | Adjacent Tissue |
|---|---|---|---|---|---|---|---|---|---|
| **192** | 49 | F | antrum corpus border | diffuse | **+** | 11 | pT3N1M0 | 22 | chronic gastritis |
| **308** | 51 | M | antrum corpus border | mixed | **+** | 12 | pT1N3M0 | 9 | chronic gastritis |
| **305** | 59 | F | antrum and corpus | diffuse | **+** | 12 | pT3N2M0 | 17 | chronic gastritis |
| **42** | 72 | M | antrum corpus border | medullar (non-Lauren) | **-** | 5 | pT1N0M0 | 75 | intestinal metaplasia, atrophy |
| **45** | 82 | M | antrum | intestinal | **-** | 11 | pT4N2M0 | 2 | intestinal metaplasia, chronic gastritis |
| **271** | 72 | M | corpus | mixed | **-** | 12 | pT3N1M0 | 24 | chronic gastritis |
| **268** | 76 | M | antrum | intestinal | **+** | 11 | pT4N2M0 | 15 | intestinal metaplasia, atrophy, chronic gastritis |

**Table 2.3 Information relating to age, gender and cell origin for post-mortem organ donors who provided NTM cells used in this study.** The donors had no known underlying medical conditions.

| Sample ID | Origin | Age | Gender |
|---|---|---|---|
| **196/2** | gastric corpus | 67 | M |
| **279/22** | gastric antrum | 60 | M |
| **334/22** | gastric antrum | 52 | F |
| **261/22** | gastric antrum | 52 | F |

### 2.2.3 Gastric cancer cell culture

Human gastric adenocarcinoma cell lines AGS and MKN45 were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS), 1% penicillin-streptomycin and 1% antibiotic–antimitotic solution. This is referred to as growth media unless otherwise stated. The cells were incubated at 37°C in a humidified atmosphere with 5% $CO_2$ and were passaged or growth media was changed approximately every 48-72 hours. In all experiments cells were used between passages 18 and 28.

## 2.3 Integrated multi-omics experiments

### 2.3.1 Patient information

Gastric cancer patient-matched CAM and ATM samples as well as NTM samples were cultured in parallel for integrated genome-wide DNA methylation, gene expression and secretome profiling. A full list of primary myofibroblast cell lines used in these experiments is shown in Table 2.4. Corresponding information relating to patient age, gender and tumour clinical assessment is provided in Table 2.2 and Table 2.3. Selection of primary myofibroblast cell lines for use in integrated multi-omic experiments was based on trends observed in a previous

bioinformatics analysis of comparative gene expression profiles in 12 patient-matched CAM and ATM samples (Dr Helen Smith, manuscript in preparation). This analysis identified two distinct groups of patients according to prognosis score that was determined using the semi-quantitative scoring system (Supplementary Table S2.2). Patients with prognosis score below 9 were classified as group A (good prognosis) and patient with prognosis score above 9 were classified as group B (bad prognosis). Dr Smith defined a subset of differentially expressed genes in CAM vs ATM that followed the patient prognosis scores and based on expression of these genes she noticed the patients within bad prognosis group (group B) could be further divided into two independent subgroups: subgroup B with detected *H*. *pylori* infection and subgroup B with undetected *H*. *pylori* infection. In this study 3 patients from bad prognosis subgroup with detected *H*. *pylori* infection were chosen for more detailed investigation.

### 2.3.2 Myofibroblast cell culture

In order to generate myofibroblast conditioned media 500 x $10^3$ of selected myofibroblast cell lines were seeded in 75$cm^2$ flasks. Growth media was replaced after 24 hours and cells were incubated under normoxic (21% $O_2$) or hypoxic (1% $O_2$) conditions for 48 hours. After this initial incubation, cells were washed three times in 1x PBS to remove serum-derived proteins and 13ml of freshly prepared serum-free DMEM supplemented with 1% penicillin-streptomycin, 1% antibiotic-antimycotic and 1% non-essential amino acid solution was added and conditioned for 24 hours under normoxic or hypoxic conditions. Conditioned media and cell lysates for DNA and RNA extraction were collected after 24 hours for use in Illumina Infinium HumanMethylation450k and HumanHT-12v4 Expression arrays as well as LC-MS/MS secretome analysis.

**Table 2.4 List of samples used for integrated multi-omic experiments.**

| Sample type | Patient ID | Sample ID | Sample label | Treatment |
|---|---|---|---|---|
| CAM | 192 | 192/1 | C1N | normoxia |
| | | | C1H | hypoxia |
| ATM | | 192/2 | A1N | normoxia |
| | | | A1H | hypoxia |
| CAM | 308 | 308/1 | C2N | normoxia |
| | | | C2H | hypoxia |
| ATM | | 308/22 | A2N | normoxia |
| | | | A2H | hypoxia |
| CAM | 305 | 305/1 | C3N | normoxia |
| | | | C3H | hypoxia |
| ATM | | 305/22 | A3N | normoxia |
| | | | A3H | hypoxia |
| NTM | 279 | 279/22 | N1N | normoxia |
| | | | N1H | hypoxia |
| NTM | 261 | 261/22 | N2N | normoxia |
| | | | N2H | hypoxia |
| NTM | 334 | 334/22 | N3N | normoxia |
| | | | N3H | hypoxia |

### 2.3.3 Conditioned media preparation for secretome profiling

Conditioned media from normoxic and hypoxic myofibroblast cells were centrifuged at 800g for 7 minutes to remove cell debris and supernatants were aliquoted and stored at -80°C prior to LC-MS/MS secretome analysis, which was performed by the Centre for Proteome Research (University of Liverpool).

## 2.3.4 DNA extraction for Illumina Infinium HumanMethylation450k array

Genomic DNA was purified using a standard phenol/chloroform extraction method. Briefly, myofibroblast cells were washed with 1x PBS and directly lysed in 75cm$^2$ flasks by adding 2ml DNA Lysis Buffer (400mM Tris-HCl, pH 8, 10mM EDTA, 1% SDS, 150mM NaCl). The cell lysates ware collected in 15ml Falcon tubes and stored in -80°C prior to further processing. Lysates were then thawed and heated for 10 minutes at 60°C on a shaking platform. Proteinase K (Qiagen, cat. no. 19131) was then added to a final concentration of 100µg/ml and samples were incubated overnight at 37°C on a shaking platform before being incubated at 70°C for 30 minutes. RNase A (Qiagen, cat. no. 19101) was then added to a final concentration of 100µg/ml and samples were incubated at 37°C for 1 hour. After treatment with RNase A, Proteinase K (final concentration of 100µg/ml) was again added to samples, which were then incubated for 2 hours at 37°C on a shaking platform. After protein digestion, 2ml phenol/$CHCl_3$/isoamyl alcohol (50:49:1) was added to samples, which were mixed by repeated inversions for 15 minutes, prior to centrifugation at full speed for 5 minutes. The aqueous phase was then transferred to new Falcon tubes and 2ml of chloroform was added and mixed by inverting the tubes, then centrifuged at full speed for 5 minutes. The aqueous phase was transferred to new Falcon tubes and 2.5 volumes of absolute ethanol were added. Samples were left overnight at -20°C for DNA precipitation. The next day samples were centrifuged at full speed for 15 minutes and DNA pellets were washed twice in 70% ethanol and centrifuged at full speed for 10 minutes. DNA pellets were left to dry before being rehydrated in 150-200µl TE Buffer (10mM Tris pH 7.5/1mM EDTA). DNA sample purity and extent of degradation were determined by spectrophotometry at 260/280 nm and gel electrophoresis on 1% agarose gel, respectively. DNA quantity was assessed using PicoGreen fluorimetry (Life Technologies, cat. no. Q-33130). DNA sample quality information is shown in Supplementary Figure S2.1 and Table S2.1. All DNA samples were normalized to the concentration of 50ng/µl in 25µl TE Buffer and prepared according to

the Gen-Probe Life Sciences (Molecular Genetic Services, Manchester) sample requirements. Before the DNA samples were sent to Molecular Genetic Services (Gen-Probe Life Sciences, Manchester) for DNA methylation analysis with Illumina Infinium HumanMethylation450k array they were randomized with respect to patient origin and treatment to avoid potential batch effects.

### 2.3.5 RNA extraction for Illumina Human HT-12 v4 array

Total RNA was purified using miRNeasy Mini Kit (Qiagen, cat. no. 217004) and a modified version of the manufacturer's instructions was used. Briefly, myofibroblast cells were washed with 1x PBS and directly lysed in 75cm$^2$ flasks by adding 1ml QIAzol Lysis Reagent. Cell lysates ware collected into Eppendorf tubes and homogenized by vortexing for 1 minute. The homogenized lysates were stored at -80°C prior to further processing. Total RNA was extracted by phase separation after addition of 140µl chloroform to the homogenized lysates. Homogenates were centrifuged at 12,000 rcf and 4°C for 15 minutes. The upper aqueous phase was collected and mixed with 1.5 volumes of absolute ethanol. Then the samples were transferred to the RNeasy Mini spin column and centrifuged at 8,000 rcf for 15 seconds, followed by a 15 minutes on-column DNase I digestion (Qiagen, cat. no. 79254). The RNeasy Mini spin column membrane was washed twice with 500µl Buffer RPE and centrifugation at 8,000 rcf for 15 seconds and 4 minutes, respectively. Columns were then transferred to new collection tubes and centrifuged at full speed for 3 minutes to eliminate any possible carryover of Buffer RPE. RNA was then eluted in 50µl RNase-free water and tested for degradation and purity using the Agilent 2100 Bioanalyzer and RNA 6000 Nano kit (Agilent Technologies, cat. no. 5067-1512). Samples RIN values and 260/280 ratios are shown in Supplementary Table S2.1. All RNA samples were normalized to the concentration of 40ng/µl in 25µl RNase-free water and prepared according to the Gen-Probe Life Sciences (Molecular Genetic Services, Manchester) sample requirements. Samples were randomized with respect to patient origin and

treatment to minimize batch processing effects prior to sending to Molecular Genetic Services (Gen-Probe Life Sciences, Manchester) for gene expression analysis using Illumina HumanHT-12 v4 array.

## 2.4 Myofibroblast conditioned media preparation

### 2.4.1 Standard conditioned media

To prepare CAM, ATM or NTM conditioned media (CM), 1.5 x $10^6$ myofibroblast cells were seeded in 75$cm^2$ tissue culture flasks and left to attach for 24 hours. The next day the cells were washed three times in 1x PBS to get rid of any serum-derived factors. Then growth media was replaced with 15ml freshly prepared serum-free DMEM supplemented with 1% penicillin-streptomycin, 1% antibiotic-antimycotic and 1% non-essential amino acid solution and incubated for 24 hours at 37 °C in a humidified atmosphere with 5% $CO_2$. The next day CM was collected and centrifuged at 800g for 7 minutes to get rid of cell debris. The freshly prepared supernatants were immediately used for cancer cell migration and proliferation assays.

### 2.4.2 Hypoxic conditioned media

In order to generate CAM or NTM hypoxic conditioned media (CAM-hypoxic-CM or NTM-hypoxic-CM) and CAM or NTM normoxic conditioned media (CAM-ctrl-CM or NTM-ctrl-CM), 5 x $10^5$ of selected myofibroblast cell lines were seeded in 75$cm^2$ flasks and left to attached for 24 hours in growth media. The next day the media was replaced and cells were subjected to either normoxia (21% $O_2$) or hypoxia (1% $O_2$) for 48 hours incubation in serum-supplemented DMEM. After 48 hours the cells were washed three times in 1x PBS to get rid of any serum-derived factors and 13ml

freshly prepared serum-free DMEM supplemented with 1% penicillin-streptomycin, 1% antibiotic-antimycotic and 1% non-essential amino acid solution was added to the cells. The serum-free media was incubated with the cells for 24 hours in normal (21% $O_2$) and hypoxic (1% $O_2$) conditions. The following day myofibroblast ctrl-CM and hypoxic-CM were collected and centrifuged at 800g for 7 minutes to get rid of cell debris. The freshly prepared CAM or NTM hypoxic-CM and ctrl-CM were immediately used for cancer cell migration and proliferation assays.

## 2.5 Cancer cell based assays

### 2.5.1 Cancer cell migration assay

The effects of myofibroblast conditioned media (CM) on gastric cancer cells migration was measured *in vitro* using trans-well Boyden chamber assay (SLS; cat. no. 354578). Briefly, 1 x $10^4$ AGS cells in 500µl of serum-free DMEM medium were added to the BioCoat Control Inserts with 8µm pore PET membrane (upper chambers). The lower chambers contained either 750µl serum-free media, or myofibroblast CM to serve as a chemoattractant. Cells were incubated at 37°C and allowed to migrate overnight. Thereafter, AGS cells were removed from the upper surface of the membrane by scrubbing with cotton swabs. Cells migrating through the membrane were fixed and detected on the lower surface using Reastain Quick-Diff Kit (Reagena; cat. no. 102164) and then examined under a bright-field microscope. Values for cancer cell migration were obtained by counting fifteen fields per membrane (20x objective) and represent an average of at least three independent membranes.

### 2.5.2 Cancer cell proliferation assay

The effects of myofibroblast conditioned media (CM) on gastric cancer cell proliferation was assessed by incorporation of 5-ethynyl-2'-deoxyuridine (EdU) (Salic and Mitchison 2008) and detected using the Click-iT EdU Alexa Fluor 488 Imaging Kit (Life Technologies; cat. no. C10337). Briefly, 1 x $10^4$ AGS cells were plated onto cover glasses in 24-well plate and left for 24 hours at 37°C to attach. The next day, growth media was replaced with serum-free DMEM supplemented with 1% penicillin-streptomycin and 1% antibiotic-antymycotic to synchronize the cells. After 24 hours cell synchronization, the media was replaced with 1ml of either serum-free media or myofibroblast CM containing 10µM EdU and cells were incubated overnight at 37°C, 5% $CO_2$. Following incubation, the manufacturer's protocol was applied to fix and permeabilize the AGS cells and detect EdU incorporation. Coverslips with cultured cells were washed three times in 1x PBS and fixed in 4% paraformaldehyde (PFA) for 15 minutes at room temperature. Then fixed cells were washed twice with 3% BSA in PBS before being incubated for 20 minutes with 0.5% Triton X-100 in PBS to permeabilize the cells. After incubation, the cells were washed twice with 3% BSA in PBS and 500µl of freshly prepared Click-iT reaction cocktail was added on each coverslip in 24-well plate. The plate was incubated in dark at room temperature for 30 minutes. Then the cells were washed with 1x PBS and nucleic acid was stained with 1x Hoechst 33342 solution for 30 minutes in dark at room temperature. Then the coverslips were washed twice in 1x PBS and left to dry before being mounted onto slides using ProLong Gold reagent (Life Technologies; cat. no. P36930).

## 2.6 Omics data processing and analysis

### 2.6.1 Illumina Infinium 450k data processing

The Illumina Infinium HumanMethylation450 BeadChip arrays were used to study genome-wide DNA methylation profiles in gastric and oesophageal stromal myofibroblasts. The Illumina 450k arrays for gastric and oesophageal myofibroblasts were performed independently. Table 2.2 provides information for gastric cancer patients who provided gastric CAM and ATM samples whereas Supplementary Table S2.3 provides information for oesophageal cancer patients who provided oesophageal CAM and ATM samples. DNA samples were processed commercially by Molecular Genetic Services (Gen-Probe Life Sciences, Manchester) who then provided raw data files (.IDAT) for further analysis.

Data processing and analysis for both 'gastric' and 'oesophageal' Illumina 450k methylation arrays were conducted as part of this project. Raw DNA methylation data was processed and analysed using Bioconductor package RnBeads version 0.99.17 (Assenov, Muller et al. 2014). RnBeads is an R package that builds upon existing software tools imported from other R/Bioconductor packages therefore providing a comprehensive analysis workflow for DNA methylation analysis. Briefly, raw data files (.IDAT) were imported directly into RnBeads. The comparison of non-CpG SNP probes present on the array (n=65) confirmed that CAMs and ATMs were sourced from the same patient as expected. The methylation level (β-value) for the interrogated CpG loci was computed by calculating the ratio of the methylated (M) probe signal intensity to the sum of both methylated (M) and unmethylated (U) probe signal intensities. β-values range from 0 to 1, corresponding to completely unmethylated and fully methylated loci, respectively. Prior to normalization, probes with *p-value>0.01* in any sample and probes overlapping with single nucleotide polymorphisms (SNPs) in the probe sequence (Chen, Lemire et al. 2013) were removed followed by *BMIQ* normalization

(Teschendorff, Marabita et al. 2013, Wu, Joubert et al. 2014) and *methylumi.noob* background subtraction (Triche, Weisenberger et al. 2013). Experimental quality control using control probe information present on the array was also performed and quality control reports are presented in *Appendix I*. Subsequently probes outside of CpG context as well as probes on sex chromosomes were removed. The Illumina Infinium HumanMethylation450 BeadChip arrays processing and analysis workflow is presented in Figure 2.1A. Downstream 'gastric' data analysis was restricted to the remaining 424383 probes whereas 'oesophageal' data analysis was restricted to 424355 probes.

### 2.6.2 Differential methylation analysis

Differential methylation analysis between sample groups was conducted using the limma method (Smyth 2005) on individual CpG loci and genomic region level, including promoter, gene, island and tiling. CpG loci and genomic regions with $|\Delta\beta|>0.2$, $p\text{-value}<0.05$ were considered differentially methylated. *Appendix II* contains lists of differentially methylated CpG sites (presented in Chapter III) and genomic regions identified in gastric (CAM vs ATM, CAM vs NTM, ATM vs NTM) and oesophageal (CAM vs ATM) comparisons. The heatmap representations for the genomic regions analysis in gastric and oesophageal CAM vs ATM comparisons are shown in Supplementary Figure S3.3 and Figure S3.4, respectively.

### 2.6.3 Illumina HT-12 Expression data processing

The Illumina HumanHT-12v4 Expression BeadChip arrays were used to quantify gene expression profiles in gastric stromal myofibroblasts. RNA samples were processed commercially by Molecular Genetic Services (Gen-Probe Life Sciences, Manchester) who then provided raw signal intensity data extracted using Illumina GenomeStudio software. The Bioconductor package lumi (version 2.18.0), which is

designed for Illumina Expression Bead array data (Du, Kibbe et al. 2008), was used to import the raw data into R. Background correction, variance stabilization transformation (Lin, Du et al. 2008), robust spline normalization and subsequent quality control were then performed using this package (*Appendix III)*. To assess the reproducibility of sample preparation, a biological replicate of one of the samples was included ($R^2$= 0.9933573). Probes with a detection *p-value* < 0.01 across all samples were considered to be non-detectable and were removed from subsequent analysis. To further reduce the number of probes, un-annotated probes were also removed, thereby restricting subsequent analysis to 18090 probes (corresponding to 13381 genes). The Illumina HumanHT-12v4 Expression BeadChip arrays processing and analysis workflow is presented in Figure 2.1B.

### 2.6.4 Differential gene expression analysis

The Bioconductor package limma (Smyth 2005, Smyth, Michaud et al. 2005) was used to identify differentially expressed transcripts between relevant sample groups (in total 6 comparisons were performed). The Benjamini and Hochberg method (Benjamini and Hochberg 1995) was used to control for false discovery rate. Probes with *p-value* < 0.05 were considered to be differentially expressed. *Appendix IV* contains lists of differentially expressed genes identified in all 6 comparisons (CAM vs ATM, CAM vs NTM, ATM vs NTM, CAM hypoxia vs CAM normoxia, ATM hypoxia vs ATM normoxia and NTM hypoxia vs NTM normoxia).

### 2.6.5 Integration of DNA methylation and gene expression data

The results from differential gene expression analysis were integrated with results from differential methylation analysis as described in details in Chapter IV (section 4.3.4). Schematic representation of bioinformatics analysis workflow is shown in Figure 2.1.

### 2.6.6 Secretome data processing

Raw mass spectrometry data was analysed using MaxQuant software (Cox and Mann 2008) with Andromeda (Cox, Neuhauser et al. 2011) as a search engine against the reference human proteome (Uniprot proteome ID UP000005640, containing 42112 reviewed canonical and isoform protein entries, accessed April 2015) and 123 potential contaminants. Protein quantification was obtained using MaxQuant intensity-based label free quantification (LFQ) algorithm. Potential contaminants identified in the mass spectrometry data were manually examined and proteins associated with Fetal Bovine Serum (FBS) or trypsin were rejected. The LFQ algorithm extracts protein intensities form the raw data and performs appropriate normalization steps to allow between sample comparisons and remove any systematic errors during data acquisition. For simplicity, protein group was reported in all cases where a protein was identified by peptides that could not be assigned to individual protein unambiguously, for example in cases of highly homologous proteins, or protein isoforms. On average, each protein was quantified with 7 unique peptides and 22% sequence coverage was achieved. To compare CAM and NTM secretomes from different microenvironmental conditions four pair-wise comparisons were done using one-way analysis of variance (ANOVA) test as implemented in DanteR software (Taverner, Karpievitch et al. 2012). Proteins with a *p-value* < 0.05 were considered to be differentially secreted between the two conditions.

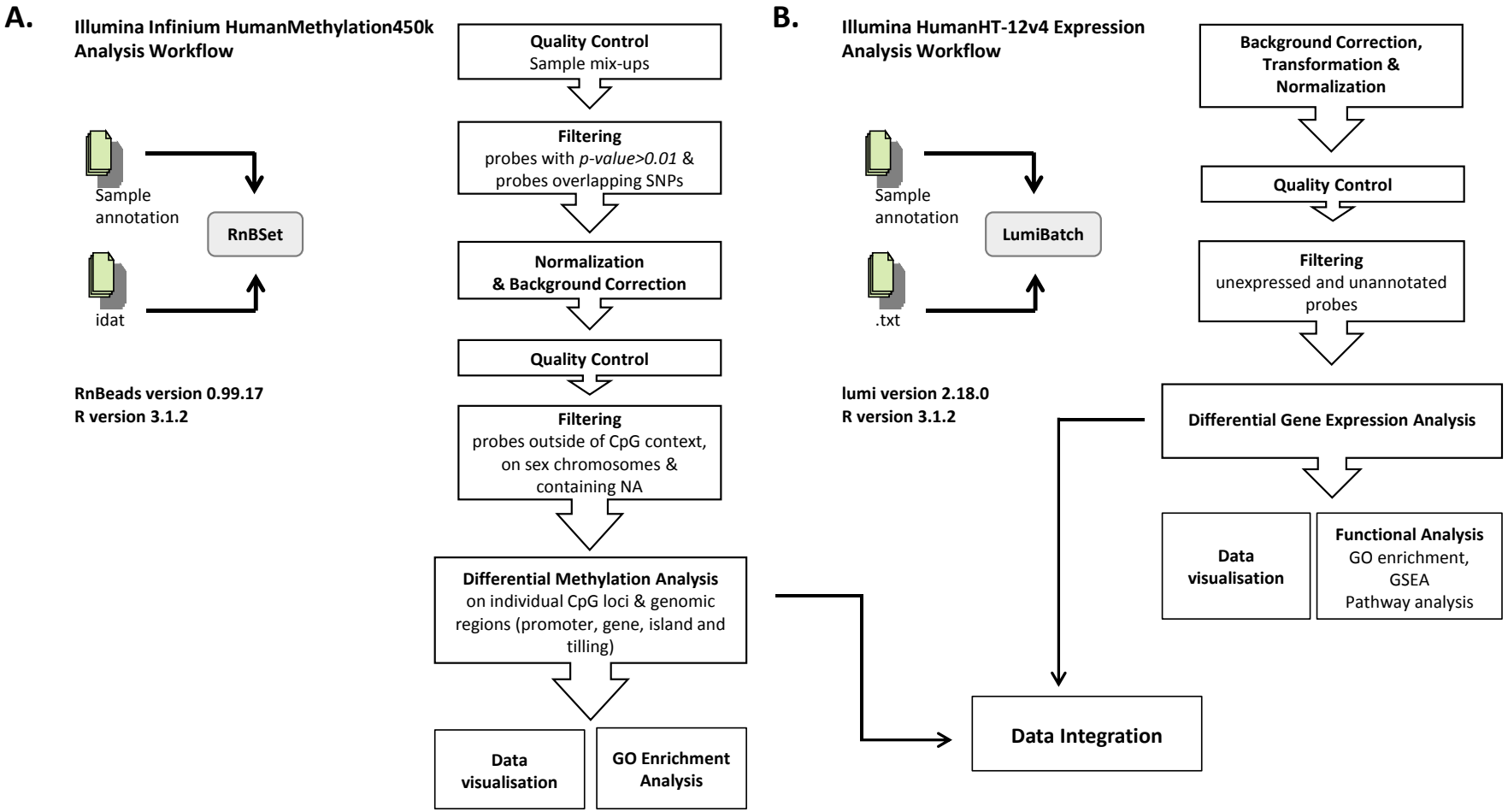


**Figure 2.1 Bioinformatics analysis workflow – integration of DNA methylation and gene expression data. A.** DNA methylation array data processing and analysis. **B.** Gene expression array data processing and analysis.

### 2.6.7 Integration of gene expression and secretome data

The results from differential gene expression analysis were integrated with results from differential secretome analysis as described in details in Chapter VI (section 6.3.4.4). Schematic representation of bioinformatics analysis workflow is shown in Figure 2.2.

### 2.6.8 R/Bioconductor

R statistical software (version 3.1.2) and Bioconductor (Gentleman, Carey et al. 2004, Huber, Carey et al. 2015) were used to process and analyse genomics, transcriptomics and proteomics data.

### 2.6.9 Ingenuity Pathway Analysis

Ingenuity Pathway Analysis (IPA) is a commercial software application for the analysis, integration and interpretation of data derived from omics experiments. IPA uses the manually curated Ingenuity Knowledge Base (IKB), which provides a comprehensive database for identifying regulators, relationships, mechanisms, functions and pathways relevant to changes observed in an analysed dataset. In this study, IPA was used to analyse and interpret data from comparative gene expression analysis and comparative secretome profiles. The Ingenuity Knowledge Base (IKB) was used to functionally annotate genes identified in several different analyses, including DNA methylation, gene expression and correlation analysis (additional details are provided in relevant result sections)

.

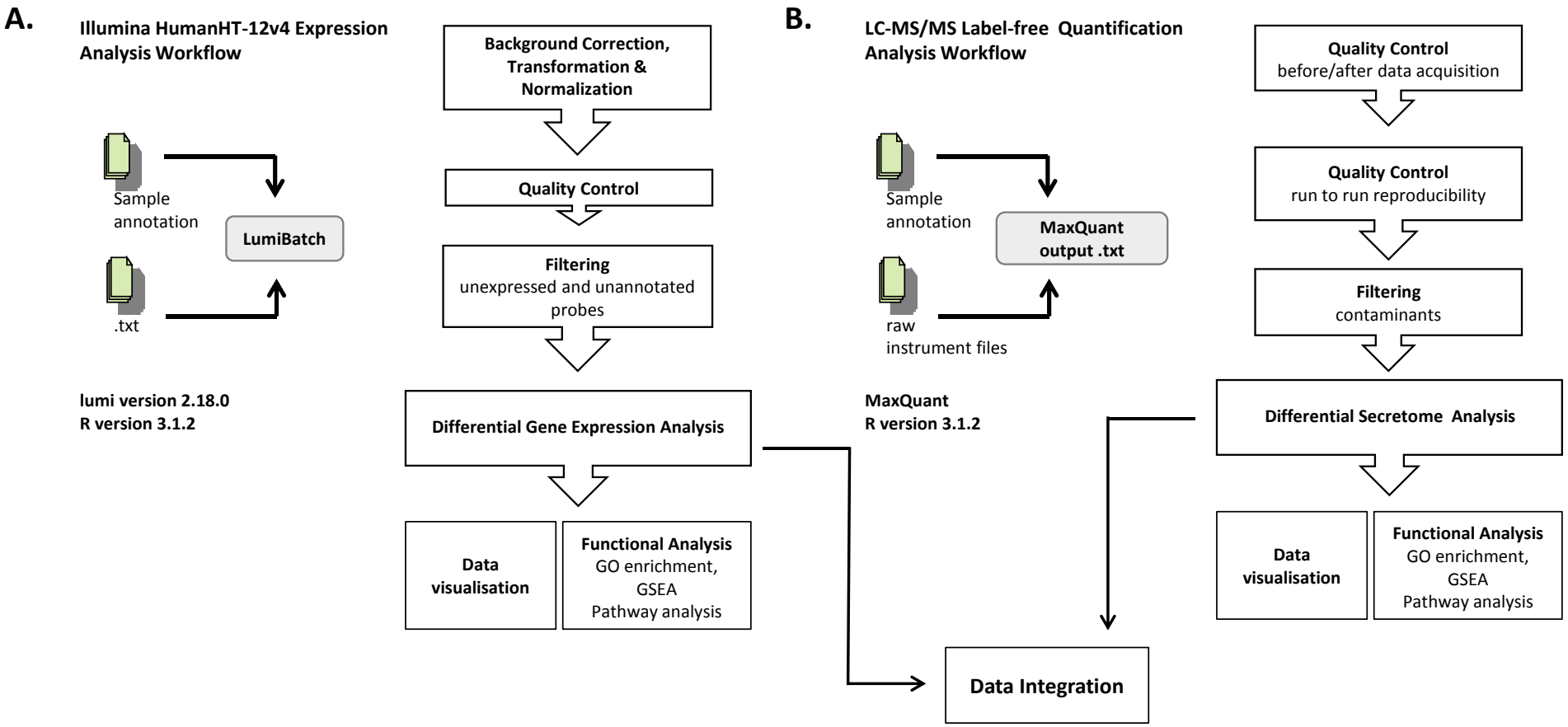


**Figure 2.2 Bioinformatics analysis workflow – integration of gene expression and secretome data. A.** Gene expression array data processing and analysis. **B.** Secretome data processing and analysis.

### 2.6.10 ConsensusPathDB

ConsensusPathDB (CPDB) (Kamburov, Pentchev et al. 2011) is a freely available database that integrates interaction networks in *Homo sapiens* that originate from 32 public resources, including KEGG (Kyoto Encyclopedia of Genes and Genomes) and Reactome. The ConsensusPathDB integrates protein-protein, biochemical, gene regulatory, genetic and drug-target interactions. In this study, ConsensusPathDB overrepresentation analysis was used to functionally annotate genes identified in several different analyses and to retrieve KEGG and Reactome pathways (additional details are provided in relevant result sections).

## 2.7 DNA methylation data analysis

### 2.7.1 Gene ontology enrichment analysis

Enrichment analysis on differentially methylated CpG loci in gastric and oesophageal CAM vs ATM comparisons was conducted using *gometh()* function from the Bioconductor package missMethyl (Maksimovic, Gordon et al. 2012, Phipson and Oshlack 2014). This function maps CpG loci to Entrez Gene IDs and tests for GO term enrichment using a hypergeometric test, taking into account the number of methylation probes per gene on the Illumina 450k array. Previously, Geeleher et al. showed that a severe bias exists when performing enrichment analysis for genome-wide DNA methylation data that occurs due to differing number of methylation probes per each gene (Geeleher, Hartnett et al. 2013). For differentially methylated CpG loci in gastric and oesophageal CAM vs ATM, the background was set to 424383 and 424355 probes, respectively.

### 2.7.2 Associating differentially methylated CpG loci with genes

The Genomic Regions Enrichment of Annotations Tool (GREAT version 3.0.0) (McLean, Bristor et al. 2010) was used to associate differentially methylated CpG loci identified in gastric and oesophageal CAM vs ATM comparisons with nearby genes and assign the distance to their transcription start sites. GREAT measures distances to transcription start site (TSS) using single TSS of a gene's canonical isoform. Briefly, lists of differentially methylated CpG loci were uploaded to GREAT and GRCh37/hg19 human genome assembly was used together with 'basal plus extension' as an association rule. The output lists of (i) CpG site to gene(s) and (ii) gene to CpG site(s) were used to integrate differentially methylated loci with gene expression profiles.

### 2.7.3 Pathway Analysis for common differentially methylated genes in gastric and oesophageal CAMs

Ingenuity Pathway Analysis (IPA) and ConsensusPathDB (CPDB) (Kamburov, Pentchev et al. 2011) overrepresentation analysis was used to assess biological relevance of common differentially methylated genes in gastric and oesophageal CAMs. The list of identified 2223 common genes was uploaded to IPA and CPDB. For IPA analysis the background list was set to Ingenuity Knowledge Base (IKB) whereas for CPDB analysis the default background was generated by CPDB based on the type of accession ID used for the input list.

## 2.8 Gene expression data analysis

### 2.8.1 Gene ontology enrichment analysis

Gene ontology (GO) enrichment analysis was performed on differentially expressed gene sets identified in CAM vs ATM comparison using GOrilla (Gene Ontology enRIchment anaLysis and visuaLizAtion tool) (Eden, Navon et al. 2009). Target genes were compared to background gene set of 13381 genes found to be expressed in myofibroblast cells. Enriched GO terms were identified using a hypergeometric model and *p-value < 0.001* threshold. The Benjamini and Hochberg method (Benjamini and Hochberg 1995) was applied to correct for multiple testing. Results were visualized using DAG (directed acyclic graph) trees in order to examine GO terms relations. The web server REVIGO (Supek, Bošnjak et al. 2011) was used to reduce redundancy within the identified enriched lists of GO terms.

### 2.8.2 Gene set enrichment analysis

Gene set enrichment analysis (GSEA v5.0) (Mootha, Lindgren et al. 2003, Subramanian, Tamayo et al. 2005) was performed on differentially expressed genes identified in CAM vs ATM and CAM vs NTM comparisons using the Molecular Signatures Database (MSigDB v5.0) and the hallmark gene set subcollection (Subramanian, Tamayo et al. 2005). The identified gene profiles were separated into two phenotypes for GSEA: CAM and ATM or CAM and NTM. For gene list ranking, multiple probes matching the same gene were sorted according to *p-value* and the probe with the lowest *p-value* was retained for the analysis. Genes were ranked using the provided signal-to-noise ranking statistic and GSEA was run using a default weighted enrichment statistics and evaluated for statistical significance by comparison to results obtained using 1 000 random permutations of each gene set. Default settings were used for all other GSEA parameters.

# 2.9 Comparative transcriptome and secretome analysis

## 2.9.1 Gene ontology enrichment analysis

Gene Ontology (GO) enrichment analysis (Khatri and Drăghici 2005) was performed on hypoxia-induced gene expression profiles identified in CAMs, ATMs and NTMs and hypoxia-induced CAM and NTM secretome profiles using GOrilla (Gene Ontology enRIchment anaLysis and visuaLizAtion tool) (Eden, Navon et al. 2009). For gene expression profiles, the target set of genes were compared to background gene set of all expressed and normalized genes from the HT-12v4 Illumina experiment whereas for secretome profiles, proteins were ranked according to their *p-value* with proteins identified as differentially secreted being at the top of the rank. The enriched GO terms in individual CAM, ATM, NTM hypoxia-induced gene expression profiles were discovered using a hypergeometric model whereas the enriched GO terms in ranked CAM and NTM differentially secreted protein profiles were discovered using mHG statistic (Eden, Lipson et al. 2007, Eden, Navon et al. 2009), $p$-value < 0.001 was applied and the results were corrected for multiple testing using the Benjamini and Hochberg method (Benjamini and Hochberg 1995). In order to examine GO terms relations the results were visualized using DAG (directed acyclic graph) trees. To reduce redundancy within the identified enriched lists of GO terms web server REVIGO (Supek, Bošnjak et al. 2011) was used.

## 2.9.2 Gene set enrichment analysis

Gene set enrichment analysis (GSEA v5.0) (Mootha, Lindgren et al. 2003, Subramanian, Tamayo et al. 2005) was used to characterize and interpret the identified unique- CAM, ATM or NTM and universal hypoxia-induced gene

expression profiles. The Molecular Signatures Database (MSigDB v5.0) (Subramanian, Tamayo et al. 2005) that contains collection of biologically predefined gene sets was used to determine statistically enriched gene sets present in the analysed gene expression data. The identified unique- CAM, ATM or NTM and universal hypoxia-induced gene expression profiles were separated into two phenotypes for GSEA: hypoxia and normoxia. For gene list ranking, multiple probes matching the same gene were sorted according to *p-value* and the probe with the lowest *p-value* was retained for the analysis. Genes were ranked using the provided signal-to-noise ranking statistic and GSEA was run using a default weighted enrichment statistics and evaluated for statistical significance by comparison to results obtained using 1 000 random permutations of each gene set. Default settings were used for all other GSEA parameters.

### 2.9.3 Ingenuity Pathway Analysis

Ingenuity Pathway Analysis (IPA) software was used to analyse and interpret the hypoxia-induced gene expression profiles identified in CAMs, ATMs and NTMs and hypoxia-induced CAM and NTM secretome profiles. Gene expression profiles were compared against IPA predefined Illumina HT-12v4 reference set whereas secretome profiles were compared against Ingenuity Knowledge Based (IKB). For gene expression data, IPA canonical pathway and downstream effects analyses were performed whereas for secretome data IPA downstream effects and upstream regulator analyses were performed. In each case *p-value* was assigned using Fisher's exact test which indicates the probability of overlap between the pathway/phenotype and input genes.

# 2.10 Pyrosequencing DNA methylation analysis

## 2.10.1 DNA extraction

DNA samples for pyrosequencing analysis were extracted from 7 patient-matched CAM and ATM samples and 4 unrelated NTM samples. DNA extraction was performed using the Wizard SV Genomic DNA Purification System (Promega, cat. no. A2360) following a modified version of the manufacturer's protocol. Briefly, myofibroblasts were seeded at the same density in 10cm$^2$ tissue culture dishes and grown to 80-90% confluence followed by 24 hours synchronisation in serum-free DMEM supplemented with 1% penicillin-streptomycin, 1% antibiotic-antimycotic and 1% nonessential amino acid solution. Cells were then washed with 1x PBS before being lysed in 150µl Wizard SV lysis buffer. Cell lysates were stored at -80°C prior to further processing. Cell lysates were thawed and transferred to Wizard SV minicolumns, which were centrifuged at 13,000g for 3 minutes. Column membranes were washed four times with Column Washed Solution (CWS) and centrifuged at 13,000g for 1 minute. Between the CWS washes, 2µl RNase A (final concentration 100µg/ml) solution was added to the membrane for 10 minutes. After the last CWS wash the membrane was dried by additional centrifugation at 13,000g for 2 minutes. DNA was then eluted in 250µl nuclease-free water and stored at -20°C. DNA quality and quantity was assessed by spectrophotometry at 260/280 nm.

## 2.10.2 Pyrosequencing analysis

For pyrosequencing DNA methylation analysis, 1µg of genomic DNA was treated with sodium bisulfite using the EZ DNA Methylation-Gold Kit (ZymoResearch, cat. no. D5005) following the manufacturer's protocol. Primers were designed using the Pyromark Assay Design Software 2.0 (Qiagen) and synthesized by Eurofins MWG

(Germany). A full list of assays, primer sequences and annealing temperatures is shown in Table 2.5. Pyrosequencing templates were prepared by PCR amplification using HotStarTaq Master Mix Kit (Qiagen, cat. no. 203603), 5μM biotinylated primer, 5-10μM non-biotinylated primer (corresponding to 1:1 or 1:2 ratio in Table 2.5), 5mM dNTPs (Qiagen, cat. no. 201900) and 3 μl (~60 ng) bisulfite-treated DNA. The PCR thermal profile consisted of initial denaturation at 95$^{o}$C for 5 min, followed by 40 cycles including 95$^{o}$C for 30 sec, annealing temperature (Table 2.5) for 30 sec, 72$^{o}$C for 30 sec. A final extension step of 72$^{o}$C for 10 min was also included. The quality of the PCR product was confirmed by 2% agarose gel electrophoresis before pyrosequencing analysis. After PCR, the biotinylated PCR product was purified and made single stranded to act as a template in a pyrosequencing reaction run. The PCR products were bound to streptavidin-coated Sepharose beads (GE Healthcare, cat. no. 17-5113-01), after which beads containing the immobilized PCR products were purified, washed, and denatured using a 0.2M NaOH solution. Thereafter, 0.5uM pyrosequencing primers were annealed to the purified single-stranded PCR product, and pyrosequencing was carried out using the Pyromark 96ID System (Qiagen). The methylation index for the analysed genomic region was calculated as the mean value of mC/(mC + C) for all examined CpG sites in the interrogated genomic region.

**Table 2.5 Nucleotide sequences of PCR primer sets and sequencing primers used for pyrosequencing DNA methylation assays;** *NA* – no CpG loci corresponding to Illumina 450k probe is interrogated by the pyrosequencing assay.

| Assay Name | Sequence 5' -> 3' | Primers ratio | Annealing temp | 450k probe ID |
|---|---|---|---|---|
| **FOXF1/ FENDRR** | R: 5'-biotin-CCCAAACCTATAACCTCCAC-3'<br>F: 5'-GGGAAAAATTTGAGAATAGATAG-3'<br>S: 5'-TTTGAGAATAGATAGGGG-3' | 1:2 | 54°C | *NA* |
| **MUC2** | R: 5'-biotin-CAAACCCTAAAACCTAATACTAAC-3'<br>F: 5'-TGTTGTTTAGATTTAGGGTTTG-3'<br>S: 5'-TTAGATTTAGGGTTTGGGA-3' | 1:2 | 54°C | cg06472341 |
| **ZNF536** | R: 5'-biotin-CAAAACTAACCTCAACAAAATC-3'<br>F: 5'-GTTTATAAGTGGTTTGTTGGG-3'<br>S: 5'-AAGTGGTTTGTTGGGAG-3' | 1:2 | 54°C | cg00386405 |
| **HOXA5** | R: 5'-GTCAAATTCCATACACTTTTATAACC-3'<br>F: 5'-biotin-GATGAAGATTTTTTAGGTTGGATA-3'<br>S: 5'-ATTCCATACACTTTTATAAC-3' | 1:2 | 57°C | cg08070327<br>cg14658493<br>cg25506432 |
| **mir802/ RUNX1** | R: 5'-CTCTACAACCTCTTATATCATTTCA-3'<br>F: 5'-biotin-GATGAATTTTTGTTATTGATTGTA-3'<br>S: 5'-AAAAATACCATATACCCATTA-3' | 1:2 | 54°C | cg00871610 |
| **NKAIN3/ ASPH** | R: 5'-biotin-AAAATAATTCTCTCTAACTCCTC-3'<br>F: 5'-GTATTTGTAGTTTTGGAGAAAG-3'<br>S: 5'-TTTGTAGTTTTGGAGAAAGT-3' | 1:2 | 52°C | cg07926952<br>cg13380112 |
| **FOXC1/ FOXF2** | R: 5'-ACCCAACTACCCTATCCC-3'<br>F: 5'-biotin-TGTGGTTTAGGATTTGTTATTAG-3'<br>S: 5'-ACCCAACTACCCTATCCC-3' | 1:1 | 55°C | cg10759602 |
| **SPON2** | R: 5'-ACCATCCCCACTCATCT-3'<br>F: 5'-biotin-TGTTTGATGTTTTGTTGTGG-3'<br>S: 5'-CATCCCCACTCATCTCA-3' | 1:1 | 53°C | cg23543318 |
| **CD47** | R: 5'-biotin-TAAACATTATTACCTATAAACACC-3'<br>F: 5'-GAATAATTTTGTTGGTGGG-3'<br>S: 5'-GGTGGGGATGTGTTGGATA-3' | 1:2 | 51°C | cg11741004 |
| **SMAD3 (p1)** | R: 5'-biotin-ATTTTCAAAAACTACTCCAAA-3'<br>F: 5'-GGGAGATTTTTGTTGTTAAA-3'<br>S: 5'-ATAAAGGGTTTAGATAT-3' | 1:1 | 52°C | cg23731272 |
| **SMAD3 (p2)** | R: 5'-CAACAACAAACCAATTAACA-3'<br>F: 5'-biotin-GAGGGGTTTGGAGTAGTT-3'<br>S: 5'-AACAACAAACCAATTAACAC-3' | 1:2 | 53°C | cg02486855 |
| **B4GALT6** | R: 5'-biotin-ACCTAAATTAATAAATCCCC-3'<br>F: 5'-TGAAGGTAAGTTTTGGTATAAG-3'<br>S: 5'-TAAGTTTTGGTATAAGG-3' | 1:2 | 52°C | cg02930996<br>cg07027513 |
| **VPS28** | R: 5'-biotin-AATTTATTCCCTAAAACCTACC-3'<br>F: 5'-GGGGAAGAGGGTAGATTT-3'<br>S: 5'-GGGGAAGAGGGTAGAT-3' | 1:1 | 51°C | cg11882377 |
| **LINE-1** | R: 5'-AACTCCCTAACCCCTTAC-3'<br>F: 5'-biotin-TAGGGAGTGTTAGATAGTGG-3'<br>S: 5'-CAAATAAAACAATACCTC-3' | 1:1 | 58°C | *NA* |

# 2.11 TaqMan gene expression analysis

## 2.11.1 RNA extraction

RNA samples for TaqMan assays were extracted from 6 patient-matched CAM and ATM samples. Total RNA was purified using miRNeasy Mini Kit (Qiagen, cat. no. 217004) according to modified version of manufacturer's instructions. Briefly, the cells were seeded at the same density in $10cm^2$ tissue culture dishes and grown to 80-90% confluence. Cells were then washed with 1x PBS before being lysed in 700ul QIAzol Lysis Reagent. Cell lysates ware collected into Eppendorf tubes and homogenized by vortexing for 1 minute. Homogenized lysates were then stored in -80°C prior to further processing. Total RNA was extracted by phase separation after addition of 140µl chloroform to homogenized lysates. Homogenates were centrifuged at 12,000 rcf, 4°C for 15 minutes and the upper aqueous phase was collected and mixed with 1.5 volumes of absolute ethanol. Samples were then transferred to the RNeasy Mini spin column and centrifuged at 8,000 rcf for 15 seconds followed by a 15 minutes on-column DNase I digestion (Qiagen, cat. no. 79254). The RNeasy Mini spin column membrane was washed twice with 500µl Buffer RPE and spin sequentially at 8,000 rcf for 15 seconds and 4 minutes. The column was then transferred to the new collection tube and centrifuged at full speed for 3 minutes to eliminate any possible carryover of Buffer RPE. Finally, RNA was eluted in 50µl RNase-free water. RNA samples quality and quantity was assessed by spectrophotometry at 260/280 nm and samples were stored at -80°C.

## 2.11.2 TaqMan qPCR analysis

TaqMan gene expression assays were used to quantify mRNA levels of target genes in patient-matched CAM and ATM samples. The QuantiTect Reverse Transcription Kit (Qiagen, cat. no. 205311) was used for cDNA synthesis. Briefly, following

removal of genomic DNA contamination, the reverse transcription reaction utilised an optimized mix of oligo-dT and random primers to convert 900ng of total RNA to cDNA in a total reaction of 20µl. The latter was then diluted with 80µl $ddH_2O$ to give 100µl cDNA working stock. This was subsequently used for the qPCR assays selected for the target genes. qPCR was performed on the StepOne system (Applied Biosystems). The amplification mixture was contained 7.5µl of 2x TaqMan Universal Master Mix II (Life Technologies, cat. no. 4440042), 0.75µl of 20x TaqMan probe and primers, 1.25µl 10x ACTB, 2µl of the cDNA and 3.5µl $ddH_2O$, giving a final volume of 15µl. Samples were analysed in triplicate, and experiments were repeated at least 2 times with independent RNA cell lysates. TaqMan assays were either designed using Oligo 7.0 software (Molecular Biology Insights Inc, USA) and synthesized by Eurofins MWG (Germany) or purchased as predesigned assays from Life Technologies (UK). A full list of assays, nucleotide sequences and PCR product sizes are shown in Table 2.6. The amplification mixtures were processed using standard conditions (50°C for 2 minutes and 95°C for 10 minutes followed by 45-50 cycles at 95°C for 15 seconds and 60°C or 61°C for 1 minute). β-actin was used as the endogenous control. The comparative ΔΔCt method was used to compute relative levels of target gene expression by subtracting the Ct values of the endogenous control (β-actin) and comparing values to a calibrator sample, where the calibrator sample = 1.0 and other samples were expressed as ***n***-fold relative to the calibrator.

**Table 2.6 List of TaqMan assays and nucleotide sequences used in this study.**

| Gene | TaqMan Assay ID | Amplicon Length (bp) | Forward Primer *5' -> 3'* | Reverse Primer *5' -> 3'* | Probe *5'-> 3'* |
|---|---|---|---|---|---|
| **DEPTOR** | Hs00224437_m1 | 82 | Proprietary sequence Life Technologies | Proprietary sequence Life Technologies | **(FAM)** |
| **ZNF536** | Hs01100020_m1 | 70 | Proprietary sequence Life Technologies | Proprietary sequence Life Technologies | **(FAM)** |
| **SULF1** | Hs00290918_m1 | 65 | Proprietary sequence Life Technologies | Proprietary sequence Life Technologies | **(FAM)** |
| **SULF2** | Hs01016476_m1 | 85 | Proprietary sequence Life Technologies | Proprietary sequence Life Technologies | **(FAM)** |
| **HOXA5** | Hs00430330_m1 | 127 | Proprietary sequence Life Technologies | Proprietary sequence Life Technologies | **(FAM)** |
| **FOXF1** | Hs00230962_m1 | 69 | Proprietary sequence Life Technologies | Proprietary sequence Life Technologies | **(FAM)** |
| **SMAD3** | Hs00969210_m1 | 87 | Proprietary sequence Life Technologies | Proprietary sequence Life Technologies | **(FAM)** |
| **SPON2** | Hs00202813_m1 | 104 | Proprietary sequence Life Technologies | Proprietary sequence Life Technologies | **(FAM)** |
| **FENDRR v1** | custom | 92 | AAGTGAAATACATGTAGATGGGAT | TGTGCCAAACTGAGTAAACC | CACCCTCTCTGGTCTTCAGTTTCTCA **(FAM)** |
| **FENDRR v2** | custom | 111 | GCTTCTGTCCAAGGCACT | CAAGCTTGCTAACTTCTTTGC | AGCCTACTCGTCAAAAGCCCGA **(TAMRA)** |
| **ACTB (TAMRA)** | custom | 128 | GGCACCCAGCACAATGAAG | CATACTCCTGCTTGCTGATCCA | CTCCTCCTGAGCGCAAGTACTCCGTG **(TAMRA)** |
| **ACTB (VIC)** | 4326315E | 171 | Proprietary sequence Life Technologies | Proprietary sequence Life Technologies | **(VIC)** |

# Chapter III

## Identification of Genome-Wide DNA Methylation Signatures in Gastric and Oesophageal Cancer-Associated Myofibroblasts

## 3.1 Introduction

While it is well established that epigenetic regulation plays a key role in cancer progression (Jones and Baylin 2002), little is known about the pattern or consequence of epigenetic programming in cancer stromal cells.

It was reported that cancer associated myofibroblasts (CAMs) isolated from the site of gastric tumours exhibit a global reduction in DNA methylation, when compared to adjacent tissue myofibroblasts (ATMs) (Jiang, Gonda et al. 2008). However, in contrast to cancer cells, CAM global DNA hypomethylation does not appear to be associated with chromosomal instability, as CAMs are non-neoplastic and continue to exhibit senescence *in vitro*. Other reports focus on gene – specific promoter methylation patterns in stromal and cancer cells, as reported for prostate tumour-associated stromal cells (Hanson, Gillespie et al. 2006, Rodriguez-Canales, Hanson et al. 2007) and breast tumour-associated stromal cells (Fiegl, Millinger et al. 2006). Also, it is not known to what extent common mechanisms or patterns of DNA methylation operate in CAMs derived from different tumours.

Significantly, primary CAMs have been shown to retain several cancer promoting properties following isolation (Orimo, Gupta et al. 2005). In particular, an ability to enhance cancer cell migration and proliferation, is commonly retained in low passage isolated gastric CAMs (Holmberg, Quante et al. 2012, Hu, Wang et al. 2013). This phenotype was also reconfirmed for all primary low passage myofibroblast cell lines used in this study (Supplementary Figure S3.1 and Figure S3.2). Given that these properties of isolated CAMs appear to be inherent, we hypothesise that CAMs may have been epigenetically programmed by cancer cells.

To assess genome-wide DNA methylation changes in stromal myofibroblasts, Illumina Infinium HumanMethylation450 BeadChip arrays were performed on a collection of primary gastric and oesophageal patient-matched CAM and ATM samples. The Illumina 450k array interrogates over 480,000 CpG loci distributed across the genome at single base resolution, providing coverage of 98.9% UCSC RefGenes with a global average of 17.2 probes per gene region (Bibikova, Barnes et al. 2011).

The Illumina 450k arrays for gastric and oesophageal myofibroblasts were performed independently by the Sanderson and Varro groups, respectively. Data processing and analysis for both Illumina 450k methylation arrays was conducted as part of this project.

## 3.2 Aims

- To compare global DNA methylation status of patient-matched CAM and ATM samples isolated from either gastric or oesophageal cancers
- To identify differential changes in DNA methylation in gastric and oesophageal myofibroblasts purified from different tissue microenvironments: CAM (cancer), ATM (adjacent tissue) and NTM (normal tissue)
- To identify genomic loci which DNA methylation patterns might serve as proxy for gastric CAM identification
- To investigate the extent to which common DNA methylation changes occur in gastric and oesophageal CAMs

# 3.3 Results

## 3.3.1 Comparison of global DNA methylation between CAMs and ATMs from gastric and oesophageal cancers

To assess the relative global DNA methylation status of CAMs and ATMs isolated from the site of gastric or oesophageal tumours, Illumina 450k probes that passed the filtering steps (as described in Methods section 2.6.1) were used to compute mean β-values (as define in Methods section 2.6.1). Analysis for gastric myofibroblasts was restricted to 424383 CpG sites whereas analysis for oesophageal myofibroblasts was restricted to 424355 CpG sites. Comparison of mean β-values in CAMs and ATMs from gastric and oesophageal cancers shows that in both cancer types, mean global DNA methylation is statistically lower in CAMs compared to mean global DNA methylation observed in corresponding ATMs (*Wilcoxon test p-value* $< 2.2 \times 10^{-16}$) (Figure 3.1).

## 3.3.2 Identification of differentially methylated loci in gastric CAMs, ATMs and NTMs

Genome–wide differential DNA methylation analysis between gastric (i) CAM vs ATM, (ii) CAM vs NTM and (iii) ATM vs NTM were performed at both individual CpG loci and genomic regions. Figure 3.2 summarises the results from differential DNA methylation analysis at the individual CpG loci, while results from differential DNA methylation analysis at genomic regions for CAM vs ATM is provided in Supplementary Figure S3.3.

This analysis identified numerous loci showing consistent differential DNA methylation in CAMs compared to ATMs (Figure 3.2A) or NTMs (Figure 3.2B). In total, 5688 differentially methylated CpG loci were identified in CAMs compared to ATMs, including 3404 hypomethylated and 2284 hypermethylated CpG loci. These differentially methylated loci were distributed throughout the genome with no obvious chromosomal bias (Figure 3.12). In the CAM vs NTM comparison, a total of 8104 CpG loci (including 4147 hypomethylated and 3957 hypermethylated in CAMs) were identified as differentially methylated whereas in the ATM vs NTM comparison only 4988 CpG loci were identified as differentially methylated (including 2078 hypomethylated and 2910 hypermethylated in ATMs).

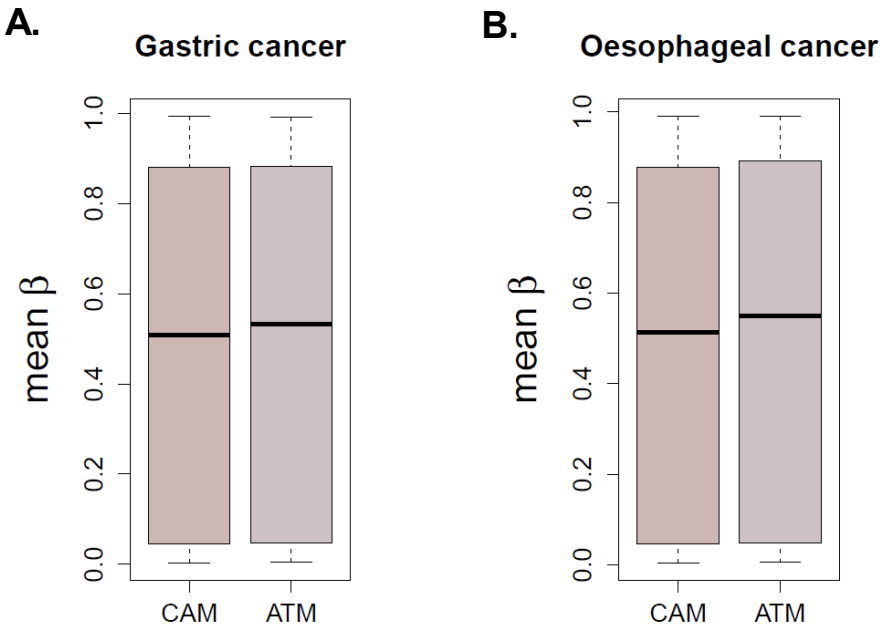


**Figure 3.1 Global DNA methylation of gastric and oesophageal patient-matched CAMs and ATMs.** Boxplots represent mean β-value for CAMs (n=3) and ATMs (n=3) isolated from: **A.** gastric cancer (mean β for 424383 CpG sites) and **B.** oesophageal cancer (mean β for 424355 CpG sites); *Wilcoxon test p-value* < $2.2 \times 10^{-16}$.

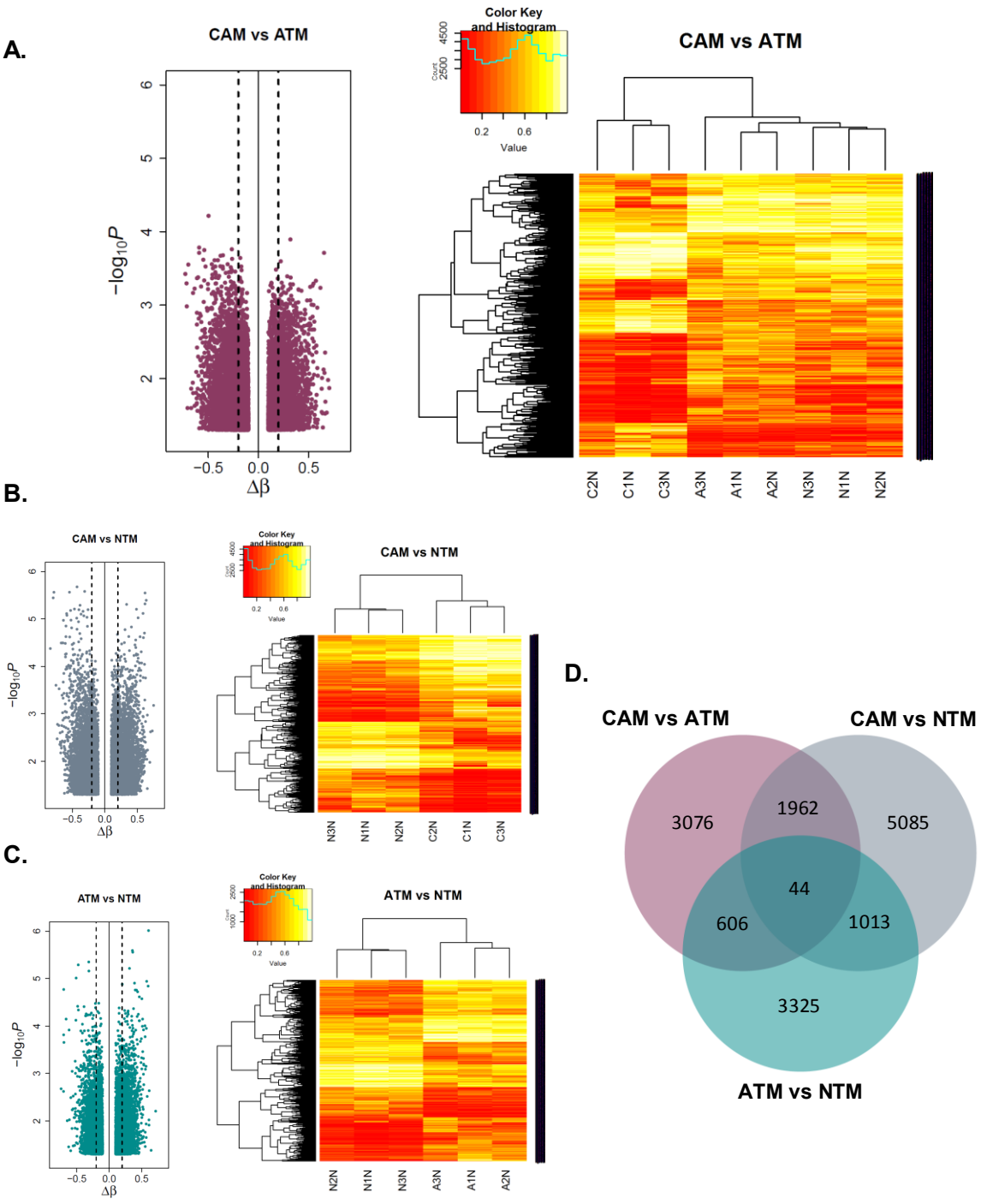


**Figure 3.2 Differentially methylated CpG loci in gastric myofibroblasts purified from different tissue microenvironments.** Volcano plots represent differentially methylated CpG loci |Δβ| > 0.1; dashed lines |Δβ| > 0.2, *p-value <0.05* identified in **A.** CAM vs ATM, **B.** CAM vs NTM and **C.** ATM vs NTM comparisons. Heatmaps represent differentially methylated CpG loci identified in respective comparisons |Δβ|>0.2, *p-value <0.05*. Heatmap shown in **A.** represents CpG loci identified in CAMs vs ATMs with projection including related values observed in NTMs. **D.** Venn diagram shows the relative overlap of identified differentially methylated CpG loci |Δβ| > 0.2, *p-value <0.05* in CAM vs ATM, CAM vs NTM and ATM vs NTM comparisons.

#### 3.3.2.1 Characterization of differentially methylated loci in gastric CAMs compared to patient-matched ATMs

As matched pairs of CAMs and ATMs were isolated from individual patients, these samples represent the closest matched controls that can be used for comparison of tumour-induced differential DNA methylation patterns.

In this analysis differentially methylated CpG loci were compared to the CpG distribution in the Illumina 450k array. The array covers 98.9% of annotated RefSeq genes and shows a wide distribution of probes among CpG islands, shores (2kb flanking the islands), shelves (2kb flanking the shores) and seas (regions outside the previous three categories) (Bibikova, Barnes et al. 2011). Hypomethylated and hypermethylated loci in CAMs were preferentially located in CpG shores rather than islands (Figure 3.3A). The comparison with overall distribution of differentially methylated loci relative to RefSeq genes showed that hypomethylated CpG loci are overrepresented in promoters, gene bodies and intergenic regions (Figure 3.3B). Notably, a significant proportion of differentially methylated loci in CAMs were located in enhancer regions (36.36% of all identified loci whereas only 21.9% of all probes on the array are annotated to enhancer regions), differentially methylated regions (DMRs) associated with reprograming (rDMRs; 5.43% of all identified loci whereas only 2.62% of all probes on the array are annotated to rDMRs) and cancer (cDMRs; 2.6% of identified loci whereas only 1.37% of all probes on the array are annotated to cDMRs).

Alteration of DNA methylation in one CpG locus has the potential to regulate multiple genes. Equally, expression of any gene may be regulated by methylation changes in multiple CpG loci. To associate identified differentially methylated loci with genes, the Genomic Regions Enrichment of Annotations Tool (GREAT version 3.0.0) (McLean, Bristor et al. 2010) was used. The analysis revealed that

26 CpG loci (including 19 hypomethylated and 7 hypermethylated) were not associated with any genes, 1807 CpG loci (including 1119 hypomethylated and 688 hypermethylated) were associated with only one gene, 3834 CpG loci (including 2250 hypomethylated and 1584 hypermethylated) were associated with 2 genes, 18 CpG loci were associated with 3 genes (including 15 hypomethylated and 3 hypermethylated), 1 hypermethylated CpG loci was associated with 4 genes, 1 hypomethylated CpG loci was associated with 5 genes and 1 hypermethylated loci was associated with 6 genes (Figure 3.4A). The analysis of the distance to transcription start site (TSS) of differentially methylated CpG loci associated with genes showed that most of the CpG loci are in region 50–500kb upstream of TSS and that hypomethylated CpG loci are slightly overrepresented in each TSS distance category (Figure 3.4B).

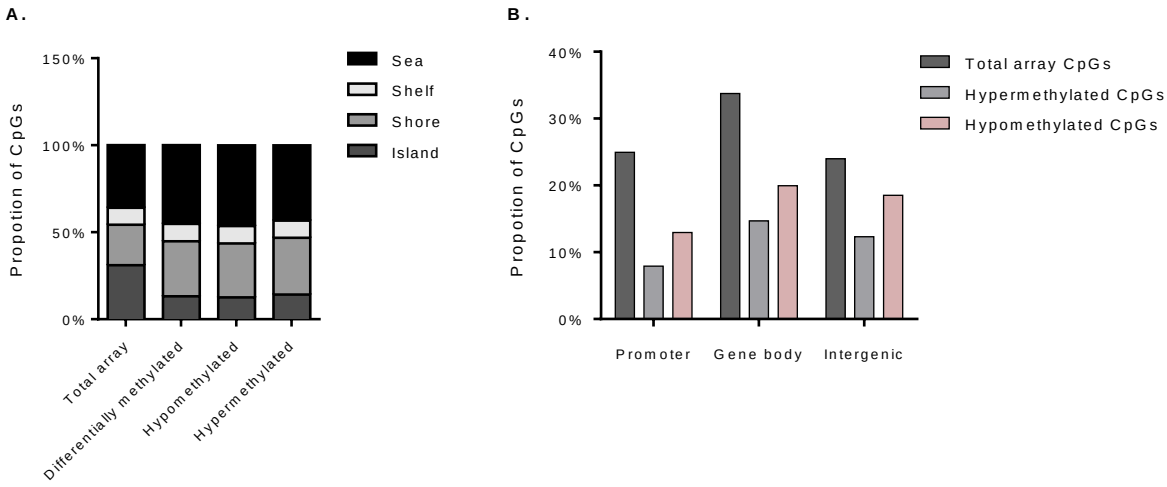


**Figure 3.3 Distribution of differentially methylated CpG loci identified in gastric CAMs vs ATMs. A.** in CpG islands, shores, shelves and sea regions or **B.** relative to RefSeq gene promoters, gene bodies and intergenic regions.

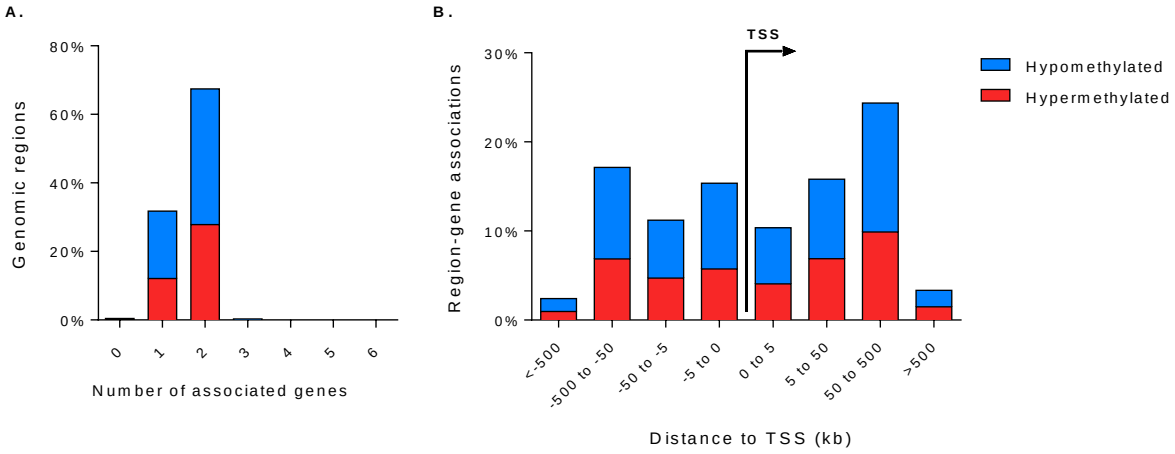


**Figure 3.4 Genomic region and gene associations of differentially methylated CpG loci in gastric CAMs vs ATMs. A.** Number of associated genes per identified CpG loci given in percentages (total 5688 CpG loci, including 3404 hypomethylated and 2284 hypermethylated in CAMs). **B.** Distance to transcription start site (TSS) of identified CpG loci that were associated with genes; *blue* – hypomethylated CpG loci in CAMs; *red* – hypermethylated sites in CAMs.

### 3.3.2.2 Technical validation of identified differentially methylated CpG loci

Pyrosequencing assays were used to validate genomic loci identified to be differentially methylated by Illumina 450k arrays in comparative analysis of gastric CAMs, ATMs and NTMs. Figure 3.5 shows correlations between β-values and [%] methylation assessed by pyrosequencing in the 9 samples (including 3 CAMs, 3 ATMs and 3 NTMs) from the array experiment. Correlations between the two types of DNA methylation assays for all 12 CpG loci interrogated were very good ($R^2$ = *0.8177 - 0.9921, p-value = 7.08852 x* $10^{-3}$ *– 1.43492 x* $10^{-7}$) thus increasing confidence in comparative differential DNA methylation trends identified in this study.

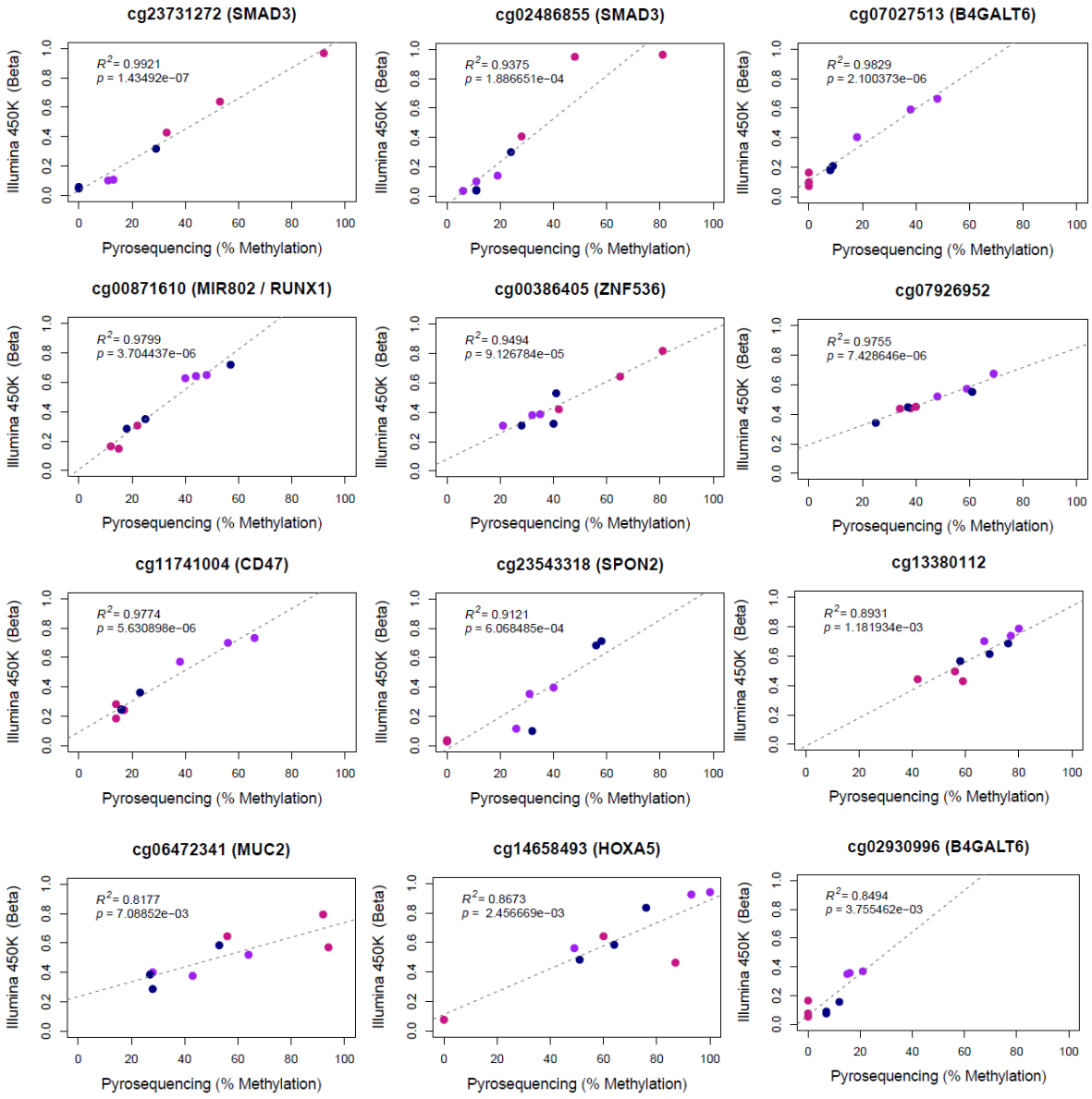


**Figure 3.5 Correlations between Illumina 450k array data and pyrosequencing analysis in gastric myofibroblasts;** *magenta–CAMs, purple–ATMs, navy–NTMs.* Representative data for 12 single CpG site identified as differentially methylated by Ilumina 450k array and validated by pyrosequencing analysis are shown.

### 3.3.2.3 Gene ontology enrichment analysis of differentially methylated loci identified in gastric CAMs compared to ATMs

It was recently suggested that popular gene-set enrichment and pathway analysis methods developed for use with gene expression data can yield spurious results when used for DNA methylation array data (Geeleher, Hartnett et al. 2013, Harper, Peters et al. 2013), as these tools do not account for the differential number of methylation probes per gene on the Illumina 450k array. Therefore, to investigate the potential biological relevance of identified differentially methylated CpG loci in gastric CAMs, gene ontology (GO) enrichment analysis was performed using *gometh()* function from the Bioconductor package missMethyl (Maksimovic, Gordon et al. 2012, Phipson and Oshlack 2014), as this function was designed explicitly for Illumina450k data, taking into account the number of probes per gene on the array. As the background a total of 424383 CpG sites were used.

The enrichment analysis of: (i) GO biological processes (BP) identified 120 GO terms, (ii) GO molecular functions (MF) identified 32 GO terms and (iii) GO cellular component (CC) identified 16 GO terms with *p-value < 0.01*. Table 3.1 shows a list of all GO terms with *FDR p-value < 0.05*. Other interesting GO biological process terms with *p-value < 0.01* include: cell differentiation (GO:0030154, 741 genes, $p = 9 \times 10^{-5}$), chemotaxis (GO:0006935, 176 genes, $p = 3.91 \times 10^{-4}$), transmembrane transport (GO:0055085, 276 genes, $p = 3.93 \times 10^{-4}$), regulation of Rab protein signal transduction (GO:0032483, 124 genes, $p = 1.33 \times 10^{-3}$), regulation of GTPase activity (GO:0043087, 124 genes $p = 1.33 \times 10^{-3}$), regulation of calcium ion transport (GO:0051924, 53 genes, $p = 1.89 \times 10^{-3}$), detection of glucose (GO:0051594, 4 genes, $p = 4.53 \times 10^{-3}$), synaptic transmission (GO:0007268, 178 genes, $p = 4.76 \times 10^{-3}$), regulation of cholesterol biosynthetic process (GO:0045540, 7 genes, $p = 7.24 \times 10^{-3}$), extracellular matrix organization (GO:0030198, 96 genes, $p = 8.18 \times 10^{-3}$).

**Table 3.1 Gene ontology (GO) enrichment analysis for differentially methylated CpG loci identified in gastric CAMs compared to ATMs.** *BP - biological process; CC - cellular component; N - number of genes in the GO term; DE - number of genes that are differentially methylated*

| GO ID | GO Term | Ontology | N | DE | *p-value* | FDR |
|---|---|---|---|---|---|---|
| GO:0007399 | nervous system development | BP | 1938 | 527 | 5.78E-10 | 1.12E-05 |
| GO:0022008 | neurogenesis | BP | 1327 | 374 | 6.56E-08 | 2.16E-04 |
| GO:0048699 | generation of neurons | BP | 1252 | 356 | 9.99E-08 | 2.42E-04 |
| GO:0030182 | neuron differentiation | BP | 1147 | 328 | 2.42E-07 | 5.21E-04 |
| GO:0007156 | homophilic cell adhesion via plasma membrane adhesion molecules | BP | 135 | 60 | 4.02E-07 | 7.80E-04 |
| GO:0048666 | neuron development | BP | 917 | 268 | 1.71E-06 | 2.55E-03 |
| GO:0048731 | system development | BP | 3783 | 866 | 2.14E-06 | 2.96E-03 |
| GO:0098742 | cell-cell adhesion via plasma-membrane adhesion molecules | BP | 187 | 70 | 4.32E-06 | 4.93E-03 |
| GO:0098609 | cell-cell adhesion | BP | 188 | 70 | 4.84E-06 | 5.21E-03 |
| GO:0032501 | multicellular organismal process | BP | 6240 | 1266 | 5.11E-06 | 5.22E-03 |
| GO:0007275 | multicellular organismal development | BP | 4319 | 961 | 8.46E-06 | 8.20E-03 |
| GO:0048468 | cell development | BP | 1843 | 467 | 8.92E-06 | 8.23E-03 |
| GO:0031175 | neuron projection development | BP | 793 | 234 | 1.24E-05 | 1.09E-02 |
| GO:0044707 | single-multicellular organism process | BP | 6019 | 1225 | 1.62E-05 | 1.37E-02 |
| GO:0048812 | neuron projection morphogenesis | BP | 616 | 192 | 2.00E-05 | 1.61E-02 |
| GO:0061564 | axon development | BP | 561 | 176 | 3.14E-05 | 2.34E-02 |
| GO:0048856 | anatomical structure development | BP | 4536 | 994 | 3.78E-05 | 2.71E-02 |
| GO:0007409 | axonogenesis | BP | 539 | 170 | 4.07E-05 | 2.80E-02 |
| GO:0071944 | cell periphery | CC | 4513 | 950 | 2.85E-09 | 2.77E-05 |
| GO:0005886 | plasma membrane | CC | 4425 | 930 | 4.47E-09 | 2.89E-05 |
| GO:0044425 | membrane part | CC | 6002 | 1169 | 4.03E-08 | 1.95E-04 |
| GO:0097458 | neuron part | CC | 920 | 265 | 6.67E-08 | 2.16E-04 |
| GO:0031224 | intrinsic component of membrane | CC | 5058 | 979 | 7.85E-08 | 2.17E-04 |
| GO:0016021 | integral component of membrane | CC | 4940 | 948 | 1.08E-06 | 1.91E-03 |
| GO:0042995 | cell projection | CC | 1503 | 384 | 1.26E-06 | 2.04E-03 |
| GO:0016020 | membrane | CC | 8174 | 1572 | 2.29E-06 | 2.96E-03 |
| GO:0043005 | neuron projection | CC | 750 | 220 | 2.56E-06 | 3.11E-03 |
| GO:0044459 | plasma membrane part | CC | 2248 | 495 | 2.29E-05 | 1.78E-02 |
| GO:0044295 | axonal growth cone | CC | 13 | 11 | 4.19E-05 | 2.80E-02 |

### 3.3.2.4 Identification of CpG loci that might serve as proxies for gastric CAM identification

Genome-wide and gene-specific methylation patterns may ultimately provide important signatures for distinguishing tumour derived myofibroblasts (CAMs) from non–tumour derived myofibroblasts (ATMs and/or NTMs). Identification of these genomic regions and DNA methylation patterns in CAMs may translate into new therapeutic strategies and prove useful as biomarkers for diagnosis and prognosis or tumour stratification. Therefore, CpG loci that distinguish CAMs from non-tumour derived myofibroblasts following differential methylation analysis at individual CpG loci in gastric CAM vs ATM ($|\Delta\beta|>0.2$) and CAM vs NTM ($|\Delta\beta|>0.2$) were compared (Figure 3.2D) and the overlap of 2006 CpG loci from these comparisons was selected to assess their methylation status. The analysis showed that these CpG loci have CAM-specific patterns of methylation as hypomethylated loci in CAMs were consistently found to be hypermethylated in both ATMs and NTMs. Conversely, hypermethylated loci in CAMs were consistently found to be hypomethylated in both ATMs and NTMs (Figure 3.6A). These loci are distributed across the genome (Figure 3.12). Notably, less stringent differential methylation analysis using a $|\Delta\beta|>0.1$ cut-off identified 3824 loci that may be used as proxies to distinguish CAMs from both ATMs and NTMs (data not shown). Figure 3.6B shows CpG loci that were selected in the same fashion from CAM vs ATM, CAM vs NTM and ATM vs NTM $|\Delta\beta|>0.1$ comparisons. As these CpG loci clearly show distinct patterns of DNA methylation in all studied myofibroblast populations, it is possible that these genomic regions may prove useful in distinguishing between different types of gastric myofibroblasts and different stages, or extents of tumour reprogramming.

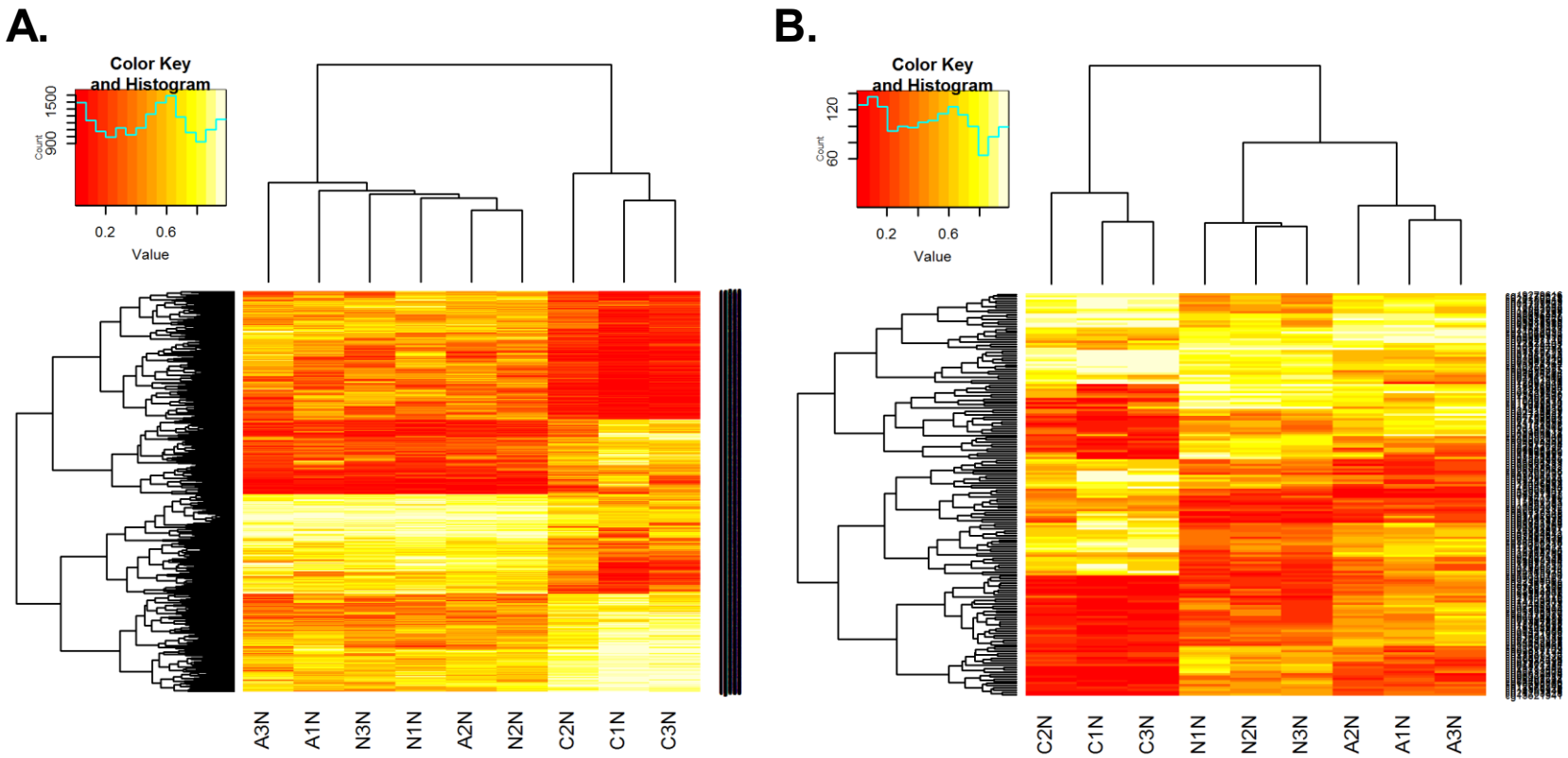


**Figure 3.6 Heatmap representations of CpG loci that may serve as proxies for gastric CAM, ATM and NTM identification. A.** CpG loci that have CAM–specific methylation pattern and were identified in CAM vs ATM and CAM vs NTM $|\Delta\beta|>0.2$ comparisons. **B.** CpG loci that have CAM- , ATM- and NTM- specific methylation pattern and were identified in CAM vs ATM, CAM vs NTM and ATM vs NTM $|\Delta\beta|>0.1$ comparisons.

#### 3.3.2.4.1 Characterization of identified CpG loci that might serve as proxies for gastric CAM identification

The complement of CpG loci which DNA methylation status was found to distinguish CAMs from non-tumour derived myofibroblasts (ATMs and NTMs) were compared to the CpG distribution in the Illumina 450k array and subjected to GREAT analysis in order to provide insight into their genomic localisation.

Hypomethylated loci were preferentially located in shelves rather than islands while hypermethylated loci were preferentially located in islands rather than shelves (Figure 3.7A). GREAT analysis revealed that 11 CpG loci (including 7 hypomethylated and 4 hypermethylated) were not associated with any genes, 609 CpG loci were associated with only one gene (including 350 hypomethylated and 259 hypermethylated), 1382 CpG loci were associated with 2 genes (including 782 hypomethylated and 600 hypermethylated), 3 CpG loci were associated with 3 genes (including 2 hypomethylated and 1 hypermethylated) and 1 hypermethylated CpG loci was associated with 6 genes (Figure 3.7B). The analysis of the distance to transcription start site (TSS) of the identified pool of CpG loci associated with genes showed that most of these CpG loci are located 50–500kb upstream of TSS and that hypomethylated CpG loci are slightly overrepresented in each TSS distance category (Figure 3.7C).

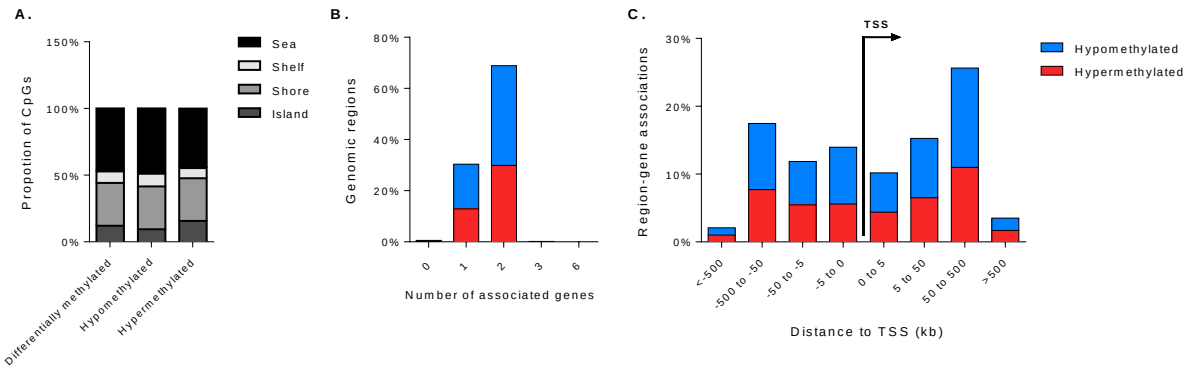


**Figure 3.7 Genomic region and gene associations of CpG proxies that might be used to distinguish gastric CAMs from ATMs and NTMs. A.** Distribution of identified CpG loci in CpG islands, shores, shelves and sea regions. **B.** Number of associated genes per identified CpG proxy given in percentages (total 2006 CpG loci, including 1141 hypomethylated and 865 hypermethylated). **C.** Distance to transcription start site (TSS) of identified CpG proxies; *blue* – hypomethylated CpG proxies; *red* – hypermethylated CpG proxies.

## 3.3.3 Identification of differentially methylated loci in oesophageal CAMs and ATMs

For comparative purposes genome–wide differential DNA methylation analysis was performed between oesophageal CAMs and patient-matched ATMs at both individual CpG loci and genomic regions. Figure 3.8 summarises the results from differential DNA methylation analysis at the individual CpG loci, while results from differential DNA methylation analysis at genomic regions is provided in Supplementary Figure S3.4.

In total, 3611 differentially methylated CpG loci were identified in oesophageal CAMs compared to ATMs, including 2826 hypomethylated and 785 hypermethylated CpG loci. Overall, differentially methylated loci were distributed throughout the genome.

#### 3.3.3.1 Characterization of identified differentially methylated loci in oesophageal CAMs compared to ATMs

Differentially methylated CpG loci were compared to the CpG distribution in the Illumina 450k array and subjected to GREAT analysis in order to assign genomic localisation.

In the case of oesophageal CAMs, hypomethylated loci were preferentially located in shelves and seas rather than islands whereas hypermethylated loci were preferentially located in CpG islands and shores (Figure 3.9A). The GREAT analysis of these data revealed that 45 CpG loci (including 44 hypomethylated and 1 hypermethylated) were not associated with any genes, 1206 CpG loci (including 967 hypomethylated and 239 hypermethylated) were associated with only one gene, 2351 CpG loci (including 1809 hypomethylated and 542 hypermethylated) were associated with 2 genes and 9 CpG loci (including 6 hypomethylated and 3 hypermethylated) were associated with 3 genes (Figure 3.9B). The analysis of the distance to transcription start site (TSS) of the differentially methylated CpG loci associated with genes showed that most of the CpG loci are located 50–500kb upstream of TSS and that hypomethylated CpG loci are overrepresented in each TSS distance category (Figure 3.9C).

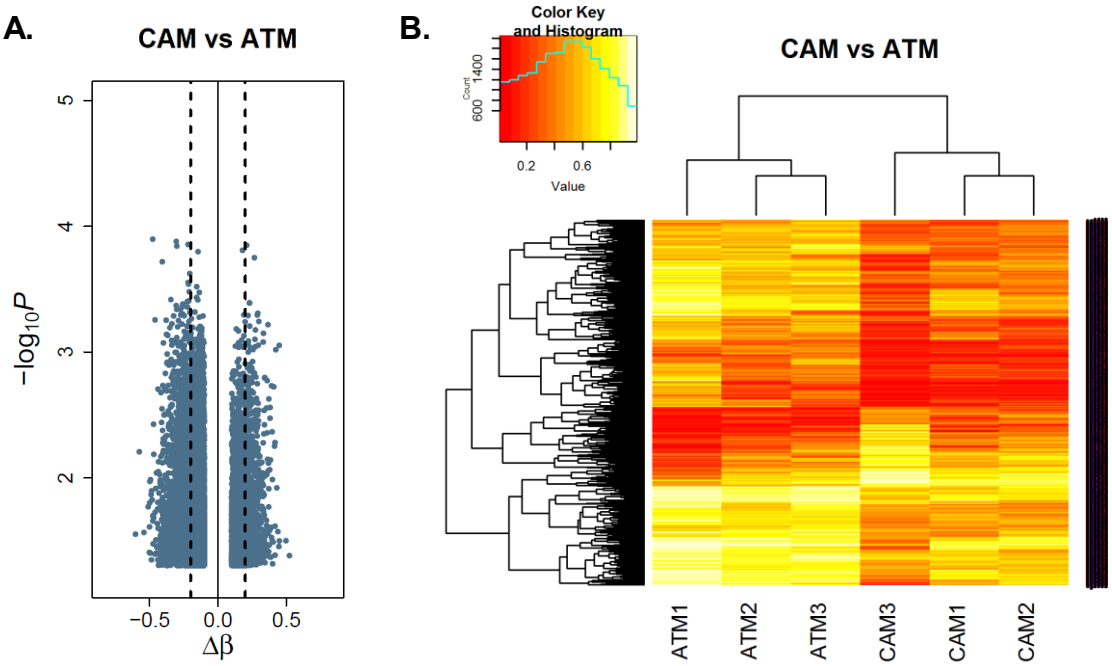


**Figure 3.8 Differentially methylated CpG loci identified in oesophageal CAM vs ATM comparison. A.** Volcano plot represents differentially methylated CpG sites |Δβ| > 0.1; dashed lines |Δβ| > 0.2, *p-value <0.05.* **B.** Heatmap represents differentially methylated CpG sites |Δβ|>0.2, *p-value <0.05.*

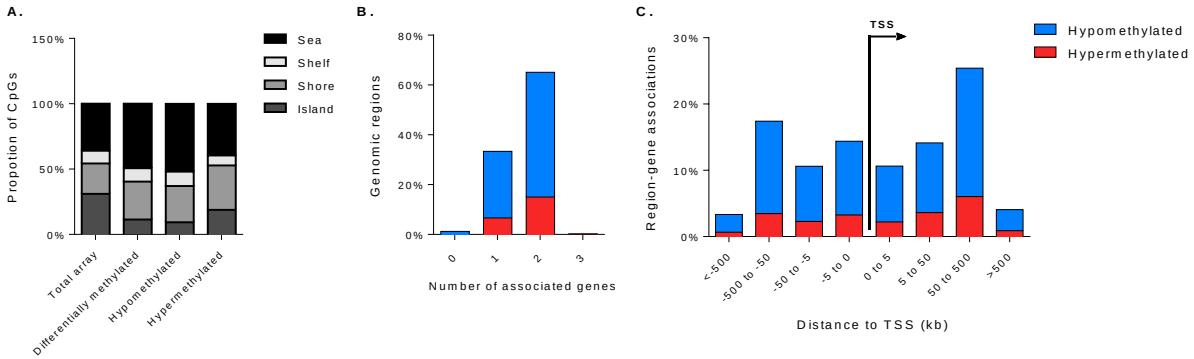


**Figure 3.9 Genomic region and gene associations of differentially methylated CpG loci in oesophageal CAMs vs ATMs. A.** Distribution of differentially methylated CpG loci in CpG islands, shores, shelves and seas. **B.** Number of associated genes per identified CpG site given in percentages (total 3611 sites, including 2826 hypomethylated and 785 hypermethylated in CAMs). **C.** Distance to transcription start site (TSS) of identified CpG loci that were associated with genes; *blue* – hypomethylated CpG sites in CAMs; *red* – hypermethylated sites in CAMs.

### 3.3.3.2 Gene ontology enrichment analysis for differentially methylated loci in oesophageal CAMs compared to ATMs

To investigate the potential biological relevance of identified differentially methylated CpG loci in oesophageal CAMs, gene ontology (GO) enrichment analysis was again performed using *gometh()* function from the Bioconductor package missMethyl (Maksimovic, Gordon et al. 2012, Phipson and Oshlack 2014), using a total of 424355 CpG sites as background.

The enrichment analysis of: (i) GO biological processes (BP) identified 116 GO terms, (ii) GO molecular functions (MF) identified 19 GO terms and (iii) GO cellular component (CC) identified 15 GO terms with *p-value < 0.01*. Table 3.2 shows list of all GO terms with *FDR p-value < 0.05*. Other interesting GO biological process terms with *p-value < 0.01* include: cell communication (GO:0007154, 741 genes, $p = 4.13 \times 10^{-5}$), regulation of signalling (GO:0023051, 415 genes, $p = 2.9 \times 10^{-4}$), regulation of signal transduction (GO:0009966, 371 genes, $p = 6.22 \times 10^{-4}$), synaptic transmission (GO:0007268, 124 genes, $p = 2.84 \times 10^{-3}$), cell differentiation (GO:0030154, 48 genes, $p = 2.88 \times 10^{-3}$), regulation of cytokine biosynthetic process (GO:0042035, 18 genes, $p = 4.20 \times 10^{-3}$) and negative regulation of secretion by cell (GO:1903531, 26 genes, $p = 5.18 \times 10^{-3}$).

**Table 3.2 Gene ontology (GO) enrichment analysis for differentially methylated CpG loci identified in oesophageal CAMs compared to ATMs.** *BP - biological process; MF - molecular function; CC - cellular component; N - number of genes in the GO term; DE - number of genes that are differentially methylated*

| GO ID | GO Term | Ontology | N | DE | *p-value* | FDR |
|---|---|---|---|---|---|---|
| GO:0007156 | homophilic cell adhesion via plasma membrane adhesion molecules | BP | 135 | 53 | 3.85E-10 | 7.46E-06 |
| GO:0098609 | cell-cell adhesion | BP | 188 | 57 | 1.43E-07 | 9.21E-04 |
| GO:0098742 | cell-cell adhesion via plasma-membrane adhesion molecules | BP | 187 | 57 | 1.29E-07 | 9.21E-04 |
| GO:0044707 | single-multicellular organism process | BP | 6017 | 822 | 1.21E-06 | 5.89E-03 |
| GO:0032501 | multicellular organismal process | BP | 6238 | 843 | 1.89E-06 | 7.32E-03 |
| GO:0022610 | biological adhesion | BP | 1303 | 227 | 6.61E-06 | 2.13E-02 |
| GO:0007155 | cell adhesion | BP | 1297 | 224 | 1.60E-05 | 3.45E-02 |
| GO:0048731 | system development | BP | 3783 | 574 | 2.50E-05 | 4.85E-02 |
| GO:0005509 | calcium ion binding | MF | 645 | 123 | 1.34E-05 | 3.25E-02 |
| GO:0071944 | cell periphery | CC | 4511 | 614 | 1.25E-05 | 3.25E-02 |

### 3.3.4 Identification of conserved DNA methylation patterns in gastric and oesophageal CAMs

To identify patterns of DNA methylation that are retained in isolated gastric and oesophageal CAMs the results from differential methylation analysis in gastric and oesophageal CAM vs ATM comparisons ($|\Delta\beta|>0.2$, *p-value* $< 0.05$) were compared at both individual CpG loci and gene level.

#### 3.3.4.1 Conserved methylation pattern at the individual CpG loci in gastric and oesophageal CAMs

Analysis of conserved differentially methylated CpG loci identified 230 susceptible CpG sites that are differentially methylated in both gastric and oesophageal CAMs when compared to patient-matched ATMs (Figure 3.10A). Notably, 172/230 of the identified CpG loci were changed in the same direction in gastric and oesophageal CAMs (Figure 3.10B). The conserved gastric and oesophageal CAM susceptible loci are distributed across the genome (Figure 3.12) and 65.22% of these loci are associated with 2 genes and 33.5% are associated with only one gene.

Many studies have shown that methylation changes at only one CpG loci can perturb gene expression and introduce functional changes (Sohn, Park et al. 2010, Claus, Lucas et al. 2012). Figure 3.11 shows representative CpG loci that have universal methylation pattern in gastric and oesophageal CAMs compared to their patient-matched ATMs. Expression changes of genes associated with these CpG loci have been linked to gastric and gastroesophageal cancer (*EHMT2, FOXF1, HS3ST6, PCDHGA8, PLAU, RGMA*), extracellular matrix disruption (*PLAU*), extracellular matrix organization (*COL9A3*) and hedgehog signalling pathway (*DHH, FOXF1)*.

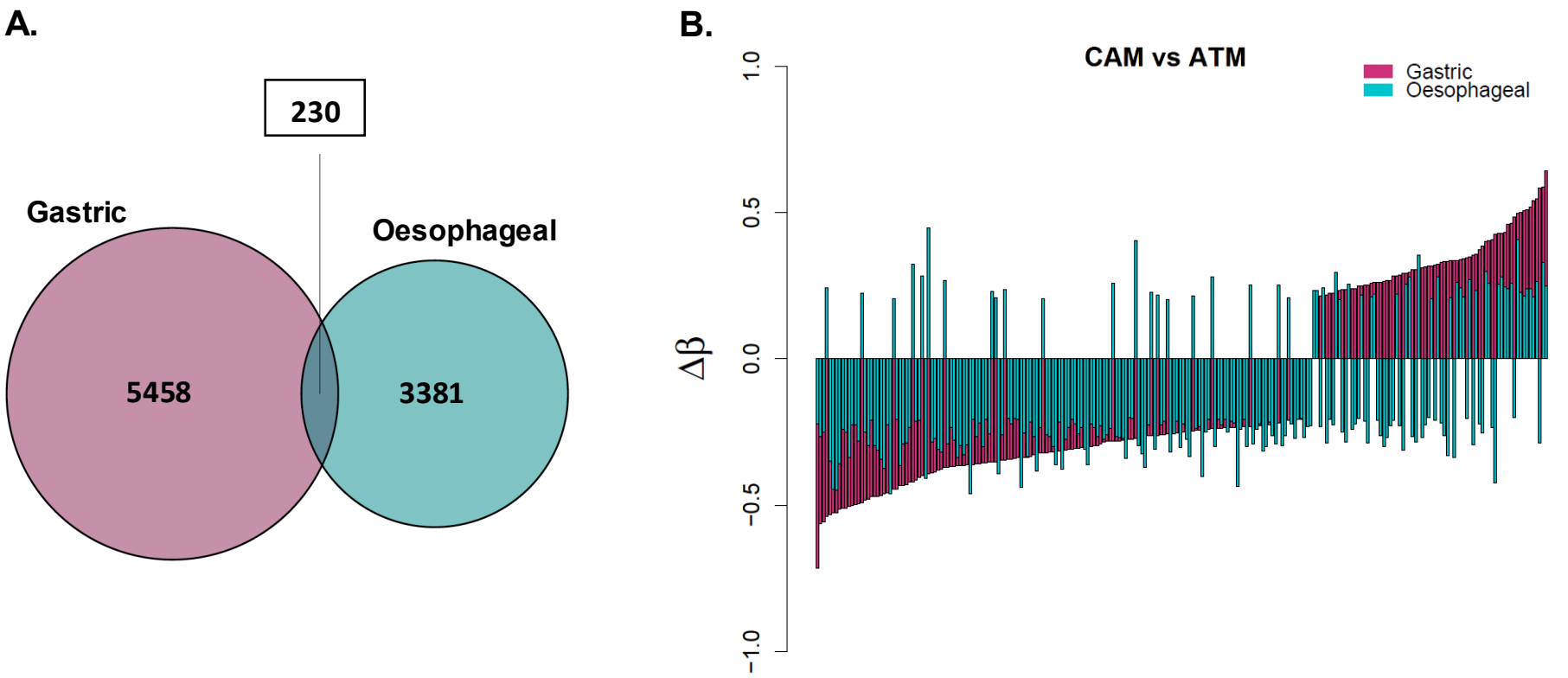


**Figure 3.10 Overlap of differentially methylated CpG sites identified in gastric and oesophageal CAMs vs ATMs.** **A.** Area-proportional Venn diagram representation of differentially methylated CpG loci identified in gastric and oesophageal CAMs**. B.** Bar chart representation of CAM susceptible loci differentially methylated in both gastric and oesophageal CAMs; |Δβ|>0.2, *p-value < 0.05*.

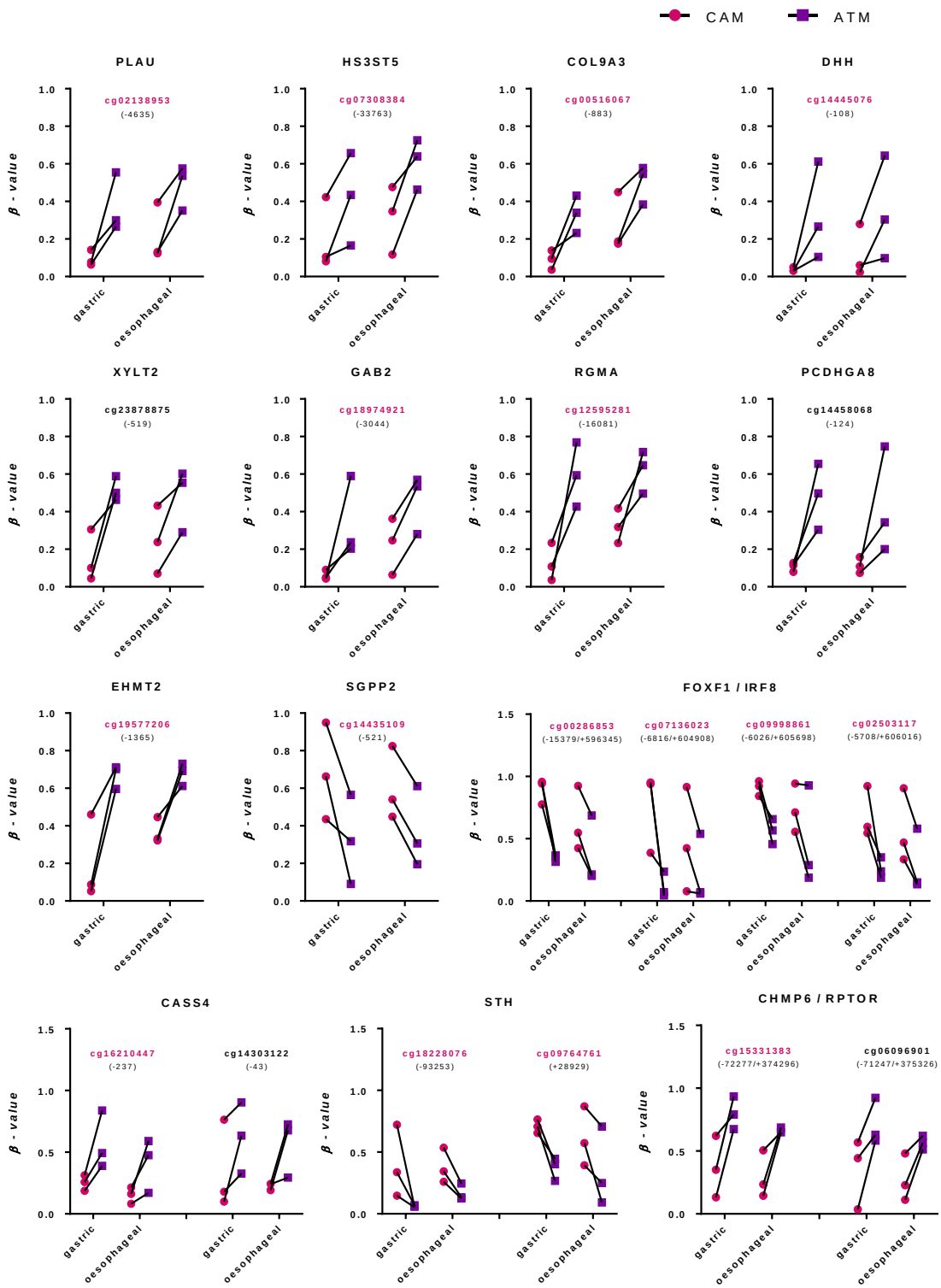


**Figure 3.11 Representative conserved DNA methylation patterns in gastric and oesophageal patient-matched CAM and ATM samples**. Probes highlighted in magenta are also identified as proxies for gastric CAMs. Numbers in brackets indicate the distance to transcription start site (TSS) of a given gene (indicated at the top of each plot); *magenta – CAMs, purple – ATMs;* |Δβ|>0.2, *p-value <0.05.*

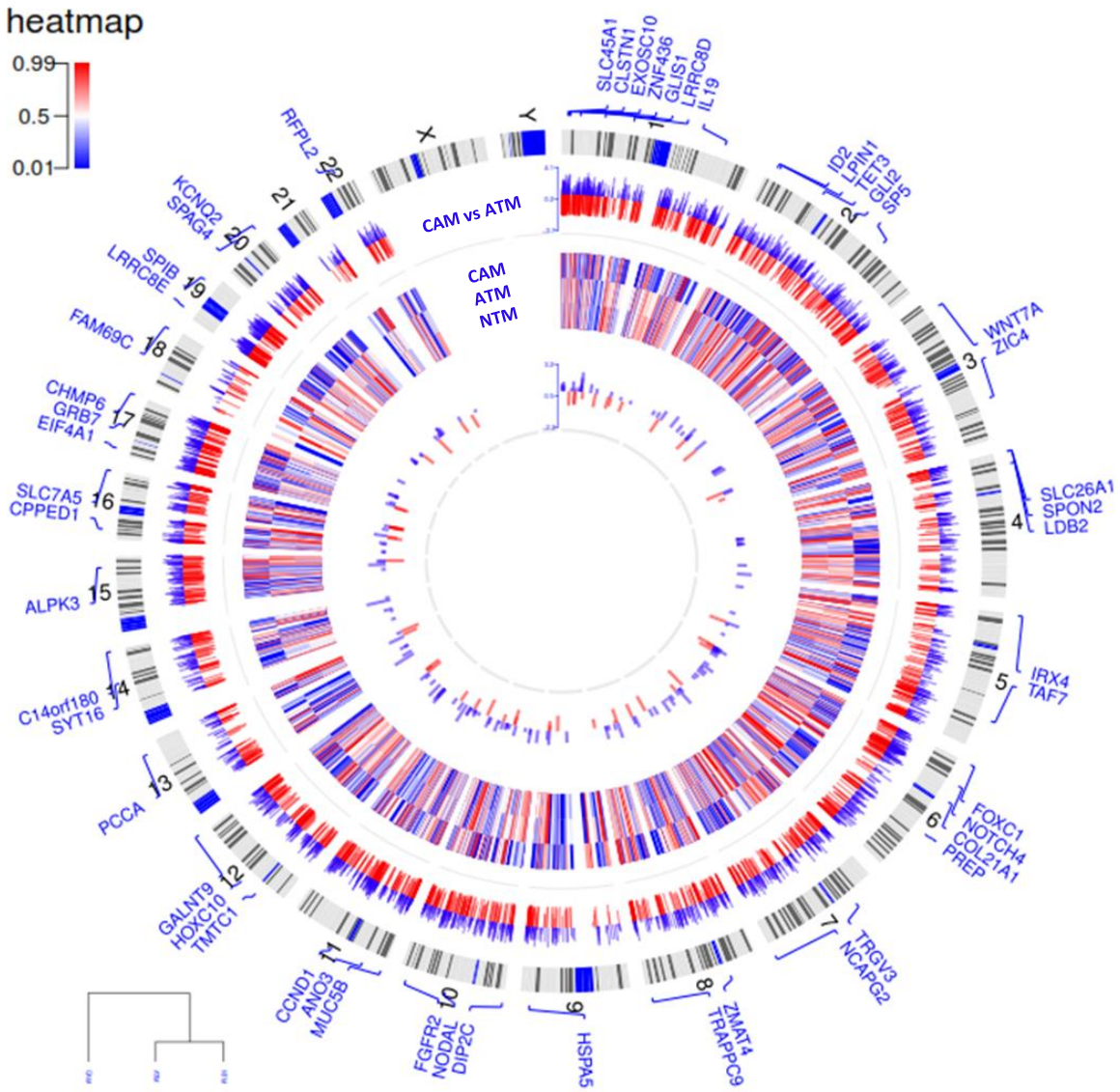


**Figure 3.12 Genome-wide overview of DNA methylation changes in stromal myofibroblasts.** The outer ring represents human ideograms. The first tract shows differentially methylated CpG loci between gastric CAMs and patient-matched ATMs. The heatmap represents CpG loci that might serve as proxies for gastric CAM identification. The inner track represents conserved methylation patterns identified in gastric and oesophageal CAMs compared to patient-matched ATMs; *red* - hypermethylated loci in CAMs, *blue* - hypomethylated loci in CAMs; |Δβ|>0.2, *p-value <0.05.*

### 3.3.4.2 Identification of common differentially methylated genes in gastric and oesophageal CAMs

To identify common differentially methylated genes in gastric and oesophageal CAMs, differentially methylated CpG loci in both sets of CAMs were assigned to genes as described in sections 3.3.2.1 and 3.3.3.1 and respective gene lists were compared. Differentially methylated CpG loci from gastric CAMs were assigned to 5918 genes whereas differentially methylated oesophageal CAM CpG loci were assigned to 4105 genes. Comparison of these gene lists identified 2223 common genes, which may be regulated by DNA methylation in both gastric and oesophageal CAMs (Figure 3.13).

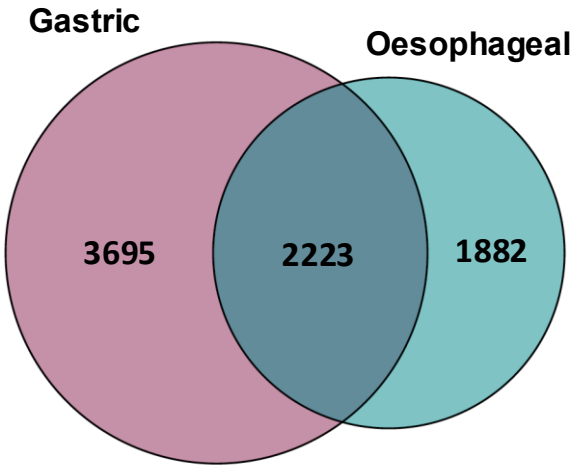


**Figure 3.13 Differentially methylated genes in gastric and oesophageal CAMs.** Area-proportional Venn diagram of genes that were associated with differentially methylated CpG loci identified in gastric and oesophageal CAM vs ATM comparisons; $|\Delta\beta|>0.2$, *p-value <0.05.*

#### 3.3.4.2.1 Universal pathways and processes affected by DNA methylation changes in both gastric and oesophageal CAMs

To investigate the extent to which DNA methylation changes in gastric and oesophageal CAMs may affect common pathways and processes, the 2223 commonly differentially methylated genes were subjected to Ingenuity Pathway Analysis (IPA) and ConsensusPathDB (Kamburov, Pentchev et al. 2011) over-representation analysis. KEGG and Reactome pathways with *p-value < 0.01* are listed in Table 3.3.

**Table 3.3 KEGG and Reactome pathways affected by DNA methylation changes in gastric and oesophageal CAMs.**

| Source | Pathway | *p-value* | FDR |
|---|---|---|---|
| KEGG | Hippo signalling pathway | 1.28E-06 | 3.24E-04 |
| | Type I diabetes mellitus | 8.69E-05 | 9.13E-03 |
| | PI3K-Akt signalling pathway | 1.09E-04 | 9.13E-03 |
| | Glycosaminoglycan biosynthesis - heparan sulfate / heparin | 2.50E-04 | 1.14E-02 |
| | Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate | 2.55E-04 | 1.14E-02 |
| | Signalling pathways regulating pluripotency of stem cells | 2.70E-04 | 1.14E-02 |
| | Inflammatory bowel disease (IBD) | 3.26E-04 | 1.17E-02 |
| | Cell adhesion molecules (CAMs) | 4.05E-04 | 1.27E-02 |
| | Pathways in cancer | 8.66E-04 | 2.42E-02 |
| | Rap1 signalling pathway | 1.21E-03 | 3.05E-02 |
| | Circadian entrainment | 2.33E-03 | 5.33E-02 |
| | Allograft rejection | 3.28E-03 | 6.41E-02 |
| | MAPK signalling pathway | 3.40E-03 | 6.41E-02 |
| | Long-term depression | 3.56E-03 | 6.41E-02 |
| | Proteoglycans in cancer | 4.26E-03 | 7.15E-02 |

**Table 3.3 (*Continued*) KEGG and Reactome pathways affected by DNA methylation changes in gastric and oesophageal CAMs.**

| Source | Pathway | *p-value* | FDR |
|---|---|---|---|
| KEGG | Hedgehog signalling pathway | 6.30E-03 | 9.13E-02 |
| | Tight junction | 6.56E-03 | 9.13E-02 |
| | Intestinal immune network for IgA production | 7.00E-03 | 9.13E-02 |
| | Phagosome | 7.21E-03 | 9.13E-02 |
| | Wnt signalling pathway | 7.25E-03 | 9.13E-02 |
| | Inflammatory mediator regulation of TRP channels | 9.73E-03 | 1.17E-01 |
| Reactome | Developmental Biology | 1.39E-06 | 7.14E-04 |
| | Transcriptional regulation of pluripotent stem cells | 1.04E-05 | 7.14E-04 |
| | Glycosaminoglycan metabolism | 2.12E-05 | 7.14E-04 |
| | POU5F1 (OCT4), SOX2, NANOG repress genes related to differentiation | 7.78E-05 | 2.53E-03 |
| | Axon guidance | 1.22E-04 | 3.82E-03 |
| | HS-GAG biosynthesis | 6.43E-04 | 1.95E-02 |
| | Diseases of glycosylation | 1.51E-03 | 4.36E-02 |
| | Signalling by FGFR mutants | 1.53E-03 | 4.36E-02 |
| | Regulation of Rheb GTPase activity by AMPK | 2.10E-03 | 5.81E-02 |
| | IRS-related events triggered by IGF1R | 3.06E-03 | 8.21E-02 |
| | Fatty acid, triacylglycerol, and ketone body metabolism | 3.52E-03 | 9.17E-02 |
| | IGF1R signalling cascade | 4.63E-03 | 1.14E-01 |
| | Signalling by Type 1 Insulin-like Growth Factor 1 Receptor (IGF1R) | 4.63E-03 | 1.14E-01 |
| | Chondroitin sulfate/dermatan sulfate metabolism | 5.26E-03 | 1.26E-01 |
| | Translocation of GLUT4 to the plasma membrane | 5.42E-03 | 1.27E-01 |
| | Signalling by FGFR1 mutants | 5.56E-03 | 1.27E-01 |
| | Regulation of lipid metabolism by Peroxisome proliferator-activated receptor alpha (PPARalpha) | 6.84E-03 | 1.48E-01 |
| | Chondroitin sulfate biosynthesis | 6.84E-03 | 1.48E-01 |
| | Signalling by NOTCH | 7.24E-03 | 1.53E-01 |
| | Energy dependent regulation of mTOR by LKB1-AMPK | 8.36E-03 | 1.65E-01 |
| | EPH-Ephrin signalling | 9.22E-03 | 1.65E-01 |
| | Hedgehog ,off, state | 9.34E-03 | 1.65E-01 |
| | Hemostasis | 9.64E-03 | 1.65E-01 |

Significantly, IPA analysis showed that these commonly differentially methylated genes are involved in digestive organ tumour ($p = 1.67 \times 10^{-17}$), expression of RNA ($p = 3.46 \times 10^{-12}$), transcription ($p = 7.36 \times 10^{-12}$), gastroesophageal cancer ($p = 7.94 \times 10^{-11}$), tumorigenesis of the tissue ($p = 3.13 \times 10^{-11}$), cell movement ($p = 1.83 \times 10^{-10}$), migration of cells ($p = 2.18 \times 10^{-9}$), proliferation of cells ($p = 1.28 \times 10^{-8}$), invasion of cells ($p = 3.43 \times 10^{-6}$), generation of fibroblasts ($p = 1.69 \times 10^{-5}$) and growth of tumour ($p = 4.97 \times 10^{-5}$).

The GO biological process (BP) terms enriched in gastric and oesophageal differentially methylated CpG loci as described in sections 3.3.2.3 and 3.3.3.2 respectively were compared to get further insight into common biological processes affected by DNA methylation changes in both gastric and oesophageal CAMs. A list of 120 GO BP terms (*p-value < 0.01*) identified in the gastric CAM vs ATM comparison (|Δβ|>0.2*, p-value < 0.05*) and 116 GO BP terms (*p-value < 0.01*) identified in the oesophageal CAM vs ATM comparison (|Δβ|>0.2*, p-value < 0.05*) were compared, resulting in the identification of 32 unifying GO biological processes (Figure 3.14) which are presented in Table 3.4.

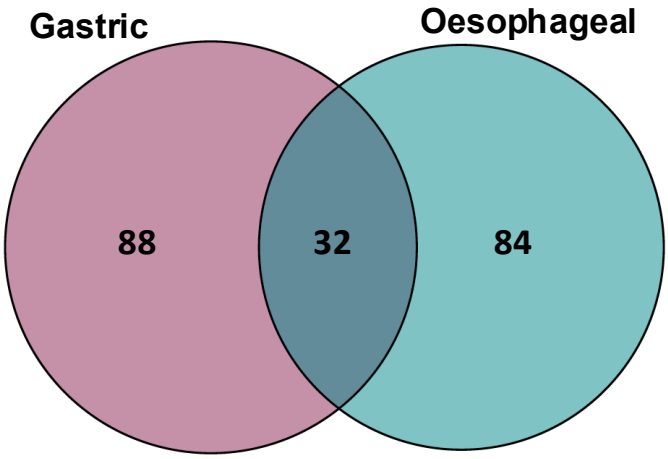


**Figure 3.14 Comparison of gene ontology (GO) biological process (BP) terms identified for gastric and oesophageal differentially methylated CpG loci** in respective CAM vs ATM comparisons (|Δβ|>0.2, *p-value <0.05).* Area-proportional Venn diagram; GO_BP *p-value<0.01*.

**Table 3.4 Common gene ontology (GO) biological process (BP) terms associated with gastric and oesophageal differentially methylated loci in CAMs compared to patient–matched ATMs.** *N - number of genes in the GO term; DE - number of genes that are differentially methylated*

| | | Gastric | | | | Oesophageal | | | |
|---|---|---|---|---|---|---|---|---|---|
| **GO ID** | **GO Term** | **N** | **DE** | ***p-value*** | **FDR** | **N** | **DE** | ***p-value*** | **FDR** |
| GO:0007399 | nervous system development | 1938 | 527 | 5.78E-10 | 1.12E-05 | 1938 | 315 | 7.33E-03 | 1.00 |
| GO:0007156 | homophilic cell adhesion via plasma membrane adhesion molecules | 135 | 60 | 4.02E-07 | 7.80E-04 | 135 | 53 | 3.85E-10 | 7.46E-06 |
| GO:0098609 | cell-cell adhesion | 188 | 70 | 4.84E-06 | 5.21E-03 | 188 | 57 | 1.43E-07 | 9.21E-04 |
| GO:0032501 | multicellular organismal process | 6240 | 1266 | 5.11E-06 | 5.22E-03 | 6238 | 843 | 1.89E-06 | 7.32E-03 |
| GO:0007275 | multicellular organismal development | 4319 | 961 | 8.46E-06 | 8.20E-03 | 4319 | 633 | 8.55E-05 | 9.75E-02 |
| GO:0030154 | cell differentiation | 3302 | 741 | 9.00E-05 | 5.40E-02 | 3302 | 480 | 2.88E-03 | 9.67E-01 |
| GO:0022610 | biological adhesion | 1303 | 317 | 1.45E-04 | 7.41E-02 | 1303 | 227 | 6.61E-06 | 2.13E-02 |
| GO:0032502 | developmental process | 5100 | 1087 | 1.44E-04 | 7.41E-02 | 5100 | 726 | 4.67E-05 | 6.47E-02 |
| GO:0032012 | regulation of ARF protein signal transduction | 448 | 124 | 1.33E-03 | 3.68E-01 | 448 | 87 | 1.52E-03 | 7.55E-01 |
| GO:0032015 | regulation of Ran protein signal transduction | 448 | 124 | 1.33E-03 | 3.68E-01 | 448 | 87 | 1.52E-03 | 7.55E-01 |
| GO:0032483 | regulation of Rab protein signal transduction | 448 | 124 | 1.33E-03 | 3.68E-01 | 448 | 87 | 1.52E-03 | 7.55E-01 |

**Table 3.4 (*Continued*) Common gene ontology (GO) biological process (BP) terms associated with gastric and oesophageal differentially methylated loci in CAMs compared to patient–matched ATMs.** *N - number of genes in the GO term; DE - number of genes that are differentially methylated*

| | | Gastric | | | | Oesophageal | | | |
|---|---|---|---|---|---|---|---|---|---|
| **GO ID** | **GO Term** | **N** | **DE** | ***p-value*** | **FDR** | **N** | **DE** | ***p-value*** | **FDR** |
| GO:0032485 | regulation of Ral protein signal transduction | 448 | 124 | 1.33E-03 | 3.68E-01 | 448 | 87 | 1.52E-03 | 7.55E-01 |
| GO:0043087 | regulation of GTPase activity | 448 | 124 | 1.33E-03 | 3.68E-01 | 448 | 87 | 1.52E-03 | 7.55E-01 |
| GO:0035020 | regulation of Rac protein signal transduction | 453 | 125 | 1.39E-03 | 3.68E-01 | 453 | 88 | 1.37E-03 | 7.55E-01 |
| GO:0044699 | single-organism process | 12294 | 2281 | 1.52E-03 | 3.93E-01 | 12292 | 1484 | 2.37E-04 | 2.09E-01 |
| GO:0032487 | regulation of Rap protein signal transduction | 450 | 124 | 1.62E-03 | 4.08E-01 | 450 | 87 | 1.78E-03 | 8.18E-01 |
| GO:0032489 | regulation of Cdc42 protein signal transduction | 452 | 124 | 1.64E-03 | 4.09E-01 | 452 | 87 | 1.77E-03 | 8.18E-01 |
| GO:0043547 | positive regulation of GTPase activity | 421 | 116 | 1.98E-03 | 4.69E-01 | 421 | 80 | 3.84E-03 | 9.98E-01 |
| GO:0007268 | synaptic transmission | 691 | 178 | 4.76E-03 | 7.96E-01 | 691 | 124 | 2.84E-03 | 9.67E-01 |
| GO:0023052 | signalling | 5438 | 1063 | 9.34E-03 | 1.00E+00 | 5436 | 726 | 7.94E-05 | 9.75E-02 |
| GO:0044700 | single organism signalling | 5436 | 1062 | 9.94E-03 | 1.00E+00 | 5434 | 725 | 9.15E-05 | 9.85E-02 |

## 3.4 Discussion

This study is the first to examine genome-wide DNA methylation patterns at individual CpG resolution in primary gastric and oesophageal patient-matched CAM and ATM samples. Data presented in this Chapter provide strong evidence for CAM-specific DNA methylation signatures and show that tumour promoting properties of CAMs may in part be due to epigenetic programming.

Comparative analysis of global DNA methylation between CAMs and patient-matched ATMs, isolated from gastric and oesophageal cancers showed that the mean global DNA methylation of CAMs is statistically lower than the mean global DNA methylation of corresponding ATMs which is in agreement with previous report that stromal gastric CAMs are characterised by global reduction in DNA methylation (Jiang, Gonda et al. 2008). Although this trend parallels the overall loss of DNA methylation that has been well documented in cancer cells from different tumours, it should be noted that net changes in global DNA methylation are subtle, and phenotypic cancer-induced signatures actually involve conserved gene-specific DNA methylation changes. The cause of genomic DNA demethylation in tumour cells is currently under intense investigation. There is evidence to suggest that global hypomethylation may be due to deficiency of circulating methyl donors in cancer tissue (Kim, Fawaz et al. 1998), a phenotype which can be reversed by physiological intakes of folic acid (Pufulete, Al-Ghnaniem et al. 2005). Study using a transgenic mouse model that is etiologically and histologically well matched with human gastric cancers showed beneficial effect of dietary folic acid supplementation, which prevented global loss of DNA methylation and markedly reduced gastric dysplasia and mucosal inflammation (Gonda, Kim et al. 2012). The study suggested that folic acid may protect against the loss of global DNA methylation both in the dysplastic gastric epithelial cells and in gastric stromal myofibroblasts.

Data from other projects in our laboratory indicates that cancer cells induce aerobic glycolysis (Warburg effect) in stromal myofibroblasts, which results in autophagic destruction of mitochondria in these cells. This phenomenon is well documented in breast cancer (Pavlides, Whitaker-Menezes et al. 2009, Martinez-Outschoorn, Sotgia et al. 2014). Superoxide signalling was proposed to play an important role in epigenetic processes such as DNA methylation, histone methylation and acetylation (Afanas'ev 2015). Mitochondria are known to alter DNA methylation by affecting folate metabolism leading to subsequent generation of S-adenosylmethionine (SAM), an important methyl donor. In addition, Smiraglia et al. showed that alterations in mitochondrial DNA copy number play an important role in regulating methylation of several nuclear genes that may contribute to tumorigenesis (Smiraglia, Kulawiec et al. 2008). Given these observations, it is possible that the global loss of DNA methylation in gastric and oesophageal CAMs might result from mitochondrial impairment in these cells.

DNA methylation is a robust biomarker, more stable than RNA or proteins therefore identification of DNA methylation signatures in various human diseases, including cancer, may offer promise for the development of new approaches for diagnosis, prognosis and tumour stratification. In fact, DNA methylation has been proposed as a molecular marker for tumour microenvironment characterization and cell typing within tumours (Jeschke, Collignon et al. 2015). A large study in breast cancer demonstrated that differences in DNA methylation do not only originate from tumour cells but also from cells in the tumour microenvironment, such as tumour-inflicting lymphocytes (Dedeurwaerder, Desmedt et al. 2011). Data from this study have revealed distinct DNA methylation signatures in CAMs isolated from gastric and oesophageal cancers, which may serve as potential diagnostic biomarkers. In addition, identified universal patterns of DNA methylation present in gastric and oesophageal CAMs may provide new insights into the molecular mechanisms of myofibroblast programing, which promote tumour growth and metastasis in different tissues.

The genes targeted by DNA methylation in gastric and oesophageal CAMs have significant overlap with many of them being involved in digestive organ tumours, tumorigenesis, cell movement, migration and proliferation of cells, generation of fibroblasts and growth of tumours. Notably, most of signalling pathways targeted by these differentially methylated genes are either commonly dysregulated in human cancers such as Hippo signalling, PI3K-Akt signalling, MAPK signalling pathways; or are involved in transcriptional regulation of pluripotent stem cells and developmental processes (Hedgehog signalling, Wnt signalling and Notch signalling pathways). Interestingly, glycosaminoglycan metabolism and biosynthesis, an important component of extracellular matrix involved in cell signalling, cell function and cancer progression (Afratis, Gialeli et al. 2012), has been commonly targeted by DNA methylation changes in gastric and oesophageal CAMs together with fatty acid, triacylglycerol, and ketone body metabolism. It is of particular interest, as previous gene expression study conducted in our laboratory predicted an alternative type of 'reverse Warburg effect' where gastric CAMs are characterised by up-regulation of fatty acid metabolism resulting in increased ketone body production, which can be then supplied to cancer cells. ATMs do not exhibit this phenotype, probably due to the distance to cancer cells (Dr Helen Smith, manuscript in preparation). This provides a strong evidence for epigenetic programing of stromal myofibroblasts by cancer cells.

Finally, most enriched biological processes targeted by DNA methylation changes in gastric and oesophageal CAMs included cell adhesion, cell differentiation, and developmental processes, signalling, regulation of signal transduction and GTPase activity. It was suggested that gastric stromal myofibroblasts have a neuroendocrine-like phenotype characterized by $Ca^{2+}$-dependent regulated secretion, dense-core secretory vesicles and expression of neuroendocrine marker proteins; loss of the phenotype is associated with advanced cancer (Balabanova, Holmberg et al. 2014). Interestingly, synaptic transmission and regulation of calcium transport were one of the significantly enriched biological processes targeted by DNA methylation changes in both gastric and oesophageal CAMs.

In summary, this study is the first to examine genome-wide DNA methylation patterns at individual CpG resolution in primary gastric and oesophageal patient-matched CAM and ATM samples. Identified widespread alterations of DNA methylation in gastric and oesophageal CAMs can provide potential clues as to the molecular mechanism of cancer programing of stromal cells. In addition, multiple genomic loci identified as proxies for gastric CAM identification may prove useful as biomarkers for improved diagnosis and prognosis of gastric cancer.

# Chapter IV

## Identification of Genes that Exhibit Correlated Changes in Gene Expression and DNA Methylation in Gastric Myofibroblasts

# 4.1 Introduction

Growing evidence suggests that tumour derived myofibroblasts (CAMs) have distinct gene expression signatures in comparison to non-tumour derived myofibroblasts (ATMs or NTMs). In breast cancer, significant differences have been described between patient-matched CAMs and ATMs (Bauer, Su et al. 2010, Al-Rakan, Colak et al. 2013, Peng, Zhao et al. 2013) or between CAMs and NTMs (Singer, Gschwantler-Kaulich et al. 2008). In non-small cell lung cancer, patient-matched CAMs and ATMs show differential expression of genes relating to the TGF-β signalling pathway (Navab, Strumpf et al. 2011). In colon cancer, comparison of patient-matched CAMs and ATMs showed functional heterogeneity of CAMs that is associated with specific prognostic gene signatures (Herrera, Islam et al. 2013). Gene expression studies in stromal cells from oral squamous cell carcinoma identified two distinct CAM subtypes with differential tumour-promoting abilities (Costea, Hills et al. 2013) and showed that CAM expression profiles reflect the stage of tumour progression (Lim, Cirillo et al. 2011).

Importantly, previous comparative gene expression study conducted in our laboratory showed that CAMs exhibit distinct global gene expression profiles compared to adjacent tissue myofibroblasts (ATMs) or normal tissue myofibroblasts (NTMs) (Dr Helen Smith, manuscript in preparation). In addition, more detailed analysis of these data showed that CAMs can be divided into two distinct groups of patients according to prognosis score, which was determined using the semi-quantitative scoring system (Supplementary Table S2.2). Patients with a prognosis score below 9 were classified as group A (good prognosis) and patient with a prognosis score above 9 were classified as group B (bad prognosis). CAMs from these prognosis related patient subgroups were found to have distinct gene expression profiles (Dr Helen Smith, manuscript in preparation). To identify potential epigenetically regulated gene expression profiles between CAMs and ATMs the expression profile needs to be consistent across all samples therefore for

this study only patients from the bad prognosis subgroup were selected for more detailed investigation (as described in Methods section 2.3.1).

To assess the extent to which differences in DNA methylation (identified in Chapter III) may regulate CAM/ATM - specific gene expression profiles Illumina HumanHT-12v4 Expression BeadChip array studies were performed on a collection of primary gastric patient-matched CAM and ATM samples. The Illumina HumanHT-12 array targets more then 47 000 probes, therefore providing genome-wide transcriptional coverage of well-characterized genes, gene candidates and splice variants. In this study, samples for both genome-wide DNA methylation and gene expression analysis were prepared in parallel to minimize technical artifacts.

## 4.2 Aims

- To identify differentially expressed genes in gastric myofibroblasts isolated from different tissue microenvironments: CAM (cancer), ATM (adjacent tissue) and NTM (normal tissue)
- To investigate how differential gene expression signatures may contribute to the tumor - promoting phenotype of gastric CAMs
- To correlate differential DNA methylation and gene expression signatures observed for gastric CAMs and ATMs
- To establish the extent to which differential DNA methylation may regulate gene expression in gastric CAMs

# 4.3 Results

## 4.3.1 Identification of differential gene expression profiles in gastric CAMs, ATMs and NTMs

Genome–wide expression analysis of gastric myofibroblasts isolated from different tissue microenvironments (CAMs, ATMs and NTMs) was performed using the Illumina HumanHT-12v4 Expression BeadChip arrays. The analysis revealed 13381 expressed genes.

Differential gene expression analysis between gastric (i) CAM vs ATM, (ii) CAM vs NTM and (iii) ATM vs NTM identified 1215 (including 574 upregulated and 641 downregulated genes in CAMs), 987 (including 508 upregulated and 479 downregulated genes in CAMs) and 713 (including 407 upregulated and 306 downregulated genes in ATMs) differentially expressed genes, respectively (Figure 4.1).

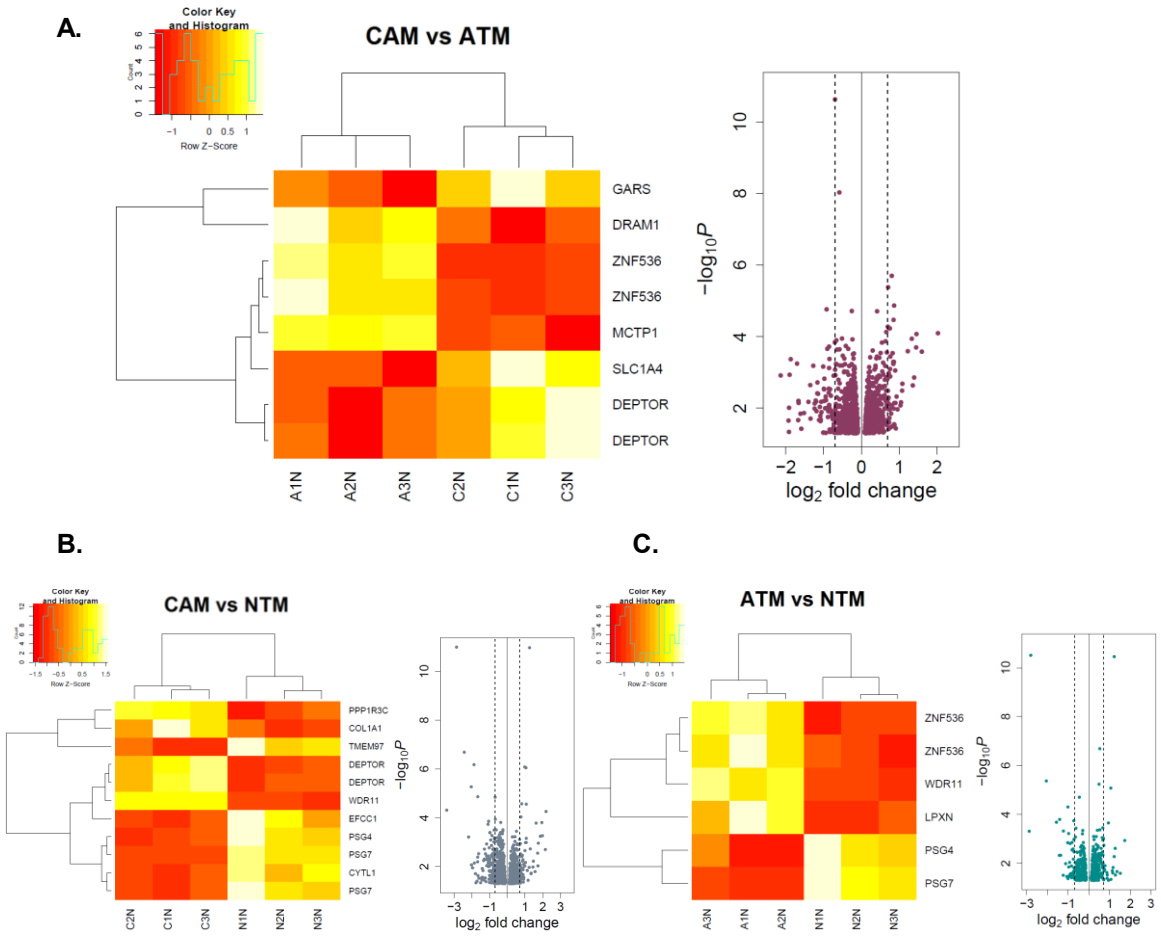


**Figure 4.1 Differential gene expression signatures in gastric myofibroblasts purified from different tissue microenvironments. A.** CAM vs ATM **B.** CAM vs NTM **C.** ATM vs NTM. Heatmaps represent differentially expressed genes in respective comparisons *FDR p-value < 0.05;* volcano plots represent differentially expressed genes *p-value <0.05;* dashed lines 1.6 fold change.

### 4.3.2 Technical validation of identified differentially expressed genes

To validate the results from the differential gene expression analysis, several candidate genes from CAM vs ATM and CAM vs NTM comparisons were selected (Table 4.1) and analysed by TaqMan qPCR assays (Figure 4.2). These genes were selected on the basis of statistical significance (*ZNF536, DEPTOR*), differential expression in both CAM vs ATM and CAM vs NTM comparisons (*ZNF536, DEPTOR, SULF1, SULF2, SPON2*) or the magnitude of observed differences in expression levels in respective comparisons (*HOXA5, SPON2, SULF1, SULF2, DEPTOR, ZNF536*). Triplicate reactions were conducted on each of 9 samples (including 3 CAMs, 3 ATMs and 3 NTMs) from the array experiment, and the data were analysed using the comparative ΔΔCt method (as describe in Methods section 2.11.2). The relative quantification of validated candidate genes in gastric CAM, ATM and NTM samples confirmed the expression patterns observed by Illumina HT-12 array analysis thus increasing confidence in the identified comparative differential gene expression trends.

**Table 4.1 List of Illumina HT-12 probes identified as differentially expressed in CAM vs ATM and CAM vs NTM comparisons that were validated by TaqMan qPCR assays.** *NA* - not identified as differentially express in given comparison

| | | | CAM vs ATM | | | CAM vs NTM | | |
|---|---|---|---|---|---|---|---|---|
| **Illumina HT-12 probe ID** | **EntrezID** | **Gene Symbol** | **$Log_2$ FC** | **FC** | ***p-value*** | **$Log_2$ FC** | **FC** | ***p-value*** |
| ILMN_2150586 | 9745 | ZNF536 | -0.697 | 0.617 | 2.32E-11 | -0.167 | 0.890 | 2.09E-02 |
| ILMN_1772155 | | | -0.582 | 0.668 | 9.24E-09 | *NA* | *NA* | *NA* |
| ILMN_1756685 | 64798 | DEPTOR | 0.806 | 1.749 | 1.99E-06 | 1.053 | 2.075 | 8.88E-07 |
| ILMN_2172755 | | | 0.710 | 1.635 | 4.15E-06 | 0.982 | 1.976 | 8.23E-07 |
| ILMN_1702363 | 23213 | SULF1 | 1.051 | 2.072 | 6.87E-03 | 2.185 | 4.548 | 5.51E-05 |
| ILMN_1667460 | 55959 | SULF2 | 0.872 | 1.831 | 7.18E-03 | 0.862 | 1.818 | 2.67E-02 |
| ILMN_2345142 | | | 0.516 | 1.430 | 2.00E-02 | 0.567 | 1.482 | 3.61E-02 |
| ILMN_1676099 | 10417 | SPON2 | 0.662 | 1.582 | 6.27E-03 | 1.025 | 2.034 | 1.04E-03 |
| ILMN_1753613 | 3202 | HOXA5 | *NA* | *NA* | *NA* | 0.928 | 1.902 | 2.79E-02 |

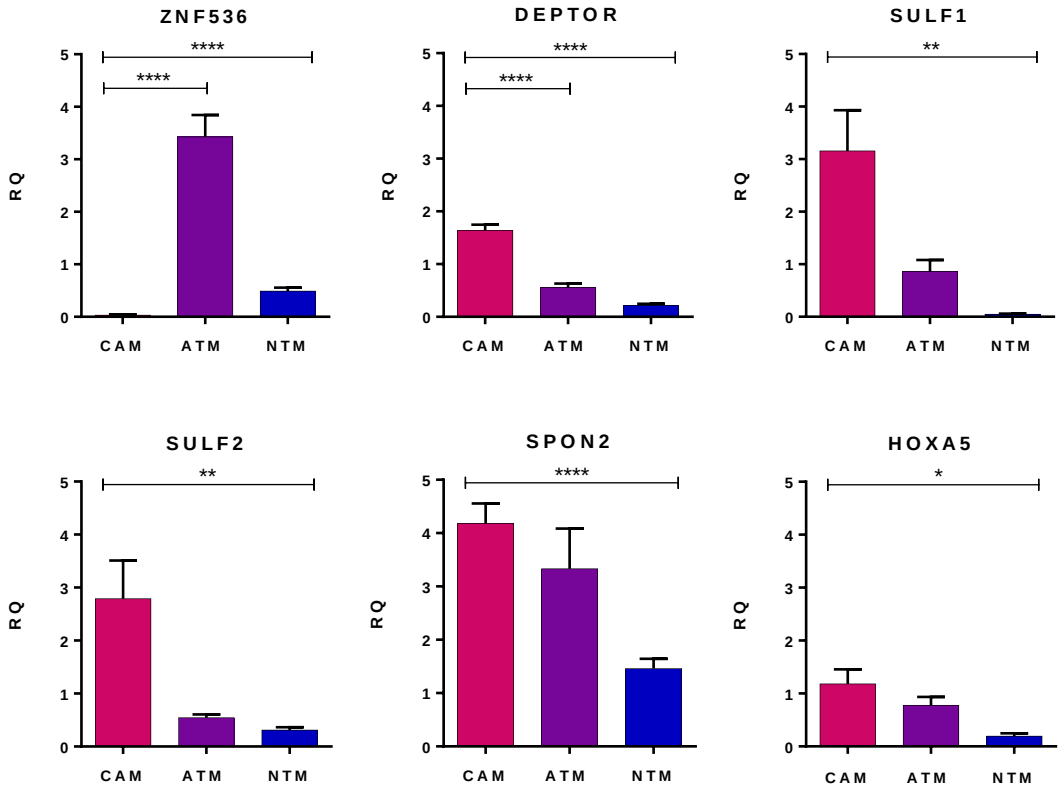


**Figure 4.2 Quantitative PCR validations of genes identified as differentially expressed in CAM vs ATM and CAM vs NTM comparisons** (as shown in Table 4.1). Each TaqMan assay was done in triplicates for CAM (n=3), ATM (n=3) and NTM (n=3) samples. The comparative ΔΔCt method was used and samples were normalized to calibrator (as describe in Methods section 2.11.2). Error bars represent SEM; CAM vs ATM *t-test* ****$p<0.0001$; CAM vs NTM *t-test* ****$p<0.0001$, **$p<0.01$, *$p<0.05$.

### 4.3.3 Potential biological significance of the identified changes in gene expression profiles

Gene ontology (GO), gene set enrichment (GSE) and pathway analysis were performed in order to investigate how the identified differential gene expression profiles may contribute to CAM tumour-promoting phenotype.

#### 4.3.3.1 Gene ontology enrichment analysis

Gene ontology (GO) enrichment analysis was performed on differentially expressed genes identified in CAM vs ATM comparison using Gene Ontology enRIchment anaLysis and visuaLizAtion tool (GOrilla) (Eden, Navon et al. 2009). All 13381 genes expressed in myofibroblast cells were used as a reference set.

The enrichment analysis of: (i) GO biological processes (BP) identified 96 GO terms, (ii) GO molecular functions (MF) identified 11 GO terms and (iii) GO cellular component (CC) identified 16 GO terms with *p-value < 0.001*. List of GO biological process (BP) terms with *FDR p-value < 0.05* is presented in Table 4.2 whereas results from GO cellular component (CC) enrichment analysis are represented as a DAG (directed acyclic graph) tree in Figure 4.3.

**Table 4.2 Gene ontology (GO) biological process (BP) enrichment for differentially expressed genes identified in gastric CAM vs ATM comparison;** *FDR p-value <0.05*.

| GO ID | GO biological process (BP) term | *p-value* | FDR |
|---|---|---|---|
| GO:0044699 | single-organism process | 6.91E-08 | 4.32E-04 |
| GO:0042221 | response to chemical | 4.45E-08 | 5.55E-04 |
| GO:0051239 | regulation of multicellular organismal process | 1.89E-07 | 7.86E-04 |
| GO:0032496 | response to lipopolysaccharide | 1.69E-06 | 3.52E-03 |
| GO:0044763 | single-organism cellular process | 2.88E-06 | 4.50E-03 |
| GO:0048513 | organ development | 3.63E-06 | 5.03E-03 |
| GO:0065008 | regulation of biological quality | 1.28E-05 | 1.33E-02 |
| GO:0071310 | cellular response to organic substance | 1.20E-05 | 1.37E-02 |
| GO:2000026 | regulation of multicellular organismal development | 1.53E-05 | 1.47E-02 |
| GO:1901700 | response to oxygen-containing compound | 1.78E-05 | 1.59E-02 |
| GO:0006984 | ER-nucleus signalling pathway | 2.51E-05 | 2.09E-02 |
| GO:0051241 | negative regulation of multicellular organismal process | 3.23E-05 | 2.24E-02 |
| GO:0006082 | organic acid metabolic process | 3.41E-05 | 2.24E-02 |
| GO:0070555 | response to interleukin-1 | 3.20E-05 | 2.35E-02 |
| GO:2000378 | negative regulation of reactive oxygen species metabolic process | 4.95E-05 | 2.69E-02 |
| GO:0032680 | regulation of tumour necrosis factor production | 4.91E-05 | 2.79E-02 |
| GO:0048519 | negative regulation of biological process | 9.25E-05 | 3.61E-02 |
| GO:0032963 | collagen metabolic process | 9.58E-05 | 3.63E-02 |
| GO:0044707 | single-multicellular organism process | 9.10E-05 | 3.67E-02 |
| GO:0071347 | cellular response to interleukin-1 | 1.07E-04 | 3.71E-02 |
| GO:0019752 | carboxylic acid metabolic process | 8.96E-05 | 3.73E-02 |
| GO:0001525 | angiogenesis | 1.07E-04 | 3.81E-02 |
| GO:0042493 | response to drug | 1.26E-04 | 4.26E-02 |
| GO:0030199 | collagen fibril organization | 1.36E-04 | 4.48E-02 |
| GO:0034976 | response to endoplasmic reticulum stress | 1.41E-04 | 4.52E-02 |

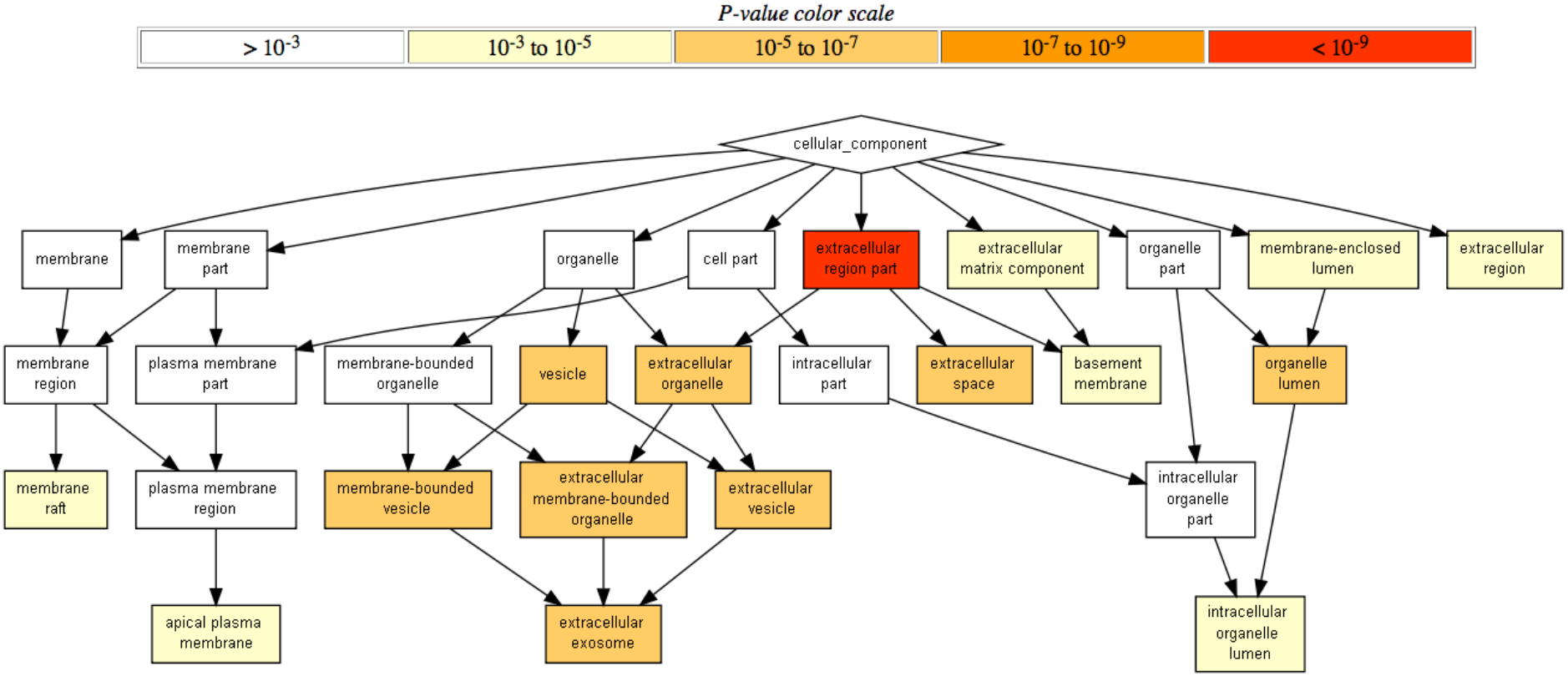


**Figure 4.3 Gene ontology (GO) cellular component (CC) enrichment for differentially expressed genes identified in gastric CAM vs ATM comparison.** DAG tree represent GO_CC terms relations; colors represent GO term *p-value* as shown in *p-value color scale* legend on top.

Other interesting GO terms with *p-value < 0.001* include: regulation of transcription factor import into nucleus (GO:0042990, $p = 1.79 \times 10^{-4}$), morphogenesis of a branching structure (GO:0001763, $p = 3.92 \times 10^{-4}$), multicellular organismal metabolic process (GO:0044236, $p = 3.83 \times 10^{-4}$), monosaccharide biosynthetic process (GO:0046364, $p = 3.61 \times 10^{-4}$), positive regulation of cell proliferation (GO:0008284, $p = 7.27 \times 10^{-4}$), response to lipid hydroperoxide (GO:0006982, $p = 8.35 \times 10^{-4}$), sulfur compound binding (GO:1901681, $p = 1.06 \times 10^{-5}$) and proteoglycan binding (GO:0043394, $p = 4.03 \times 10^{-4}$).

From this analysis, the extracellular region part (GO:0044421, $p = 1.37 \times 10^{-11}$) was found to be the most enriched GO_CC term. Interestingly, extracellular exosome (GO:0070062, $p = 1.6 \times 10^{-7}$) was one of the most enriched GO_CC terms (Figure 4.3). Exosome database, ExoCarta (Mathivanan, Fahner et al. 2012) (data released 29 May 2012) was used to check what molecules might be transferred via exosomes to tumour cells and other cells in the vicinity of tumour–programed stromal myofibroblasts. The analysis revealed that 20% of all differentially expressed genes identified in CAM vs ATM comparison have been found as components of human exosomes. Analysis of these genes, using Ingenuity Knowledge Base, showed that they are involved in tumorigenesis of tissue ($p = 1.91 \times 10^{-26}$), neoplasia of epithelial tissue ($p = 6.1 \times 10^{-26}$), digestive organ tumour ($p = 4.39 \times 10^{-23}$), cell movement of tumour cell lines ($p = 1.21 \times 10^{-19}$), migration of tumour cell lines ($p = 1.67 \times 10^{-19}$), angiogenesis ($p = 1.74 \times 10^{-16}$), proliferation of cells ($p = 3.66 \times 10^{-16}$), differentiation of cells ($p = 6.40 \times 10^{-16}$), invasion of cells ($p = 2.08 \times 10^{-14}$), cell movement of tumour cells ($p = 1.06 \times 10^{-13}$), invasion of tumour cell lines ($p = 5.59 \times 10^{-13}$), progression of tumour ($p = 3.33 \times 10^{-12}$), growth of tumour ($p = 1.73 \times 10^{-11}$) and chemotaxis ($p = 5.93 \times 10^{-11}$).

#### 4.3.3.2 Gene set enrichment analysis

Gene set enrichment analysis (GSEA) (Mootha, Lindgren et al. 2003, Subramanian, Tamayo et al. 2005) was performed on differentially expressed genes identified in CAM vs ATM and CAM vs NTM comparisons using hallmark gene set subcollection from Molecular Signatures Database (MSigDB v5.0). The analysis identified 14 gene sets in CAM vs ATM comparison and 8 gene sets in CAM vs NTM comparison with *p-value < 0.05*. Table 4.3 shows list of all significantly enriched hallmark gene sets in CAM vs ATM phenotype with projection of these gene sets in CAM vs NTM phenotype.

Notably, the most enriched gene set in CAM vs ATM from the entire MSigDB collection was HELLER_SILENCED_BY_METHYLATION_DN (Figure 4.4A), which is a curated gene set (Heller, Schmidt et al. 2008). Figure 4.4 shows the GSEA analysis results for gene sets in which gene members are reported to be epigenetically regulated (Heller, Schmidt et al. 2008).

**Table 4.3 GSEA results showing a list of significantly enriched hallmark gene sets in CAM and ATM phenotypes.** The right-hand column represents a CAM vs NTM projection of gene sets identified in CAM vs ATM (left-hand column) comparison; *NES* – normalized enrichment score; *NA* – gene set not identified as significantly enriched in CAM vs NTM comparison.

| | | **CAM vs ATM** | | | | **CAM vs NTM** | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Phenotype** | **HALLMARK GENE SET NAME** | **SIZE** | **NES** | ***p-value*** | **FDR** | **SIZE** | **NES** | ***p-value*** | **FDR** |
| CAM | HALLMARK_UNFOLDED_PROTEIN_RESPONSE | 26 | 2.98 | 0 | 0 | *NA* | *NA* | *NA* | *NA* |
| CAM | HALLMARK_MTORC1_SIGNALING | 38 | 2.60 | 0 | 0 | *NA* | *NA* | *NA* | *NA* |
| ATM/NTM | HALLMARK_TNFA_SIGNALING_VIA_NFKB | 28 | -2.38 | 0 | 0 | 26 | -2.21 | 0 | 7.48E-04 |
| ATM/NTM | HALLMARK_KRAS_SIGNALING_UP | 20 | -2.33 | 0 | 6.36E-04 | 19 | -1.61 | 3.30E-02 | 7.22E-02 |
| CAM | HALLMARK_GLYCOLYSIS | 27 | 2.14 | 0 | 1.56E-03 | 16 | 1.02 | 4.40E-01 | 4.27E-01 |
| CAM | HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION | 28 | 1.85 | 4.99E-03 | 1.25E-02 | 29 | 1.99 | 0 | 1.03E-02 |
| ATM/NTM | HALLMARK_ADIPOGENESIS | 25 | -1.88 | 1.64E-03 | 1.47E-02 | 21 | -1.18 | 2.56E-01 | 3.12E-01 |
| ATM/NTM | HALLMARK_INFLAMMATORY_RESPONSE | 21 | -1.88 | 3.26E-03 | 1.91E-02 | 18 | -2.01 | 0 | 4.33E-03 |
| ATM/NTM | HALLMARK_COMPLEMENT | 17 | -1.81 | 3.38E-03 | 2.06E-02 | 17 | -1.82 | 6.79E-03 | 2.15E-02 |
| ATM/NTM | HALLMARK_APOPTOSIS | 20 | -1.76 | 1.82E-02 | 2.29E-02 | 18 | -1.10 | 3.40E-01 | 3.72E-01 |
| ATM | HALLMARK_INTERFERON_GAMMA_RESPONSE | 18 | -1.76 | 2.02E-02 | 2.62E-02 | *NA* | *NA* | *NA* | *NA* |
| ATM | HALLMARK_XENOBIOTIC_METABOLISM | 23 | -1.71 | 1.46E-02 | 2.79E-02 | *NA* | *NA* | *NA* | *NA* |
| CAM | HALLMARK_HYPOXIA | 30 | 1.60 | 3.67E-02 | 5.08E-02 | 18 | 1.19 | 2.44E-01 | 4.05E-01 |
| CAM | HALLMARK_UV_RESPONSE_DN | 24 | 1.63 | 2.93E-02 | 5.15E-02 | *NA* | *NA* | *NA* | *NA* |

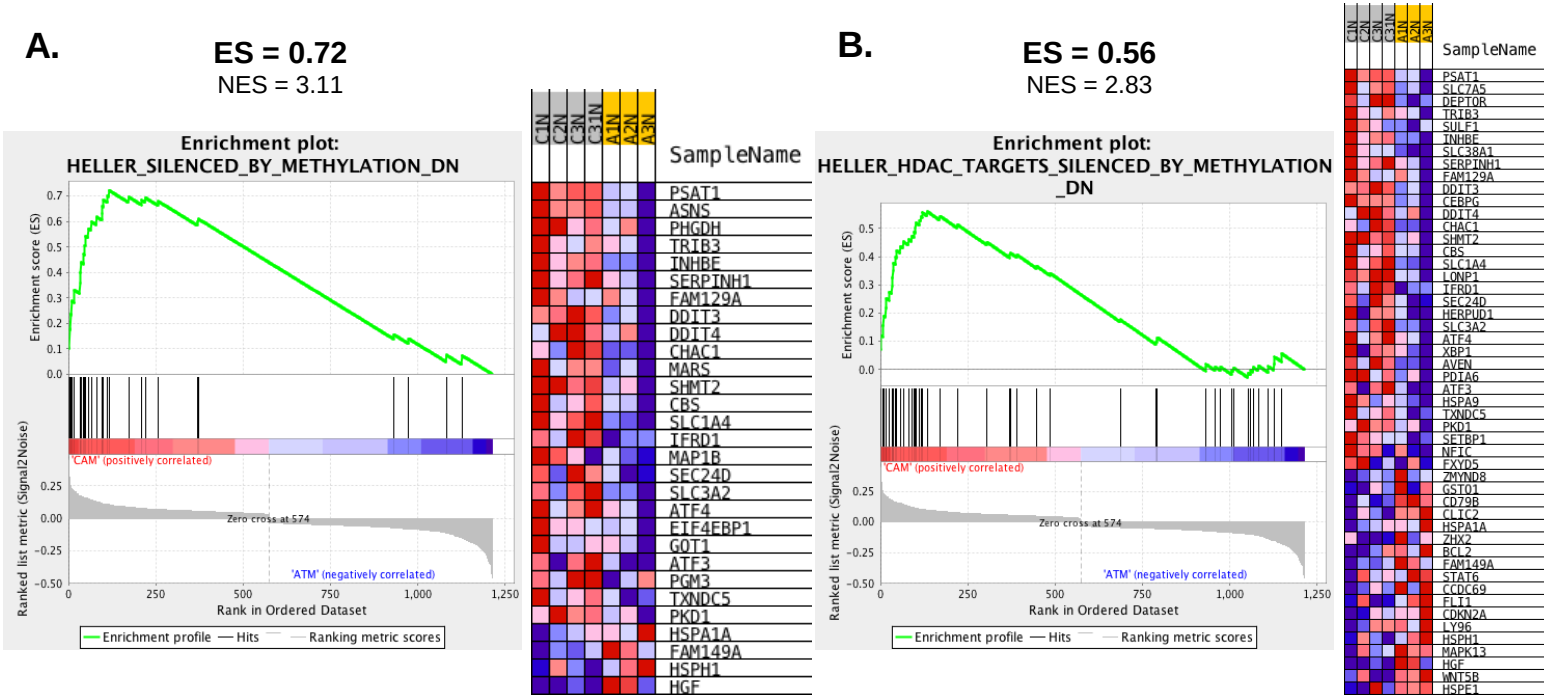


**Figure 4.4 GSEA enrichment plots for gene sets enriched in CAM and ATM phenotypes, which gene members are reported to be epigenetically regulated. A.** Members of this gene set are reported to be downregulated after treatment with the DNA hypomethylating agent (5-aza-2'-deoxycytidine); these genes are upregulated in CAMs compared to ATMs. **B.** Members of this gene set are reported to be downregulated after treatment with histone deacetylase inhibitor (TSA); these genes are upregulated in CAMs compared to ATMs. The green curves show the enrichment score reflecting the degree to which each gene set member/gene (black vertical line) is represented at the top or bottom of the ranked gene list. The horizontal heatmap underneath indicates the relative abundance (red/high to blue/low) of the genes specifically enriched in CAMs as compared to ATMs. Corresponding heatmaps on the right represent the expression values of the genes in the gene set (*red* – high expression, *dark blue* - low expression); *FDR p-value < 0.01*; *ES* - enrichment score, *NES* – normalized enrichment score.

#### 4.3.3.3 Pathway analysis

Differentially expressed genes in CAM vs ATM comparison were subjected to Ingenuity Pathway Analysis (IPA) and ConsensusPathDB (Kamburov, Pentchev et al. 2011) over-representation analysis. For Ingenuity Pathway Analysis, CAM vs ATM gene expression profile was compared against a predefined Illumina HT-12v4 reference set, whereas for ConsensusPathDB an over-representation analysis reference set of all 13381 genes expressed in myofibroblast cells was used.

*Unfolded protein response* ($p = 3.77 \times 10^{-2}$) was the most significant canonical pathway reported by IPA (Figure 4.5) confirming the results obtained from GSEA analysis (Table 4.3). The second most significant canonical pathway was *Role of IL-17A in Arthritis* ($p = 4.01 \times 10^{-2}$) followed by *Dendritic Cell Maturation* ($p = 7.64 \times 10^{-2}$).

KEGG (Kyoto Encyclopedia of Genes and Genomes) and Reactome pathways with *p-value < 0.01* identified by ConsensusPathDB over-representation analysis are presented in Figure 4.6.

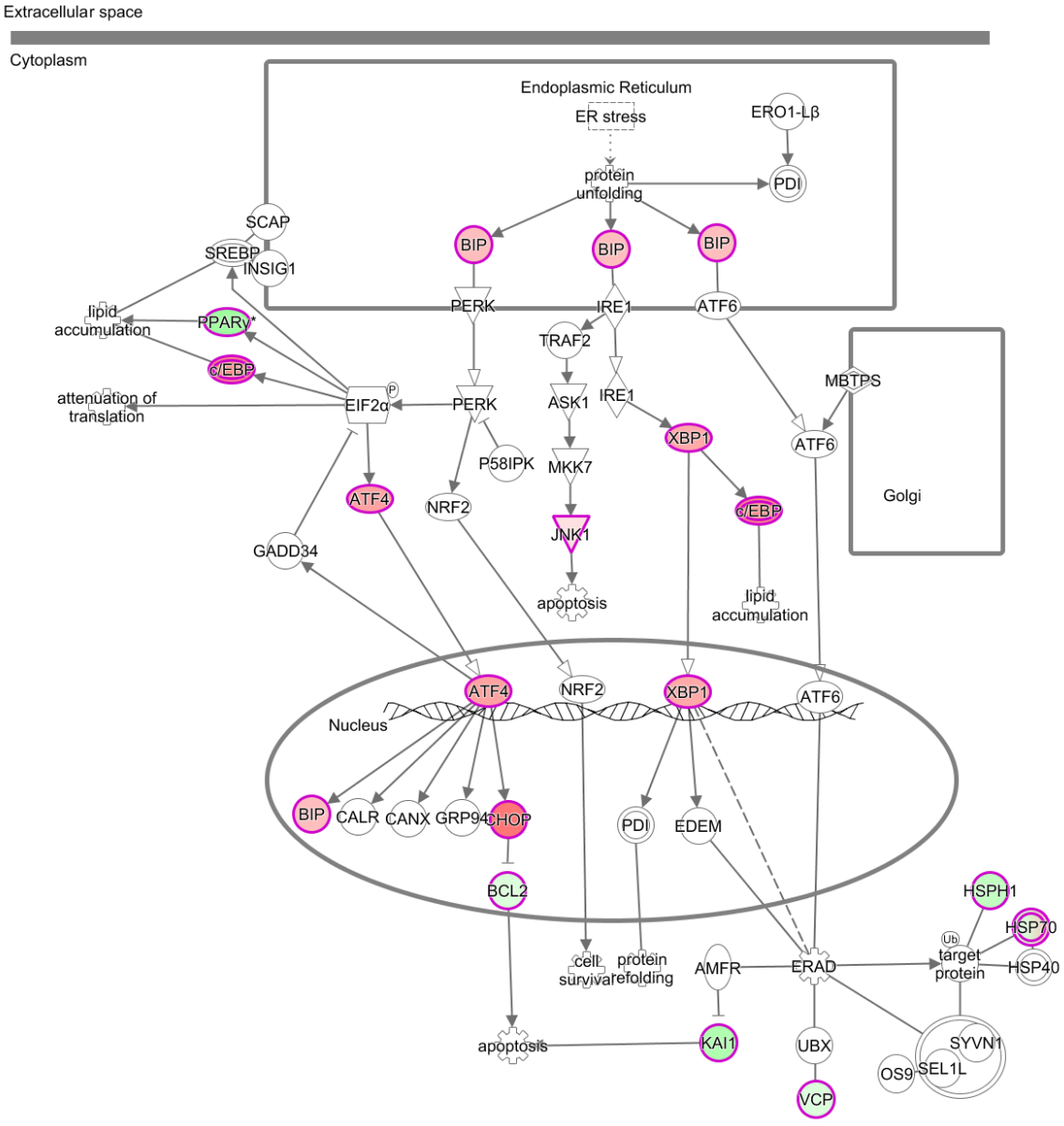


**Figure 4.5 Unfolded protein response pathway -** the most overrepresented Ingenuity Canonical Pathway in CAM vs ATM comparison; *red* – upregulated genes in CAMs, *green* – downregulated genes in CAMs.

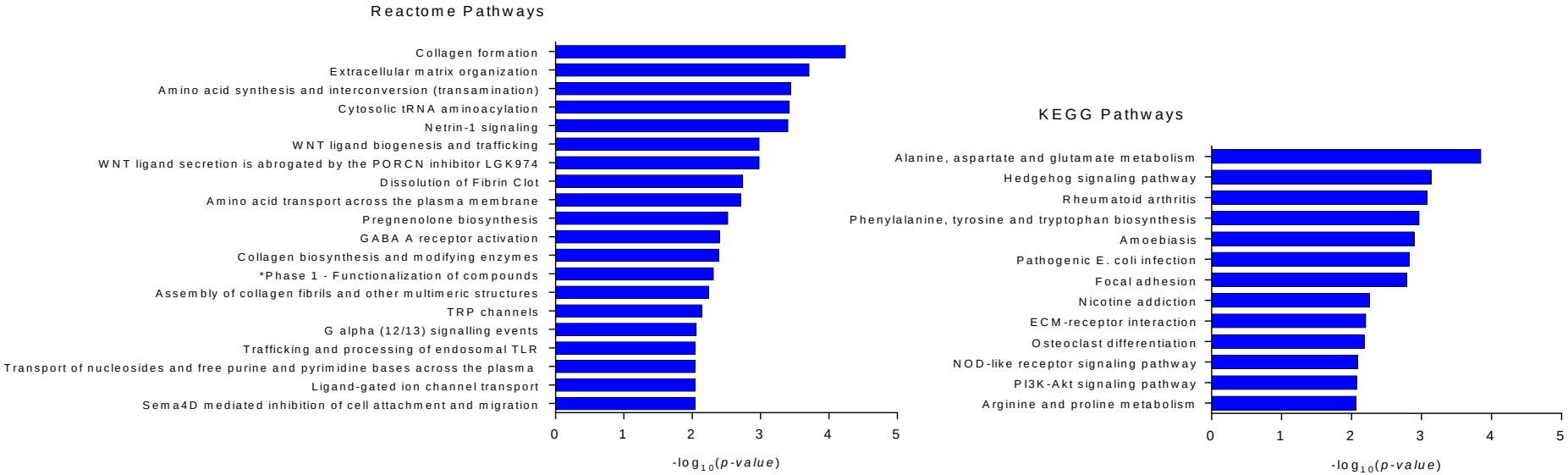


**Figure 4.6 Reactome and KEGG pathways overrepresented in CAM vs ATM comparison;** *p-value<0.01*.

**Defective CYP27A1 causes Cerebrotendinous xanthomatosis (CTX); Defective CYP1B1 causes Glaucoma; Defective CYP4F22 causes Ichthyosis, congenital, autosomal recessive 5 (ARCI5); Defective CYP11B2 causes Corticosterone methyloxidase 1 deficiency (CMO-1 deficiency); Defective CYP24A1 causes Hypercalcemia, infantile (HCAI); Defective CYP2U1 causes Spastic paraplegia 56, autosomal recessive (SPG56); Defective MAOA causes Brunner syndrome (BRUNS); Defective CYP7B1 causes Spastic paraplegia 5A, autosomal recessive (SPG5A) and Congenital bile acid synthesis defect 3 (CBAS3); Defective FMO3 causes Trimethylaminuria (TMAU); Defective CYP21A2 causes Adrenal hyperplasia 3 (AH3); Defective CYP11A1 causes Adrenal insufficiency, congenital, with 46,XY sex reversal (AICSR); Defective CYP2R1 causes Rickets vitamin D-dependent 1B (VDDR1B); Defective CYP17A1 causes Adrenal hyperplasia 5 (AH5); Defective CYP11B1 causes Adrenal hyperplasia 4 (AH4); Defective CYP19A1 causes Aromatase excess syndrome (AEXS); Defective CYP26B1 causes Radiohumeral fusions with other skeletal and craniofacial anomalies (RHFCA); Defective CYP26C1 causes Focal facial dermal dysplasia 4 (FFDD4); Defective CYP27B1 causes Rickets vitamin D-dependent 1A (VDDR1A); Defective TBXAS1 causes Ghosal hematodiaphyseal dysplasia (GHDD)*

### 4.3.4 Integration of differentially methylated loci with differential gene expression profiles in gastric CAMs compared to ATMs

Gene symbols were used as common identifiers to integrate DNA methylation and gene expression data. Firstly, identified differentially methylated CpG loci in gastric CAM vs ATM comparison were associated with genes as described in Chapter III (section 3.3.2.1). In summary, 5688 CpG loci were associated with 5917 genes. Secondly, identified differentially expressed probes in the CAM vs ATM comparison were mapped to gene symbols. For genes with multiple differentially expressed probes, the probe with the lowest *p-value* was used. In total, 1215 differentially expressed genes were identified. Then, both gene lists were compared (Figure 4.7A) and 419 genes showing changes in both expression and DNA methylation were subjected to further analysis (Figure 4.7B and Figure 4.7C).

To identify the relationship between changes in DNA methylation and gene expression, *p-values* for each differentially methylated loci were compared to corresponding *p-value* from gene expression data. This analysis generated four groups of genes, which segregate according to the status of both DNA methylation and gene expression, identifying 144 genes that are hypomethylated and upregulated, 119 genes that are hypermethylated and upregulated, 129 genes that are hypomethylated and downregulated and 106 genes that are hypermethylated and downregulated (Figure 4.6B).

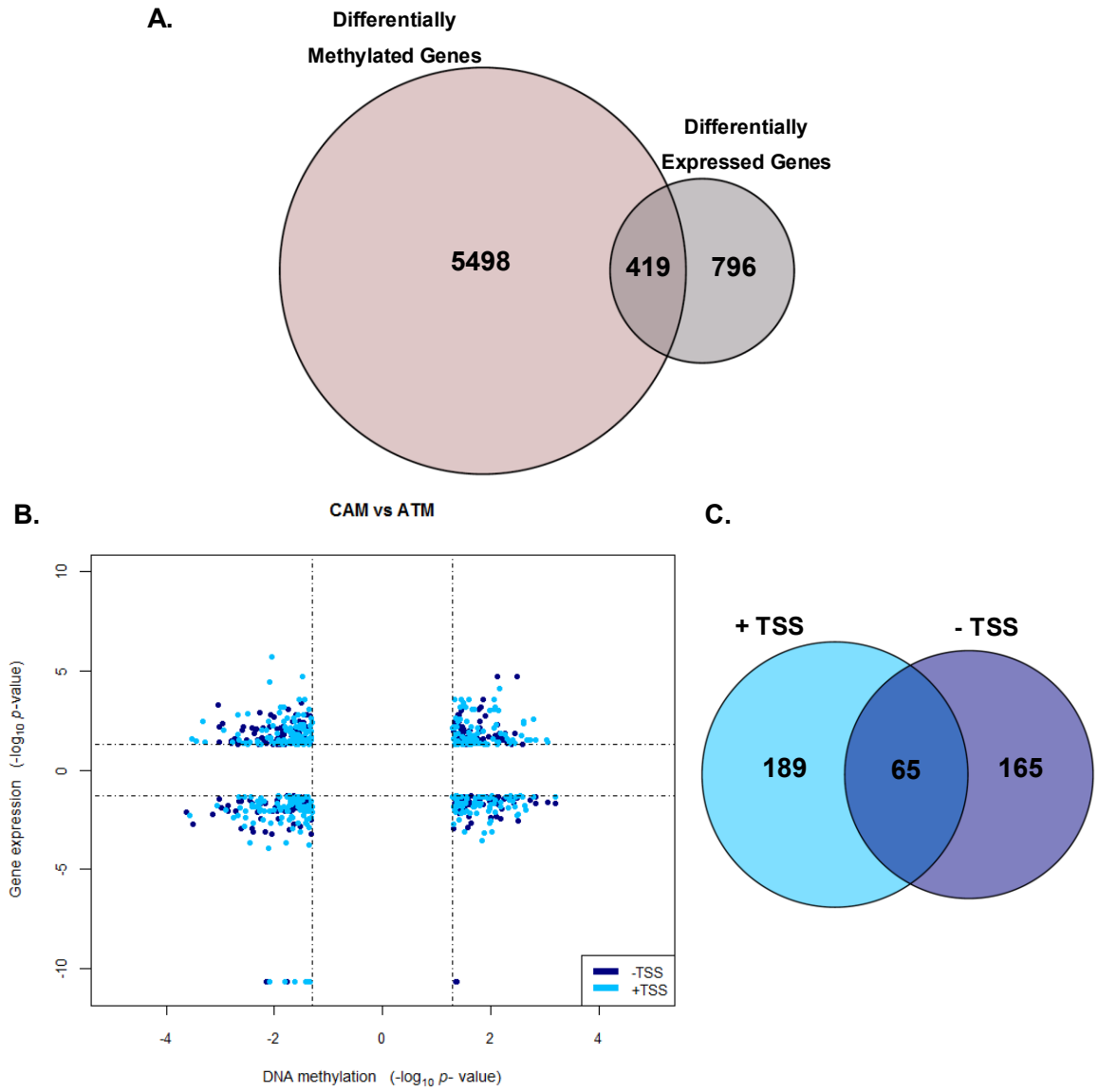


**Figure 4.7 Identification of a subset of genes, which expression may be regulated by DNA methylation changes in gastric CAMs. A.** Overlap of genes identified as differentially methylated (|Δβ| > 0.2, *p-value < 0.05)* and differentially expressed (*p-value < 0.05)* in gastric CAM vs ATM comparisons. **B.** Quadrant plot for genes showing changes in expression and DNA methylation in CAMs compared to ATMs. Vertical and horizontal dashed lines indicate *p-value < 0.05*. The four quadrants shown are: (i) top left – hypomethylated and upregulated in CAMs, (ii) top right – hypermethylated and upregulated in CAMs, (iii) bottom left – hypomethylated and downregulated in CAMs, (iv) bottom right – hypermethylated and downregulated in CAMs; *navy* - CpG sites downstream of transcription start site (-TSS); *blue* – CpG sites upstream of transcription start site (+TSS). **C.** Venn diagram representing differentially expressed genes that are regulated by changes in DNA methylation upstream (+TSS) and/or downstream (-TSS) of the transcription start site.

### 4.3.4.1 Analysis of the relationship between DNA methylation and gene expression in gastric CAMs

The relationship between DNA methylation and gene expression is complex. It is important to consider genomic location of differentially methylated loci when trying to relate these data to gene expression changes. It is generally accepted that methylation in promoter region is inversely correlated with gene expression, whereas methylation in gene-body is positively correlated with gene expression (Jones 1999, Jones 2012).

Therefore, to focus the analysis on promoter and gene-body regions, identified differentially methylated loci were classified according to distance to transcription start site (TSS) of the associated gene (as described in Chapter III section 3.3.2.1). In summary, the analysis revealed that 230 genes have the potential to be regulated by altered DNA methylation downstream of TSS (promoter region) and 254 genes may be regulated by altered DNA methylation upstream of TSS (gene-body region); 65 genes may be regulated by simultaneous methylation changes downstream and upstream of TSS (Figure 4.7C). An example of a gene where expression may be regulated by simultaneous methylation changes downstream and upstream of TSS is shown in Figure 4.8.

Genes with inverse correlation between DNA methylation and gene expression in promoter region or genes with positive correlation between gene expression and DNA methylation in gene-body were selected for further analysis.

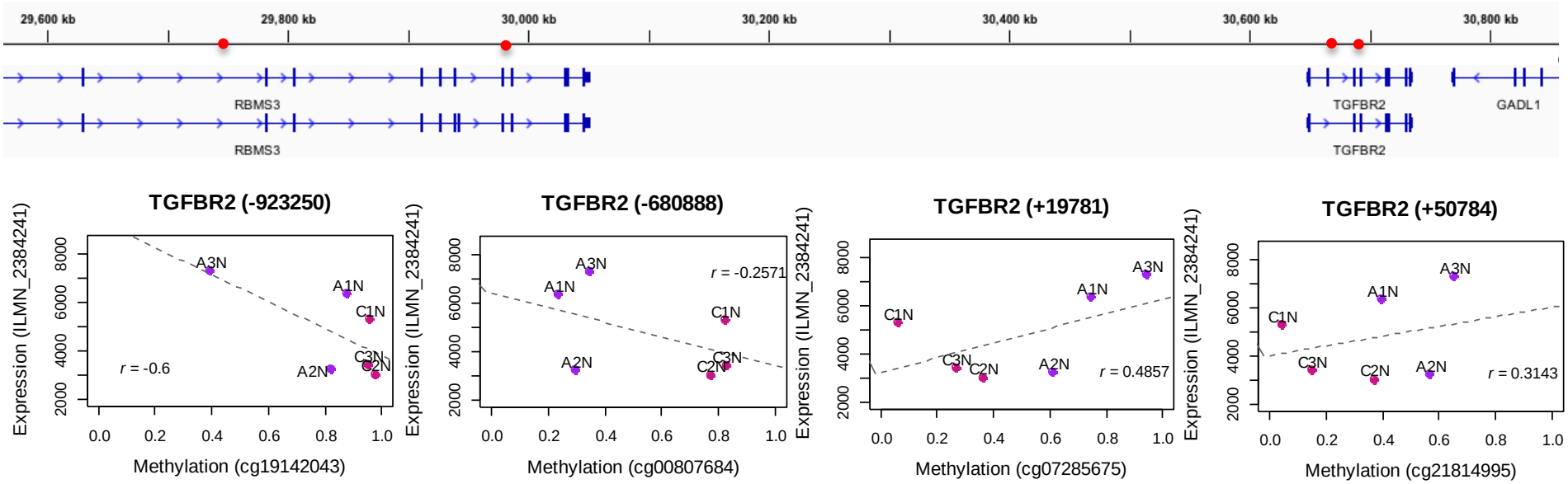


**Figure 4.8 Simultaneous negative correlation between gene expression and DNA methylation in promoter region and positive correlation in gene body of *TGFBR2* in patient-matched CAM and ATM samples.** Individual sample methylation levels (β-value, *X-axis*) are plotted against individual sample gene expression levels (intensity, *Y-axis*) for 4 CpG loci (marked with red dots) associated with *TGFBR2* regulation; *magenta – CAMs, purple – ATMs*; numbers in brackets indicate the distance to transcription start site of *TGFBR2*; *r - Spearman's rank correlation rho*.

### 4.3.4.2 Identification of genes exhibiting correlated changes in promoter methylation and gene expression

This analysis identified 124 genes with coordinated changes in promoter methylation and gene expression, including 69 hypomethylated and upregulated genes (Table 4.4) and 55 hypermethylated and downregulated genes (Table 4.5) in CAMs compared to ATMs. Representative data showing negative correlations between promoter methylation and gene expression are shown in Figure 4.9.

Ingenuity Knowledge Base and ConsensusPathDB (Kamburov, Pentchev et al. 2011) over-representation analyses were performed to identify potential biological functions and processes that might be affected by the observed changes in the subsets of 69 hypomethylated-induced and 55 hypermethylated-repressed genes in gastric CAMs.

Interestingly, hypermethylated and downregulated genes in CAMs were found to be involved in gastrointestinal adenocarcinoma ($p = 1.31 \times 10^{-3}$), gastrointestinal tract cancer ($p = 1.99 \times 10^{-3}$), gastro-oesophageal carcinoma ($p = 8.48 \times 10^{-3}$) and gastric cancer ($p = 1.06 \times 10^{-2}$) whereas hypomethylated and upregulated genes in CAMs are not associated with any of these types of cancer.

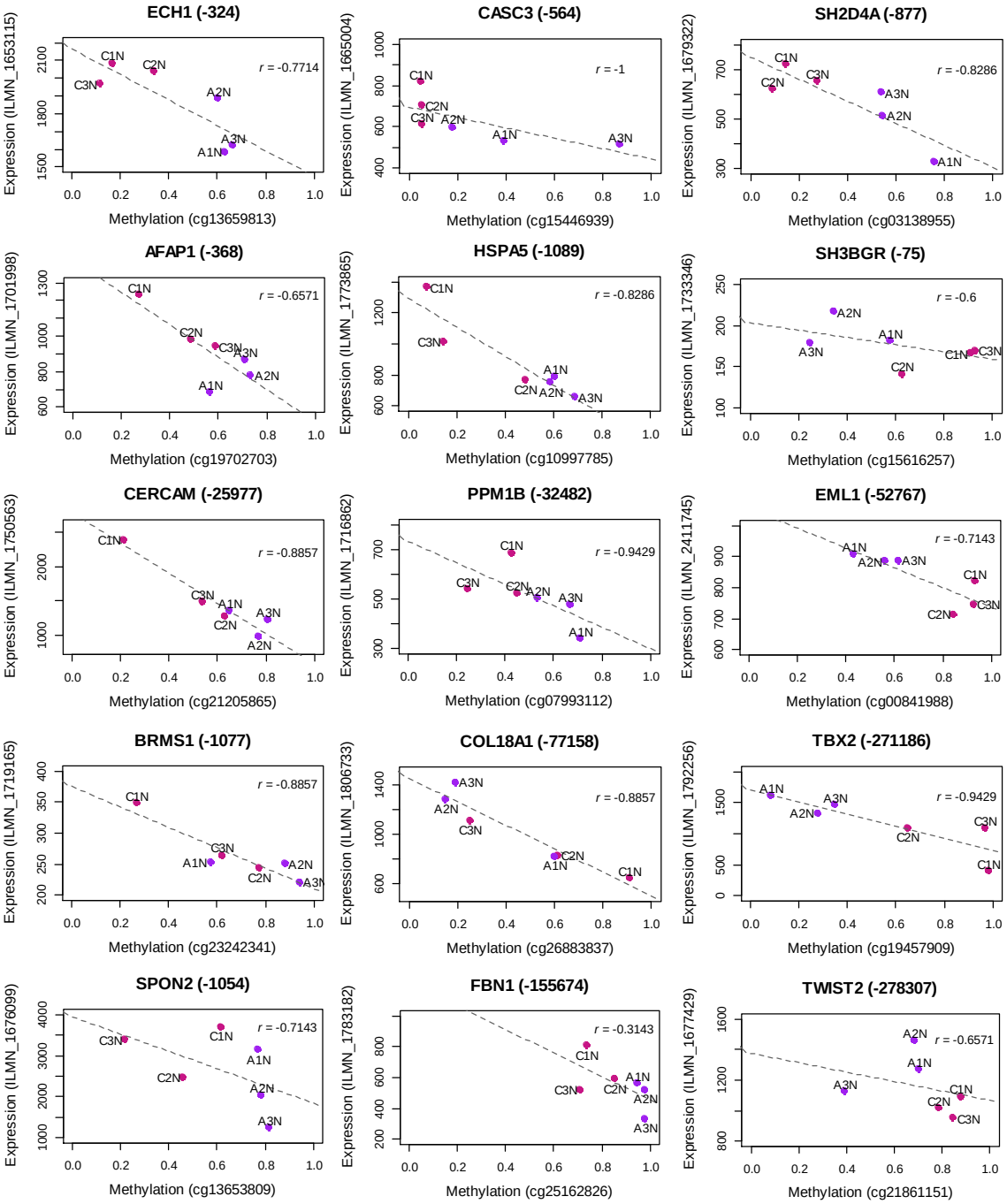


**Figure 4.9 Negative correlations identified between gene expression and promoter methylation in patient-matched CAM and ATM samples.** Individual sample methylation level (β-value, *X-axis*) is plotted against corresponding gene expression level (intensity, *Y-axis*); *magenta – CAMs, purple – ATMs*. Representative data for 15 genes are shown; numbers in brackets indicate the distance to transcription start site of a given gene (indicated at the top of each plot); *r - Spearman's rank correlation rho*.

#### 4.3.4.2.1 Analysis of hypomethylated genes that are upregulated in CAMs compared to ATMs

The analysis of 69 hypomethylated-induced genes in CAMs using Ingenuity Knowledge Base showed that these genes are involved in cell spreading ($p = 2.71 \times 10^{-4}$), transport of amino acids ($p = 9.23 \times 10^{-4}$), communication of breast cancer/melanoma cell lines ($p = 3.61 \times 10^{-3}$), secretion of neurotransmitter ($p = 4.88 \times 10^{-3}$), size of tumour cell lines ($p = 8.71 \times 10^{-3}$), dystrophy of muscle ($p = 8.98 \times 10^{-3}$), cell viability of brain cancer cell lines ($p = 1.57 \times 10^{-2}$), development of cardiovascular system ($p = 1.77 \times 10^{-2}$), cell movement ($p = 2.61 \times 10^{-2}$), secretion of molecule ($p = 3.34 \times 10^{-2}$), production of reactive oxygen species ($p = 3.38 \times 10^{-2}$) and transport of molecule ($p = 3.78 \times 10^{-2}$).

Complementary gene ontology (GO) analysis revealed that these genes are part of endoplasmic reticulum lumen (GO:0005788, $p = 6.5 \times 10^{-4}$), extracellular vesicular exosome (GO:0070062, $p = 9.13 \times 10^{-3}$); take part in glycoprotein binding (GO:0001948, $p = 3.396 \times 10^{-3}$), and the following biological processes: exocytosis (GO:0006887, $p = 2.293 \times 10^{-3}$), carboxylic acid transport (*GO:0046942,* $p = 3.492 \times 10^{-3}$), organic substance transport (GO:0071702, $p = 3.53 \times 10^{-3}$), morphogenesis of a branching epithelium (GO:0061138, $p = 6.074 \times 10^{-3}$), neurotransmitter transport (GO:0006836, $p = 3.676 \times 10^{-3}$) and membrane organization (GO:0061024, $p = 5.677 \times 10^{-3}$).

**Table 4.4 Genes that are hypomethylated and transcriptionally induced in CAMs compared to ATMs.** Table shows the top 10 transcriptionally changed genes that contain differentially methylated loci within their promoter region.

| | | | Expression | | | Methylation | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Gene Symbol** | **HT-12 probe** | **450k probe** | **$Log_2$ FC** | **FC** | ***p - value*** | **CAM β** | **ATM β** | ***p - value*** | ***Spearman correlation*** |
| **SULF1** | ILMN_1702363 | cg10791884 | 1.051 | 2.072 | 6.87E-03 | 0.050 | 0.258 | 9.70E-04 | -0.600 |
| | ILMN_1702363 | cg18545695 | 1.051 | 2.072 | 6.87E-03 | 0.055 | 0.394 | 1.42E-02 | -0.600 |
| **SLC6A9** | ILMN_1714445 | cg25387812 | 0.847 | 1.799 | 5.05E-04 | 0.597 | 0.891 | 8.99E-04 | -0.829 |
| **SGCG** | ILMN_1659649 | cg00748494 | 0.815 | 1.759 | 3.23E-02 | 0.060 | 0.271 | 8.89E-03 | -0.886 |
| | ILMN_1659649 | cg11067829 | 0.815 | 1.759 | 3.23E-02 | 0.331 | 0.603 | 1.40E-02 | -0.943 |
| | ILMN_1659649 | cg07126559 | 0.815 | 1.759 | 3.23E-02 | 0.279 | 0.557 | 2.27E-02 | -0.543 |
| **SPARC** | ILMN_1796734 | cg19939793 | 0.669 | 1.590 | 7.83E-03 | 0.055 | 0.322 | 6.04E-03 | -0.543 |
| **SPON2** | ILMN_1676099 | cg13558774 | 0.662 | 1.582 | 6.27E-03 | 0.107 | 0.653 | 2.69E-03 | -0.886 |
| | ILMN_1676099 | cg23543318 | 0.662 | 1.582 | 6.27E-03 | 0.032 | 0.288 | 8.08E-03 | -0.600 |
| | ILMN_1676099 | cg13653809 | 0.662 | 1.582 | 6.27E-03 | 0.431 | 0.788 | 3.64E-02 | -0.714 |
| | ILMN_1676099 | cg23127323 | 0.662 | 1.582 | 6.27E-03 | 0.174 | 0.622 | 3.73E-02 | -0.600 |
| **COL5A1** | ILMN_1706505 | cg13478045 | 0.622 | 1.539 | 1.60E-03 | 0.358 | 0.777 | 2.53E-02 | -0.829 |
| | ILMN_1706505 | cg00753924 | 0.622 | 1.539 | 1.60E-03 | 0.084 | 0.292 | 3.93E-02 | -0.257 |
| **SEL1L3** | ILMN_1797822 | cg25546651 | 0.614 | 1.530 | 1.02E-02 | 0.039 | 0.356 | 2.21E-02 | -0.543 |
| **SYNM** | ILMN_1712075 | cg26942432 | 0.515 | 1.429 | 2.36E-02 | 0.625 | 0.953 | 1.33E-02 | -0.600 |
| **CHAC1** | ILMN_1739241 | cg07891971 | 0.477 | 1.392 | 3.87E-04 | 0.116 | 0.317 | 3.16E-02 | -0.543 |
| **CERCAM** | ILMN_1750563 | cg21205865 | 0.458 | 1.374 | 1.90E-02 | 0.460 | 0.743 | 2.08E-02 | -0.886 |

## 4.3.4.2.2 Analysis of hypermethylated genes that are downregulated in CAMs compared to ATMs

The analysis of 55 hypermethylated-repressed genes in CAMs using Ingenuity Knowledge Base showed that these genes are involved in formation of membrane ruffles *($p = 1.23 \times 10^{-3}$)*, formation of actin filaments ($p = 1.88 \times 10^{-3}$), angiogenesis ($p = 3 \times 10^{-3}$), migration of connective tissue cells ($p = 4.64 \times 10^{-3}$), cell movement of tumour cell lines ($p = 7.6 \times 10^{-3}$), metabolism of reactive oxygen species ($p = 9.74 \times 10^{-3}$), cell movement ($p = 1.52 \times 10^{-2}$), morphology of cells ($p = 1.66 \times 10^{-2}$), advance malignant tumour ($p = 1.7 \times 10^{-2}$), formation of lamellipodia ($p = 1.86 \times 10^{-2}$), migration of cells ($p = 1.92 \times 10^{-2}$), differentiation of connective tissue cells ($p = 2.4 \times 10^{-2}$), proliferation of tumour cell lines ($p = 4 \times 10^{-2}$) and expression of RNA ($p = 4.74 \times 10^{-2}$).

Gene ontology (GO) analysis revealed that these genes are involved in positive regulation of growth (GO:0045927*, $p = 2.749 \times 10^{-4}$*), cell activation (GO:0001775*, $p = 1.077 \times 10^{-3}$*), regulation of response to stimulus (GO:0048583*, $p = 2 \times 10^{-3}$*), morphogenesis of a branching structure (GO:0001763*, $p = 3.399 \times 10^{-3}$*), regulation of developmental process (GO:0050793*, $p = 4.48 \times 10^{-3}$*), nervous system development (GO:0007399*, $p = 5.39 \times 10^{-3}$*), regulation of signal transduction (GO:0009966*, $p = 5.627 \times 10^{-3}$*), cell fate commitment (GO:0045165*, $p = 6.616 \times 10^{-3}$*) and embryo development (GO:0009790*, $p = 8.25 \times 10^{-3}$*).

**Table 4.5 Genes that are hypermethylated and transcriptionally repressed in CAMs compared to ATMs.** Table shows the top 10 transcriptionally changed genes that contain differentially methylated loci within their promoter region.

| | | | Expression | | | Methylation | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Gene Symbol** | **HT-12 probe** | **450k probe** | **$Log_2$ FC** | **FC** | ***p-value*** | **CAM β** | **ATM β** | ***p - value*** | ***Spearman correlation*** |
| **STOM** | ILMN_1696419 | cg14215970 | -0.943 | 0.520 | 4.72E-03 | 0.791 | 0.590 | 2.51E-02 | -0.257 |
| **SEPP1** | ILMN_1785071 | cg12887985 | -0.830 | 0.563 | 4.60E-02 | 0.870 | 0.620 | 3.25E-03 | -0.657 |
| **MGAT3** | ILMN_1853824 | cg03318904 | -0.806 | 0.572 | 1.18E-02 | 0.446 | 0.181 | 6.23E-03 | -0.029 |
| **TBX2** | ILMN_1792256 | cg19457909 | -0.726 | 0.605 | 2.75E-02 | 0.866 | 0.235 | 2.85E-02 | -0.943 |
| **WNT2B** | ILMN_1740269 | cg04571584 | -0.710 | 0.611 | 4.87E-02 | 0.912 | 0.577 | 1.49E-03 | -0.429 |
| **ZNF536** | ILMN_2150586 | cg00386405 | -0.697 | 0.617 | 2.32E-11 | 0.626 | 0.358 | 4.25E-02 | -0.886 |
| | ILMN_2150586 | cg05804948 | -0.697 | 0.617 | 2.32E-11 | 0.322 | 0.014 | 4.42E-02 | -0.886 |
| **SPRY1** | ILMN_2329914 | cg00137840 | -0.590 | 0.664 | 2.73E-02 | 0.910 | 0.691 | 7.61E-03 | -0.257 |
| **TGFBR2** | ILMN_2384241 | cg00807684 | -0.587 | 0.666 | 2.38E-02 | 0.809 | 0.292 | 8.59E-04 | -0.257 |
| | ILMN_2384241 | cg19142043 | -0.587 | 0.666 | 2.38E-02 | 0.962 | 0.700 | 2.02E-02 | -0.600 |
| **FAT1** | ILMN_3247578 | cg03199366 | -0.516 | 0.699 | 2.79E-02 | 0.717 | 0.456 | 1.05E-02 | -0.314 |
| **ABR** | ILMN_1672878 | cg19807420 | -0.452 | 0.731 | 4.74E-02 | 0.879 | 0.531 | 1.73E-02 | -0.429 |
| | ILMN_1672878 | cg05856321 | -0.452 | 0.731 | 4.74E-02 | 0.843 | 0.356 | 2.97E-02 | -0.429 |

### 4.3.4.3 Identification of genes exhibiting correlated changes in gene-body methylation and gene expression

This analysis identified 152 genes with coordinated changes in gene-body methylation and gene expression, including 79 hypomethylated and downregulated genes (Table 4.6) and 73 hypermethylated and upregulated genes (Table 4.7) in CAMs compared to ATMs. Representative data showing positive correlations between gene-body methylation and gene expression are shown in Figure 4.10.

Ingenuity Knowledge Base and ConsensusPathDB (Kamburov, Pentchev et al. 2011) over-representation analyses were performed to identified potential biological functions and processes that may be affected by observed changes in the subsets of 79 hypomethylated-repressed and 73 hypermethylated-induced genes in gastric CAMs.

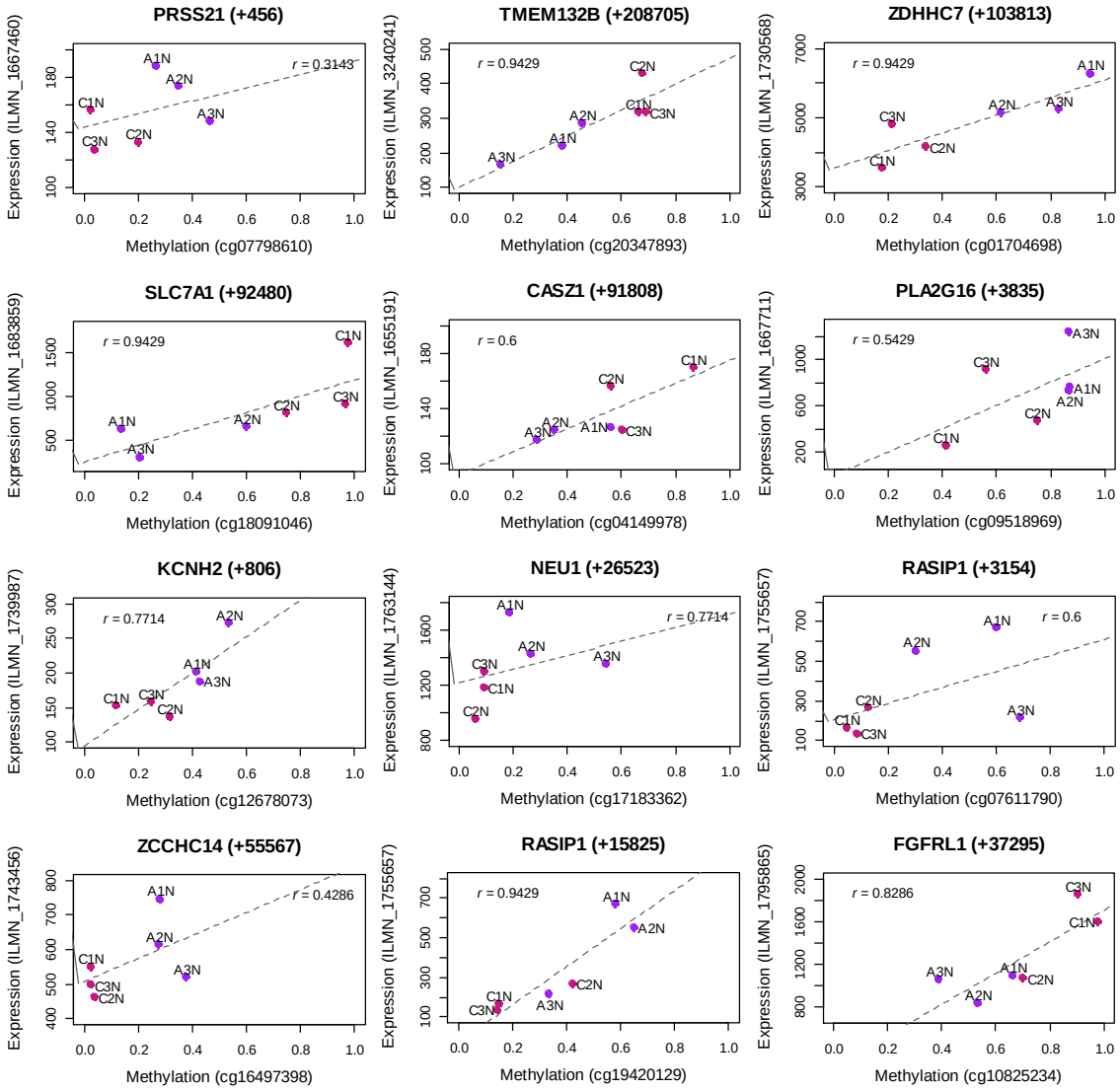


**Figure 4.10 Positive correlations identified between gene expression and gene-body methylation in patient-matched CAM and ATM samples.** Individual sample methylation level (β-value, *X-axis*) is plotted against corresponding gene expression level (intensity, *Y-axis*); *magenta – CAMs, purple – ATMs*. Representative data for 12 genes are shown; numbers in brackets indicate the distance to transcription start site of a given gene (indicated at the top of each plot); *r - Spearman's rank correlation rho*.

#### 4.3.4.3.1 Analysis of hypomethylated genes that are downregulated in CAMs compared to ATMs

Analysis of the subgroup of 79 hypomethylated-repressed genes in CAMs using Ingenuity Knowledge Base showed that these genes are involved in invasion of embryonic cell lines ($p = 3.69 \times 10^{-5}$), invasion of cells ($p = 5.90 \times 10^{-4}$), proliferation of breast cancer cell lines ($p = 4.35 \times 10^{-3}$), cell death ($p = 4.60 \times 10^{-3}$), formation of actin filaments ($p = 6.78 \times 10^{-3}$), transformation of tumour cell lines ($p = 6.84 \times 10^{-3}$), transport of molecule ($p = 1.26 \times 10^{-2}$), migration of cells ($p = 1.26 \times 10^{-2}$), proliferation of cells ($p = 1.33 \times 10^{-2}$), differentiation of cells ($p = 1.48 \times 10^{-2}$), angiogenesis ($p = 1.66 \times 10^{-2}$) and cell survival ($p = 2.80 \times 10^{-2}$).

Gene ontology (GO) analysis of this subset of genes indicates that these genes are involved in regulation of cell adhesion (GO:0030155, $p = 1.32 \times 10^{-4}$), organ development (GO:0048513, $p = 3.09 \times 10^{-4}$), tissue morphogenesis (GO:0048729, $p = 8.94 \times 10^{-4}$), positive regulation of growth (GO:0045927, $p = 1.492 \times 10^{-3}$) and reactive oxygen species metabolic process (GO:0072593, $p = 2.392 \times 10^{-3}$).

**Table 4.6 Top 10 genes that are hypomethylated in gene-body and transcriptionally repressed in CAMs compared to ATMs.**

| | | | Expression | | | Methylation | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Gene Symbol** | **HT-12 probe** | **450k probe** | **$Log_2$ FC** | **FC** | ***p - value*** | **CAM β** | **ATM β** | ***p - value*** | ***Spearman correlation*** |
| **AKR1B1** | ILMN_1701731 | cg26925463 | -1.145 | 0.452 | 3.96E-03 | 0.276 | 0.760 | 1.76E-03 | 0.771 |
| **LIMCH1** | ILMN_2139761 | cg03742238 | -1.118 | 0.461 | 2.15E-02 | 0.027 | 0.308 | 1.13E-02 | 0.600 |
| **ABCC4** | ILMN_2194009 | cg15896117 | -0.978 | 0.508 | 1.50E-03 | 0.667 | 0.891 | 2.81E-03 | 0.657 |
| **RSPO3** | ILMN_1681983 | cg10997634 | -0.922 | 0.528 | 9.93E-03 | 0.542 | 0.858 | 1.00E-02 | 0.657 |
| **CD248** | ILMN_1726589 | cg08924469 | -0.884 | 0.542 | 1.77E-02 | 0.624 | 0.840 | 3.01E-02 | 0.829 |
| **LAMA5** | ILMN_1773567 | cg04632997 | -0.853 | 0.554 | 8.12E-03 | 0.487 | 0.717 | 4.99E-02 | 0.886 |
| **RASIP1** | ILMN_1755657 | cg02804819 | -0.852 | 0.554 | 8.39E-03 | 0.107 | 0.478 | 6.39E-03 | 0.486 |
| | ILMN_1755657 | cg19420129 | -0.852 | 0.554 | 8.39E-03 | 0.240 | 0.522 | 1.71E-02 | 0.943 |
| | ILMN_1755657 | cg07611790 | -0.852 | 0.554 | 8.39E-03 | 0.087 | 0.531 | 1.84E-02 | 0.600 |
| **RFTN1** | ILMN_1800787 | cg01336231 | -0.847 | 0.556 | 1.22E-02 | 0.143 | 0.624 | 5.15E-03 | 0.543 |
| **PPARG** | ILMN_1800225 | cg07676920 | -0.733 | 0.602 | 1.23E-03 | 0.675 | 0.901 | 2.22E-02 | 0.943 |
| **DENND2A** | ILMN_1666503 | cg11176169 | -0.732 | 0.602 | 3.29E-02 | 0.342 | 0.554 | 2.14E-02 | 0.257 |

## 4.3.4.3.2 Analysis of hypermethylated genes that are upregulated in CAMs compared to ATMs

Analysis of 73 hypermethylated-induced genes in CAMs using Ingenuity Knowledge Base showed that these genes are involved in metabolism of amino acids ($p = 5.59 \times 10^{-3}$), metabolism of heparan sulfate proteoglycan ($p = 7.63 \times 10^{-3}$), O-glycosylation of peptide ($p = 1.14 \times 10^{-2}$), infection of cells ($p = 1.37 \times 10^{-2}$), transport of amino acids ($p = 1.75 \times 10^{-2}$), folding of protein ($p = 1.88 \times 10^{-2}$), cancer of secretory structure ($p = 2.25 \times 10^{-2}$), metabolism of protein ($p = 3.92 \times 10^{-2}$) and proliferation of cells ($p = 3.92 \times 10^{-2}$).

Gene ontology (GO) analysis of this subset of genes revealed that they are associated with extracellular vesicular exosomes (GO:0070062, $p = 3.465 \times 10^{-3}$); take part in the following processes: cellular amino acid biosynthetic process (GO:0008652, $p = 2.28 \times 10^{-4}$), one-carbon metabolic process (GO:0006730, $p = 2.93 \times 10^{-4}$), cellular amino acid metabolic process (GO:0006520, $p = 1.212 \times 10^{-3}$), cell differentiation (GO:0030154, $p = 2.198 \times 10^{-3}$), organic acid biosynthetic process (GO:0016053, $p = 4.91 \times 10^{-3}$), organic acid metabolic process (GO:0006082, $p = 5.374 \times 10^{-3}$) and cell morphogenesis (GO:0000902, $p = 6.721 \times 10^{-3}$).

**Table 4.7 Top 10 genes that are hypermethylated in gene-body and transcriptionally induced in CAMs compared to ATMs.**

| | | | Expression | | | Methylation | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Gene Symbol** | **HT-12 probe** | **450k probe** | **$Log_2$ FC** | **FC** | ***p - value*** | **CAM β** | **ATM β** | ***p - value*** | ***Spearman correlation*** |
| **PSAT1** | ILMN_1692938 | cg13612583 | 2.029 | 4.080 | 7.99E-05 | 0.846 | 0.592 | 6.85E-03 | 0.886 |
| **SLC7A5** | ILMN_1720373 | cg26569315 | 1.230 | 2.346 | 2.55E-04 | 0.746 | 0.422 | 3.46E-02 | 0.657 |
| **KLF2** | ILMN_1735930 | cg19280540 | 0.902 | 1.869 | 7.14E-03 | 0.819 | 0.613 | 1.73E-02 | 0.257 |
| | ILMN_1735930 | cg02668248 | 0.902 | 1.869 | 7.14E-03 | 0.860 | 0.657 | 4.65E-02 | 0.600 |
| **MTHFD1L** | ILMN_1772521 | cg00945443 | 0.897 | 1.862 | 8.83E-04 | 0.683 | 0.469 | 8.79E-03 | 0.886 |
| **SULF2** | ILMN_1667460 | cg14286048 | 0.872 | 1.831 | 7.18E-03 | 0.530 | 0.226 | 4.59E-02 | 0.429 |
| **SLC7A1** | ILMN_1683859 | cg15955521 | 0.828 | 1.775 | 2.89E-04 | 0.811 | 0.608 | 8.49E-03 | 0.943 |
| | ILMN_1683859 | cg18091046 | 0.828 | 1.775 | 2.89E-04 | 0.897 | 0.314 | 4.62E-02 | 0.943 |
| **SHMT2** | ILMN_1661264 | cg13433012 | 0.759 | 1.693 | 8.02E-04 | 0.457 | 0.106 | 2.15E-02 | 0.714 |
| **TARS** | ILMN_1685480 | cg15902864 | 0.675 | 1.596 | 2.75E-04 | 0.542 | 0.226 | 2.48E-02 | 0.886 |
| **SPON2** | ILMN_1676099 | cg08462122 | 0.662 | 1.582 | 6.27E-03 | 0.772 | 0.428 | 4.95E-02 | 0.771 |
| **FGFRL1** | ILMN_1795865 | cg20518497 | 0.633 | 1.551 | 6.58E-04 | 0.833 | 0.625 | 8.26E-03 | 0.886 |
| | ILMN_1795865 | cg10825234 | 0.633 | 1.551 | 6.58E-04 | 0.862 | 0.529 | 2.92E-02 | 0.829 |
| | ILMN_1795865 | cg01321174 | 0.633 | 1.551 | 6.58E-04 | 0.778 | 0.571 | 3.66E-02 | 0.086 |

## 4.4 Discussion

This study is the first to integrate genome-wide DNA methylation and gene expression profiles in patient-matched gastric CAM and ATM samples in order to assess the extent to which DNA methylation may contribute to CAM/ATM-specific gene expression profiles.

In this study, the correlation analysis between DNA methylation and gene expression allowed identification of a number of novel epigenetically regulated candidate genes, which may contribute to the ability of stromal myofibroblasts to exert tumour-promoting properties. Notably, promoter DNA hypermethylation in CAMs emerged as a novel regulatory mechanism for transcriptional repression of genes involved in: the pathology of gastrointestinal cancers, cell activation, regulation of response to stimulus, regulation of developmental process and cell fate commitment. Promoter DNA hypomethylation in CAMs, on the other hand, emerged as an important regulatory mechanism for transcriptional activation of genes involved in secretion and transport of molecules, including extracellular vesicular exosomes. Importantly, the enrichment analysis of genes positively correlated with gene-body methylation in CAMs largely confirmed these trends supporting the hypothesis that tumour-promoting properties of CAMs may well be epigenetically programed.

Comparative differential gene expression profiles in gastric myofibroblasts identified in this study confirm previous observations in gastric and other cancer types that tumour derived myofibroblasts (CAMs) and non-tumour derived myofibroblasts (ATMs and NTMs) have distinct gene expression signatures. Significantly, the most over-represented pathway in CAMs compared to ATMs was unfolded protein response (UPR) pathway. Activation of this pathway is a feature of specialized secretory cells and has been implicated in many diseases including

cancer, inflammatory disease, neurodegenerative disorders and metabolic disorders (Hetz, Chevet et al. 2013). Recent studies show that this pathway is also important in many other physiological processes that are not directly linked with protein folding, including lipid and cholesterol metabolism, energy control, inflammation and cell differentiation (Rutkowski and Hegde 2010, Wang and Kaufman 2012). The chaperone protein HSPA5 (also known as BIP or GRP78) is a key regulator of the endoplasmic reticulum (ER) stress response, critical for protein quality control and activation of the UPR pathway. Misfolded proteins bind and sequester HSPA5, and the resulting reduction in free HSPA5 activates signalling pathways to induce transcription of HSPA5 and other chaperones involved in UPR (Wang and Kaufman 2014). This is of particular significance as expression of *HSPA5* and other components of UPR pathway (*CHAC1, XPOT, PSAT1, SLC7A5, TARS, SLC1A4, EIF4EBP1, ATF3*) appear to be regulated by altered DNA methylation in gastric CAMs. In particular, promoter hypomethylation of *HSPA5* is associated with transcriptional activation of *HSPA5* in CAMs (Figure 4.8). Notably, upregulation of *HSPA5* and other components of UPR pathway (*CHAC1, ATF3*) were also observed in a previous gene expression study on 12 patient-matched gastric CAM and ATM samples (Supplementary Figure S4.1).

Extracellular region (GO:0044421) and extracellular exosome (GO:0070062) were the most overrepresented GO cellular component terms in CAM vs ATM comparison. Strikingly, the correlation analysis revealed that promoter DNA hypomethylation may be an important mechanism for transcriptional activation of genes associated with these GO_CC terms (*YWHAZ, SPON2, SEMA5A, EHD4, CPM, HSPA5, TRAP1, SLC3A2, DHRS4, TTYH3, FBN1, ECH1, GNG4, COL5A1, RAP1GDS1, SLC1A4, GOT1 and VPS28;* examples in Figure 4.9 and Table 4.3). In addition, a large proportion of these genes (*SPON2, SEMA5A, HSPA5, SLC3A2, ECH1, GOT1*) were also found to be upregulated in CAMs in a previous gene expression study on 12 patient-matched gastric CAM and ATM samples (Supplementary Figure S4.1) supporting the idea that transcriptional regulation of

these consistently changed genes may also be controlled by tumour-induced epigenetic programing.

These data are of particular interest as there is mounting evidence to show that exosomes are important players in tumour-stroma communication. They carry biologically active molecules, such as microRNAs, mRNAs and proteins, which can be horizontally transferred to, and function in, recipient cells. Exosomes can function in an autocrine or paracrine manner to promote tumour-induced immune suppression, angiogenesis, and formation of pre-metastatic niche (Minciacchi, Freeman et al. 2015). Janowska-Wieczorek et al. showed that exosomes secreted by activated platelets induce expression of proangiogenic and invasive factors in lung cancer cells therefore promoting angiogenesis and metastasis (Janowska-Wieczorek, Wysoczynski et al. 2005). Luga et al. reported that CAM-secreted exosomes promote breast cancer cell motility and metastasis by mobilizing autocrine Wnt-planar cell polarity (PCP) signalling in cancer cells (Luga, Zhang et al. 2012). Several studies have reported that components of PCP pathway are upregulated in cancer and are involved in cancer progression (Wang 2009). This pathway was also identified in CAM vs ATM comparison by the Ingenuity Pathway Analysis (Supplementary Figure S4.2). Notably, 20% of all differentially expressed genes in CAM vs ATMs were identified as exosome components using ExoCarta (Mathivanan, Fahner et al. 2012). Largely, these genes are involved in tumorigenesis, cell movement, migration, angiogenesis, proliferation, invasion and tumour growth.

The relationship between DNA methylation and gene expression is complex. Generally, the correlation between promoter-associated DNA methylation and gene expression is modest (Bock 2012). However, this may not be surprising, given that gene expression can also be regulated by other epigenetic factors, such as histone modifications and non-coding RNAs.

DNA methylation also controls expression of non-protein coding genes, including microRNAs (miRNAs) and long non–coding RNAs (lncRNAs), which can in turn impact global gene expression profiles (Lopez-Serra and Esteller 2012, Li, Zhang et al. 2015). Hence, alteration of DNA methylation patterns in CAMs can be responsible for aberrant regulation of microRNAs, which may consequently alter the expression of their target genes. In recent years, many studies have shown that microRNAs play an important role in the tumour-promoting properties of CAMs. Differential microRNAs expression profiles between patient-matched CAM and ATM samples were reported in breast cancer (Zhao, Sun et al. 2012), bladder cancer (Enkelmann, Heinzelmann et al. 2011) and gastric cancer (Yang, Yang et al. 2014). Recent work in gastric cancer showed that miR-149 modulates prostaglandin E2 (PGE2) and interleukin 6 (IL-6) signalling in the crosstalk between tumour cells and stroma. The study suggested that *H. pylori* infection induces secretion of PGE2 from tumour cells and promotes the methylation of miR-149 in CAMs. This subsequently induces the expression of its target gene IL-6 in stromal myofibroblasts (Li, Shan et al. 2015).

In summary, data presented in this Chapter suggest that DNA methylation is involved in an epigenetic regulation of biological pathways and processes involved in the tumour–promoting function of gastric CAMs. In particular, promoter DNA hypomethylation emerged as regulatory mechanism for transcriptional activation of genes involved in secretion and transport of molecules, including extracellular vesicular exosomes while promoter DNA hypermethylation emerged as regulatory mechanism for transcriptional repression of genes involved in pathology of gastrointestinal cancers and regulation of developmental processes.

# Chapter V

## Validation of DNA Methylation Patterns that Correlate with CAM-specific Gene Expression Profiles

## 5.1 Introduction

Combined genome–wide DNA methylation and gene expression profiling of gastric myofibroblasts has identified several interesting trends. To further validate observed differences in DNA methylation between gastric tumour derived myofibroblasts (CAMs) and non-tumour derived myofibroblasts (ATMs and NTMs) a targeted set of pyrosequencing assays were performed. In total 20 genomic CpG loci were chosen for verification in a set of 7 independent gastric patient-matched CAM and ATM samples and 4 unrelated gastric NTM samples. These 20 genomic loci selected are associated with the regulation of 15 genes reported to be involved in a range of functions, including development of the digestive system (*ASPH, FOXC1, FOXF1, HOXA5, RUX1, SMAD3*), chemotaxis of neuronal cell lines (*CD47, SMAD3*), transcription of DNA (*ASPH, FOXC1, FOXF1, HOXA5, RUX1, SMAD3, ZNF536*), cell migration (*CD47, FOXC1, FOXF1, MUC2, RUX1, SMAD3, SPON2, VPS28*), development of connective tissue (*CD47, HOXA5, RUNX1, SMAD3*), proliferation of cells (*ASPH, B4GALT6, CD47, FOXC1, FOXF1, HOXA5, MUC2, RUNX1, SMAD3, VPS28*), angiogenesis (*FOXC1, FOXF1, HOXA5, RUNX1, SMAD3*) and differentiation of cells (*CD47, FOXC1, FOXF1, HOXA5, RUNX1, SMAD3, ZNF536*). Finally, to assess whether differential DNA methylation in gene promoter regions effect transcriptional regulation of neighboring genes additional qPCR studies were also performed on selection of adjacent gene targets.

## 5.2 Aims

- To validate DNA methylation patterns and gene expression profiles identified in genome-wide studies in independent gastric CAM, ATM and NTM samples
- To verify if differently methylated CpG sites identified in gastric CAMs, ATMs and NTMs are concordant within methylation patterns in a border genomic region
- To investigate whether locus specific changes in DNA methylation in gastric CAMs may also affect expression of neighbouring genes

# 5.3 Results

## 5.3.1 Validation of CAM differentially methylated CpG loci by targeted pyrosequencing analysis

Pyrosequencing is a widely used method for the exact simultaneous quantification of the DNA methylation status of multiple CpG loci within a region of up to 350bp (Tost and Gut 2007). Representative pyrograms of assays performed in this study are shown in Figure 5.1.

In total 13 pyrosequencing assays were designed (Methods Table 2.5) to investigate multiple CpG loci, including the 20 CpG loci selected from genome-wide DNA methylation profiling with the Illumina 450k array. These genomic loci were selected on the basis of their differential methylation in both CAM vs ATM and CAM vs NTM comparisons (*FOXF1, FOXC1, MUC2, SMAD3, SPON2, ZNF536, HOXA5*), their localization within CpG islands (*FOXF1, FOXC1, HOXA5*) and their position downstream of the annotated transcription start site (TSS) of genes known to be implicated in cancer progression (*HOXA5, B4GALT6, CD47, MUC2, VPS28, ASPH, SPON2, SMAD3, ZNF536*). Other important criterion for target validation was the number of differentially methylated probes found to be associated with target genes in CAM vs ATM comparison. These include *FOXF1* (23), *FOXC1* (11), *ZNF536* (10), *HOXA5* (7), *SMAD3* (4), *SPON2* (4), *ASPH* (3) and *B4GALT6* (2).

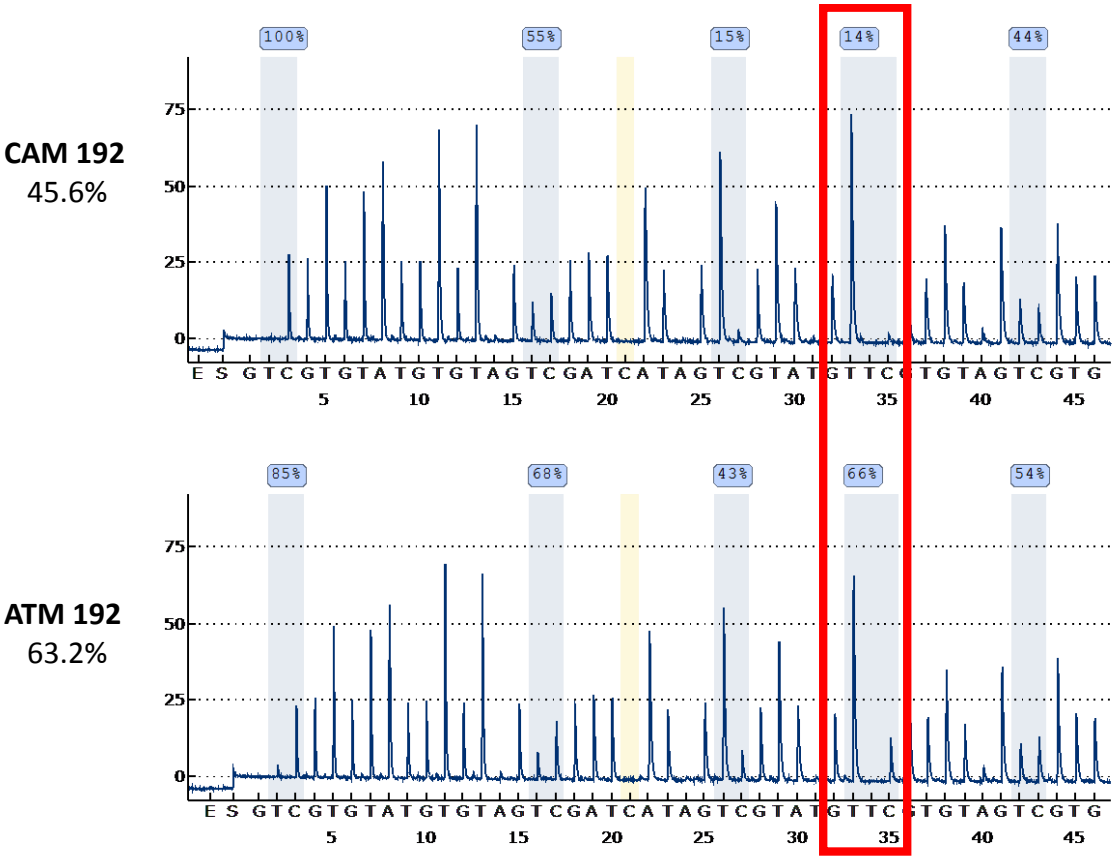


**Figure 5.1 Representative pyrograms from comparative pyrosequencing DNA methylation assays.** The analysis of five CpG sites in the *CD47* promoter region is shown as an example. *X*-axes show the nucleotide dispensation order while *Y*-axes show signal intensity levels. The percentage of methylation for individual CpG positions is shown above the respective positions (highlighted in blue). The C dispensation at position 21 (highlighted in yellow) controls for the full bisulfite conversion of DNA. The top pyrogram represents one CAM sample, whereas the bottom pyrogram represents the patient-matched ATM sample (patient 192). Percentage values next to pyrograms represent methylation index, the overall methylation level of the interrogated region. Red box highlights the CpG loci identified to be differentially methylated following analysis of CAM vs ATM Illumina 450k array data.

### 5.3.2 Pyrosequencing analysis identifies concordant DNA methylation changes within interrogated regions of gastric CAMs

A quantitative DNA methylation analysis was performed on 13 genomic regions in a set of 7 independent gastric patient-matched CAM and ATM samples and 4 unrelated gastric NTM samples. Biological replicates of each assay were performed to ensure reproducibility of sample preparation (Supplementary Table S5.1).

Significantly, pyrosequencing results confirmed the existence of distinct DNA methylation changes between CAMs and non-tumour derived myofibroblasts (ATMs and/or NTMs) in all of the 13 interrogated promoter regions thereby providing confidence that DNA methylation changes identified by the Illumina 450k arrays are concordant within a border region, which also includes neighbouring CpG sites (Figure 5.2, Figure 5.3B,C, Figure 5.4B,C, Figure 5.5B,C and Figure 5.6B,C).

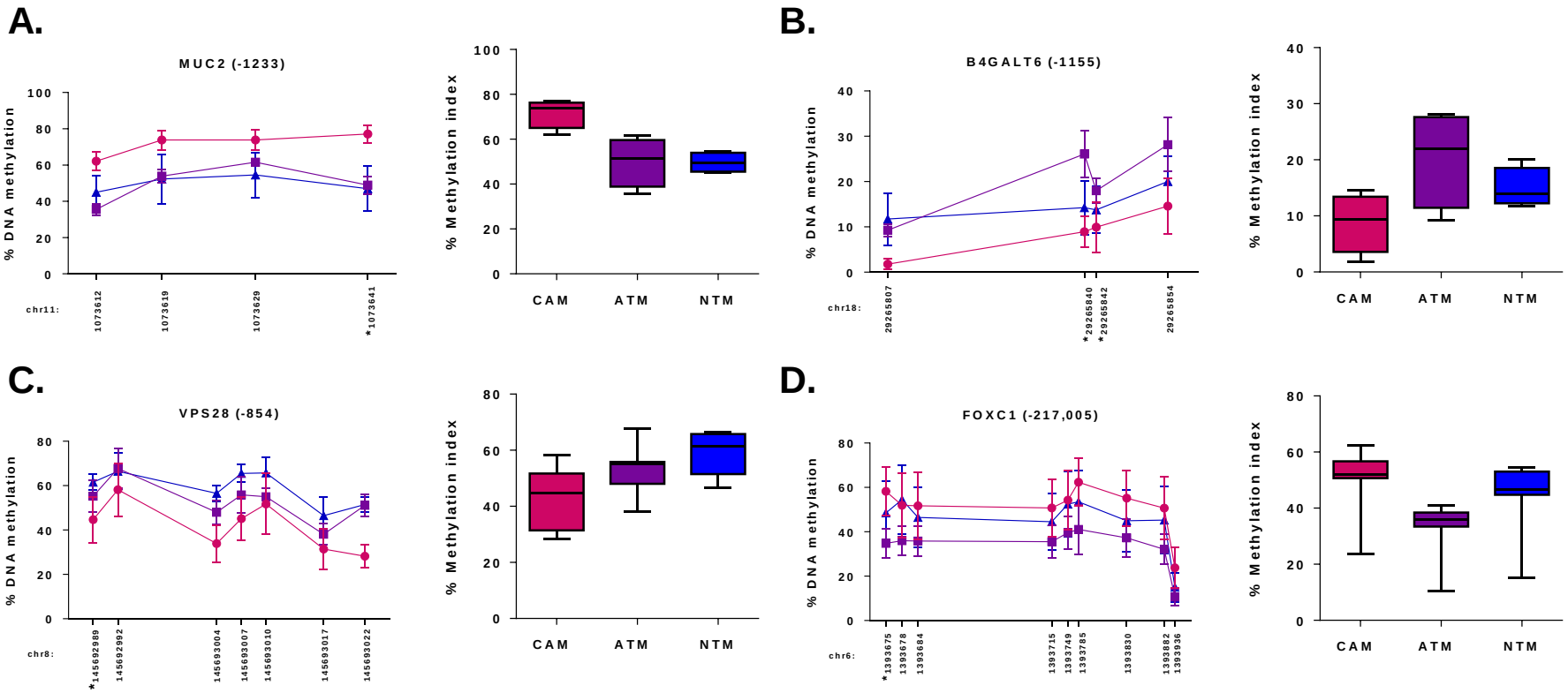


**Figure 5.2 Methylation changes at CpG loci identified in gastric CAMs and ATMs by Illumina 450k array are concordant within a broader local genomic region.** Pyrosequencing assays were performed to interrogate the promoter regions of 8 genes. Gene names are indicated at the top of each plot. Numbers in bracket indicate distance to the annotated transcription start site (TSS) of a given gene. Mean methylation values are plotted for each examined CpG site in the interrogated regions of gastric CAMs (n=7), ATMs (n=7) and NTMs (n=4). *X*-axis indicates chromosomal position of examined CpG loci. Positions marked with * correspond to the Illumina 450k probe; *magenta - CAMs, purple - ATMs, nave - NTMs.* Error bars represent SEM. Boxplots represent methylation distribution and mean for all CpG sites in the interrogated promoter region in CAMs (n=7), ATMs (n=7) and NTMs (n=4).

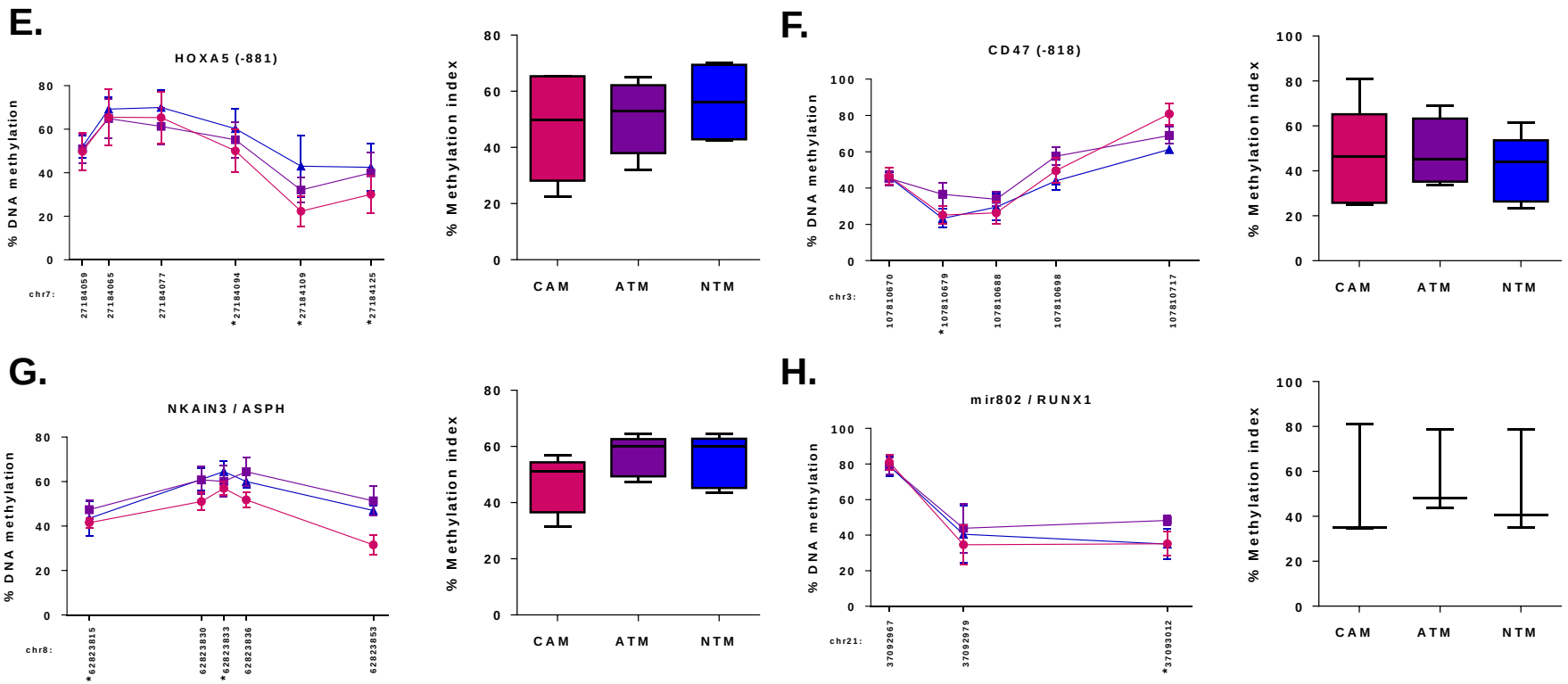


**Figure 5.2 (*Continued*) Methylation changes at CpG loci identified in gastric CAMs and ATMs by Illumina 450k array are concordant within a broader local genomic region.** Pyrosequencing assays were performed to interrogate the promoter regions of 8 genes. Gene names are indicated at the top of each plot. Numbers in bracket indicate distance to the annotated transcription start site (TSS) of a given gene. Mean methylation values are plotted for each examined CpG site in the interrogated regions of gastric CAMs (n=7), ATMs (n=7) and NTMs (n=4). *X*-axis indicates chromosomal position of examined CpG loci. Positions marked with * correspond to the Illumina 450k probe; *magenta - CAMs, purple - ATMs, nave - NTMs.* Error bars represent SEM. Boxplots represent methylation distribution and mean for all CpG sites in the interrogated promoter region in CAMs (n=7), ATMs (n=7) and NTMs (n=4).

### 5.3.3 Locus specific changes in CAM DNA methylation also affect expression of neighbouring genes

Four differentially methylated genomic regions spanning more than 600bp and up to 526,426bp identified by Illumina 450k array analysis were chosen for more detailed pyrosequencing and qPCR analyses, in order to assess if differential methylation levels also correlate with expression of neighbouring genes. The pyrosequencing and qPCR analyses were performed on a set of 7 independent gastric patient-matched CAM and ATM samples and 4 unrelated gastric NTM samples.

#### 5.3.3.1 Selection of target regions

Genomic regions selected for further investigation are associated with the regulation of five genes (*SMAD3, SPON2, ZNF536, FOXF1* and *FENDRR*), all of which have been previously implicated in cancer and tumour-stroma communication and were also found to be differentially methylated/expressed in this study.

#### 5.3.3.2 Promoter hypermethylation may repress *SMAD3* expression in gastric CAMs

SMAD3 is an intracellular protein that functions as a signal transducer and transcriptional modulator. It is activated by TGF-β (transforming growth factor - β) and is thought to have a role in carcinogenesis (Pruitt, Brown et al. 2014).

The *SMAD3* gene is located on chromosome 15 and the longest isoform spans 129,339 bp. Genome-wide DNA methylation analysis identified several differentially

methylated CpG loci between CAMs and non-tumour derived myofibroblasts (ATMs and NTMs) that are associated with *SMAD3* regulation, including 3 CpGs in the promoter region (Figure 5.3A). Interestingly, these 3 CpGs are classified as reprogramming-specific differentially methylated regions (rDMRs) by Illumina annotation. In addition, *SMAD3* was found to be downregulated in CAMs when compared to ATMs from gastric (previous gene expression study, Dr Helen Smith) and oesophageal (collaborative work with the Varro group, University of Liverpool) cancer (Supplementary Figure S5.1).

To assess whether *SMAD3* expression in gastric CAMs is regulated by promoter DNA methylation, two pyrosequencing assays were designed in *SMAD3* promoter region and SMAD3 TaqMan assay were performed on an independent set of patient-matched gastric CAM and ATM samples, which were cultured and processed in parallel. The pyrosequencing assays cover 224bp region and are spanning 10 CpG sites, including 2 CpGs that were identified by Illumina 450k array (Figure 5.3B).

This pyrosequencing analysis confirmed that the *SMAD3* promoter region is hypermethylated in CAMs when compared to either ATMs or NTMs (Figure 5.3B, Figure 5.3C), whereas qPCR analysis showed that *SMAD3* expression is downregulated in CAMs compared to ATMs (Figure 5.3D) confirming the previous observations in gastric and oesophageal CAMs (Supplementary Figure S5.1). Notably, DNA methylation levels in the *SMAD3* promoter region were found to be very similar in ATMs and NTMs (Figure 5.3C) therefore this genomic region might be used as a biomarker for gastric CAMs. Taken together, these experimental data provide a strong indication that *SMAD3* expression may be repressed by promoter DNA hypermethylation in gastric CAMs. Supplementary Figure S5.2A shows the observed correlation between *SMAD3* expression and promoter methylation assessed by qPCR and pyrosequencing analysis.

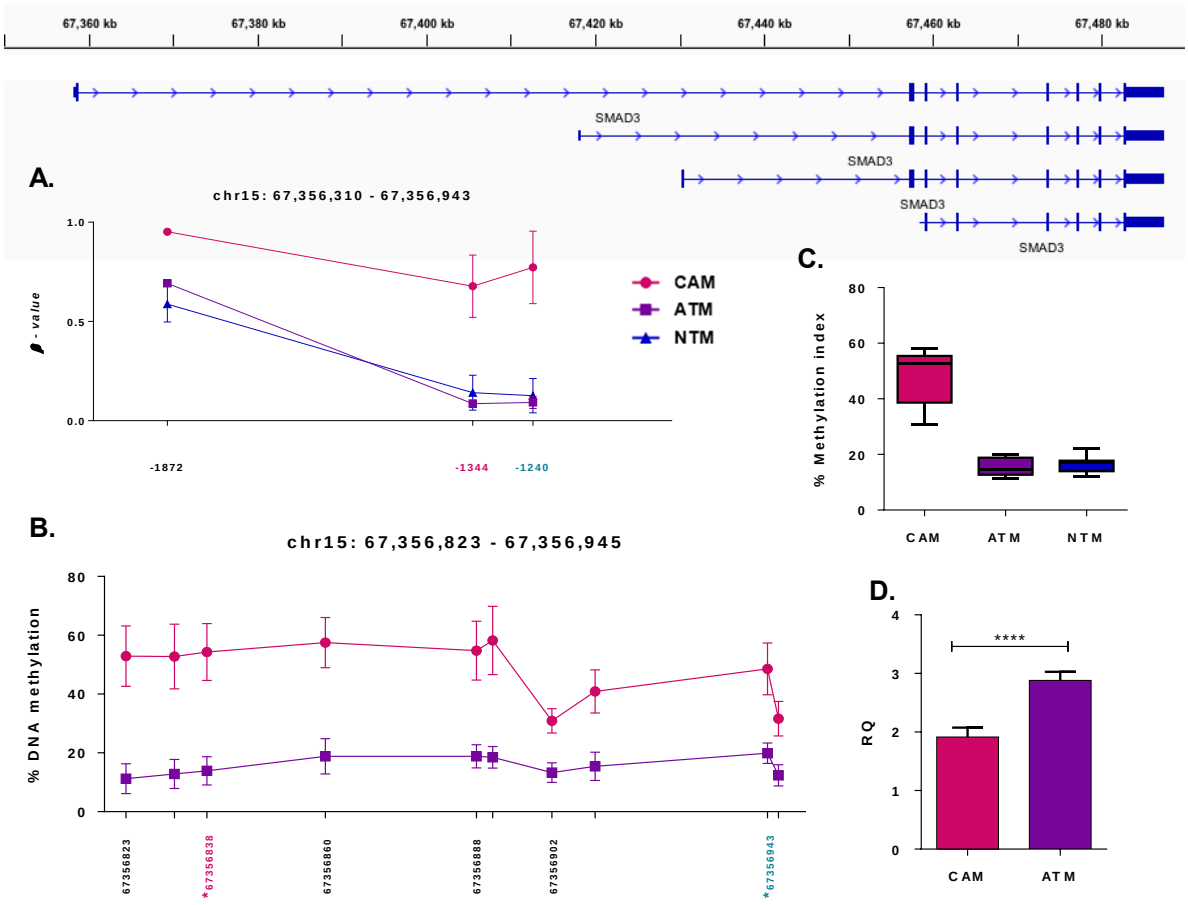


**Figure 5.3 DNA methylation in the *SMAD3* promoter region regulates *SMAD3* expression in gastric CAMs and ATMs. A.** Differentially methylated CpG loci identified by Illumina 450k array in the *SMAD3* promoter region. Mean β values (n=3) for probes identified as differentially methylated in CAM vs ATM and CAM vs NTM comparisons are plotted. The *X*-axis indicates distance of Illumina 450k probes to *SMAD3* transcription start site. Highlighted positions (in magenta and blue) are within the genomic region examined by pyrosequencing assays. **B.** Pyrosequencing analysis of the *SMAD3* promoter region in patient-matched CAM (n=7) and ATM (n=7) samples. Methylation means for 10 individual CpG sites in the interrogated promoter region are plotted. The *X*-axis indicates the chromosomal position of examined CpG sites. Positions marked with * correspond to the Illumina 450k probes. **C.** The overall methylation level of the *SMAD3* promoter region interrogated by pyrosequencing analysis. Boxplots represent methylation distribution and mean for 10 CpG sites in CAM (n=7), ATM (n=7) and NTM (n=4) samples. **D.** Quantitative PCR analysis of *SMAD3* gene expression in CAM (n=6) and ATM (n=6) samples; *t-test* ****$p<0.0001$. Error bars represent SEM.

#### 5.3.3.3 Promoter hypomethylation may induce *SPON2* expression in gastric CAMs

SPON2 is an extracellular matrix protein that has been implicated to have a role in axon guidance, cell adhesion and cellular response to lipopolysaccharide (Smith, Aleksic et al. 2012). It has been reported as a component of extracellular exosomes (Mathivanan, Fahner et al. 2012).

The *SPON2* gene is located on chromosome 4 and the longest isoform spans 42,030 bp. Genome-wide DNA methylation analysis identified several differentially methylated CpG loci between gastric CAMs and non-tumour derived myofibroblasts (ATMs and NTMs) that are associated with *SPON2* regulation, including 5 CpGs in the promoter region (Figure 5.4A). Notably, the genomic region associated with *SPON2* regulation was also differentially methylated in oesophageal CAMs compared to their patient-matched ATMs, including 3 CpGs in the promoter region (Supplementary Figure S5.3). Genome-wide expression analysis showed that *SPON2* is upregulated in gastric CAMs compared to both ATMs and NTMs (Table 4.1 and Figure 4.2). Supplementary Figure S5.4 shows simultaneous negative correlation between *SPON2* expression and promoter methylation and positive correlation in the gene body of SPON2 in patient-matched CAMs and ATMs, whereas Supplementary Figure S5.5 shows negative correlation between *SPON2* expression and promoter methylation in CAMs, ATMs and NTMs.

To assess whether *SPON2* expression in gastric CAMs is regulated by promoter DNA methylation, a pyrosequencing assay was designed in the *SPON2* promoter region and SPON2 TaqMan assay were performed on an independent set of patient-matched gastric CAM and ATM samples, which were cultured and processed in parallel. The pyrosequencing assay covers 117bp and is spanning 7 CpG sites, including 1 CpG that was identified by Illumina 450k array (Figure 5.4B).

This pyrosequencing analysis confirmed that the *SPON2* promoter region is hypomethylated in CAMs when compared to either ATMs or NTMs (Figure 5.4B and Figure 5.4C) and qPCR analysis showed that *SPON2* expression is upregulated in CAMs compared to ATMs (Figure 5.4D) confirming the previous observations (Supplementary Figure S5.4 and Figure S5.5). Interestingly, pyrosequencing data from this study show that the extent of *SPON2* promoter DNA methylation gradually changes in gastric stromal myofibroblasts, with low levels in CAMs, intermediate levels in patient-matched ATMs and highest levels being observed in NTMs (Figure 5.4C). Significantly, these trends show a good negative correlation with observed patterns of *SPON2* gene expression in these cells (Figure 4.2). Taken together, these data provide a strong indication that *SPON2* expression may be regulated by cancer-induced differential promoter DNA methylation in gastric CAMs, ATMs and NTMs. Supplementary Figure S5.2B shows the observed correlation between *SPON2* expression and promoter methylation assessed by qPCR and pyrosequencing analysis.

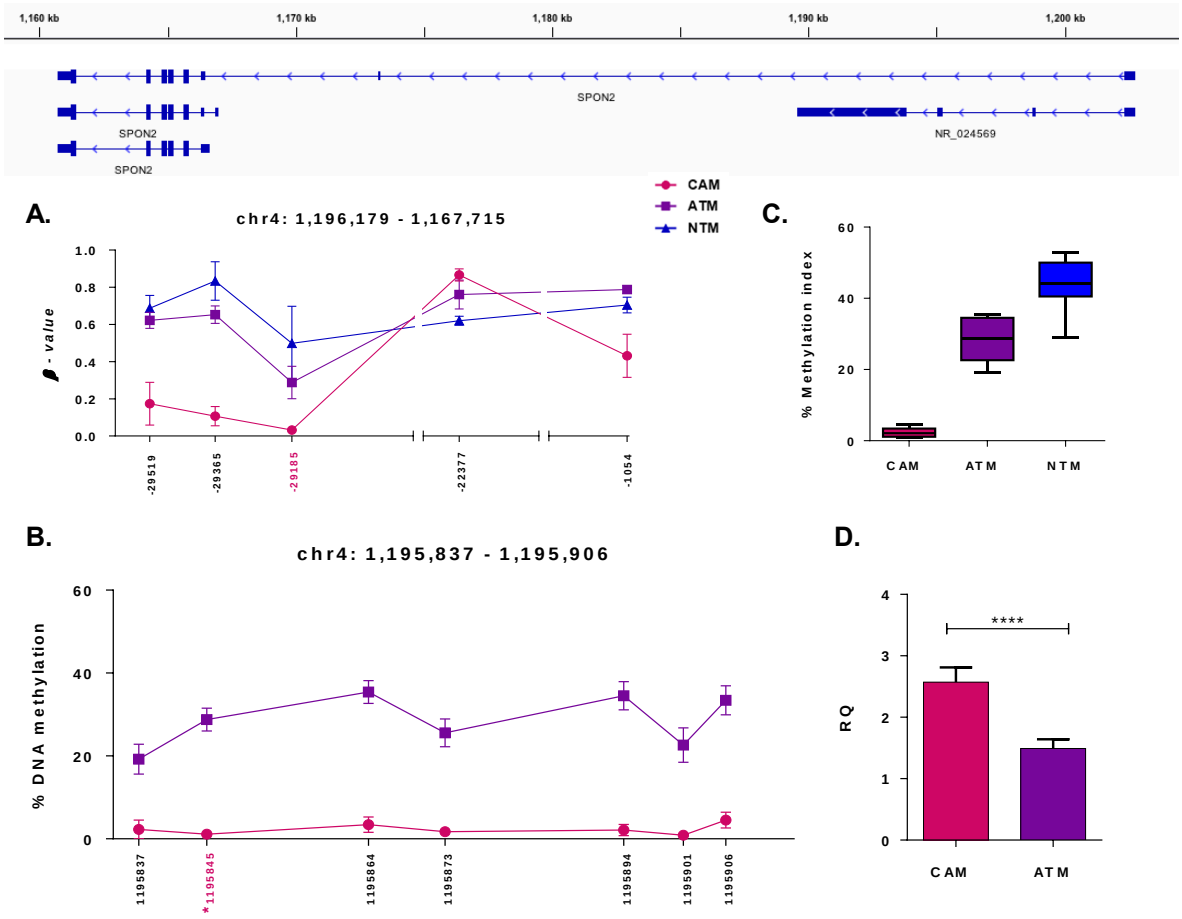


**Figure 5.4 DNA methylation in the *SPON2* promoter region correlates with *SPON2* gene expression profiles in gastric CAMs and ATMs. A.** Differentially methylated CpG sites identified by Illumina 450k array in the *SPON2* promoter region. Mean β values (n=3) for probes identified as differentially methylated in CAM vs ATM and CAM vs NTM comparisons are plotted. The *X*-axis indicates distance of Illumina 450k probes to *SPON2* transcription start site. Position highlighted in magenta is within the genomic region examined by pyrosequencing assay. **B.** Pyrosequencing analysis of the *SPON2* promoter region in patient-matched CAM (n=7) and ATM (n=7) samples. Methylation means for 7 individual CpG sites in the interrogated promoter region are plotted. The *X*-axis indicates the chromosomal position of examined CpG sites. Position marked with * corresponds to the Illumina 450k probe highlighted in magenta. **C.** The overall methylation level of the *SPON2* promoter region interrogated by pyrosequencing analysis. Boxplots represent methylation distribution and mean for 7 CpG sites in CAM (n=7), ATM (n=7) and NTM (n=4) samples. **D.** Quantitative PCR analysis of *SPON2* gene expression in CAM (n=6) and ATM (n=6) samples; *t-test* ****$p<0.0001$*. Error bars represent SEM.

#### 5.3.3.4 Promoter hypermethylation may repress *ZNF536* expression in gastric CAMs

ZNF536 is a zinc finger protein that may be involved in transcriptional regulation, negative regulation of neuron differentiation and negative regulation of retinoic acid receptor signalling (Qin, Ren et al. 2009). Gene ontology enrichment analysis performed in Chapter IV showed that ZNF536 may have a role in advance malignant tumour, regulation of developmental process and regulation of signal transduction.

The *ZNF536* gene is located on chromosome 19. Genome-wide DNA methylation analysis identified several differentially methylated CpG loci between CAMs and non-tumour derived myofibroblasts (ATMs and NTMs) that are associated with *ZNF536* regulation, including 4 CpGs in the promoter region (Figure 5.5A). Interestingly, some of these CpGs are classified as cancer-specific differentially methylated regions (cDMR) by Illumina annotation. Genome-wide expression analysis showed that *ZNF536* is downregulated in CAMs compared to both ATMs and NTMs (Table 4.1, Figure 4.2). Supplementary Figure S5.6 shows simultaneous negative correlation between *ZNF536* expression and promoter methylation and positive correlation in gene body of *ZNF536* in patient-matched CAMs and ATMs, whereas Supplementary Figure S5.7 shows correlations in CAMs, ATMs and NTMs.

To assess whether *ZNF536* expression in gastric CAMs is regulated by promoter DNA methylation, a pyrosequencing assay was designed in *ZNF536* promoter region and ZNF536 TaqMan assay were performed on an independent set of patient-matched gastric CAM and ATM samples, which were cultured and processed in parallel. The pyrosequencing assay covers 77bp and is spanning 3 CpG sites, including 1 CpG that was identified by Illumina 450k array (Figure 5.5B).

This pyrosequencing analysis again confirmed the pattern of DNA methylation identified by Illumina 450k array in CAMs and non-tumour derived myofibroblasts (ATMs and NTMs; Figure 5.5B and Figure 5.5C). Also, qPCR analysis showed that *ZNF536* expression is downregulated in CAMs compared to ATMs (Figure 5.5D), thus confirming previous observations (Supplementary Figure S5.6 and Figure S5.7). Notably, DNA methylation of CpG loci located in the promoter region 43bp downstream of *ZNF536* transcription start site (TSS) was hypermethylated in CAMs compared to both ATMs and NTMs (Figure 5.5C). DNA methylation pattern at this locus was very similar in ATMs and NTMs thus this locus may represent a potential biomarker/signature component for gastric CAMs. Taken together, these experimental data provide a strong indication that *ZNF536* expression may be regulated by promoter DNA methylation, in particular the methylation at CpG located 43bp downstream of *ZNF536* TSS, in gastric CAMs, ATMs and NTMs. Supplementary Figure S5.2C shows the observed correlation between *ZNF536* expression and methylation at the CpG located 43bp downstream of *ZNF536* TSS assessed by qPCR and pyrosequencing analysis.

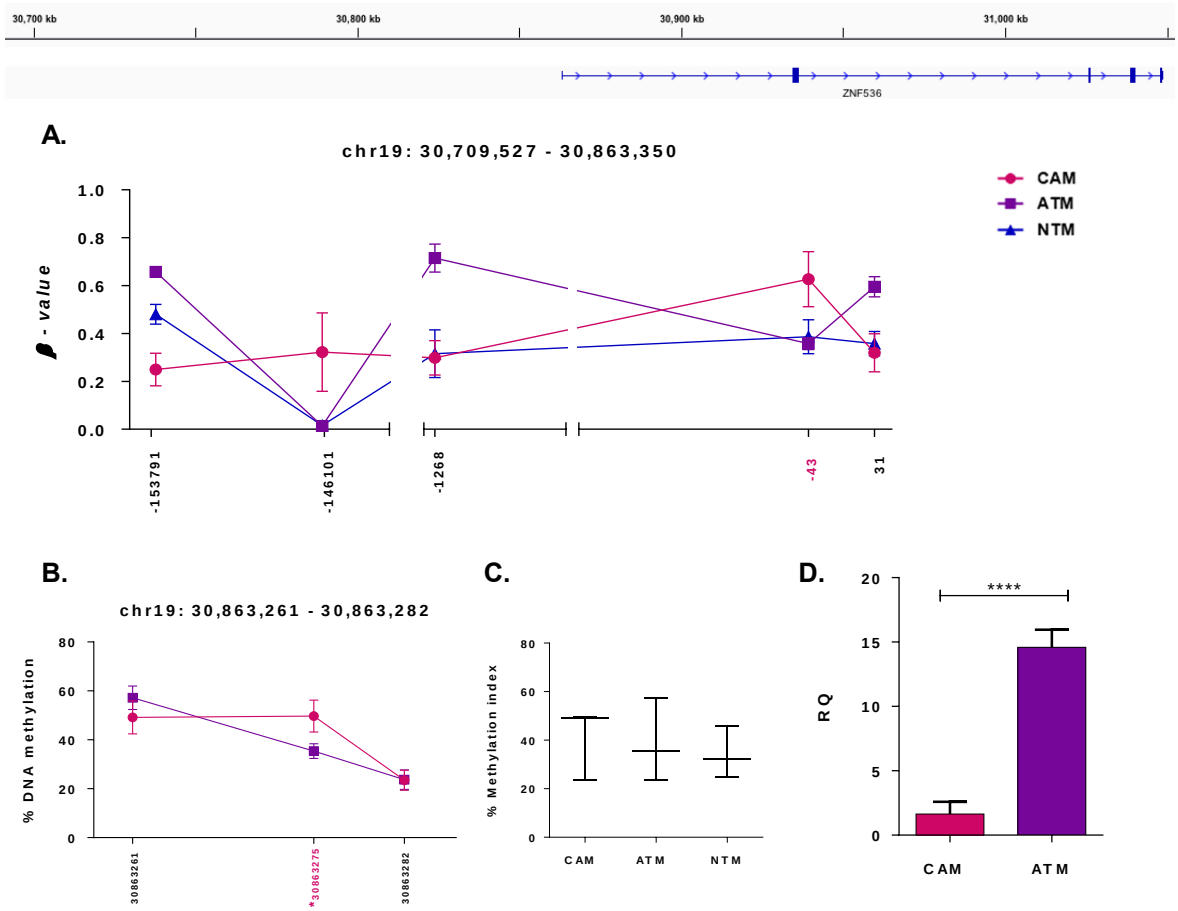


**Figure 5.5 DNA methylation levels within the *ZNF536* promoter region correlate with *ZNF536* gene expression profiles in gastric CAMs and ATMs. A.** Differentially methylated CpG sites identified by Illumina 450k array in the *ZNF536* promoter region. Mean β values (n=3) for probes identified as differentially methylated in CAM vs ATM and CAM vs NTM comparisons are plotted. The *X*-axis indicates distance of Illumina 450k probes to *ZNF536* transcription start site. Position highlighted in magenta is within the genomic region examined by pyrosequencing assay. **B.** Pyrosequencing analysis of the *ZNF536* promoter region in patient-matched CAM (n=7) and ATM (n=7) samples. Methylation means for 7 individual CpG sites in the interrogated promoter region are plotted. The *X*-axis indicates chromosomal position of examined CpG sites. Position marked with * corresponds to the Illumina 450k probe highlighted in magenta. **C.** The overall methylation level of the *ZNF536* promoter region interrogated by pyrosequencing analysis. Boxplots represent methylation distribution and mean for 3 CpG sites in CAM (n=7), ATM (n=7) and NTM (n=4) samples. **D.** Quantitative PCR analysis of *ZNF536* gene expression in CAM (n=6) and ATM (n=6) samples; *t-test* ****$p<0.0001$. Error bars represent SEM.

### 5.3.3.5 DNA hypermethylation may repress the expression of *FOXF1* and *FENDRR* in gastric CAMs

As mentioned in Chapter III (section 3.3.2), genome–wide differential DNA methylation analysis between CAMs and non-tumour derived myofibroblasts (ATMs and NTMs) was performed at individual CpG loci and genomic regions. Tiling analysis is one of the region-level analysis methods used to identify differentially methylated regions over the whole genome with a window size of 5 kilobases (Supplementary Figure S4.1D). A region on chromosome 16 (ch16: 86,012,085 – 86,538,511) spanning 526,426bp was identified as one of the largest differentially methylated regions in gastric CAM vs ATM and CAM vs NTM comparisons (Figure 5.6A and Supplementary Figure S5.8). Notably, a smaller part of this region (ch16: 86,528,753 – 86,538,425) spanning 9,673bp was also identified as being differentially methylated in a parallel oesophageal CAM vs ATM comparison (Supplementary Figure S5.9 and Figure 3.11). Interestingly, many of the CpG loci within this region are classified as cancer-specific differentially methylated regions (cDMR) by Illumina annotation. Differential DNA methylation within this region may regulate the expression of *FOXF1* and several long non-coding RNAs (lncRNAs), such as *LOC146513, LINC00917* and *FENDRR*.

FOXF1 belongs to the forkhead family of transcription factors; it may play a role in embryonic development however the specific function of this gene has not yet been determined (Pruitt, Brown et al. 2014). *FENDRR* is lncRNA that is transcribed bidirectionally with *FOXF1* on the opposite strand (Pruitt, Brown et al. 2014). Significantly, *FENDRR* was recently found to play a role in the progression of gastric cancer (Xu, Huang et al. 2014).

To assess whether *FOXF1* and *FENDRR* expression in gastric CAMs might be regulated by DNA methylation within this region, a pyrosequencing assay was

designed in *FOXF1* promoter region (that overlaps with *FENDRR*) and TaqMan assays for FOXF1 and FENDRR (splice variant 1 and 2) were performed on an independent set of patient-matched gastric CAM and ATM samples, which were cultured and processed in parallel. The pyrosequencing assay covers 104bp and is spanning 8 CpG sites, none of which were present on the Illumina 450k array (Figure 5.6B).

This analysis confirms that the FOXF1 promoter region is hypermethylated in CAMs when compared to either ATMs or NTMs (Figure 5.6B and Figure 5.6C). In addition, qPCR analysis showed that FOXF1 and FENDRR expression are both downregulated in CAMs compared to ATMs (Figure 5.6D). Notably, the pyrosequencing analysis revealed that the FOXF1 promoter shows a gradual change in DNA methylation status (CAMs>ATMs>NTMs) in gastric stromal myofibroblasts (Figure 5.6C). Taken together, these data provide a strong indication that FOXF1 and FENDRR expression may be regulated by DNA methylation within this region in gastric CAMs and ATMs. Supplementary Figure S5.2D shows the observed correlation between FOXF1 expression and promoter methylation assessed by qPCR and pyrosequencing analysis.

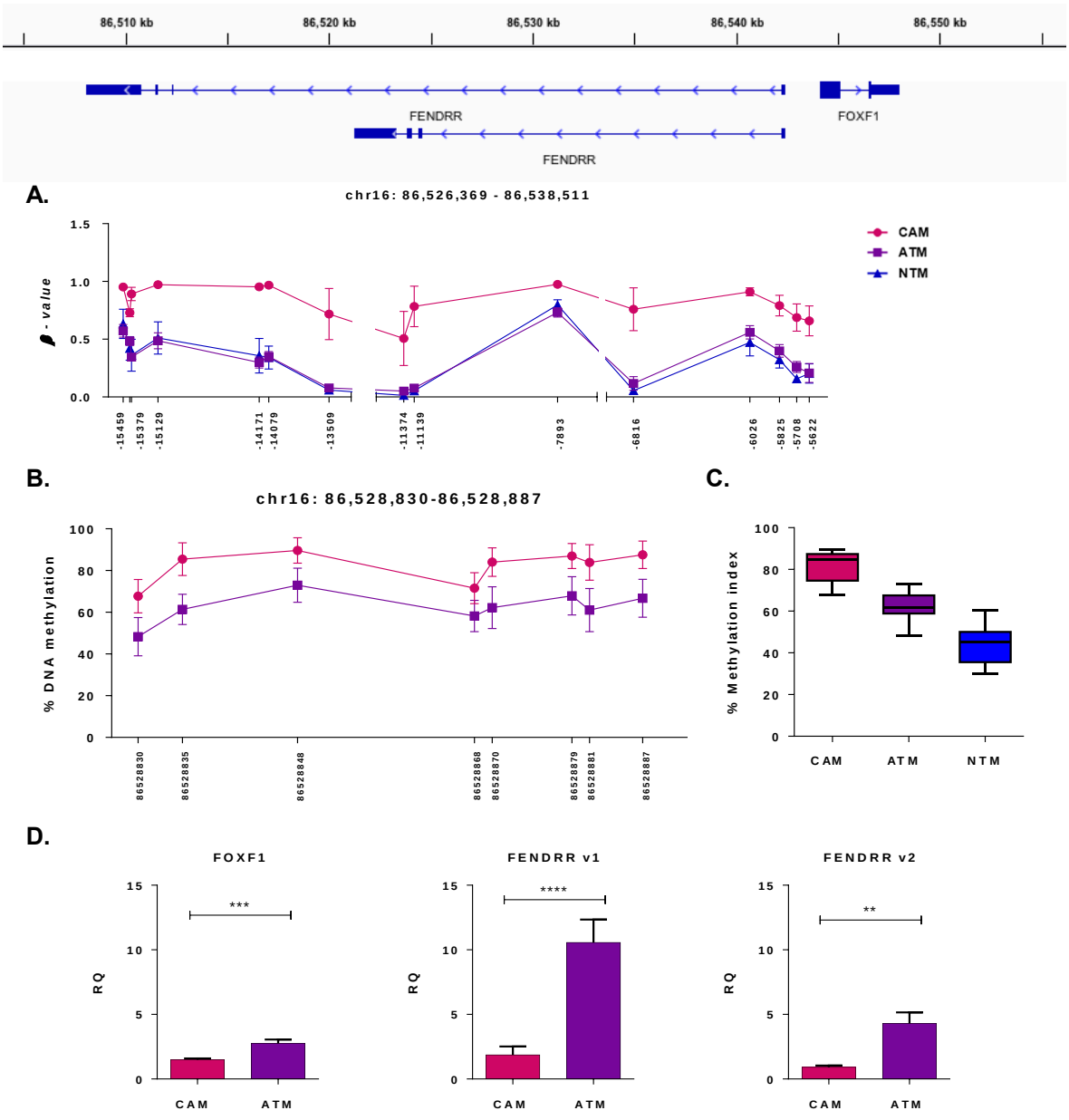


**Figure 5.6 DNA methylation pattern in the genomic region associated with regulation of *FOXF1* and *FENDRR* expression in gastric CAMs and ATMs. A.** Differentially methylated CpG sites identified by Illumina 450k array in the region downstream of the *FOXF1* transcription start site. Mean β values (n=3) for probes identified as differentially methylated in CAM vs ATM and CAM vs NTM comparisons are plotted. The *X*-axis indicates distance of Illumina 450k probes to *FOXF1* transcription start site. **B.** Pyrosequencing analysis of the *FOXF1* promoter region in patient-matched CAM (n=7) and ATM (n=7) samples. Methylation means for 8 individual CpG sites in the interrogated promoter region are plotted. The *X*-axis indicates chromosomal position of examined CpG sites. **C.** The overall methylation level of the *FOXF1* promoter region interrogated by pyrosequencing assay. Boxplots represent methylation distribution and mean for 8 CpG sites in CAM (n=7), ATM (n=7) and NTM (n=4) samples. **D.** Quantitative PCR analysis of *FOXF1* and *FENDRR* gene expression in CAM and ATM samples; *t-test FOXF1 (n=5) ***p=0.0008; FENDRR v1 (splice variant 1; n=4) ****p<0.0001; FENDRR v2 (splice variant 2; n=4) **p=0.0028*. Error bars represent SEM.

## 5.4 Discussion

In order to further investigate correlated DNA methylation and gene expression profiles identified in earlier Chapters further pyrosequencing and qPCR studies were performed on an additional set of myofibroblast samples derived from gastric cancer patients, which were not included in our initial array profiling studies (Chapters III and IV). Data from these extended studies provide further strong evidence that CAM-specific promoter DNA methylation patterns may regulate the expression of associated genes. In addition, pyrosequencing analysis confirmed the existence of DNA methylation changes within border genomic regions, spanning several neighbouring CpG sites, between gastric CAMs and non-tumour derived myofibroblasts (ATMs and NTMs).

The transforming growth factor-β (TGF-β) signalling pathway plays an important role in both development, and in the initiation and progression of cancer. It exerts various context-dependent effects on both epithelial cancer cells and stromal cells (Ikushima and Miyazono 2010, Calon, Tauriello et al. 2014). This study strongly suggests that DNA methylation may confer a mechanism for the transcriptional regulation of key TGF-β signalling components in gastric CAMs and ATMs. In particular, promoter hypermethylation may be associated with *SMAD3* downregulation in gastric CAMs compared to their patient-matched ATMs. Notably, the degree of *SMAD3* promoter methylation was very similar in ATMs and NTMs, thus this genomic region might be used as a proxy for gastric CAMs identification. SMAD3 is an important intracellular mediator of TGF-β signalling. Upon TGF-β stimulation SMAD3 is activated by receptor-mediated phosphorylation and forms complexes with other SMAD proteins, which then are translocated to the nucleus where they, in cooperation with other transcription factors, co-activators and co-repressors, regulate the transcription of TGF-β-responsive genes. The activity and stability of SMAD proteins are carefully regulated by a plethora of post-translational modifications (Heldin and Moustakas 2012). Significantly, a previous study in *Smad3*-null mice showed accelerated wound healing and an

impaired local inflammatory response (Ashcroft, Yang et al. 1999) consistence with CAM tumour-promoting properties. In addition, genome-wide DNA methylation analysis identified several differentially methylated CpG loci between CAMs and ATMs that are associated with *TGFBR2* (the type II TGF-β receptor) regulation. In particular, the integration analysis (presented in Chapter IV) revealed that *TGFBR2* is downregulated in CAMs compared to ATMs (*TGFBR2* downregulation is also reported in previous gene expression study on 12 patient-matched CAMs and ATMs, Supplementary Figure S4.1) and its transcriptional repression might be mediated by simultaneous hypermethylation in promoter region and hypomethylation in gene-body (Figure 4.8). TGFBR2 is critical for initiation of the TGF-β signalling cascade therefore loss of this protein abrogates TGF-β signalling. Bhowmick et al. generated mice in which TGFBR2 was selectively ablated in stromal fibroblasts and demonstrated that upon loss of TGFβ responsiveness in stromal fibroblasts, these mice spontaneously developed neoplastic lesions which were accompanied by stromal expansion (Bhowmick, Chytil et al. 2004). Interestingly, components of TGF-β signalling pathway, including receptors (TGFBR2) and transcription factors (SMAD4), are often inactivated and silenced in cancer (Grady and Markowitz 2002) indicating that CAMs might acquire some of the cancer-like phenotype. Together, these data show that cancer-induced DNA methylation changes in CAMs might affect the expression of key TGF-β signalling components resulting in a possible weakened TGFβ-TGFβR2-SMAD3 signalling axis in gastric CAMs compared to their patient-matched ATMs.

Long non-coding RNAs (lncRNAs) have recently emerged as important regulators of a wide variety of biological processes, including epigenetic silencing, transcriptional regulation, RNA processing, and RNA modification (Mercer, Dinger et al. 2009). Their aberrant expression has been reported in various cancers, including gastric cancer (Fang, Pan et al. 2015, Gu, Chen et al. 2015). DNA methylation is one of the mechanisms regulating the expression of lncRNAs (Lujambio, Portela et al. 2010). A recent study in breast cancer showed that aberrant DNA methylation patterns in the promoters of non-coding RNAs (microRNAs and lncRNAs) are more frequent

than aberrant DNA methylation patterns in promoters of protein-coding genes (Li, Zhang et al. 2015). The authors also predicted the potential biological functions of these aberrantly methylated non-coding RNAs and suggested that non-coding RNAs and protein-coding genes cooperatively mediate pathway dysregulation during the development and progression of breast cancer (Li, Zhang et al. 2015).

In this study, the genomic region associated with the regulation of *FENDRR* and *FOXF1* was investigated in detail in order to assess if there is a correlation between differential DNA methylation and expression of these genes in gastric CAMs and ATMs. *FENDRR* (FOXF1 adjacent non-coding developmental regulatory RNA) is lncRNA transcribed bidirectionally with *FOXF1* on the opposite strand. It binds to polycomb repressive complex 2 (PRC2) to epigenetically regulate the expression of its target gene (Grote, Wittler et al. 2013). FOXF1 (forkhead box-F1) is a transcription factor that plays a key role in regulating embryonic development (Mahlapuu, Ormestad et al. 2001) and mesenchymal-epithelial interaction in lung and gut morphogenesis (Mahlapuu, Enerback et al. 2001). It has been suggested that FOXF1 may function as a tumour suppressor as it is inactivated by DNA methylation in breast cancer (Lo, Lee et al. 2010) and its low expression was also reported in prostate cancer (Watson, Doggett et al. 2004). The experimental data presented in this Chapter showed that DNA hypermethylation within the interrogated region may lead to transcriptional repression and reduced expression of *FENDRR* and *FOXF1* in gastric CAMs. Interestingly, this genomic region is commonly hypermethylated in gastric cancer (Supplementary Figure S5.10 and Figure S5.11) suggesting that CAMs might also acquire some of the cancer-like DNA methylation patterns. Supplementary Figure S5.10 shows gastric cancer cell lines data for the *FOXF1* promoter region interrogated by pyrosequencing in AGS and MKN45 cells, whereas Supplementary Figure S5.11 represents stomach adenocarcinoma data from the Cancer Genome Atlas (TCGA) corresponding to the wider differentially methylated region identified by Illumina 450k array in CAMs compared to either ATMs or NTMs (as shown in Figure 5.6A).

Significantly, Xu et al. showed that *FENDRR* regulates gastric cancer metastasis and is downregulated in gastric cancer cells compared to normal gastric epithelial cells. Low expression of *FENDRR* was significantly correlated with poor prognosis and aggressive tumour characteristics, such as greater invasion depth, higher tumour stage, and lymphatic metastasis (Xu, Huang et al. 2014). Saito et al. demonstrated that FOXF1 regulates tumour-promoting properties of lung cancer-associated fibroblasts (CAFs), including upregulation of α-SMA, secretion of paracrine growth factors, induction of contractile and migration-stimulatory properties, and *in vivo* pro-tumorigenic effects. Using *FOXF1* gain-and loss-of-function fibroblasts, they showed that FOXF1 controls the expression of hepatocyte growth factor (HGF) and fibroblast growth factor-2 (FGF-2), well-established CAM-derived stimulators of tumour growth. The expression of these factors was upregulated in *FOXF1* transduced cells and downregulated in the cells where *FOXF1* was silenced (Saito, Micke et al. 2010). Interestingly, *HGF* was found to be downregulated in CAMs compared to ATMs in a gene-wide expression study presented in Chapter IV which is in agreement with lower *FOXF1* expression in these cells. Together, these data suggest that cancer-induced DNA hypermethylation within the interrogated region in gastric CAMs may affect the expression of *FENDRR* and *FOXF1* resulting in possible more aggressive cancer phenotype.

The correlation analysis of the genome-wide DNA methylation and gene expression profiles in gastric CAMs compared to ATMs revealed that promoter hypomethylation in gastric CAMs may be an important mechanism for transcriptional activation of genes associated with extracellular region and exosomes. In this Chapter the *SPON2* and *VPS28* promoter regions were analysed by pyrosequencing. SPON2 and VPS28 were both identified as components of the extracellular vesicular exosomes as reported in ExoCarta database (Mathivanan, Fahner et al. 2012). SPON2 is an extracellular matrix protein that has been implicated in axon guidance, cell adhesion and cellular response to lipopolysaccharide, whereas VPS28 is part of the endosomal complexes that function in the transport and sorting of proteins into subcellular vesicles. SPON2

was detected in exosomes derived from prostate cancer cells, whereas VPS28 was detected in exosomes derived from multiple cancer types, including bladder, colorectal, melanoma, ovarian and prostate cancer. Pyrosequencing data from this study confirmed that *SPON2* and *VPS28* promoter regions are hypomethylated in gastric CAMs compared to either ATMs or NTMs (Figure 5.4 and Figure 5.2C). In addition, the qPCR analysis revealed that promoter hypomethylation may transcriptionally induce *SPON2* expression in gastric CAMs. Notably, the methylation pattern of *SPON2* promoter shows inverse correlation with the expression of *SPON2* in CAMs, ATMs and NTMs with *SPON2* being hypomethylated-induced in CAMs and hypermethylated-repressed in NTMs (Figure 5.4C, Figure 5.4D and Figure 4.2). This observation was also confirmed at the protein level (Chapter VI section 6.3.4). Comparative CAM vs NTM secretome profiling revealed that SPON2 protein was upregulated 3 fold in the CAM secretome compared to the NTM secretome.

Gene ontology enrichment analysis of differentially methylated loci identified in gastric CAMs vs ATMs revealed that many of these differentially methylated loci might regulate genes involved in nervous system development and neurogenesis (Table 3.1). ZNF536 is a zinc finger protein that is specifically expressed in the brain and negatively regulates neuron differentiation by repressing retinoic acid-induced gene transcription. Overexpression of ZNF536 results in an inhibition of neuronal differentiation, while depletion or mutation of the *ZNF536* gene results in enhanced differentiation (Qin, Ren et al. 2009). Genome-wide DNA methylation analysis identified several differentially methylated CpG loci between CAMs and non-tumour derived myofibroblasts (ATMs and NTMs) that are associated with regulation of this gene. Pyrosequencing analysis in *ZNF536* promoter region showed that hypermethylation at CpG located 43bp downstream of *ZNF536* transcription start site may repress *ZNF536* expression in gastric CAMs as shown by qPCR analysis. Together, these observations suggest that gastric CAMs might exhibit some of the reported neuronal phenotypes.

In summary, the experimental data obtained using pyrosequencing and qPCR analysis performed on a larger set of patient CAM and ATM samples show for the first time that cancer-induced DNA methylation changes in gastric stromal myofibroblasts may have a functional role in regulating the expression of associated genes. In some instances these DNA methylation changes may prove to be useful signatures, which may be used in combination as molecular biomarkers for gastric CAMs identification, or patient stratification. In addition, the pyrosequencing analysis confirmed the DNA methylation patterns identified by Illumina 450k array in CAMs and non-tumour derived myofibroblasts (ATMs and NTMs).

# Chapter VI

## Effects of Hypoxia on Gene Expression and Secretion of Pro-tumorigenic Factors in Gastric Stromal Myofibroblasts

## 6.1 Introduction

Tumour hypoxia is another important microenvironmental factor contributing to aspects of tumour progression including tissue invasion and metastatic spread. As tumours grow and increase size sub-populations of both cancer cells and stromal cells are exposed to hypoxia. Reduced oxygen availability activates hypoxia-inducible factors (HIFs), which turn on transcription of target genes, which encode proteins that play important roles in many aspects of cancer biology (Kaelin and Ratcliffe 2008, Casazza, Di Conza et al. 2014). A lot of effort has been devoted to studying the consequences of tumour hypoxia on cancer cells but relatively little is known about hypoxia-induced effects on the stromal component of the tumour microenvironment, including CAMs. Previously, it was proposed that hypoxia may have a role in modulating the differentiation and activity of these cells (Giaccia and Schipani 2010). Data presented in previous Chapters provide strong evidence for the existence of robustly maintained CAM-specific DNA methylation signatures. However, the extent to which exposure to hypoxia may contribute to or change myofibroblast epigenetic programing remains unexplored.

Hypoxia is strongly associated with tumour progression, metastasis and resistance to therapy. Comito et al. demonstrated that stromal fibroblasts synergize with hypoxic oxidative stress to enhance melanoma aggressiveness. In this context, hypoxia-mediated oxidative stress was found to be mandatory for the activation of dermal fibroblasts and secretion of cytokines and pro-migratory factors, as well as actively promoting the invasion and chemotaxis of melanoma cells (Comito, Giannoni et al. 2012).

In this study, we attempt to address some of the outstanding questions relating to the modulation of stromal myofibroblast activity under hypoxic conditions. In particular: Does hypoxia induce a more aggressive CAM phenotype?; Do NTMs

become more aggressive or CAM-like under hypoxic conditions?; If so do DNA methylation patterns also change?; Does hypoxia induce common or selective changes in CAM, ATM and NTM gene expression profiles?; and finally, does hypoxia have a differential effect on the composition of CAM and NTM secretome profiles?

In order to address these questions, an integrated multi-omics approach was used to obtain complementary data from primary gastric patient-matched CAM and ATM and unrelated NTM samples cultured under hypoxic or normoxic conditions. Firstly, the relative ability of secreted factors derived from hypoxic CAMs and NTMs to induce the migration and/or proliferation of AGS gastric cancer cells was measured. Genome-wide hypoxia-induced DNA methylation and gene expression profiles of CAMs, ATMs and NTMs used in these studies were explored using Illumina Infinium HumanMethylation450 BeadChip and Illumina HumanHT-12v4 Expression BeadChip arrays. Samples for both DNA methylation and gene expression were prepared in parallel in order to minimize technical artefacts. Finally, conditioned media from hypoxic CAMs and hypoxic NTMs were profiled by LC-MS/MS to identify myofibroblast secretome signatures that may contribute to hypoxia-induced CAM and NTM phenotype/properties.

## 6.2 Aims

- To determine the relative ability of conditioned media (CM) from hypoxic CAMs or hypoxic NTMs to enhance cancer cell migration and proliferation
- To identify hypoxia-induced gene expression signatures in gastric CAMs, ATMs and NTMs
- To determine whether hypoxia - induced changes in CAMs, ATMs and NTMs gene expression profiles may be regulated by DNA methylation
- To predict the relative biological effects of hypoxia on CAMs, ATMs and NTMs
- To identify secreted proteins which may contribute to the pro-migratory and proliferative effects of hypoxic CAMs and NTMs based on their hypoxia-induced gene expression signatures
- To identify and quantify CAM and NTM hypoxia-induced secretomes
- To correlate CAM and NTM hypoxia-induced gene expression and secretome profiles
- To identify pathways and processes that may contribute to CAM-like properties of hypoxic NTMs.

## 6.3 Results

### 6.3.1 Hypoxia enhances CAM-induced cancer cell migration and promotes CAM-like properties in NTMs

Boyden chamber migration assays and EdU cell proliferation assays were used to assess the effects of hypoxia (1% $O_2$) on the ability of CAMs or NTMs to enhance AGS cell migration or proliferation. AGS cell migration and proliferation assays were performed in parallel with fresh CAM or NTM conditioned media (CM) prepared under both hypoxic (hypoxic-CM) and normoxic (ctrl-CM) conditions (as described in Methods section 2.4.2 and section 2.5).

#### 6.3.1.1 Gastric cancer cell migration assays

CAM and NTM conditioned media prepared under hypoxic conditions (CAM/NTM-hypoxic-CM) was found to enhance AGS cell migration compared to corresponding control conditioned media prepared under normoxic conditions (CAM/NTM-ctrl-CM) (Figure 6.1A) thereby demonstrating that hypoxia induces the expression of factors that enhance gastric cancer cell migration. Significantly, the group mean data of AGS cell migration assays show that the pro-migratory effect of CAM-hypoxic-CM is higher than that observed for NTM-hypoxic-CM (Figure 6.1B).

#### 6.3.1.2 Gastric cancer cell proliferation assays

Data from this study show that both CAM and NTM conditioned media prepared under hypoxic conditions (CAM/NTM-hypoxic-CM) exert different effects on AGS cell proliferation compared to their respective control conditioned media

prepared in normal condition (CAM/NTM-ctrl-CM). Interestingly, our data shows that CAM-hypoxic-CM reduces AGS cell proliferation, whereas NTM-hypoxic-CM increases AGS cell proliferation when compared to their respective control conditioned media prepared in normal condition (CAM/NTM-ctrl-CM) (Figure 6.1C). Together these results indicate that hypoxia has a differential effect on the expression of cancer cell pro-proliferative factors in gastric CAMs and NTMs.

### 6.3.2 Identification of hypoxia-induced gene expression signatures in gastric CAMs, ATMs and NTMs

To identify hypoxia-induced gene expression signatures in gastric myofibroblasts purified from different tissue microenvironments Illumina HumanHT-12v4 Expression BeadChip arrays were perform on three patient-matched CAM and ATM samples and three independent NTM samples, all of which were exposed to either hypoxia (1% $O_2$) or normoxia (21% $O_2$) for 72 hours (as described in Methods section 2.3.2) resulting in 13381 expressed genes. Differential gene expression analysis was then performed between hypoxic and control myofibroblasts, resulting in the identification of: (i) 2467 genes that were consistently differentially expressed in hypoxic CAMs compared to control CAMs (Figure 6.2A); (ii) 2722 genes that were constantly differentially expressed in hypoxic NTMs compared to control NTMs (Figure 6.2B) and (iii) 2561 genes that were constantly differentially expressed in hypoxic ATMs compared to control ATMs (Figure 6.2C).

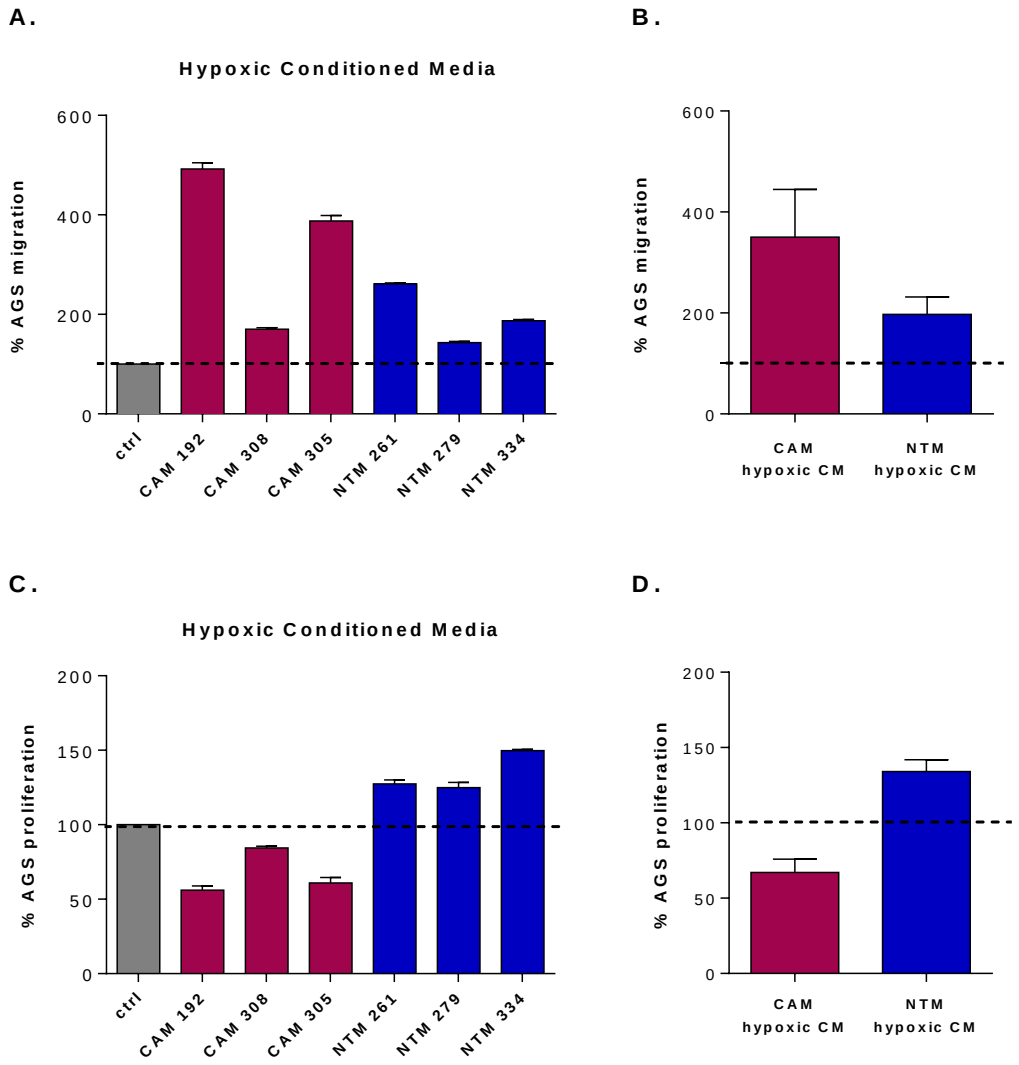


**Figure 6.1 Differential effects of CAM / NTM hypoxic conditioned media (CM) on AGS gastric cancer cell migration or proliferation. A & B.** CAM-hypoxic-CM (magenta) and NTM-hypoxic-CM (navy) induce AGS cell migration compared to respective CAM or NTM normoxic-control CM (grey). **A.** Individual patient data corrected for AGS basal migration (serum-free media) normalized to respective patient-specific control CM obtained from normoxia, *t-test p-value<0.05*; Error bars represent SEM of technical replicates. **B.** Group mean data of AGS cell migration in response to hypoxic-CM from CAMs (n=3) and NTMs (n=3). Error bars represent SEM. **C & D.** NTM-hypoxic-CM (navy) induces AGS cell proliferation whereas CAM-hypoxic-CM (magenta) reduces the ability of AGS cells to proliferate compared to respective CAM or NTM control CM (grey) obtained from normoxia. **C.** Individual patient data corrected for AGS basal proliferation (serum-free media) and normalized to respective patient-specific control CM obtained from normoxia (NTM-hypoxic-CM vs NTM-ctrl-CM *t-test p-value<0.05*). Error bars represent SEM of technical replicates. **D.** Group mean data of AGS cell proliferation in response to hypoxic-CM from CAMs (n=3) and NTMs (n=3). Error bars represent SEM.

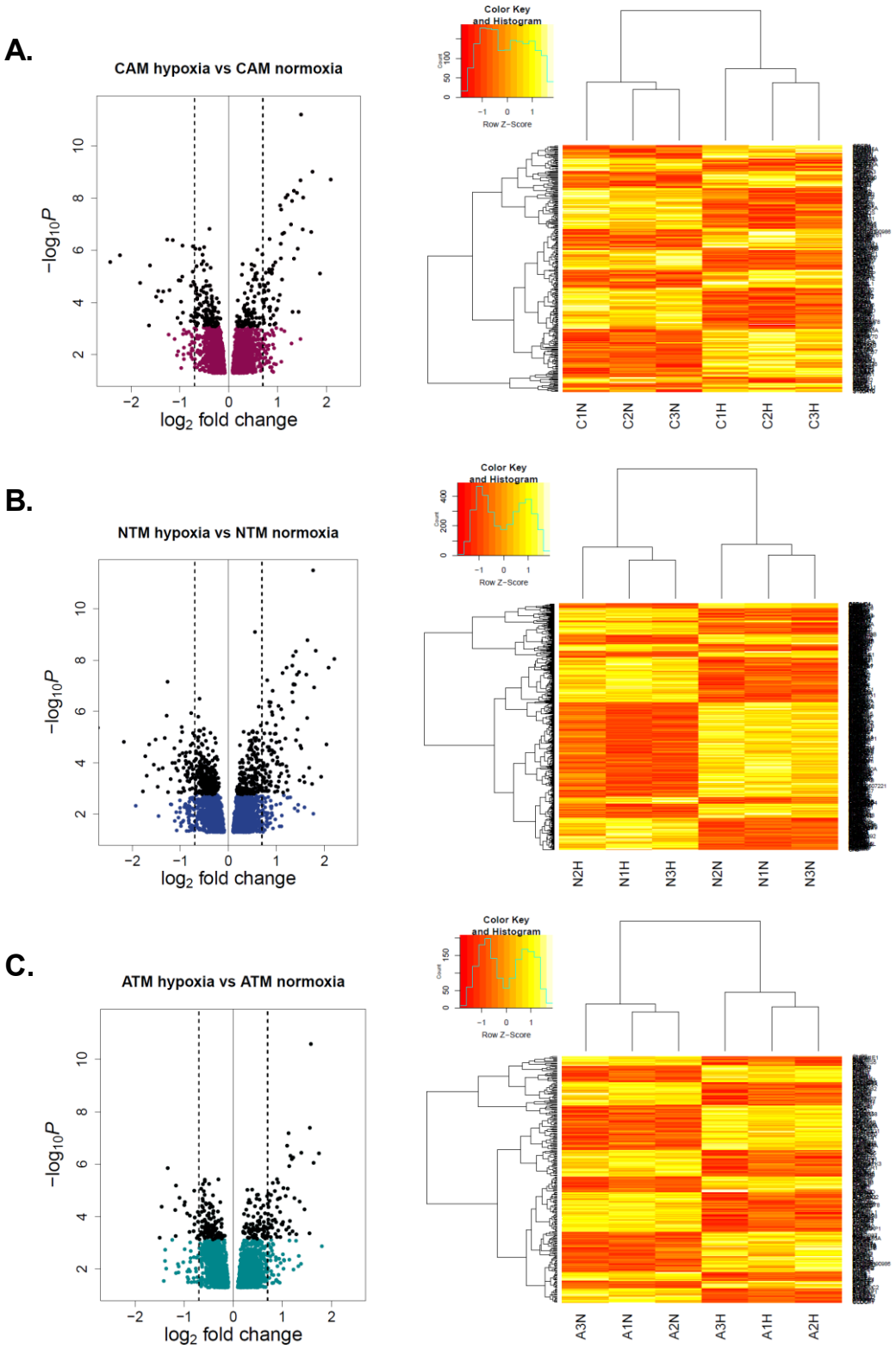


**Figure 6.2 Hypoxia-induced gene expression signatures in gastric myofibroblasts purified from different tissue microenvironments. A.** CAM hypoxia vs CAM normoxia **B.** NTM hypoxia vs NTM normoxia. **C.** ATM hypoxia vs ATM normoxia. Volcano plots represent differentially expressed genes in respective comparisons; *FDR p-value < 0.05* adjusted genes are shown in black; dashed lines 1.6 fold change. Heatmaps represent differentially expressed genes with *FDR p-value < 0.05* (shown in black in volcano plots).

### 6.3.2.1 Differential hypoxia-induced myofibroblast phenotypes are not due to induced changes in DNA methylation profiles

To assess whether hypoxia induces DNA methylation changes in gastric myofibroblast cells Illumina Infinium HumanMethylation450 BeadChip arrays were performed on the same CAM (cancer), ATM (adjacent tissue) and NTM (normal tissue) patient samples used in Illumina HT-12 array studies (section 6.3.2). Cultures of primary myofibroblasts for use in both expression and methylation arrays were prepared in parallel and were exposed to either hypoxia (1% $O_2$) or normoxia (21% $O_2$) for 72 hours (as described in Methods section 2.3.2).

Comparative genome–wide differential DNA methylation analyses performed between (i) CAM hypoxia vs CAM normoxia, (ii) ATM hypoxia vs ATM normoxia and (iii) NTM hypoxia vs NTM normoxia showed a striking similarity in DNA methylation profiles between matched hypoxia-treated and control samples (Figure 6.3). These data suggest that exposure to hypoxic conditions did not induce significant alterations in DNA methylation patterns, although differential hypoxia-induced gene expression profiles were observed in different myofibroblast populations.

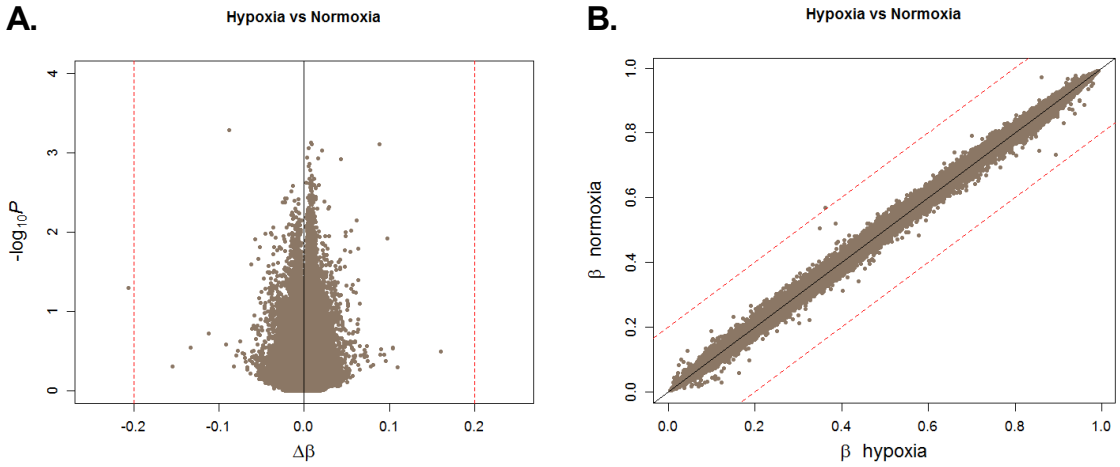


**Figure 6.3 Comparative global DNA methylation profiles of primary gastric myofibroblasts following exposure to hypoxic or normoxic conditions. A.** Volcano plot showing significance versus Δβ values for 424383 methylation probes in hypoxia vs normoxia comparison (n=9). **B.** Scatter plot representing mean β values for normoxic and hypoxic myofibroblasts (n=9). Red dashed lines represent |Δβ|>0.2.

## 6.3.3 Potential biological significance of hypoxia-induced gene expression signatures in CAMs, ATMs and NTMs

Although cells appear to react to hypoxia via a common HIF-1α mediated mechanism (Wang, Jiang et al. 1995, Semenza 1998), net functional responses may be influenced by cell - type specific variations in epigenetic programming. With respect to gastric myofibroblasts, although DNA methylation profiles do not change in response to hypoxia, different myofibroblast populations have distinct patterns of DNA methylation, which may confer functional differences in their responses to hypoxia. Therefore, in order to identify common and population specific changes, hypoxia–induced gene expression profiles for CAMs, ATMs and NTMs, were systematically compared (Figure 6.4), in order to predict both common and distinct functional changes.

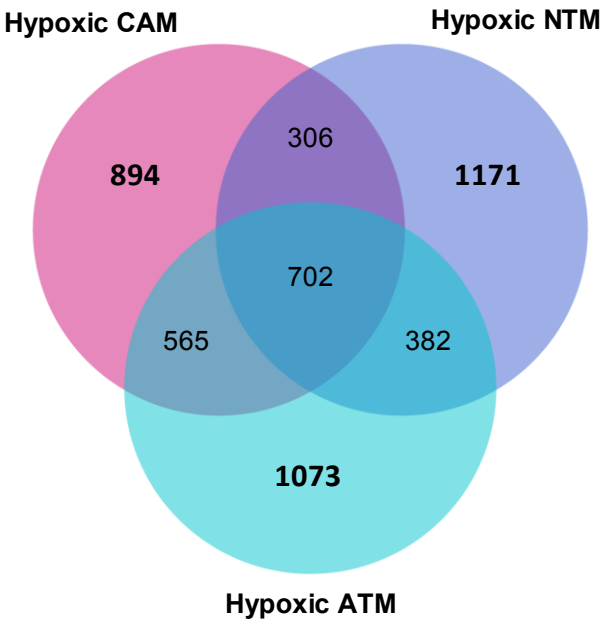


**Figure 6.4 Comparison of hypoxia-induced gene expression signatures identified in gastric CAMs, ATMs and NTMs.** Overlapping intersections represent common gene expression changes induced by hypoxia in stromal myofibroblasts. Unique CAM-, ATM- and NTM- hypoxia-induced gene signatures are shown in bold.

### 6.3.3.1 Universal hypoxia-induced gene expression signatures in CAMs, ATMs and NTMs

To establish if the overlapping gene profiles are associated with known universal hypoxia-induced biological processes, gene ontology (GO) and gene set (GSEA) enrichment analyses were performed on the universal hypoxia-induced gene subgroup identified in all three hypoxia vs normoxia comparisons (for CAMs, ATMs and NTMs). Supplementary Figure S6.1 shows that these genes are changed in the same direction in all myofibroblast types. The GO and GSEA analysis showed that various metabolic processes, characteristic for hypoxic cells, are universally altered in CAMs, ATMs and NTMs under hypoxia (Figure 6.5, Figure 6.6 and Table 6.1).

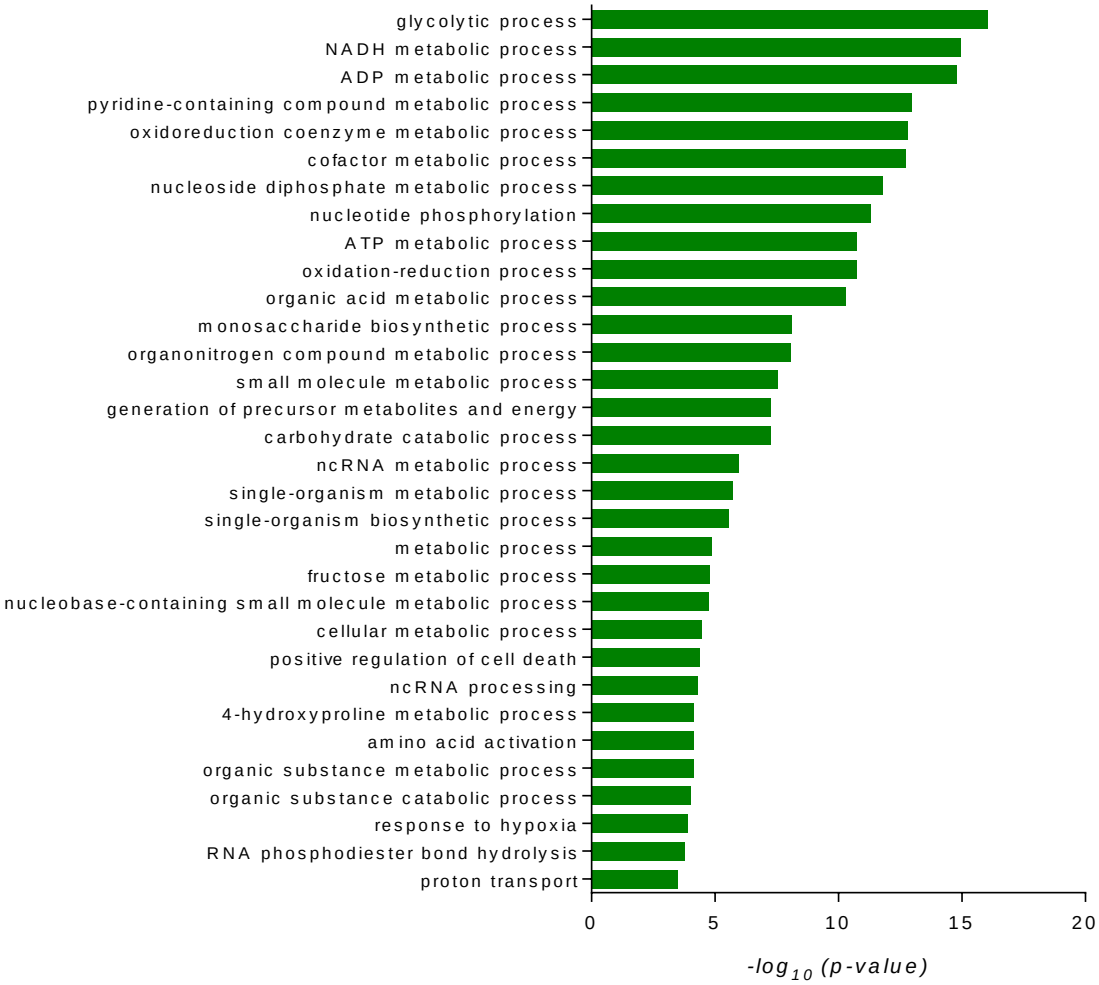


**Figure 6.5 Universal changes induced by hypoxia in gastric CAMs, ATMs and NTMs.** Gene ontology (GO) biological process (BP) terms associated with genes that are universally changed under hypoxia in the same direction in CAMs, ATMs and NTMs. Only GO terms with *FDR q-value <0.05* are shown.

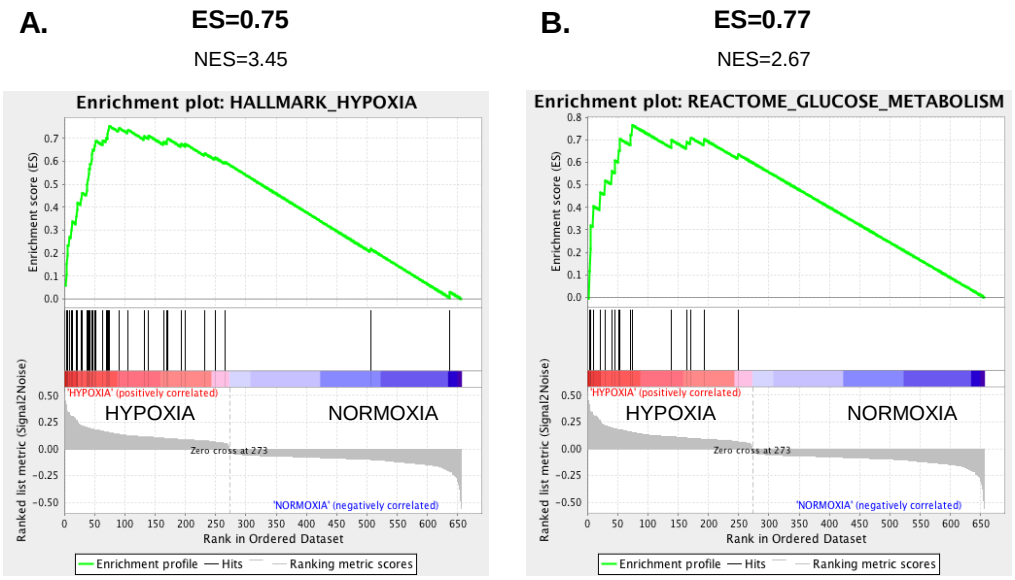


**Figure 6.6 GSEA enrichment plots** for the most significantly enriched **A.** hallmark gene set (h.all) and **B.** canonical pathway (c2.cp) gene set presented in Table 6.1. *ES* - Enrichment Score, *NES* – Normalized Enrichment Score

**Table 6.1 GSEA result summary.** Significantly enriched hallmark gene sets (h.all) and canonical pathways (c2.cp) are listed for universal CAM, ATM and NTM gene expression profiles, *FDR q-value <0.05*. *ES* - Enrichment Score, *NES* – Normalized Enrichment Score

| | Phenotype | GENE SET NAME | SIZE | ES | NES | *p-value* | FDR |
|---|---|---|---|---|---|---|---|
| h.all | Hypoxia | HALLMARK_HYPOXIA | 43 | 0.75 | 3.45 | 0 | 0 |
| | Hypoxia | HALLMARK_GLYCOLYSIS | 35 | 0.62 | 2.72 | 0 | 0 |
| | Normoxia | HALLMARK_UNFOLDED_PROTEIN_RESPONSE | 17 | -0.55 | -2.04 | 4.65E-03 | 7.75E-03 |
| | Normoxia | HALLMARK_P53_PATHWAY | 21 | -0.46 | -1.86 | 4.91E-03 | 1.28E-02 |
| | Hypoxia | HALLMARK_MTORC1_SIGNALING | 44 | 0.38 | 1.78 | 6.70E-03 | 2.27E-02 |
| | Hypoxia | HALLMARK_IL2_STAT5_SIGNALING | 23 | 0.45 | 1.69 | 1.43E-02 | 3.23E-02 |
| c2.cp | Hypoxia | REACTOME_GLUCOSE_METABOLISM | 17 | 0.77 | 2.67 | 0 | 0 |
| | Hypoxia | REACTOME_METABOLISM_OF_CARBOHYDRATES | 26 | 0.65 | 2.60 | 0 | 0 |
| | Hypoxia | KEGG_GLYCOLYSIS_GLUCONEOGENESIS | 19 | 0.69 | 2.53 | 0 | 0 |
| | Normoxia | REACTOME_METABOLISM_OF_AMINO_ACIDS_AND_DERIVATIVES | 17 | -0.55 | -2.06 | 2.43E-03 | 2.95E-03 |
| | Hypoxia | NABA_MATRISOME | 18 | 0.54 | 1.91 | 3.58E-03 | 1.06E-02 |

### 6.3.3.2 Unique CAM, ATM and NTM hypoxia-induced gene expression signatures

Unique hypoxia-induced gene expression signatures identified in gastric CAMs, ATMs and NTMs were subjected to gene ontology (GO) analysis, gene set enrichment analysis (GSEA) and Ingenuity pathway analysis (IPA), in order to identify hypoxia-induced processes that may contribute to the observed effects of CAM-hypoxic-CM and NTM-hypoxic-CM on cancer cell migration and proliferation (Figure 6.1), and also to shed light on the broader spectrum of hypoxia-induced functional changes in these cells.

#### 6.3.3.2.1 Gene ontology enrichment analysis

Gene ontology (GO) biological process (BP) enrichment analysis was performed on unique hypoxia-induced gene expression signatures identified in CAMs, ATMs and NTMs using Gene Ontology enRIchment anaLysis and visuaLizAtion tool (GOrilla) (Eden, Navon et al. 2009). All 13381 genes expressed in myofibrobalst cells were used as a reference set. The results from the respective GO analyses for CAMs, ATMs and NTMs are presented in Table 6.2.

**Table 6.2 Gene ontology (GO) biological process (BP) enrichment for unique hypoxia-induced gene expression signatures identified in CAMs, ATMs or NTMs.** GO terms with *FDR q-value <0.05* are highlighted in bold.

| | GO ID | GO biologicl process (BP) term | *p-value* |
|---|---|---|---|
| **CAM** | **GO:0006695** | **cholesterol biosynthetic process** | **1.40E-11** |
| | **GO:0008299** | **isoprenoid biosynthetic process** | **1.85E-05** |
| | GO:0048679 | regulation of axon regeneration | 2.90E-04 |
| | GO:0060011 | Sertoli cell proliferation | 4.88E-04 |
| | GO:0009448 | gamma-aminobutyric acid metabolic process | 4.88E-04 |
| | GO:0008202 | steroid metabolic process | 6.62E-04 |
| **ATM** | **GO:0044764** | **multi-organism cellular process** | **1.16E-05** |
| | GO:0044267 | cellular protein metabolic process | 6.71E-05 |
| | GO:0071360 | cellular response to exogenous dsRNA | 6.38E-05 |
| | GO:0043412 | macromolecule modification | 8.45E-05 |
| | GO:0016032 | viral process | 8.78E-06 |
| | GO:0006464 | cellular protein modification process | 5.40E-05 |
| | GO:0043122 | regulation of I-kappaB kinase/NF-kappaB signalling | 1.22E-04 |
| | GO:0019538 | protein metabolic process | 1.18E-04 |
| | GO:0044419 | interspecies interaction between organisms | 5.14E-05 |
| | GO:0010324 | membrane invagination | 1.45E-04 |
| | GO:0044260 | cellular macromolecule metabolic process | 4.51E-05 |
| | GO:0051704 | multi-organism process | 1.83E-04 |
| | GO:1903047 | mitotic cell cycle process | 2.54E-04 |
| | GO:0051017 | actin filament bundle assembly | 3.13E-04 |
| | GO:0040029 | regulation of gene expression, epigenetic | 4.09E-04 |
| | GO:0010660 | regulation of muscle cell apoptotic process | 4.47E-04 |
| | GO:0045087 | innate immune response | 5.07E-04 |
| | GO:0030522 | intracellular receptor signalling pathway | 6.80E-04 |
| | GO:0042059 | negative regulation of epidermal growth factor receptor signalling pathway | 9.07E-04 |
| | GO:2000653 | regulation of genetic imprinting | 9.99E-04 |
| | GO:0050851 | antigen receptor-mediated signalling pathway | 9.73E-04 |
| | GO:0043170 | macromolecule metabolic process | 9.03E-04 |

**Table 6.2 (*Continued*) Gene ontology (GO) biological process (BP) enrichment for unique hypoxia-induced gene expression signatures identified in CAMs, ATMs or NTMs.** GO terms with *FDR q-value <0.05* are highlighted in bold.

| | GO ID | GO biologicl process (BP) term | *p-value* |
|---|---|---|---|
| **NTM** | **GO:0034641** | **cellular nitrogen compound metabolic process** | **3.94E-06** |
| | **GO:0006807** | **nitrogen compound metabolic process** | **5.59E-06** |
| | GO:0006139 | nucleobase-containing compound metabolic process | 2.93E-05 |
| | GO:1901137 | carbohydrate derivative biosynthetic process | 8.49E-05 |
| | GO:0090304 | nucleic acid metabolic process | 8.57E-05 |
| | GO:0046483 | heterocycle metabolic process | 8.92E-05 |
| | GO:0051252 | regulation of RNA metabolic process | 1.14E-04 |
| | GO:0009058 | biosynthetic process | 1.20E-04 |
| | GO:0010648 | negative regulation of cell communication | 1.29E-04 |
| | GO:0009968 | negative regulation of signal transduction | 1.48E-04 |
| | GO:0006725 | cellular aromatic compound metabolic process | 1.51E-04 |
| | GO:1901576 | organic substance biosynthetic process | 2.77E-04 |
| | GO:1901360 | organic cyclic compound metabolic process | 3.09E-04 |
| | GO:0051171 | regulation of nitrogen compound metabolic process | 3.28E-04 |
| | GO:0048585 | negative regulation of response to stimulus | 3.39E-04 |
| | GO:0045786 | negative regulation of cell cycle | 3.76E-04 |
| | GO:0023057 | negative regulation of signalling | 4.11E-04 |
| | GO:0006750 | glutathione biosynthetic process | 4.34E-04 |
| | GO:0006575 | cellular modified amino acid metabolic process | 6.32E-04 |
| | GO:0044260 | cellular macromolecule metabolic process | 6.78E-04 |
| | GO:0006309 | apoptotic DNA fragmentation | 7.41E-04 |
| | GO:0035878 | nail development | 8.88E-04 |
| | GO:0044238 | primary metabolic process | 9.40E-04 |
| | GO:0006488 | dolichol-linked oligosaccharide biosynthetic process | 9.55E-04 |

#### 6.3.3.2.2 Gene set enrichment analysis

Unique CAM, ATM or NTM hypoxia-induced gene expression signatures were subjected to gene set enrichment analysis (GSEA) (Mootha, Lindgren et al. 2003, Subramanian, Tamayo et al. 2005) using the hallmark gene set subcollection from Molecular Signatures Database (MSigDB v5.0) to characterize biological states and processes that are unique for CAMs, ATMs or NTMs under hypoxic conditions. Results from these GSEA analyses are presented in Table 6.3 and representative GSEA enrichment plots for the most significant hallmark gene sets for each dataset are presented in Figure 6.7. In addition, Leading Edge Analyses (LEA) were performed on enriched gene sets from the complete MSigDBv5.0 collection obtained for unique CAM, ATM and NTM hypoxia-induced gene expression profiles, in order to identify genes that may have a key role in each of the three analysed phenotypes. The respective LEA results for CAMs, ATMs and NTMs are presented in Supplementary S6 (section S6.2, Table S6.1).

#### 6.3.3.2.3 Pathway analysis

The unique hypoxia-induced gene profiles from respective hypoxia vs normoxia comparisons were subjected to Ingenuity Pathway Analysis (IPA), in order to identify CAM-, ATM- and NTM- specific pathways that are changed in response to hypoxic conditions. In each case expression profiles were compared against the IPA predefined Illumina HT-12v4 reference set. Top 10 most significant pathways identified for respective unique hypoxia-induced gene profiles are presented in Figure 6.8.

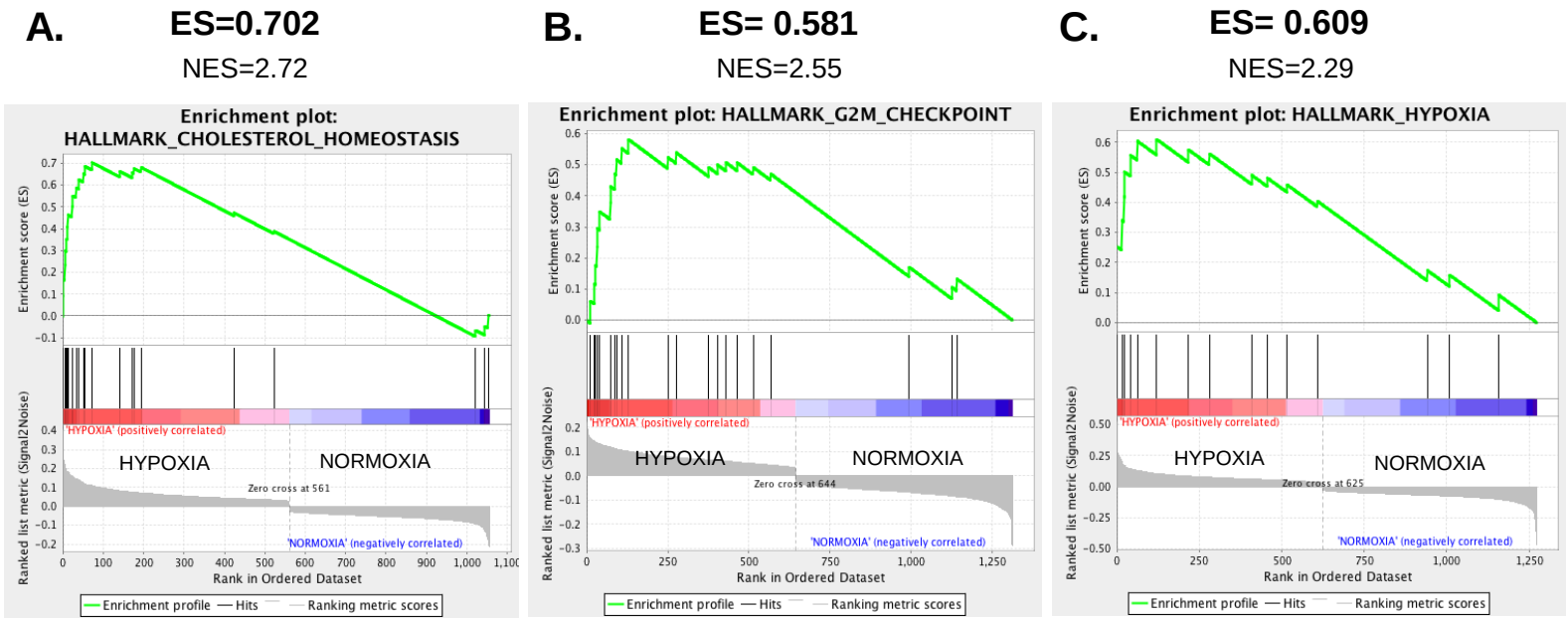


**Figure 6.7 GSEA enrichment plots** for the most significantly enriched hallmark gene sets presented in Table 6.3. The most significant hallmark gene set for **A.** CAM unique hypoxia-induced gene expression profile; **B.** ATM unique hypoxia-induced gene expression profile; **C.** NTM unique hypoxia-induced gene expression profile. *ES* - Enrichment Score, *NES* – Normalized Enrichment Score

**Table 6.3 GSEA result summary.** Significantly enriched hallmark gene sets are listed for CAM, ATM and NTM unique hypoxia-induced gene expression profiles, *FDR q-value <0.05*. *ES* - Enrichment Score, *NES* – Normalized Enrichment Score

| | Phenotype | HALLMARK GENE SET NAME | SIZE | ES | NES | *p-value* | FDR |
|---|---|---|---|---|---|---|---|
| **CAM** | Hypoxia | HALLMARK_CHOLESTEROL_HOMEOSTASIS | 23 | 0.70 | 2.72 | 0 | 0 |
| | Hypoxia | HALLMARK_MTORC1_SIGNALING | 25 | 0.61 | 2.41 | 0 | 0 |
| | Hypoxia | HALLMARK_FATTY_ACID_METABOLISM | 19 | 0.53 | 1.96 | 0 | 5.17E-03 |
| | Hypoxia | HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION | 16 | 0.53 | 1.90 | 4.62E-03 | 8.79E-03 |
| | Hypoxia | HALLMARK_E2F_TARGETS | 19 | 0.46 | 1.71 | 1.80E-02 | 3.36E-02 |
| **ATM** | Hypoxia | HALLMARK_G2M_CHECKPOINT | 23 | 0.58 | 2.55 | 0 | 0 |
| | Hypoxia | HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION | 16 | 0.62 | 2.40 | 0 | 0 |
| | Hypoxia | HALLMARK_E2F_TARGETS | 23 | 0.46 | 2.06 | 0 | 5.33E-03 |
| | Hypoxia | HALLMARK_MITOTIC_SPINDLE | 24 | 0.42 | 1.92 | 1.14E-02 | 1.49E-02 |
| | Hypoxia | HALLMARK_APICAL_JUNCTION | 26 | 0.38 | 1.73 | 1.62E-02 | 4.15E-02 |
| **NTM** | Hypoxia | HALLMARK_HYPOXIA | 16 | 0.61 | 2.29 | 0 | 0 |
| | Normoxia | HALLMARK_XENOBIOTIC_METABOLISM | 28 | -0.54 | -2.49 | 0 | 0 |
| | Normoxia | HALLMARK_UNFOLDED_PROTEIN_RESPONSE | 18 | -0.61 | -2.32 | 0 | 0 |
| | Hypoxia | HALLMARK_MITOTIC_SPINDLE | 17 | 0.58 | 2.18 | 0 | 1.84E-03 |
| | Hypoxia | HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION | 21 | 0.55 | 2.24 | 0 | 2.38E-03 |
| | Hypoxia | HALLMARK_KRAS_SIGNALING_UP | 17 | 0.49 | 1.86 | 3.85E-03 | 2.74E-02 |

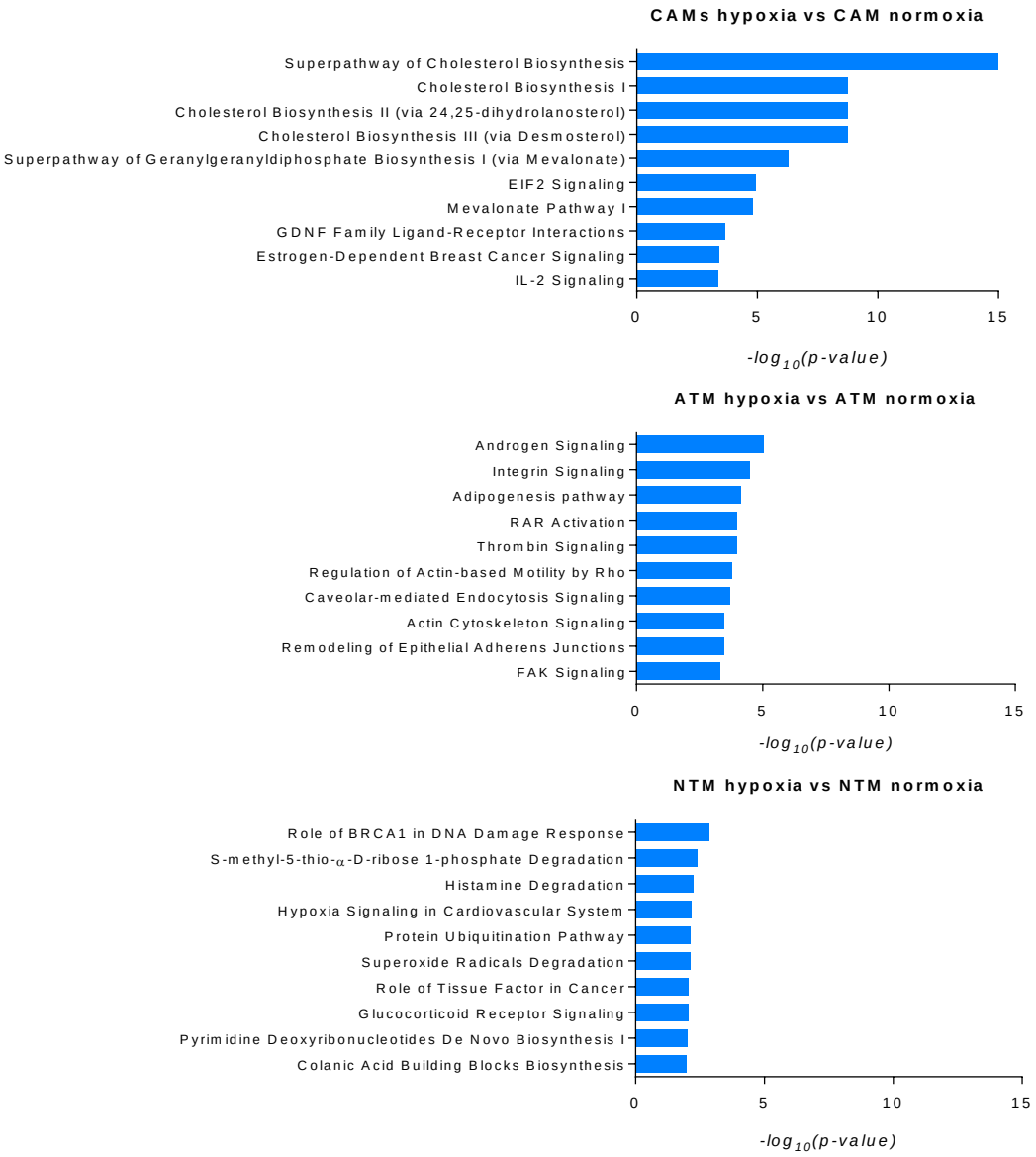


**Figure 6.8 IPA canonical pathways significantly enriched in unique CAM, ATM and NTM hypoxia-induced gene expression profiles.** Top 10 pathways in respective hypoxia vs normoxia comparisons (indicated at the top of the bar chart) are shown; *Fisher's Exact test p-value.*

### 6.3.3.3 Predicted biological effects of hypoxia on CAMs, ATMs and NTMs

To predict the biological effects of hypoxia on gastric CAMs, ATMs and NTMs the IPA downstream effects analysis was performed on differentially expressed genes identified in the respective hypoxia vs normoxia comparisons. The downstream effects analysis facilitates the prediction of functions and processes that are expected to be increased or decreased, based on the observed changes in gene expression in the respective datasets. The results from these analyses for CAMs, ATMs and NTMs are presented in Figure 6.9. Processes with an absolute *z-score* of ≥ |2| are reported irrespective of their *p-value*.

### 6.3.3.4 Prediction of factors that may be secreted by hypoxic CAMs and NTMs

To predict factors that may be secreted by hypoxic CAMs and hypoxic NTMs and have an effect on cancer cell migration and proliferation, the IPA downstream effects analysis was performed on differentially expressed genes identified in CAM hypoxia vs CAM normoxia and NTM hypoxia vs NTM normoxia comparisons. Genes annotated as being involved in cell migration and proliferation were selected and further analysed to identify genes that are known to encode extracellular proteins. The results from these respective analyses are shown in Figure 6.10.

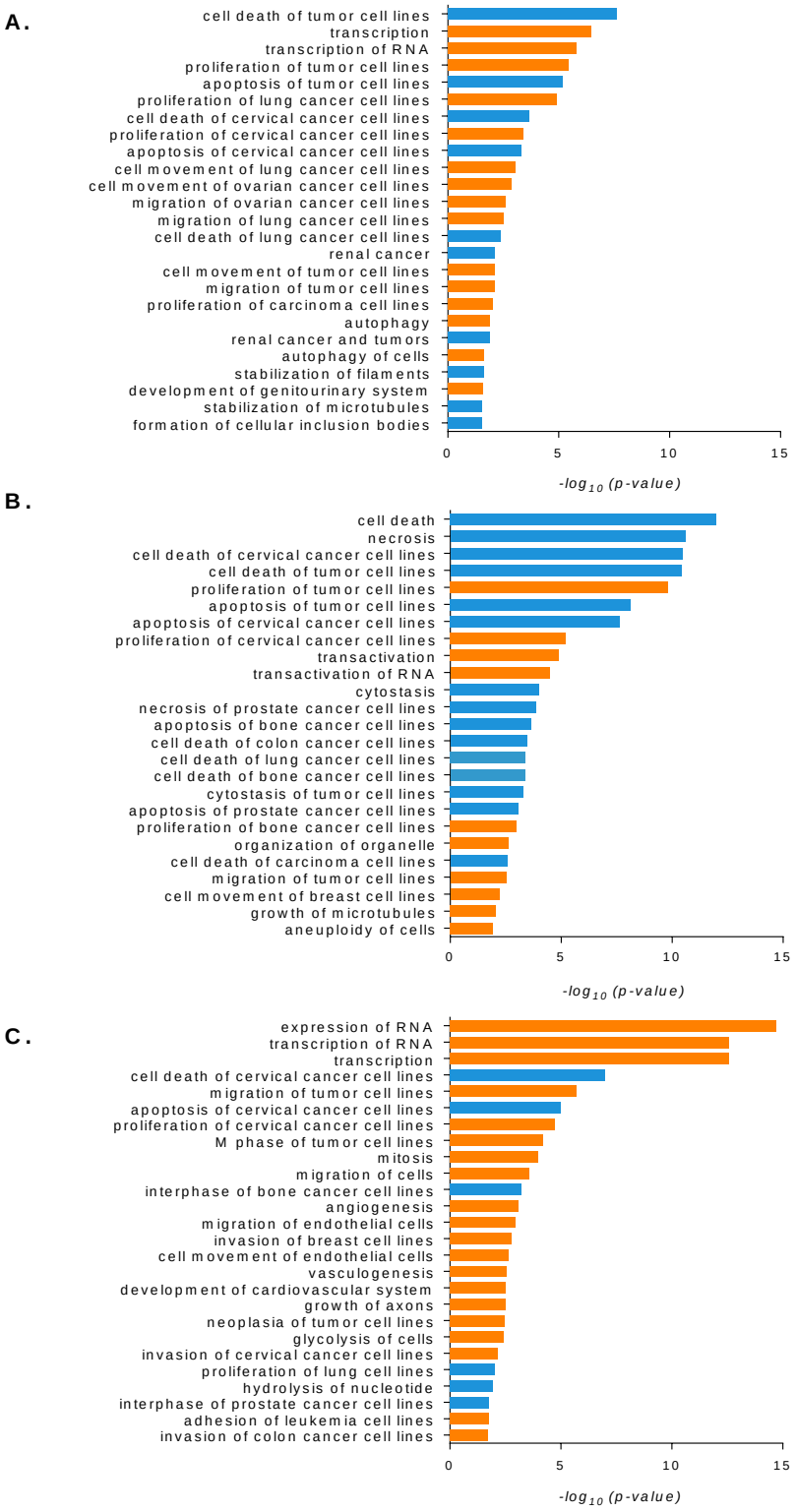


**Figure 6.9 Predicted biological effects of hypoxia on gastric CAMs, ATMs and NTMs.** **A.** CAM hypoxia vs CAM normoxia. **B.** ATM hypoxia vs ATM normoxia. **C.** NTM hypoxia vs NTM normoxia; *orange* – predicted increase (*z-score* ≥ *2*), *blue* – predicted decrease (*z-score* ≤ *-2*); *Fisher's Exact test p-value.*

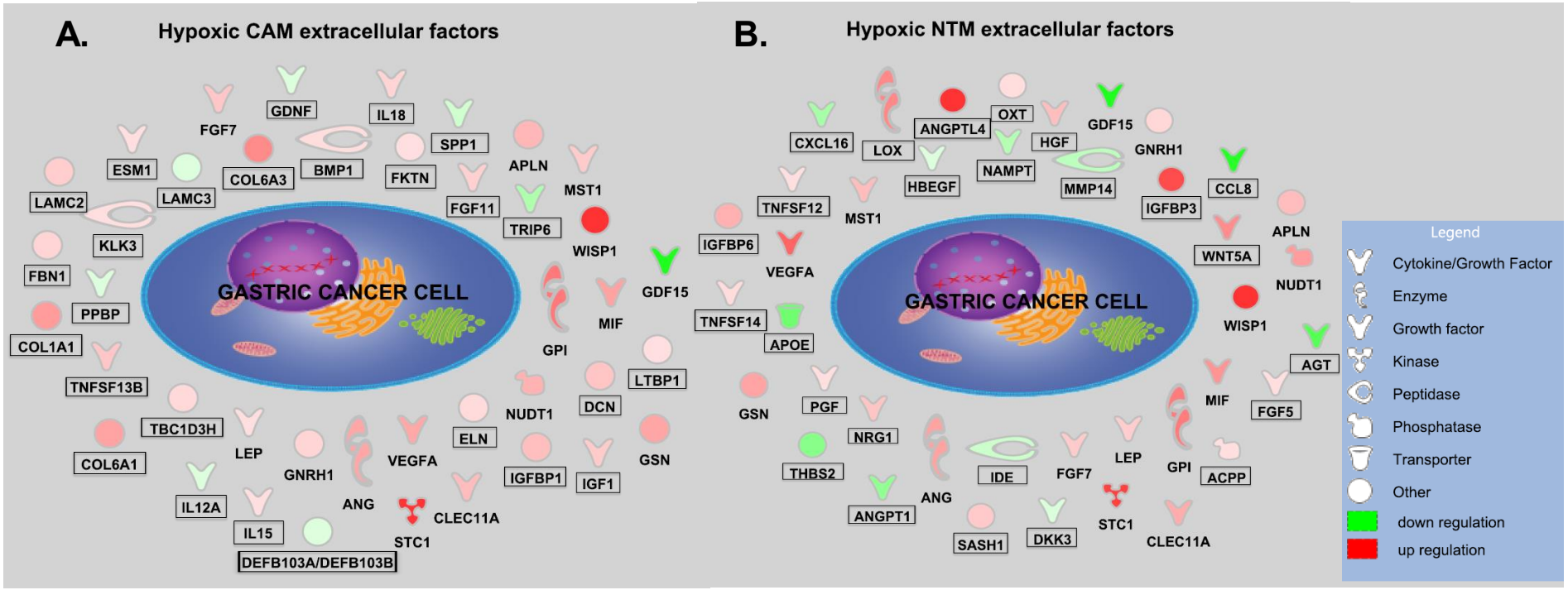


**Figure 6.10 Predicted pro-migratory and pro-proliferative factors secreted by hypoxic CAMs and hypoxic NTMs. A.** Extracellular molecules expressed by hypoxic CAMs. **B.** Extracellular molecules expressed by hypoxic NTMs. Molecules displayed in boxes are uniquely expressed in hypoxic CAM or hypoxic NTM.

### 6.3.4 Secretome analysis of CAM and NTM hypoxic conditioned media

In order to identify secretory factors released by gastric CAMs and NTMs in response to hypoxia, CAM and NTM hypoxic and control conditioned media was profiled by LC-MS/MS. In total four different conditions were examined: (i) CAM-ctrl-CM; (ii) CAM-hypoxic-CM; (iii) NTM-ctrl-CM; and (iv) NTM-hypoxic-CM. In each case CM was collected from CAMs (n=3) or NTMs (n=3) after 72 hours incubation under hypoxic (1% $O_2$) or normoxic (21% $O_2$) conditions as described in Methods section 2.4.2.

#### 6.3.4.1 Identification and quantification of CAM and NTM secreted proteins

Altogether, 1169 protein groups were identified and quantified across all 12 CM samples. The complete table of quantified proteins is presented in *Appendix V*.

A core of 702 proteins was detected in all 12 samples. These proteins were subjected to gene ontology enrichment analysis using the PANTHER Classification System (Mi, Muruganujan et al. 2013) in order to verify their cellular location. The identified proteins were significantly enriched in terms of annotation for extracellular region ($p = 4.59 \times 10^{-227}$) and extracellular vesicle ($p = 1.74 \times 10^{-200}$). The other top 10 most significantly enriched GO terms were also associated with extracellular structures confirming that the experimental strategy worked well.

In addition, the Metazoa Secretome and Subcellular Proteome Knowledge Base (MetazSecKB, released May 2014) (Meinken, Walker et al. 2015), MatrisomeDB (updated August 2014) (Naba, Clauser et al. 2012) and ExoCarta (released May 2012) (Mathivanan, Fahner et al. 2012) were also used to further annotate the identified proteins present in CAM and NTM secretomes. The MetazSecKB search identified 295 proteins that are annotated as predicted secreted. The MatrisomeDB search identified 218 extracellular matrix (ECM) proteins and associated factors whereas ExoCarta search identified 752 proteins that were detected in exosomes released by various cell types. The overlap of these database searches is shown in Figure 6.11.

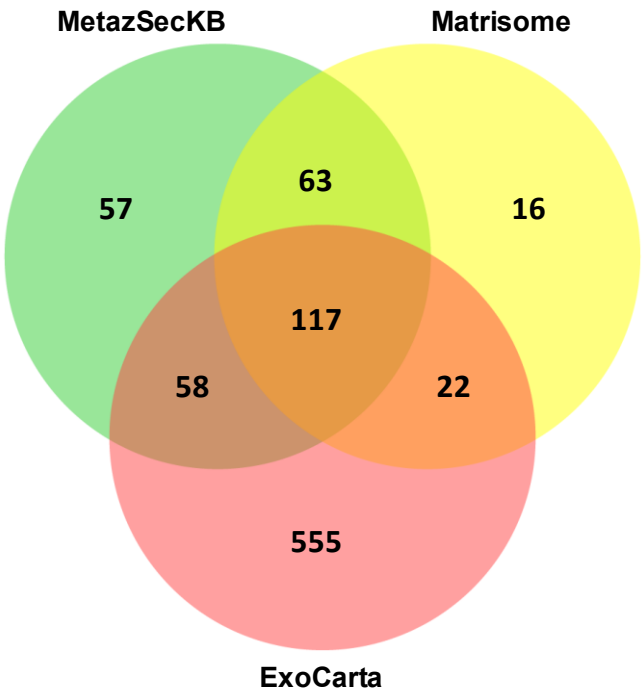


**Figure 6.11 Venn diagram representation of database searches used to classify proteins identified in CAM and NTM conditioned media obtained from normoxia and hypoxia.**

### 6.3.4.2 Hypoxia–induced secretome signatures in gastric CAMs and NTMs

To define hypoxia-induced secretome signatures statistical significance score was calculated for protein fold change in the following comparisons: (i) CAM-hypoxic-CM vs CAM-ctrl-CM, (ii) NTM-hypoxic-CM vs NTM-ctrl-CM, (iii) CAM-ctrl-CM vs NTM-ctrl-CM and (iv) CAM-hypoxic-CM vs NTM-hypoxic-CM.

The analysis identified: (i) 43 proteins that were differentially secreted by hypoxic CAMs compared to control CAMs (Figure 6.12A); (ii) 225 proteins that were differentially secreted by hypoxic NTMs compared to control NTMs (Figure 6.12B); (iii) 84 proteins differentially secreted by CAMs compared to NTMs (Figure 6.12C); and (iv) 145 proteins differentially secreted by hypoxic CAMs compared to hypoxic NTMs (Figure 6.12D). Supplementary Figure S6.3 shows which of the differentially secreted proteins identified in respective CM comparisons were annotated as being secreted or extracellular, based on MetazSecKB, Matrisome and ExoCarta searches (as describe in section 6.3.4.1).

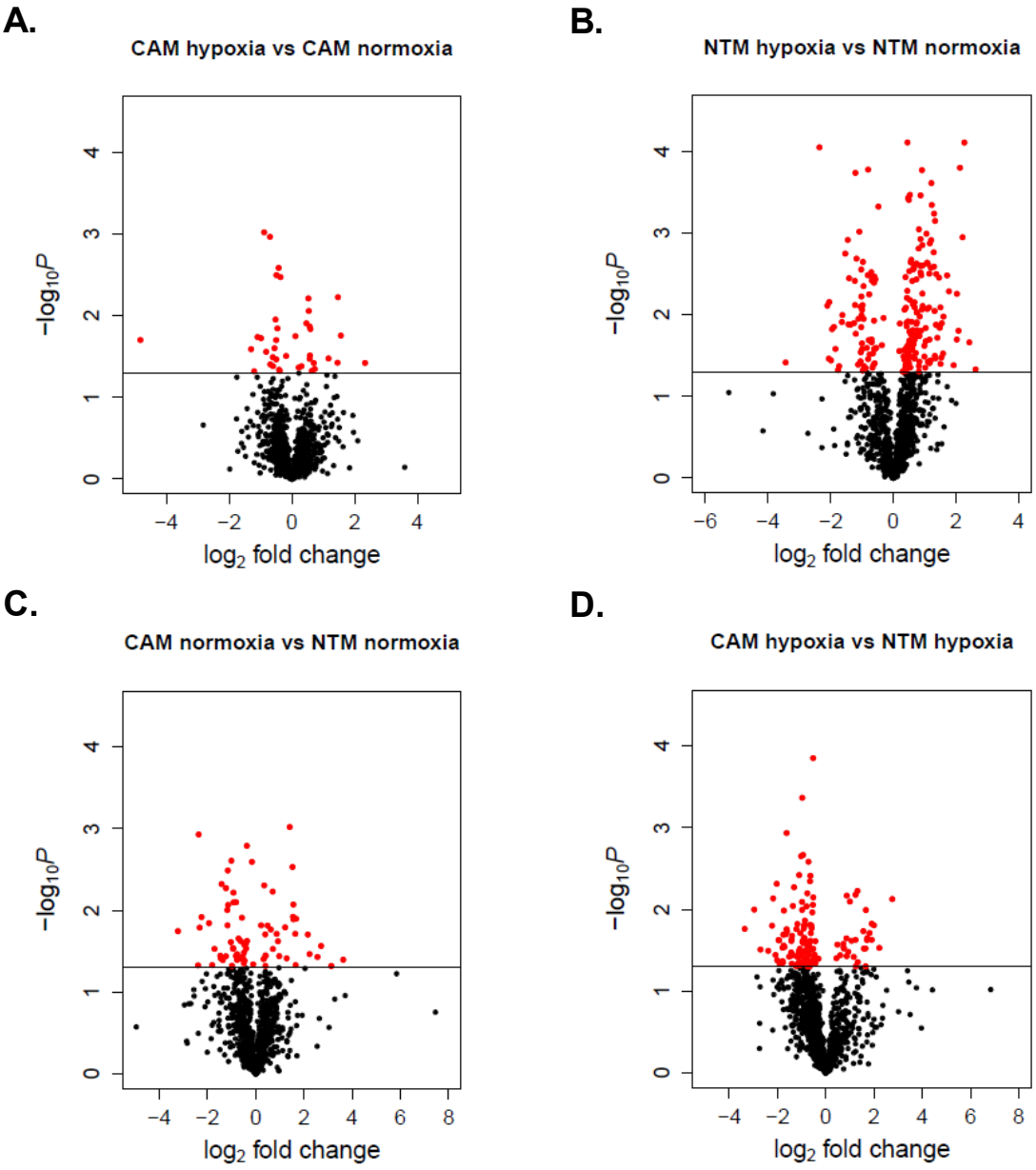


**Figure 6.12 Volcano plots of differentially secreted proteins** in the following comparisons: **A.** CAM-hypoxic-CM vs CAM-ctrl-CM **B.** NTM-hypoxic-CM vs NTM-ctrl-CM **C.** CAM-ctrl-CM vs NTM-ctrl-CM **D.** CAM-hypoxic-CM vs NTM-hypoxic-CM. Proteins highlighted in red are considered to be differentially secreted between the given conditions.

Comparative analysis of differently secreted proteins identified in CAM-hypoxic-CM and NTM-hypoxic-CM compared to respective normoxic-control CM revealed secretion of 14 proteins that are universally altered under hypoxia in both CAMs and NTMs (Figure 6.13). Significantly, this analysis confirmed some of the predictions made based on hypoxia-induced CAM and NTM gene expression profiles (Figure 6.10). In particular, gelsolin (GSN), vascular endothelial growth factor A (VEGFA), glucose-6-phosphate isomerase (GPI) were verified at protein level as universally-changed CAM and NTM secreted proteins under hypoxia (Supplementary Figure S6.4). In addition, lysyl oxidase (LOX), insulin-like growth factor-binding protein 3 (IGFBP3), insulin-like growth factor-binding protein 6 (IGFBP6), thrombospondin-2 (THBS2) and angiopoietin-related protein 4 (ANGPTL4) were also verified at protein level as NTM unique hypoxia-induced secreted proteins (Supplementary Figure S6.5 and Figure S6.6).

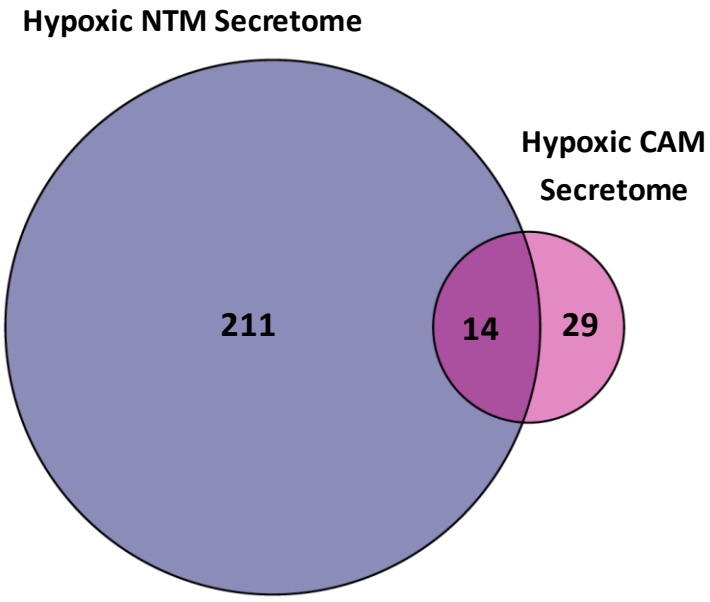


**Figure 6.13 Comparison of differentially secreted proteins** identified in CAM-hypoxic-CM and NTM-hypoxic-CM compared to their respective normoxic-control-CM.

### 6.3.4.3 Potential biological effects of the hypoxia-induced CAM and NTM secretomes

Gene ontology (GO) enrichment analysis and IPA downstream effects analysis were performed to gain insight into possible biological effects of the identified hypoxia-induced CAM and NTM secretomes.

#### 6.3.4.3.1 Gene ontology enrichment analysis

Gene ontology (GO) enrichment analysis was performed on CAM and NTM hypoxia-induced secretomes using GOrilla (Gene Ontology enRIchment anaLysis and visuaLizAtion tool) (Eden, Navon et al. 2009). Identified proteins were ranked according to their *p-value* with proteins identified as differentially secreted being at the top of the rank.

The enrichment analyses of biological processes (BP) identified 12 GO terms and 62 GO terms with *p-value < 0.001* in CAM and NTM hypoxia-induced secretomes, respectively. List of GO biological process (BP) terms with *FDR p-value < 0.01* identified in NTM hypoxia-induced secretome is shown in Table 6.4.

GO biological process (BP) terms with *p-value < 0.001* identified in CAM hypoxia-induced secretome include: regulation of cytoskeleton organisation (GO:0051493 *,* $p = 4.08 \times 10^{-4}$), regulation of actin filament-based process (GO:0032970, $p = 4.10 \times 10^{-4}$), protein modification by small protein removal (GO:0070646, $p = 5.30 \times 10^{-4}$) and single-organism metabolic process (GO:0044710, $p = 7.23 \times 10^{-4}$). Notably, extracellular exosome (GO:0070062, $p = 1.07 \times 10^{-4}$) was

one of the most enriched GO cellular component (CC) terms identified in CAM hypoxia-induced secretome.

**Table 6.4 Gene ontology (GO) biological process (BP) enrichment for differentially secreted proteins by hypoxic NTMs;** *FDR p-value < 0.01.*

| GO ID | GO biological process (BP) term | *p-value* | FDR |
|---|---|---|---|
| GO:0006090 | pyruvate metabolic process | 2.94E-08 | 1.90E-04 |
| GO:0032787 | monocarboxylic acid metabolic process | 1.99E-07 | 6.42E-04 |
| GO:0044712 | single-organism catabolic process | 3.17E-06 | 1.20E-03 |
| GO:0006082 | organic acid metabolic process | 8.00E-07 | 1.29E-03 |
| GO:0046496 | nicotinamide nucleotide metabolic process | 2.90E-06 | 1.34E-03 |
| GO:0072524 | pyridine-containing compound metabolic process | 2.90E-06 | 1.44E-03 |
| GO:0016052 | carbohydrate catabolic process | 7.63E-06 | 2.35E-03 |
| GO:0006094 | gluconeogenesis | 1.97E-05 | 5.53E-03 |
| GO:0009167 | purine ribonucleoside monophosphate metabolic process | 2.33E-05 | 6.01E-03 |
| GO:0006091 | generation of precursor metabolites and energy | 3.88E-05 | 8.96E-03 |
| GO:0042157 | lipoprotein metabolic process | 3.85E-05 | 9.22E-03 |
| GO:0008380 | RNA splicing | 4.33E-05 | 9.33E-03 |

## 6.3.4.3.2 IPA downstream effects analysis

To predict the biological effects of hypoxia-induced CAM and NTM secretomes the IPA downstream effects analysis was performed on differentially secreted proteins identified in CAM-hypoxic-CM vs CAM-ctrl-CM and NTM-hypoxic-CM vs NTM-ctrl-CM comparisons. The downstream effects analysis facilitates identification of functions and processes that are expected to increase or decrease, based on the observed changes in protein levels in the given datasets. The results from these analyses are presented in Figure 6.14. Processes with an absolute *z-score* of ≥ |2| are reported irrespective of their *p-value*.

Notably, proteins secreted by hypoxic NTMs were predicted to be regulated by HIF-1α activation (*z-score = 2.303*, $p = 4.76 \times 10^{-9}$) as revealed by IPA upstream regulator analysis. These hypoxic NTM secreted proteins positively affect angiogenesis, migration of endothelial cells and proliferation of tumour cell lines (Figure 6.15).

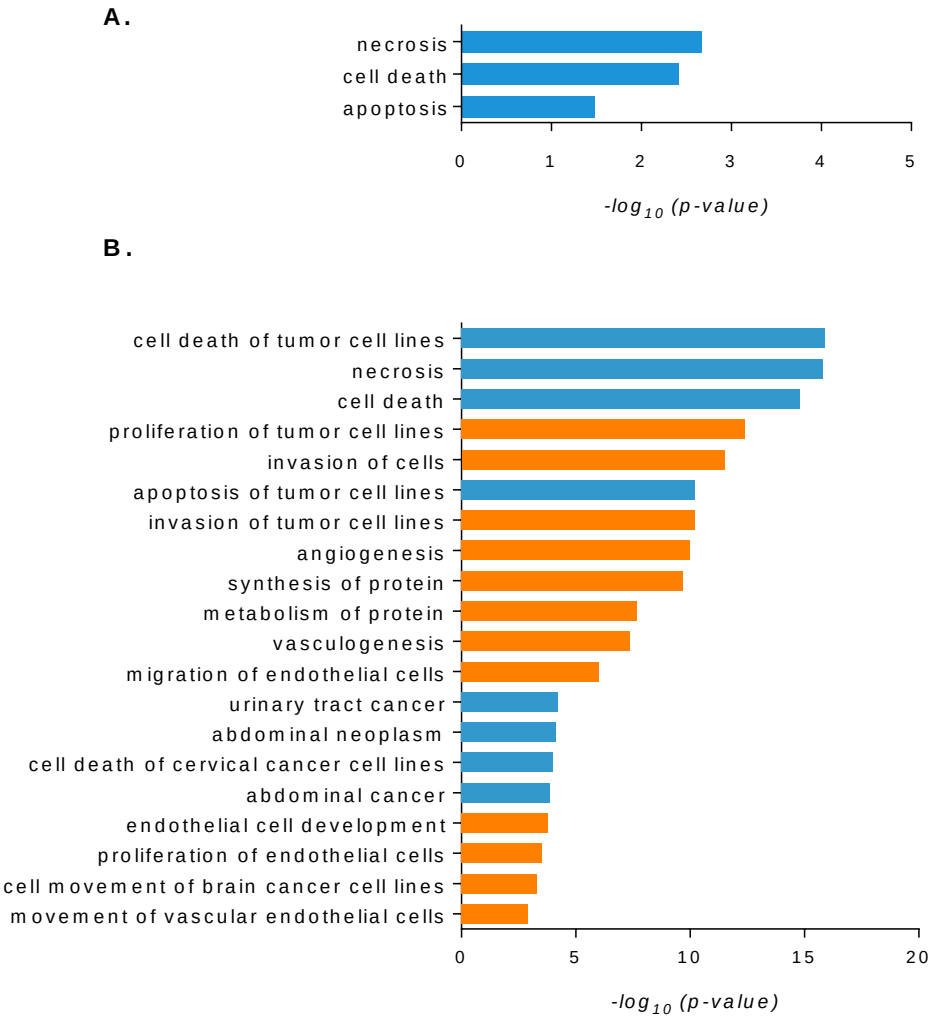


**Figure 6.14 Predicted biological effects of the hypoxia-induced CAM and NTM secretomes. A.** CAM-hypoxic-CM vs CAM-ctrl-CM **B.** NTM-hypoxic-CM vs NTM-ctrl-CM; *orange* – predicted increase (*z-score* ≥ *2*), *blue* – predicted decrease (*z-score* ≤ *-2*); *Fisher's Exact test p-value.*

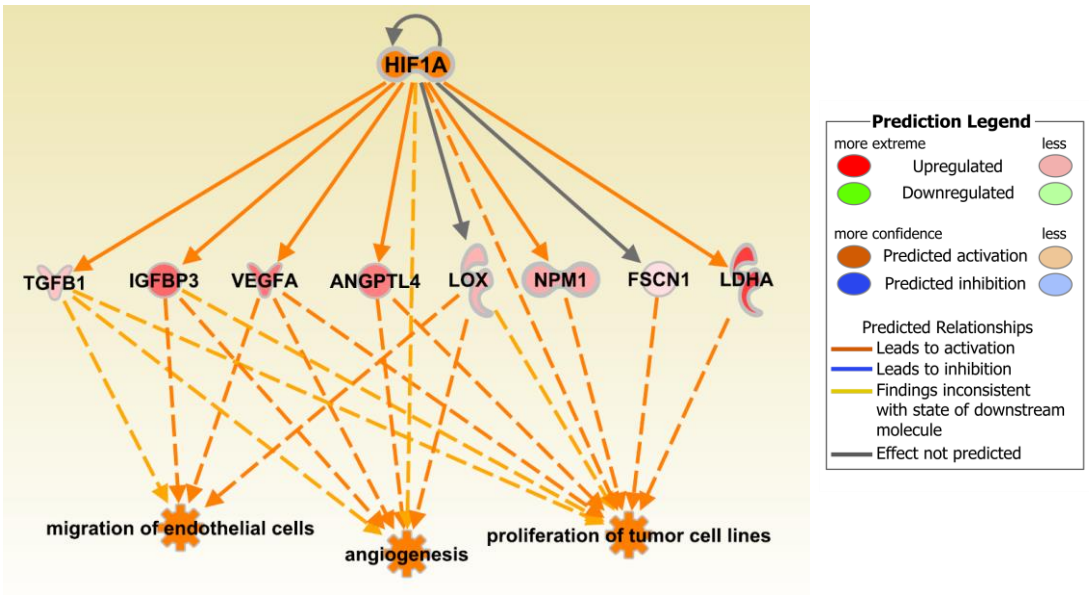


**Figure 6.15 Predicted activation of HIF-1α regulates expression of proteins secreted by hypoxic NTMs.** The hypoxic NTM secreted proteins positively affect angiogenesis, migration of endothelial cells and proliferation of tumour cell lines.

### 6.3.4.4 Integration of hypoxia-induced CAM and NTM secretomes with CAM and NTM hypoxia-induced gene expression signatures

The differentially secreted proteins identified in CAM-hypoxic-CM and NTM-hypoxic-CM compared to their respective control CM were integrated with CAM and NTM hypoxia-induced gene expression signatures in order to establish the correlation between CAM and NTM secreted proteins and their gene expression in respective comparisons.

The integrative analysis showed that a good correlation exist between respective gene expression profiles and secreted protein levels (*Pearson correlation: 0.4231 – 0.8151, $p < 2.073 \times 10^{-12}$*) (Figure 6.16).

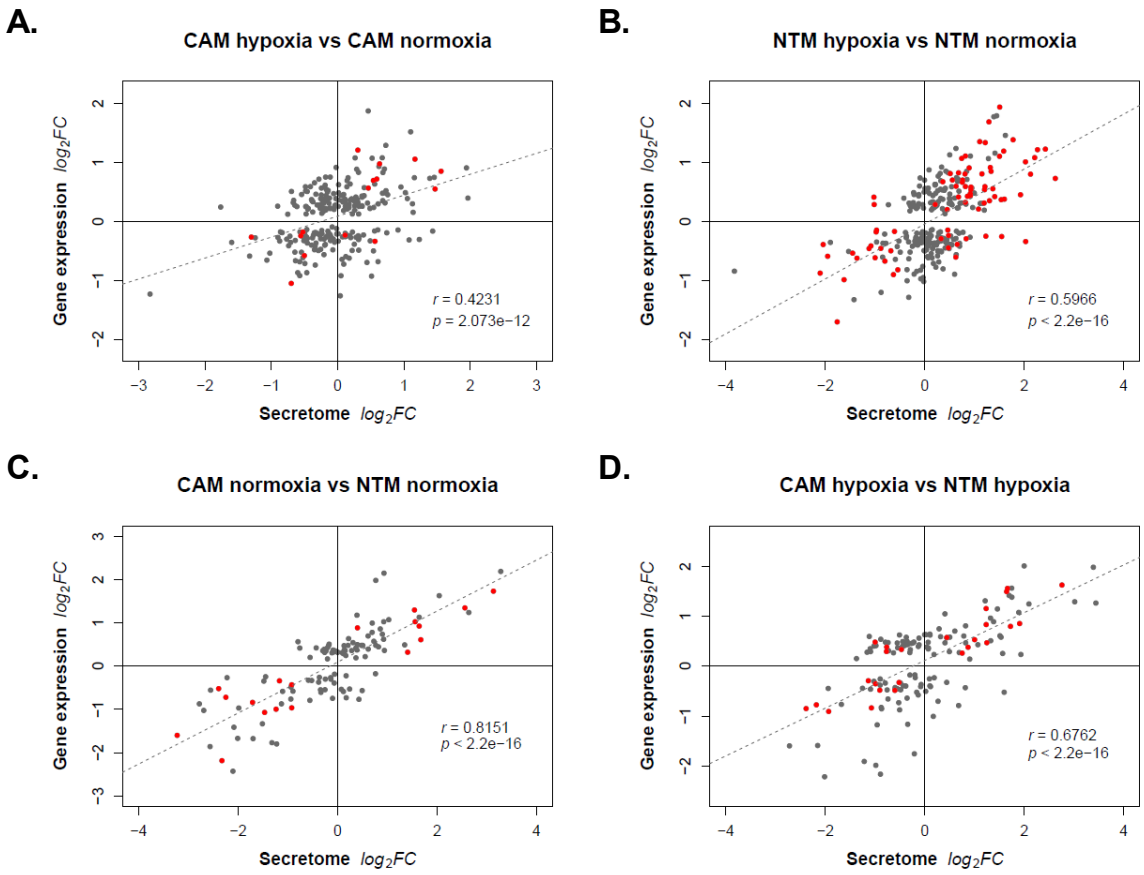


**Figure 6.16 Integration of secretome and gene expression data** for **A.** CAM hypoxia vs CAM normoxia **B.** NTM hypoxia vs NTM normoxia **C.** CAM normoxia vs NTM normoxia **D.** CAM hypoxia vs NTM hypoxia; *grey* - differentially expressed gene and corresponding protein identified in given secretome, *red* – differentially expressed gene and corresponding differentially secreted protein identified in given CM comparison, *r – Pearson correlation*.

## 6.4 Discussion

In this Chapter, an integrated multi-omics approach was used to understand modulation of stromal cells activity under hypoxia. The presented analysis of genomic and proteomic data provides an insight into the profile of changes in gene expression and protein secretion that result from exposure of normal gastric myofibroblasts (NTMs) or cancer associated gastric myofibroblasts (CAMs) to hypoxic conditions.

Data presented in previous Chapters, show that gastric CAMs increase cancer cell migration and proliferation compared to both ATMs and NTMs (Supplementary Figure S3.1 and Figure S3.2) and that DNA methylation may contribute to the regulation of CAM-specific gene expression profiles and therefore may play an important part in the ability of CAMs to exert tumour-promoting properties. Functional studies presented in this Chapter show that hypoxia induces phenotypic changes in the effects that CAMs and NTMs exert on cancer cell migration and proliferation. In particular, data from this study show that under hypoxic conditions CAMs become more aggressive and NTMs acquire more CAM-like properties. Specifically, CAM-hypoxic-CM was found to increase CAM-induced cancer cell migration, while reducing the ability of CAMs to stimulate cancer cell proliferation suggesting that hypoxia enhances CAM secretion of pro-migratory factors. In contrast, NTM-hypoxic-CM stimulated both migration and proliferation of cancer cells suggesting that hypoxia activates NTMs and enhances secretion of both pro-migratory and pro-proliferative factors by these cells. Notably, migration of cancer cells was found to be greater towards CAM-hypoxic-CM then NTM-hypoxic-CM.

Surprisingly, genome-wide DNA methylation profiling of hypoxia-treated CAMs and NTMs showed that hypoxia does not alter DNA methylation patterns of these cells.

However, it is possible that longer hypoxia exposure may be needed to confer DNA methylation changes. Therefore, to assess whether longer-term hypoxia exposure could induce changes in NTM or cancer cells global DNA methylation, a further LINE-1 pyrosequencing assay was performed (Supplementary Figure S6.7). The results from these studies showed no further changes in global DNA methylation in both NTMs and AGS cells even after 6 and 8 days hypoxia treatment, respectively. Taken together, these data suggest that presence of cancer cells, or possibly a combination of other microenvironmental factors may be needed to confer epigenetic reprograming of gastric myofibroblasts to exert tumour-promoting properties and hypoxia, as a single microenvironmental factor, is unlikely to impose CAM-like DNA methylation patterns in normal tissue myofibroblast (NTMs).

Genome-wide transcriptional profiling revealed unique transcriptional responses to hypoxia among the different populations of gastric myofibroblasts used in this study. These unique transcriptional signatures of hypoxic CAMs, hypoxic ATMs and hypoxic NTMs most likely stem from differential epigenetic backgrounds. Notably, upregulation of cholesterol biosynthesis and fatty acid metabolism emerged as a unique feature of hypoxic CAMs. Cholesterol is an essential structural component of cell membranes that modulates their fluidity and permeability. It serves also as a precursor for many signalling molecules and plays a key role in membrane trafficking, transmembrane signalling processes (Simons and Gerl 2010) and cell proliferation (Fernandez, Lobo et al. 2004, Fernandez, Martin et al. 2005). Cholesterol biosynthesis is tightly regulated at multiple levels and is typically carried out in hepatocytes, adipocytes, brain tissue and lactating breast tissue. However, activation of *de novo* lipogenesis has been frequently observed in cancer (Menendez and Lupu 2007) and enhanced cholesterol biosynthesis has been implicated in development of prostate cancer (Hager, Solomon et al. 2006).

The reduction of the oxygen level in gastric CAMs resulted in upregulation of all components of the superpathway for cholesterol biosynthesis (*Appendix VI File 1*), suggesting that hypoxia enhances production of cholesterol in gastric CAMs. Significantly, several studies have shown that adipocytes promote growth of cancer cells, including breast cancer (Klopp, Zhang et al. 2012) and prostate cancer (Tokuda, Satoh et al. 2003). Nieman et al. have shown that malignant cells induce increased lipolysis in adipocytes, which then secrete fatty acids that are taken up by the cancer cells to be used for energy production (Nieman, Kenny et al. 2011). Therefore it is possible that CAMs under hypoxia may contribute to the energy metabolism of cancer cells. Rapidly proliferating cancer cells require lipids for the generation of biological membranes and as a source of energy during times of nutrient depletion. The coupling of energy metabolism between CAMs and cancer cells may be mediated by fatty acid binding proteins, such as FABP1 and FABP3, which were upregulated in hypoxic CAMs or by exosomes that are secreted by CAMs (as demonstrated in previous Chapters), moreover their secretion is enhanced in hypoxic microenvironment (as shown in this study: GO:0070062 extracellular exosome, $p = 4.75 \times 10^{-4}$ and has also been observed in other systems). King et al. showed that breast cancer cells release greater levels of exosomes when exposed to hypoxia and that these hypoxic exosomes contained higher levels of transported microRNA compared to normoxic exosomes (King, Michael et al. 2012).

Exosomes are known to transport lipids in their membrane, or in their lumen. Exosomal lipids interact with receptors on target cell and are then internalized into endosomes where they concentrate the bioactive lipids that they carry (Record, Carayon et al. 2014). Notably, in ovarian cancer it has been shown that omental adipocytes are able to promote homing, migration and invasion of cancer cells at least in part through release of cytokines (Nieman, Kenny et al. 2011). This may explain why cancer cells preferentially migrate towards CAM-hypoxic-CM. Also, Hedgehog signalling, an important developmental pathway implicated to have an important role in cancer development, progression and metastasis, provides

another link between cholesterol and cancer. Cholesterol can be covalently bond to Hedgehog proteins (Kornberg 2011) which activate Hedgehog signalling by autocrine or paracrine mechanisms in cancer cells (Harris, Samant et al. 2011). Lastly, upregulation of cholesterol biosynthesis under hypoxia in CAMs may serve as adaptive mechanism that protects CAMs from oxidative damage. It has been reported that activation of *de novo* lipogenesis in cancer cells leads to increase membrane lipid saturation, resulting in higher levels of saturated and monounsaturated phospholipids, that potentially protects cancer cells from free radicals and chemotherapeutics (Rysman, Brusselmans et al. 2010).

The transcriptional analysis of the unique gene expression profile identified in hypoxic ATMs revealed that low oxygen concentration induces the expression of genes involved in G2/M cell cycle checkpoint, mitotic spindle assembly and integrin signalling. Significantly, upregulated aurora kinase A (AURKA), which is commonly overexpressed in many types of tumours, emerged as a key regulator of the observed hypoxic ATM phenotype in GSEA leading edge analysis (Supplementary Table S6.1). AURKA is a serine-threonine kinase that plays a critical role in regulating cell cycle and mitosis and ensures correct spindle assembly in normal cells. It was reported that overexpression of AURKA directly leads to malignant transformation and tumour formation (Wang, Zhou et al. 2006). Interestingly, it has also been shown that AURKA regulates pluripotency of embryonic stem cells, in particular cell self-renewal, reprograming and differentiation (Lee, Su et al. 2012). This may suggest that AURKA functions as a regulator of cell stemness. Importantly, Klein et al. showed that hypoxia regulates AURKA expression at the transcriptional level in hepatoma cells (Klein, Fluegel et al. 2008). Recently, it was also shown that AURKA mediates ovarian cancer cell migration and adhesion (Do, Xiao et al. 2014). Integrin signalling plays an important role in cell adhesion, survival, proliferation, differentiation and migration. Integrins are transmembrane receptors that precipitate in cell–cell and cell-extracellular matrix adhesion. They allow cells to sense chemical and physical information in their local environment

and can modulate intracellular signalling pathways and gene expression in response to that (Parsons, Horwitz et al. 2010). Presented analysis revealed that hypoxia affects ATM cells' motility in part via upregulation of talin, an integrin-actin linkage protein, and upregulation of α-actinin, an actin cross-linking protein (*Appendix VI File 2*). Collectively, these observations suggest that upregulation of AURKA and components of integrin signalling pathway under hypoxia increases ATM cells motility and cell cycle progression.

Finally, the transcriptional analysis of the unique gene expression profile identified in hypoxic NTMs showed that these genes mainly contribute to signalling pathways related to hypoxia response and xenobiotic metabolism. The hypoxic microenvironment was previously reported to mediate fibroblast to myofibroblast differentiation (Misra, Fu et al. 2010, Robinson, Neary et al. 2012). In most studies these activated fibroblasts are referred to as cancer-associated fibroblasts/myofibroblasts (CAF/CAMs) since they express myofibroblast cell markers.

The IPA downstream effect analysis predicted distinct biological effects of hypoxia among the gastric myofibroblast populations used in this study. This analysis complements the aforementioned observations. Significantly, activation of autophagy emerged as a key CAM-specific process that was enhanced in response to hypoxia, whereas processes such as organization of organelle, growth of microtubules and aneuploidy of cells appeared to be activated only in hypoxic ATMs. Interestingly, expression and transcription of RNA were the most significantly activated processes in hypoxic NTMs and the most significant processes among all studied myofibroblast populations. In addition, processes, such as mitosis, angiogenesis, invasion of cells, growth of axons, neoplasia of tumour cell lines and glycolysis of cells were predicted to be most active in hypoxic NTMs.

The secretome analysis confirmed some of the predictions made based on hypoxia-induced CAM and NTM gene expression profiles. Interestingly, lysyl oxidase (LOX), among others, was verified at the protein level as an NTM unique hypoxia-induced secreted protein. LOX is an extracellular amine oxidase that post-translationally modifies collagens and elastin in the extracellular matrix therefore catalyses the covalent crosslinking of fibers (Kagan and Li 2003). LOX is crucial for hypoxia-induced metastasis. It is highly expressed in hypoxic tumour cells, including breast cancer, where it remodels extracellular matrix in the lungs, leading to the formation of a metastatic niche (Erler, Bennewith et al. 2009). The comparative analysis of secretomes and gene expression profiles revealed that both expression and secretion of LOX is increased in CAMs compared to NTMs under normal conditions. However, when oxygen is limited LOX is induced in NTMs, both at the mRNA and protein level, but not in CAMs. Neither LOX mRNA transcript nor protein was identified as being differentially expressed/secreted in response to hypoxia in CAMs (Supplementary Figure S6.6) supporting the notion that NTMs acquire CAM-like phenotype under hypoxia.

The differential analysis of proteins secreted by hypoxic CAMs compared to control CAMs and hypoxic NTMs compared to control NTMs revealed that hypoxia has a significantly bigger impact on protein secretion in NTMs than CAMs. This is in line with aforementioned IPA downstream effect analysis, which predicted expression ($p = 2.02 \times 10^{-15}$, *z-score = 2.611*) and transcription ($p = 2.74 \times 10^{-13}$, *z-score = 3.081*) of RNA as the most significantly activated processes in hypoxic NTMs and also the most significant when compared to hypoxic ATMs or hypoxic CAMs as transcription in hypoxic CAMs was activated to a lesser extend ($p = 3.34 \times 10^{-7}$, *z-score = 2.564*). Taken together, these data may suggest that CAMs are somehow less responsive to low oxygen levels then NTMs. This notion can be supported by the GSEA results presented in Chapter IV which showed upregulation of hypoxia-induced genes in CAMs compared to both ATMs and NTMs (Table 4.2).

Significantly, proteins differentially secreted by hypoxic NTMs compared to control NTMs were predicted to be regulated by HIF-1α activation and enhance tumour cell proliferation, invasion, cell movement, angiogenesis, vasculogenesis and migration of endothelial cells. These data are consistent with the experimental data showing increased cancer cell migration and proliferation in response to NTM-hypoxic-CM compared to NTM-ctrl-CM. Proteins differentially secreted by hypoxic CAMs compared to control CAMs were predicted to reduce cell death, including necrosis and apoptosis. Taken together, these results are in line with gene expression analysis of hypoxia-induced NTM- and CAM-specific gene signatures.

In summary, this study is the first to investigate in the integrative manner the effects of hypoxia on DNA methylation, gene expression and protein secretion of gastric primary myofibroblast cells derived from cancer tissue (CAMs), pre-neoplastic tissue (ATMs) and normal tissue (NTMs). The analysis identified hypoxia-induced factors that may drive CAM-, ATM- and NTM- specific hypoxia response and therefore have differential effects on cancer cell migration and proliferation.

# Chapter VII

# Final Discussion

## 7.1 Summary and Conclusions

The tumour microenvironment plays an important role in cancer development and progression therefore much effort has been made to understand the molecular alterations within tumour stroma and their effects on cancer cells. In this study, an integrated multi-omics approach was used to understand regulation and crosstalk in the tumour microenvironment. In line with the two main objectives of this study, several key findings have emerged:

- Myofibroblasts derived from the site of gastric and oesophageal tumours (CAMs) have distinct genome-wide DNA methylation patterns compared to non-tumour derived myofibroblasts (ATMs and/or NTMs). These widespread alterations of DNA methylation in gastric and oesophageal CAMs can provide potential clues as to molecular mechanism of tumour programing of stromal cells.
- DNA methylation patterns at multiple genomic loci were identified either as proxies for gastric CAM identification or as conserved signatures in both gastric and oesophageal patient-matched CAM and ATM samples. DNA methylation at these genomic loci may be used as stromal biomarkers with potential clinical relevance for gastric/oesophageal cancer.
- DNA methylation is involved in the epigenetic regulation of biological pathways and processes involved in the tumour–promoting function of gastric CAMs. In particular, promoter DNA hypomethylation emerged as a regulatory mechanism for transcriptional activation of genes involved in secretion and transport of molecules while promoter DNA hypermethylation emerged as a regulatory mechanism for transcriptional repression of genes involved in pathology of gastrointestinal cancers and regulation of developmental processes.
- Several potential gastric CAM biomarkers (*SMAD3, SPON2, ZNF536, FOXF1, FENDRR*) were validated on an independent set of patient-matched

CAM and ATM samples and unrelated NTM samples by pyrosequencing and qPCR.

- Gastric myofibroblasts purified from different tissue microenvironments (CAMs, ATMs, NTMs) have distinct responses to hypoxia, which are in part due to differential epigenetic backgrounds.
- Presented data provide a novel insight into the profile of changes in gene expression and protein secretion that result from exposure of normal (NTMs) or cancer-associated myofibroblasts (CAMs) to hypoxic conditions. Analyses performed in this study identified hypoxia-induced factors that may drive CAM-, ATM- and NTM- specific hypoxia responses and therefore have differential effects on cancer cell migration and proliferation.

Taken together, this study identified distinct genome-wide DNA methylation signatures that are characteristic for gastric tumour derived stromal myofibroblasts (CAMs) and that distinguish these cells from non-tumour derived gastric myofibroblasts (ATMs and/or NTMs). The fact that CAMs retain their tumour-promoting properties and specific gene expression signatures in the absence of neighbouring tumour epithelium over several passages *in vitro*, suggests that tumour stroma is composed of specialized myofibroblast population, which has been epigenetically programmed by neighbouring neoplastic epithelium to function to its advantage and favour tumour growth and malignant progression (Figure 7.1).

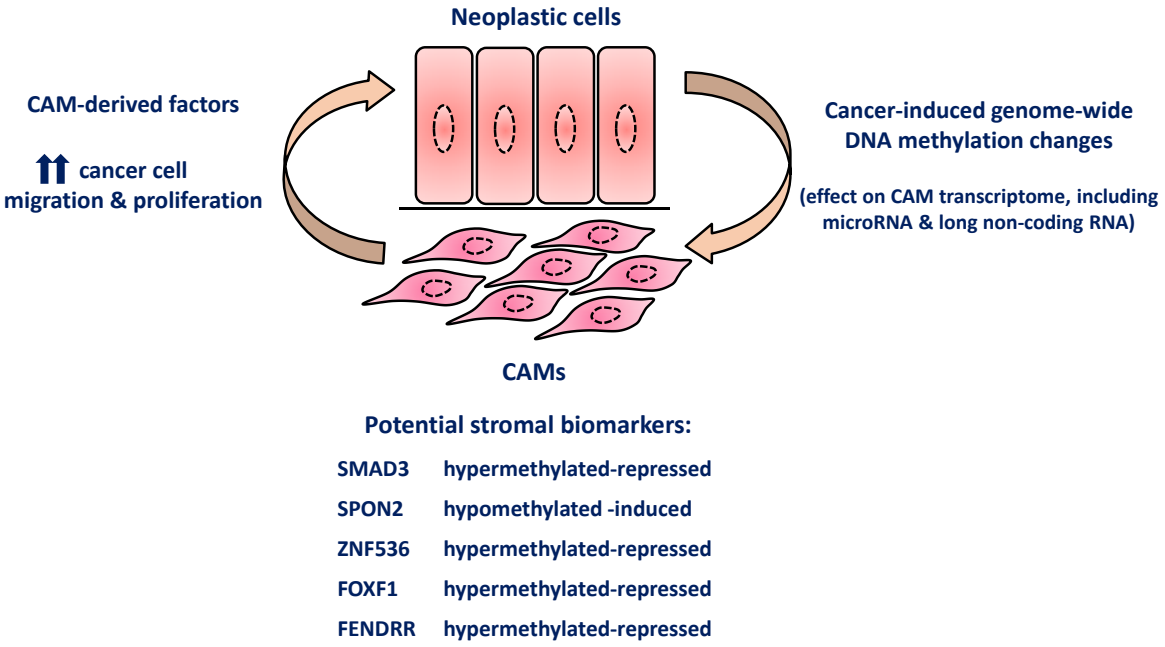


**Figure 7.1 Novel insights into molecular mechanisms underlying the tumour-promoting phenotype of gastric CAMs that have emerged from this study.**

## 7.2 DNA methylation profiling of CAMs

This study is the first to examine genome-wide DNA methylation patterns at individual CpG resolution in primary gastric and oesophageal patient-matched CAM and ATM samples. The identified widespread alterations of DNA methylation in gastric and oesophageal CAMs can provide potential clues as to the molecular mechanism of cancer programing of stromal cells and may serve as potential diagnostic biomarkers. In addition, multiple genomic loci identified as proxies for gastric CAM identification may prove useful as biomarkers for improved diagnosis and prognosis of gastric cancer.

Although, there is much knowledge about epigenetic alterations in cancer cells relatively little is known about epigenetic alterations within tumour stroma. Our knowledge of genome-wide epigenetic changes within tumour stroma is based largely on two pioneering studies (Hu, Yao et al. 2005, Jiang, Gonda et al. 2008). Hu et al. examined non-neoplastic stromal cells, including epithelial and myoepithelial cells, and stromal fibroblasts from normal breast tissue, and *in situ* and invasive breast carcinomas showing that distinct epigenetic alterations occur in all three cell types during breast tumorigenesis in a tumour stage– and cell type–specific manner (Hu, Yao et al. 2005). This study suggested that epigenetic alterations have a role in the maintenance of the abnormal cellular microenvironment in breast cancer. Significantly, Jiang et al. reported that CAMs derived from gastric tumours are characterised by global loss of DNA methylation concomitant with focal gain of DNA methylation (Jiang, Gonda et al. 2008). Although this study is of particular interest to the work presented in this thesis, it did not examine the gene-specific patterns of DNA methylation in CAMs and ATM/NTM. As such, the present study is the first to provide strong evidence for CAM-specific DNA methylation signatures that may contribute to the regulation of CAM-specific gene expression profiles and therefore show that tumour promoting properties of CAMs may in part be due to epigenetic programming.

## 7.3 Global hypomethylation in CAMs

Comparative analysis of global DNA methylation between CAMs and patient-matched ATMs, isolated from gastric and oesophageal cancers showed that the mean global DNA methylation of CAMs is statistically lower than the mean global DNA methylation of corresponding ATMs which is in agreement with the aforementioned study in gastric CAMs reported by Jiang, Gonda et al. 2008. Significantly, a recently published analysis of genome-wide DNA methylation profiling of primary lung CAMs also reported global DNA hypomethylation concomitant with focal gain of DNA methylation in CAMs compared to patient-

matched control fibroblasts derived from non-small cell lung cancer (NSCLC) (Vizoso, Puig et al. 2015). Taken together, these findings support the notion that global loss of DNA methylation concurrent with focal DNA hypermethylation can be used as a defining epigenetic biomarker of CAMs derived from different solid tumours and may underlie their tumour-promoting phenotype.

Interestingly, the global DNA hypomethylation in CAMs parallels the overall loss of DNA methylation that has been well documented in cancer cells. However, in contrast to cancer cells, CAM global DNA hypomethylation does not appear to be associated with chromosomal instability as CAMs are non-neoplastic and exhibit senescence *in vitro*. This phenomenon was also reported by Jiang, Gonda et al. 2008. They also reported lack of obvious changes in DNMTs expression in CAMs and concluded that reduced expression of DNMTs probably does not explain the global loss of DNA methylation in CAMs (Jiang, Gonda et al. 2008). Notably, Merry et al. recently demonstrated that deregulation of lncRNAs associated with DNMT1 contributes to global changes in DNA methylation patterns without any detectable mutations or changes in the expression of DNMTs (Merry, Forrest et al. 2015). They identified a subset of lncRNAs that interact with DNMT1 in colon cancer cells. One lncRNA named *DACOR1* (DNMT1-associated Colon Cancer Repressed lncRNA 1) was shown to be repressed in a panel of colon tumours and patient-derived colon cancer cell lines. Induction of *DACOR1* in colon cancer cell lines impaired growth and colony formation, and affected global gene expression. Significantly, *DACOR1* induction resulted in downregulation of cystathionine β-synthase, which is known to lead to increased levels of S-adenosyl methionine (SAM), the key methyl donor for DNA methylation (Merry, Forrest et al. 2015). These findings suggest that DNMT1, through its interaction with lncRNA *DACOR1*, indirectly regulates the cellular levels of SAM and subsequent genome-wide DNA methylation in colon tissue. Notably, accumulating studies have shown that lncRNAs play an active role in gastric cancer tumorigenesis, metastasis, prognosis and drug resistance (Fang, Pan et al. 2015, Gu, Chen et al. 2015). However, given that many lncRNAs have tissue-specific expression patterns and are poorly conserved, it remains to be

elucidated whether dysregulation of lncRNAs contributes to global DNA hypomethylation in other types of cancer and whether similar mechanism exists in CAMs derived from these tumours.

## 7.4 DNA methylation in CAMs from different tissues

Data presented in this thesis show that epigenetic changes can provide a potential mechanism for gastric CAMs to acquire their tumour-promoting phenotype. However, it is not clear whether these epigenetic changes are conserved in CAMs derived from different tumours and whether they target the same genes and/or genomic regions.

The genome-wide DNA methylation profiling of gastric and oesophageal CAMs identified 230 susceptible CpG loci that are differentially methylated in both gastric and oesophageal CAMs when compared to patient-matched ATMs. Notably, 172/230 of the identified CpG loci were changed in the same direction in gastric and oesophageal CAMs. The conserved gastric and oesophageal CAM susceptible loci are distributed across the genome and 65.22% of these loci are associated with 2 genes and 33.5% are associated with only one gene. Notably, the analysis on a gene level identified 2223 common differentially methylated genes in gastric and oesophageal CAMs. These genes were found to largely target signalling pathways that are either commonly dysregulated in human cancers or are involved in transcriptional regulation of pluripotent stem cells and developmental processes. Importantly, aberrant promoter methylation of a number of genes (e.g. *ZMIZ1, EYA4, SLC22A18AS, WIPF1, FAM49A, RUNX3, ESRRG*) found in both gastric and oesophageal CAMs were also reported in CAMs derived from lung cancer (Vizoso, Puig et al. 2015) indicating that epigenetic changes may contribute to the aberrant expression of these common genes in CAMs derived from different tumours and

ultimately disrupt the same signalling pathways. However, further studies on much bigger sample sets are needed to determine whether that is the case.

Data presented in this study (Chapter V) strongly suggests that DNA methylation may confer a mechanism for the transcriptional regulation of key TGF-β signalling components in gastric CAMs. Specifically, promoter hypermethylation may be associated with *SMAD3* downregulation in gastric CAMs. Notably, *SMAD3* was also downregulated in oesophageal CAMs; however the methylation status of *SMAD3* promoter in oesophageal CAMs needs to be examined. Interestingly, Vizoso et al. also found promoter hypermethylation-associated *SMAD3* silencing in CAMs derived from lung cancer. The *SMAD3* hypermethylation was associated with hyperresponsiveness to exogenous TGF-β1 in terms of contractility and ECM deposition. They examined the expression of a panel of wound-related ECM genes and showed that *COL1A1, EDA-FN, LOX* and *SPARC* are upregulated in CAMs compared to patient-matched control fibroblasts after 5 days of treatment with TGF-β1 (Vizoso, Puig et al. 2015). Significently, we also observed the upregulation of *COL1A1, SPARC* and *LOX* in gastric CAMs compared to ATM in our gene expression data. Collectivly, based on these observation promoter hypermethylation-associated *SMAD3* silencing may potentially be a major contributor to CAM aberrant phenotype and may underlie tumour-promoting properties of CAMs derived from different tissues. Further studies are needed to determine whether promoter hypermethylation-associated *SMAD3* silencing may be used as define epigenetic biomarker of CAMs derived from different tumours and to assess its function in cancer-specific CAM programming.

## 7.5 Stromal DNA methylation signatures

In this study, genome-wide DNA methylation profiling identified stromal DNA methylation signatures that may potentially be used in clinical practice as biomarkers for improved diagnosis and prognosis of gastric cancer.

Significantly, region on chromosome 16 that spans 526,426 bp was identified as one of the largest hypermethylated region in gastric CAMs compared to both ATMs and NTMs. Interestingly, this region is commonly hypermethylated in gastric tumours and gastric cancer cell lines (Supplementary Figure S5.10 and Figure S5.11). DNA methylation within this region may regulate the expression of lncRNA *FENDRR* (Figure 5.6). Significantly, Xu et al. reported that expression of *FENDRR* is downregulated in gastric cancer cells compared to normal gastric epithelial cells (Xu, Huang et al. 2014) which is in line with aforementioned TCGA data. Low expression of *FENDRR* was significantly correlated with poor prognosis and aggressive tumour characteristics, such as greater invasion depth, higher tumour stage, and lymphatic metastasis (Xu, Huang et al. 2014). Although, further studies are needed to correlate the hypermethylation of *FENDRR* with patient prognosis in CAM, DNA methylation of this region may potentially be used as biomarkers for improved diagnosis and prognosis of gastric cancer given that generally in clinical practice whole tumour tissue samples are profiled rather than isolated neoplastic cells. Therefore this finding may translate into clinically useful biomarker of gastric tumours that may also aid patient stratification and prognosis.

## 7.6 Limitations of the study

Although microarray technologies are widely used in biomedical research they have several limitations that should be kept in mind when interpreting the results. In this study, Illumina Infinium HumanMethylation450 BeadChip and HumanHT-12v4 Expression BeadChip arrays were used to profile genome-wide DNA methylation and gene expression of primary patient-matched CAM and ATM samples. Albeit, both Illumina 450k and HumanHT-12 arrays provide good genome-wide coverage as they allow investigation of DNA methylation levels of many previously identified CpG sites and expression levels of many well-characterised genes, they are limited to the pre-designed probe set therefore do not allow discovery of novel transcripts, including microRNAs and long non-coding RNAs as well as 'novel' methylation signatures that may be located outside of the pre-determined regions. More comprehensive next generation sequencing-based approaches, such as RNA- and bisulfite DNA sequencing address some of these limitations and can provide a better understanding of the epigenetic changes that contribute to CAM aberrant phenotype.

Cell culture conditions and cell proliferation rate may affect DNA methylation patterns (Wilson and Jones 1983, Jiang, Gonda et al. 2008). In this work, genome-wide DNA methylation analysis was performed on early-passage primary myofibroblast cultures and patient-matched CAM and ATM pairs were always cultured in parallel to reduce cell culture-induced artifacts. However, some specific limitations stemming from *in vitro* cell propagation were not addressed. Also, it is not known whether there are any differences in proliferation rate between patient-specific CAM and ATM pairs and whether these differences if present would have effect on DNA methylation changes.

## 7.7 Concluding remarks

Findings presented in this thesis shed light on the molecular alterations underlying the aberrant phenotype of gastric CAMs and identify stromal biomarkers with potential clinical relevance. However, more comprehensive studies on larger sample sets are needed to confirm the true utility of identified trends indicating a correlation between stromal DNA methylation signatures and clinicopathological features of gastric tumours. Also, further studies are needed to characterize the role of the genes that were shown to be regulated by aberrant DNA methylation in gastric CAMs. These studies may provide novel insights into the different pathways involved in gastric CAM programming, thereby improving our understanding of the complex interactions between tumour and microenvironment and potentially contribute to new diagnostic and treatment strategies.

# Supplementary Data

# *Chapter II*

## *Materials and Methods*

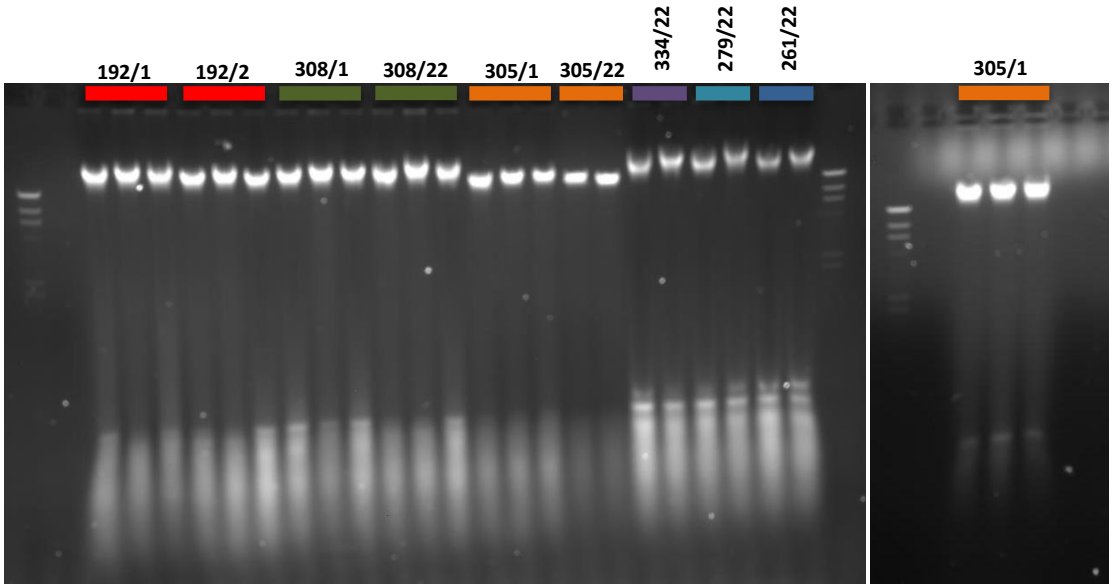


**Figure S2.1 Quality of DNA samples used for Illumina Infinium HumanMethylation450k array;** 1% agarose gel electrophoresis, lambda DNA/HindIII was used as molecular size marker (kb).

**Table S2.1 Quality control information for RNA and DNA samples used for Ilumina HumanHT-12 v4 Expression and Illumina Infinium HumanMethylation450k arrays;** *RIN - RNA Integrity Number*

| | RNA samples | | DNA samples |
|---|---|---|---|
| **Sample** | **RIN** | **RNA 260/280 Ratio** | **DNA 260/280 Ratio** |
| **C1N** | 10 | 2.09 | 2 |
| **C1H** | 10 | 2.1 | 2 |
| **A1N** | 9.9 | 2.08 | 1.99 |
| **A1H** | 9.7 | 2.06 | 1.98 |
| **C2N** | 8.1 | 2.06 | 2.01 |
| **C2H** | 7.8 | 2.11 | 2 |
| **A2N** | 10 | 2.07 | 2 |
| **A2H** | 10 | 2.12 | 1.95 |
| **C3N** | 9.9 | 2.08 | 1.98 |
| **C3H** | 9.8 | 2.07 | 1.95 |
| **A3N** | 9.7 | 2.12 | 1.96 |
| **A3H** | 9.8 | 2.08 | 1.98 |
| **N1N** | 10 | 2.05 | 2 |
| **N1H** | 9.2 | 2.06 | 1.99 |
| **N2N** | 9.9 | 2.08 | 2.02 |
| **N2H** | 9.7 | 2.06 | 2.01 |
| **N3N** | 9.8 | 2.07 | 2.01 |
| **N3H** | 9.8 | 2.05 | 1.98 |

**Table S2.2 Patient information relating to the scoring details.** Row displaying the Total represents the prognosis score and is calculated based on the sum of all the variables.

| Sz187 | Sz197 | Sz389 | Sz192 | Sz45 | Sz195 | Sz294 | Sz268 | Sz271 | Sz194 | Sz308 | Sz305 | Sz190 | Sz198 | Sz42 | Patient code | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CagA +ve | CagA +v | None | CagA +ve | None | CagA +ve | None | CagA +ve | None | *H.pylori* | *H.pylori* | CagA +ve | None | *H.pylori* | None | *H.pylori* status. H.pylori = infected CagA +ve =cagA strain detected None = No H.pylori infection | |
| F | M | M | F | M | F | F | M | M | M | M | M | F | M | M | Gender | |
| 39 | 54 | 67 | 49 | 82 | 85 | 84 | 76 | 72 | 76 | 51 | 59 | 65 | 77 | 72 | Age | |
| TOTAL | TOTAL | DISTAL | TOTAL | DISTAL | DISTAL | TOTAL | TOTAL | TOTAL | DISTAL | TOTAL | TOTAL | TOTAL | DISTAL | TOTAL | Type of resection | |
| 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | Anaemia HGB<100G/L or Blood Transfusion | |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | BMI <19 or Bodyweight loss >10KG | |
| 4 | 2 | 3 | 3 | 3 | 1 | 3 | 4 | 4 | 1 | 1 | 3 | 4 | 2 | 1 | pTNM | T |
| 2 | 0 | 1 | 2 | 2 | 1 | 0 | 2 | 1 | 0 | 2 | 2 | 0 | 0 | 0 | | N |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | | M |
| 3 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 3 | 3 | 3 | 2 | 3 | Grade | |
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | Histological type: Intestinal or diffuse | |
| 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | Lymphatic vessel invasion | |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | Vascular invasion | |
| 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | Positive margins of resection | |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | Serological tumour marker-elevation | |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | Tumour recurrence | |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | Synchron or metachron tumour | With surgical intervention |
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | | Without surgical intervention |
| 15.5.07 | 16.8.07 | 2.7.06 | 21.5.07 | 26.6.06 | 7.8.07 | 5.12.07 | 6.11.07 | 13.11.07 | 6.8.07 | 8.1.08 | 3.1.08 | 18.5.07 | 27.8.07 | 16.6.06 | Date of operative | |
| 42 | 62 | 50 | 22 | 2 | 13 | 51 | 15 | 24 | 60 | 9 | 17 | 5 | 60 | 75 | Survival (months) | |
| | | | 29.3.09 | 8.06 | 14.9.07 | 12.3.12 | 2.09 | | 1.9.12 | 19.10.08 | 1.8.09 | 14.10.07 | 10.9.12 | | Exit | |
| 11 | 6 | 7 | 11 | 11 | 9 | 7 | 11 | 12 | 4 | 12 | 12 | 13 | 7 | 5 | Total | |
| 6 | 6 | 6 | 4 | 6 | 6 | 6 | 6 | 5 | 6 | 3 | 6 | 6 | 4 | 6 | Myoscore | |

**Table S2.3 Patient information relating to age, gender, tumour location and tumour clinical assessment for oesophageal cancer patients who provided CAM and ATM cells used in this study.** Tumour staging is defined in terms of pathology of tumour (pT) where 0 defines no sign of tumour and 3 maximum size and/or extensions. Similarly, the involvement of local and metastasis to proximal lymph nodes is stated as N0-N3 and metastasis severity by M0, no metastasis and M, distal metastasis.

| Patient ID | Age | Gender | Location of Tumour | Classification | Tumour Staging | Survival (months) | Adjacent Tissue |
|---|---|---|---|---|---|---|---|
| 1. | 72 | M | esophagus | adenocarcinoma | pT3N1M0 | 18 | Barrett's eosophagus |
| 2. | 70 | F | esophagus cardia junction | adenocarcinoma | pT3N1M0 | >25 | Intestinal metaplasia |
| 3. | 64 | M | esophagus cardia junction | adenocarcinoma | pT2N3M0 | 19 | Intestinal metaplasia |

# *Chapter III*

## *Identification of Genome-Wide DNA Methylation Signatures in Gastric and Oesophageal Cancer-Associated Myofibroblasts*

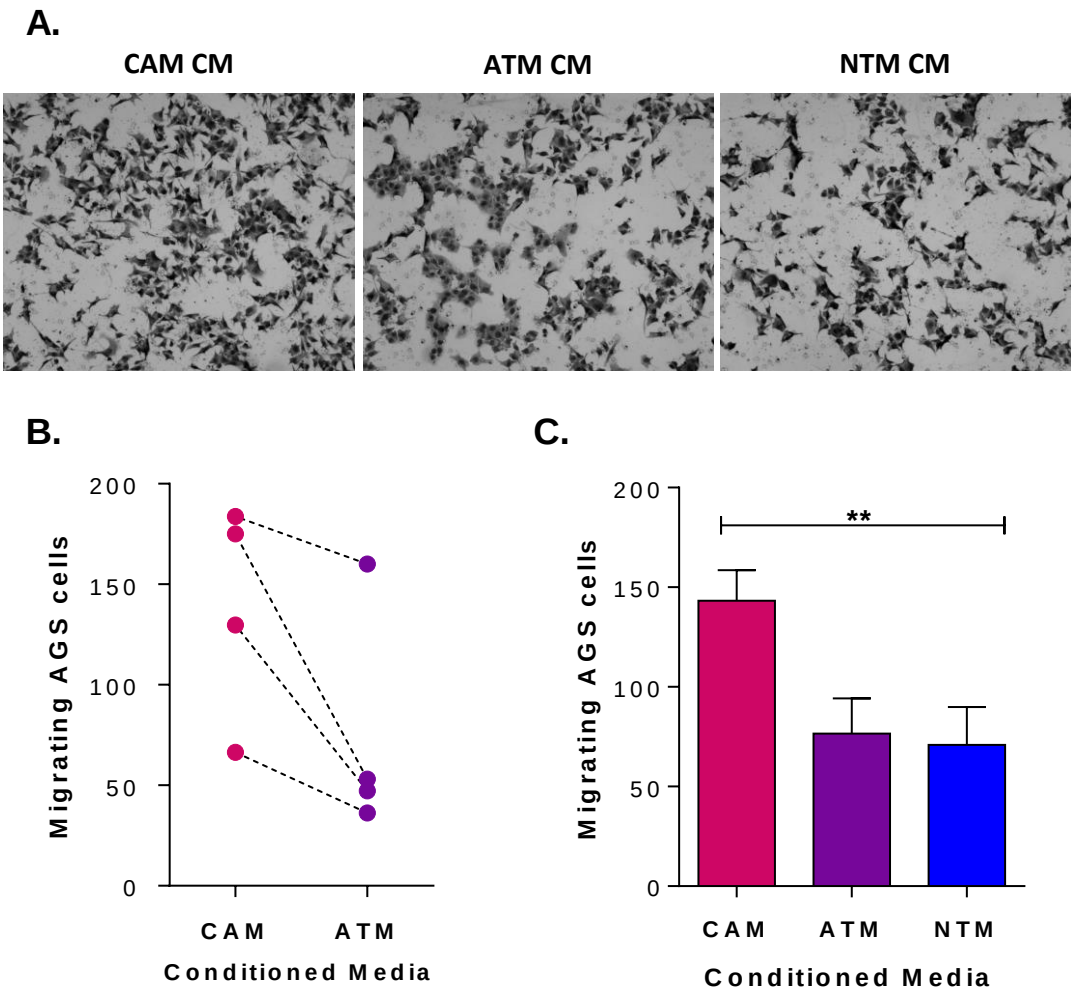


**Figure S3.1 Increased migration of gastric cancer cells in response to CAM conditioned media (CM) compared to ATM-CM and NTM-CM. A.** Representative Boyden chamber migration assays of AGS cells treated with CAM, ATM or NTM conditioned media**. B.** Individual patient-paired Boyden chamber migration assays of AGS cells treated with CAM-CM or ATM-CM. All data points were corrected for basal migration by subtracting respective serum-free media controls. **C.** Group mean data of AGS cell migration in response to CAM-CM (n=4), ATM-CM (n=4) or NTM-CM (n=2). Mean, SEM ** *ANOVA p-value < 0.01 (p=0.0093).*

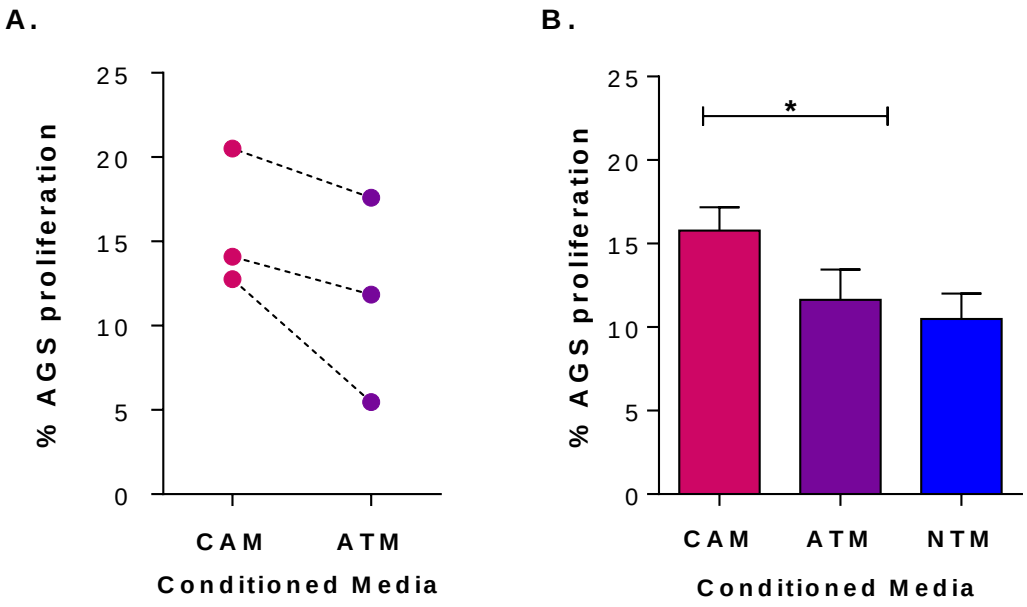


**Figure S3.2 Increased proliferation of gastric cancer cells in response to CAM conditioned media (CM) compared to ATM-CM and NTM-CM. A.** Individual patient-paired EdU proliferation assays of AGS cells treated with CAM-CM or ATM-CM. All data points were corrected for basal proliferation by subtracting respective serum-free media controls. **B.** Group mean data of AGS cell proliferation in response to CAM-CM (n=3), ATM-CM (n=3) or NTM-CM (n=2). Mean, SEM **paired t-test p-value<0.05 (p=0.0143).*

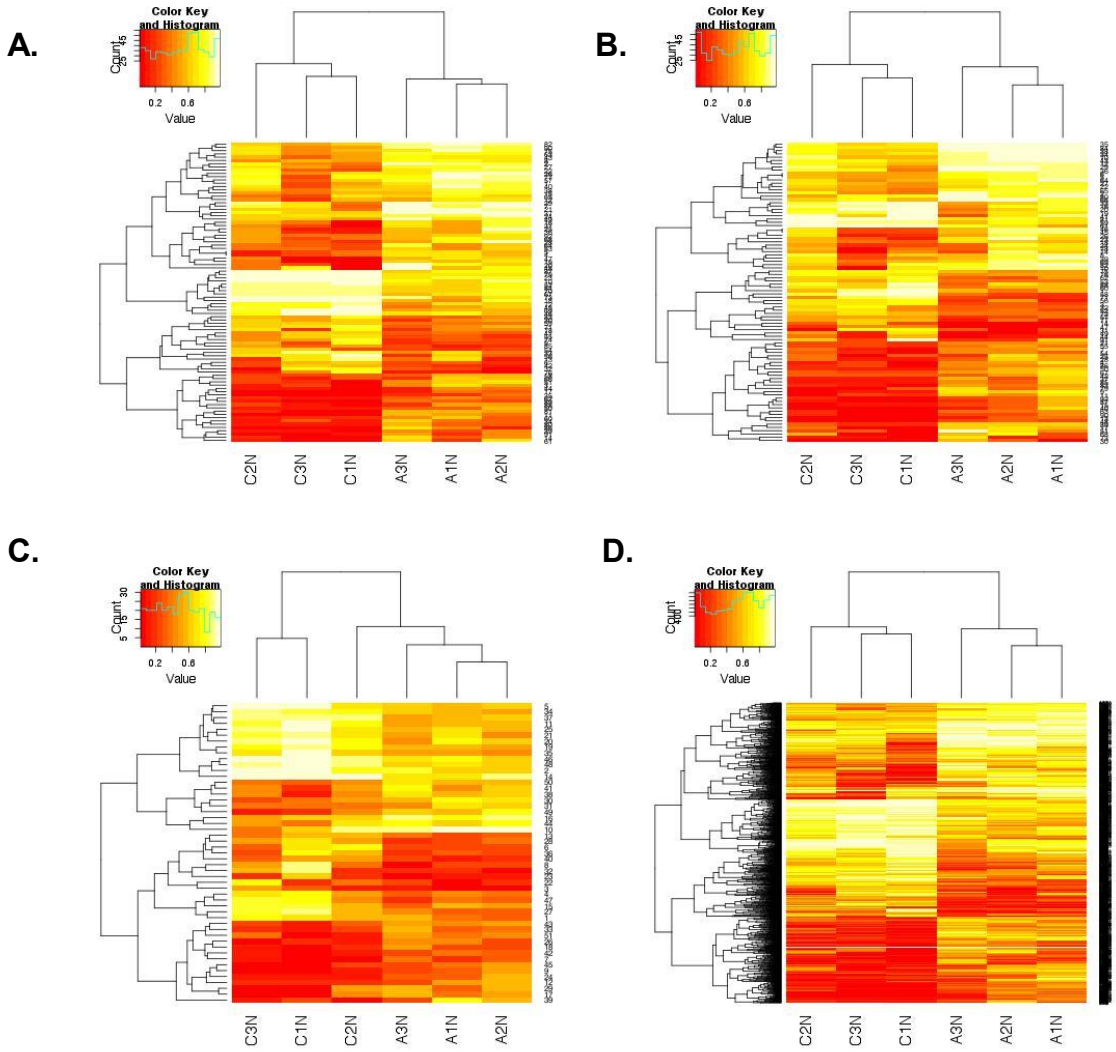


**Figure S3.3 Heatmap representation of differentially methylated regions (DMRs) identified in gastric CAMs compared to their patient-matched ATMs. A.** Promoters **B.** Genes **C.** CpG Islands **D.** Tiling regions as defined in RnBeads* (version 0.99.17)*; |Δβ| > 0.2, p-value <0.05.*

*RnBeads uses: (i) Ensembl gene definitions to define genes and promoters. A promoter is defined as the region spanning 1,500 bases upstream and 500 bases downstream of the transcription start site of the corresponding gene; (ii) UCSC Genome Browser CpG island track to define CpG islands; (iii) tiling regions are defined over the whole genome as regions with a window size of 5 kilobases.

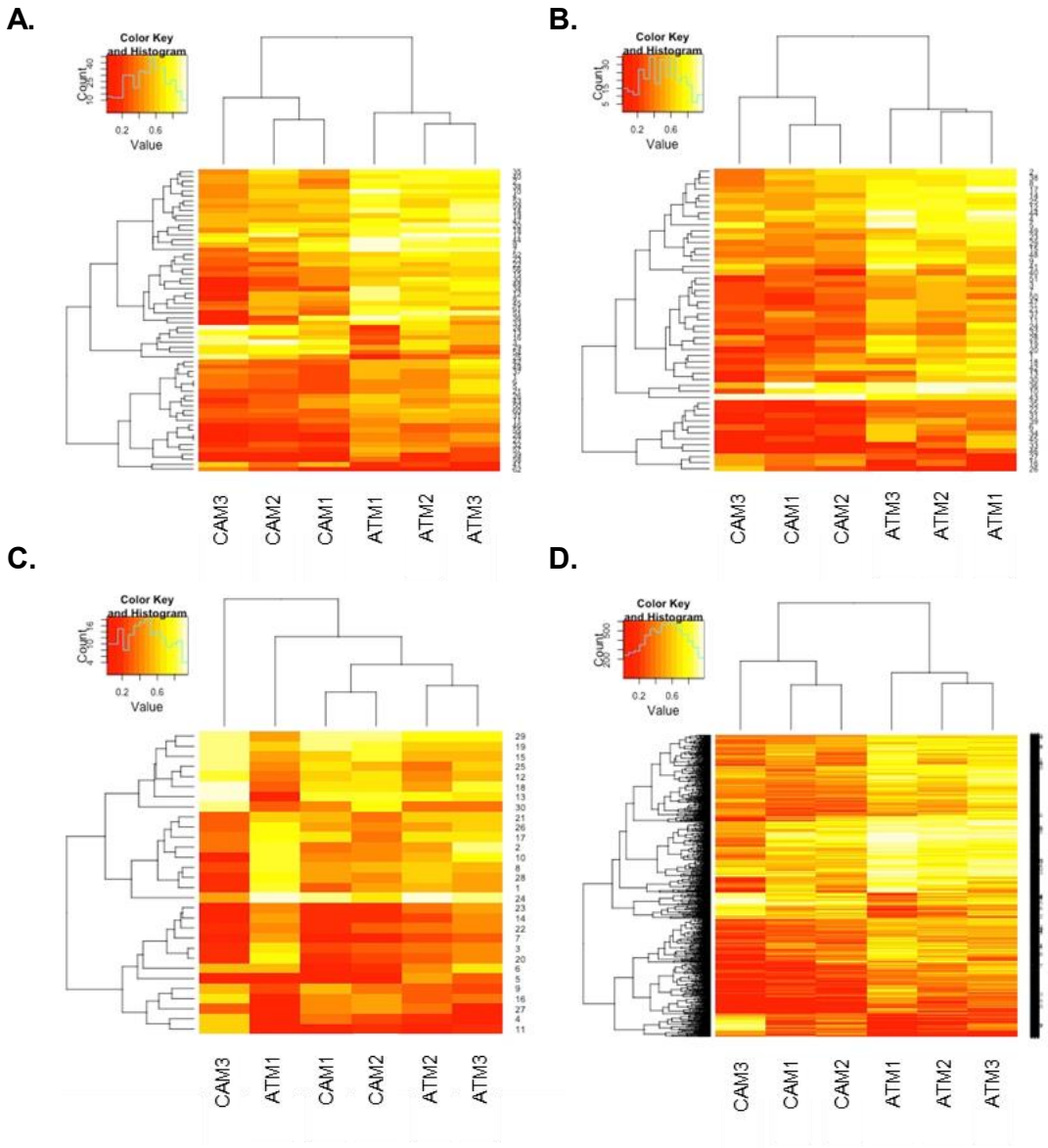


**Figure S3.4 Heatmap representation of differentially methylated regions (DMRs) identified in oesophageal CAMs compared to their patient-matched ATMs.** **A.** Promoters **B.** Genes **C.** CpG Islands **D.** Tiling regions as defined in RnBeads* (version 0.99.17)*; $|\Delta\beta| > 0.2$, p-value <0.05.*

*RnBeads uses: (i) Ensembl gene definitions to define genes and promoters. A promoter is defined as the region spanning 1,500 bases upstream and 500 bases downstream of the transcription start site of the corresponding gene; (ii) UCSC Genome Browser CpG island track to define CpG islands; (iii) tiling regions are defined over the whole genome as regions with a window size of 5 kilobases.

# *Chapter IV*

## *Identification of Gens that Exhibit Correlated Changes in Gene Expression and DNA Methylation in Gastric Myofibroblasts*

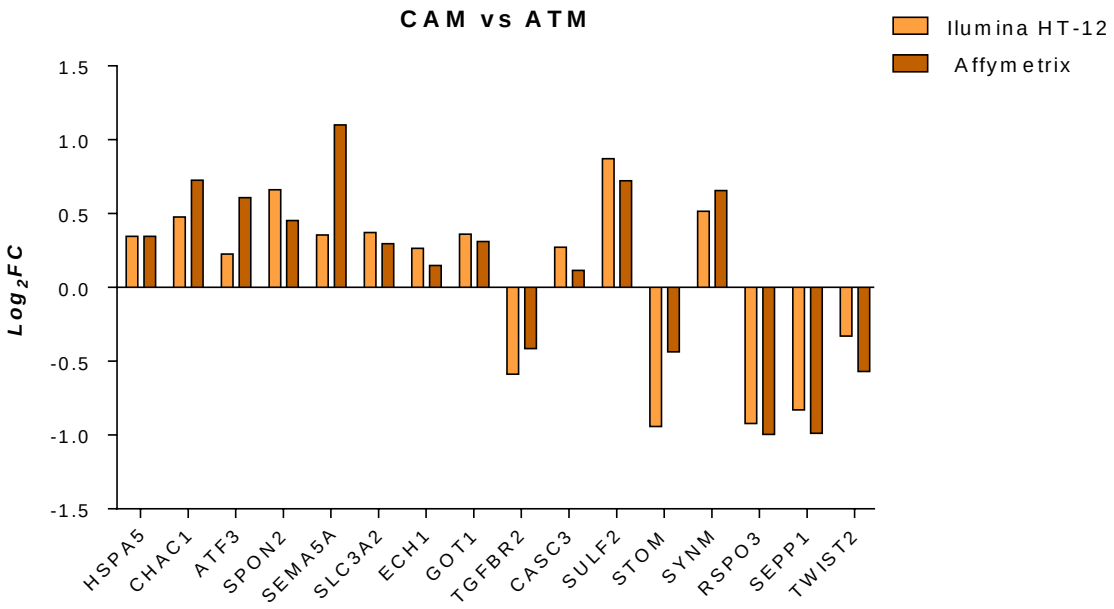


**Figure S4.1 Representative examples of genes showing consistent changes in gene expression profiles in two independent studies of gastric CAM vs ATM gene expression profiles.** These genes were identified independently by Illumina HumanHT-12v4 Expression BeadChip array (yellow) in this study on 3 patient-matched CAMs and ATMs and by Affymetrix GeneChip Human Genome U133 Plus 2.0 Array (dark orange) in previous gene expression study on 12 patient-matched CAMs and ATMs; *p-value < 0.05*.

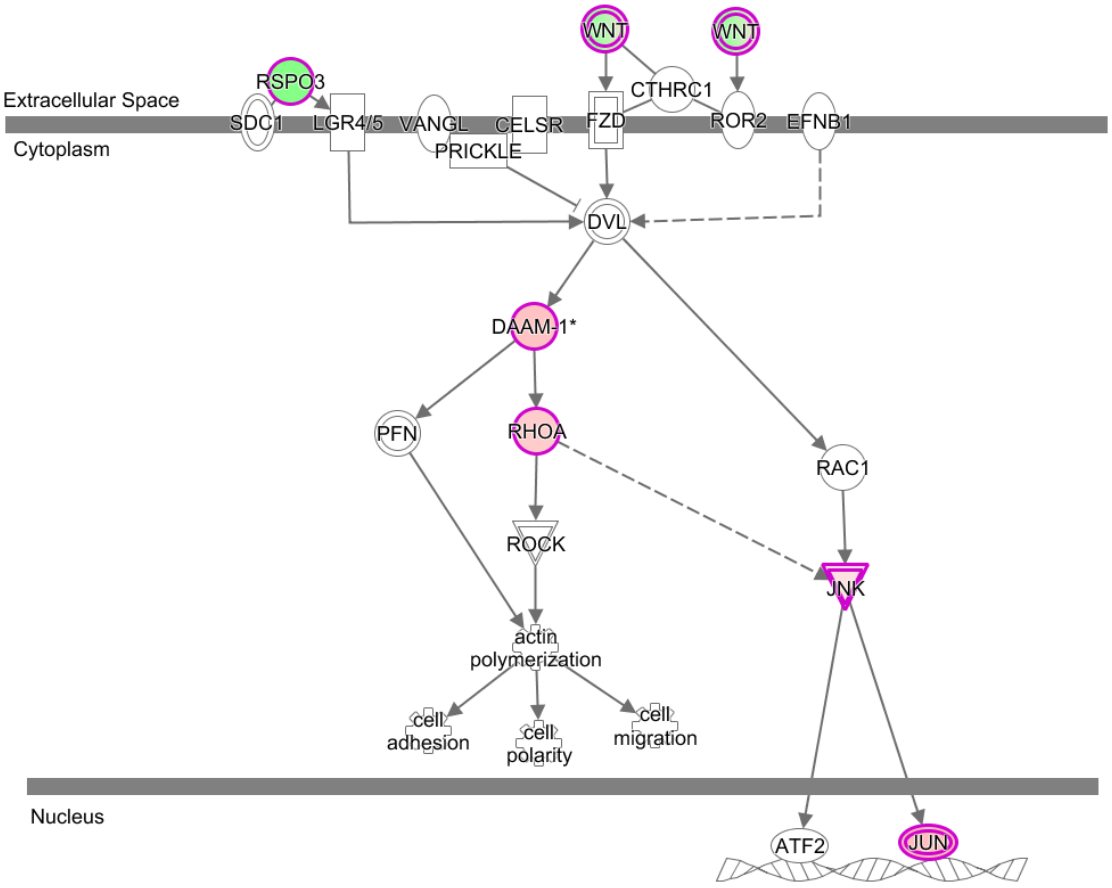


**Figure S4.2 PCP pathway** identified in CAM vs ATM comparison (*p = 1.52 x 10*$^{-2}$; *z-score = 0.333)* by Ingenuity Canonical Pathway Analysis; *red* – upregulated genes in CAMs, *green* – downregulated genes in CAMs.

# *Chapter V*

## *Validation of DNA Methylation Patterns that Correlate with CAM-specific Gene Expression Profiles*

**Table S5.1 Reproducibility of pyrosequencing assays.** Correlation $R^2$ between methylation levels for biological replicates.

| Patient | CD47 | B4GALT6 | NKAIN3 /ASPH | SMAD3 (p1) | SMAD3 (p2) | SPON2 | FOXF1 /FENDRR | ZNF536 | MUC2 | mir802 /RUX1 | HOXA5 | FOXC2 /FOXF2 | VPS28 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **42** | 0.991 | 0.875 | 0.926 | 0.987 | 0.869 | 0.994 | 0.861 | 0.846 | 0.915 | 0.991 | 0.981 | 0.709 | 0.446 |
| **45** | 0.952 | 0.853 | 0.692 | 0.994 | 0.928 | 0.891 | 0.779 | 0.950 | 0.734 | 0.059 | 0.884 | 0.326 | 0.800 |
| **192** | 0.991 | 0.969 | 0.954 | 0.997 | 0.952 | 0.974 | 0.997 | 0.888 | 0.979 | 0.987 | 0.972 | 0.961 | 0.937 |
| **268** | 0.836 | 0.814 | 0.315 | 0.969 | 0.367 | 0.641 | 0.875 | 0.828 | 0.869 | NA | 0.843 | 0.870 | 0.882 |
| **271** | 0.827 | 0.809 | 0.896 | 0.947 | 0.909 | 0.467 | 0.859 | 0.981 | 0.671 | 0.852 | 0.959 | 0.843 | 0.966 |
| **305** | 0.905 | 0.949 | 0.974 | 0.988 | 0.989 | 0.990 | 0.981 | 0.986 | 0.955 | 0.990 | 0.974 | 0.902 | 0.800 |
| **308** | 0.953 | 0.281 | 0.867 | 0.902 | 0.586 | 0.994 | 0.981 | 0.932 | 0.951 | 0.969 | 0.971 | 0.983 | 0.919 |
| **279** | 0.987 | 0.999 | 0.955 | 1.000 | 0.619 | 0.966 | 0.988 | 1.000 | 0.913 | 0.997 | 0.982 | 0.959 | 0.318 |
| **334** | NA | NA | 0.929 | NA | NA | NA | 0.978 | 0.951 | 0.957 | 0.908 | 0.987 | NA | NA |

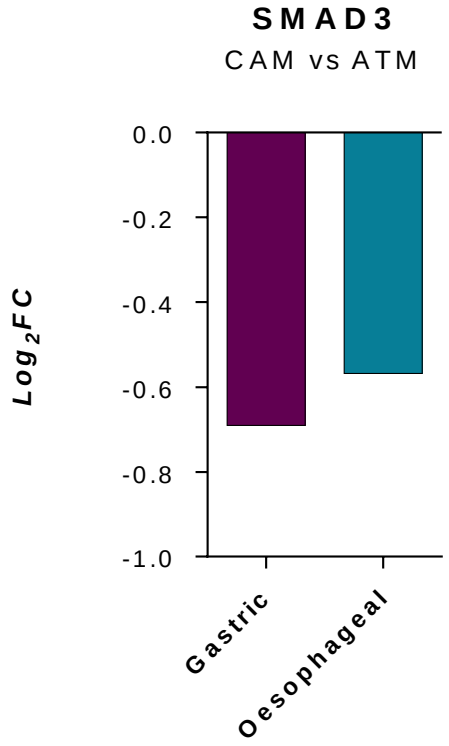


**Figure S5.1 *SMAD3* expression in gastric and oesophageal CAMs compared to their parent-matched ATMs.** *SMAD3* expression levels were detected by Affymetrix GeneChip Human Genome U133 Plus 2.0 Array (purple) in a previous gene expression study performed on 12 gastric patient-matched CAMs and ATMs, *p-value* < *0.05* (Smith et al, unpublished data) and by Illumina HumanHT-12v4 Expression BeadChip array (blue) as part of this study in 4 oesophageal patient-matched CAMs and ATMs, *p-value* < *0.05* (collaborative work with the Varro group, University of Liverpool).

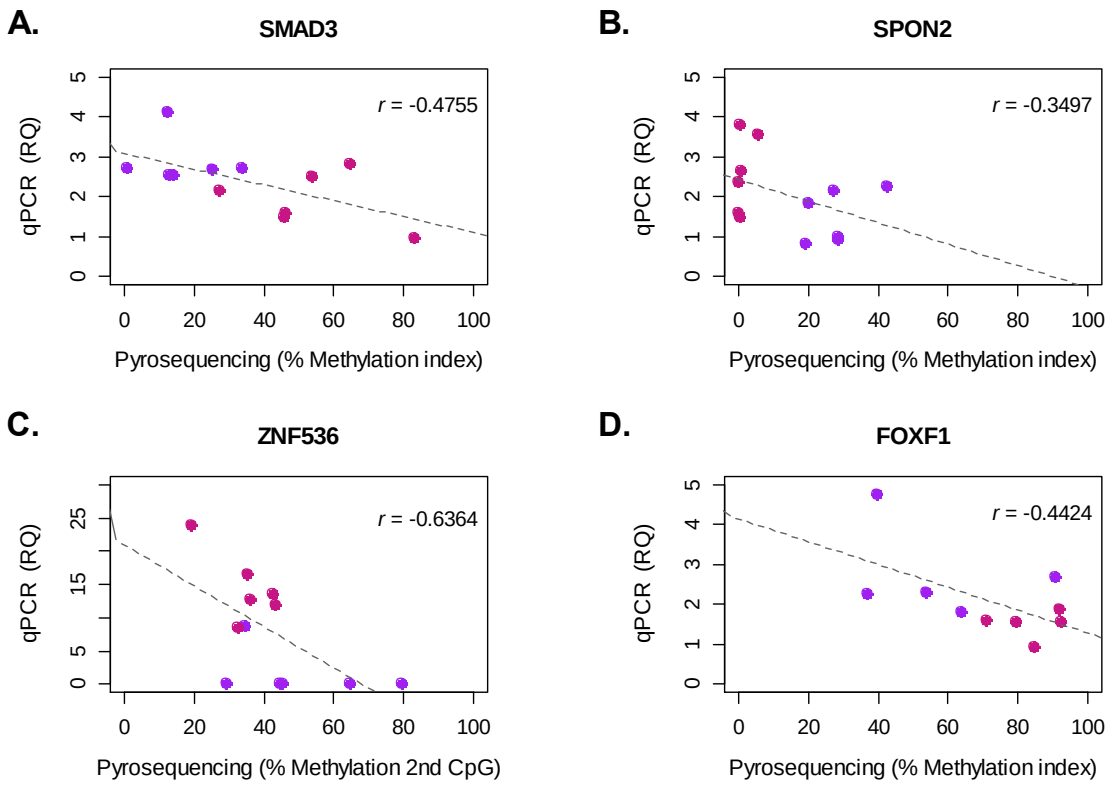


**Figure S5.2 Negative correlations between gene expression (qPCR data) and promoter methylation (pyrosequencing data) in patient-matched CAMs and ATMs.** Individual sample methylation levels (*methylation index for the interrogated region*, *X-axis*) are plotted against individual sample gene expression levels (*RQ-value relative to calibrator*, *Y-axis*). **A.** *SMAD3* (n=6) **B.** *SPON2* (n=6) **C.** *ZNF536* (n=6) **D.** *FOXF1* (n=5); *magenta – CAMs, purple – ATMs*; *r - Spearman's rank correlation rho*.

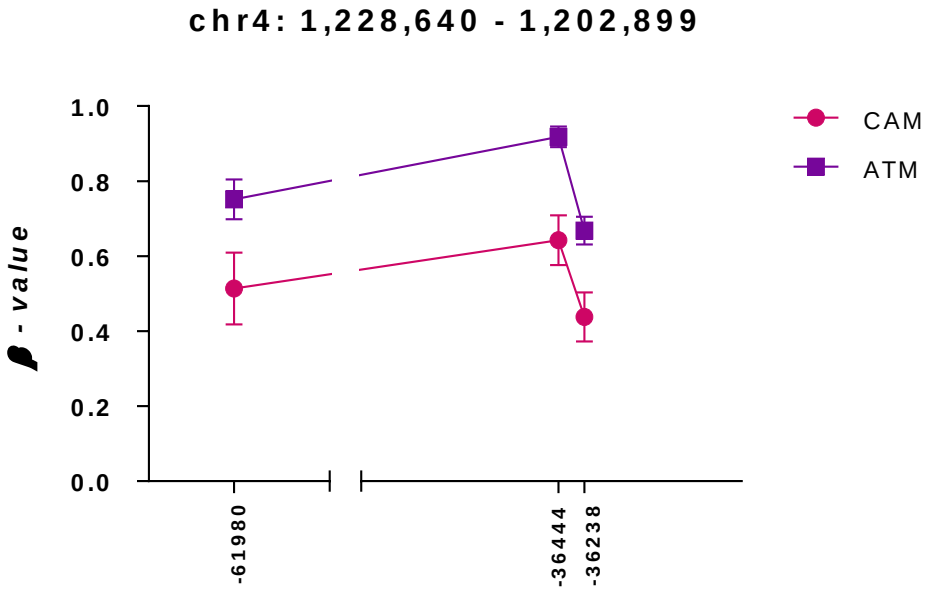


**Figure S5.3 Differentially methylated *SPON2* promoter-region identified by Illumina 450k array in oesophageal patient-matched CAM and ATM samples.** Mean β-values (n=3) for probes identified as differentially methylated in oesophageal CAM vs ATM comparison are plotted. The *X*-axis indicates distance of Illumina 450k probes to *SPON2* transcription start site. Error bars represent SEM.

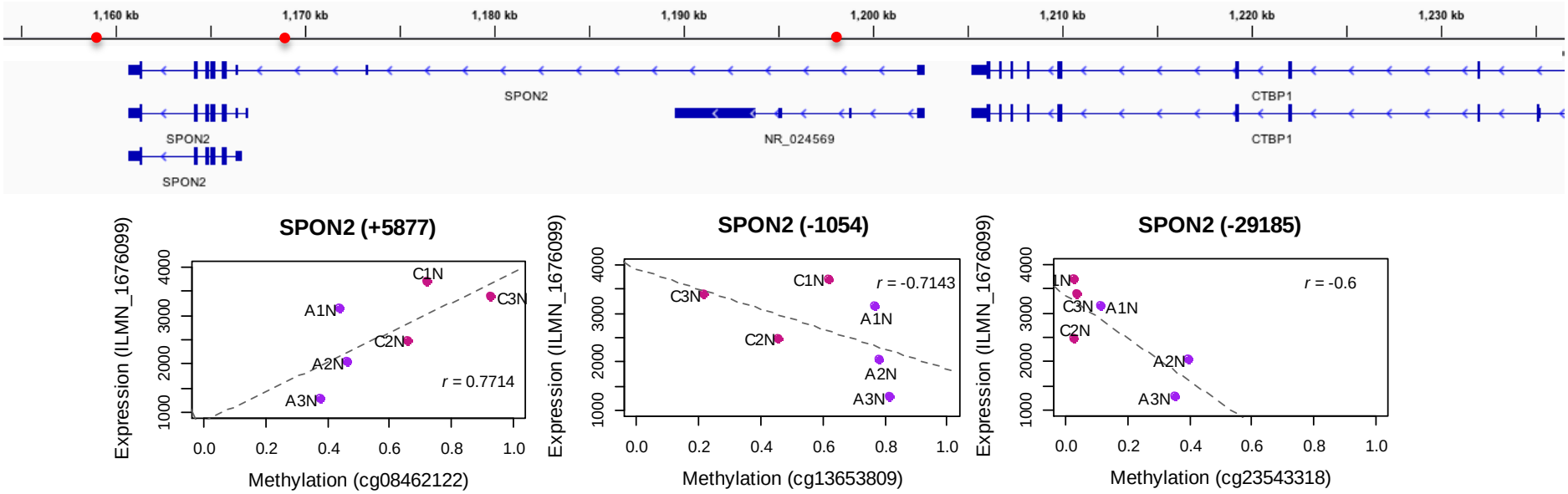


**Figure S5.4 Simultaneous negative correlation between promoter methylation and *SPON2* gene expression and positive correlation between gene-body methylation and *SPON2* gene expression levels in patient-matched CAM and ATM samples detected in genome–wide DNA methylation and gene expression studies using Illumina 450k and HT-12 arrays.** Individual sample methylation levels (β-value, *X-axis*) are plotted against individual sample gene expression levels (intensity, *Y-axis*) for 3 representative CpG loci (red dots) associated with *SPON2* regulation; *magenta – CAMs, purple – ATMs*; numbers in brackets indicate the distance to transcription start site of *SPON2*; *r - Spearman's rank correlation rho*.

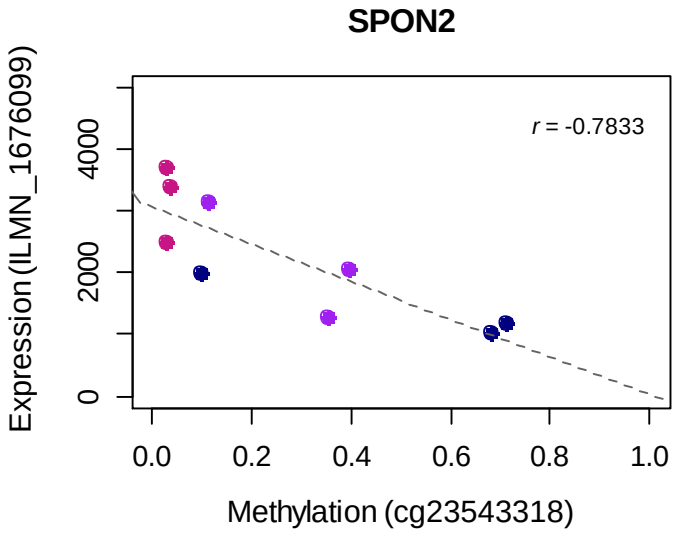


**Figure S5.5 Negative correlation between *SPON2* gene expression and promoter methylation in patient-matched CAMs and ATMs and NTMs detected in genome–wide DNA methylation and gene expression study using Illumina 450k and HT-12 arrays.** Individual sample methylation levels (β-value, *X-axis*) are plotted against individual sample gene expression levels (intensity, *Y-axis*) for CpG loci associated with *SPON2* regulation; *magenta – CAMs, purple – ATMs, navy - NTMs*; *r - Spearman's rank correlation rho*.

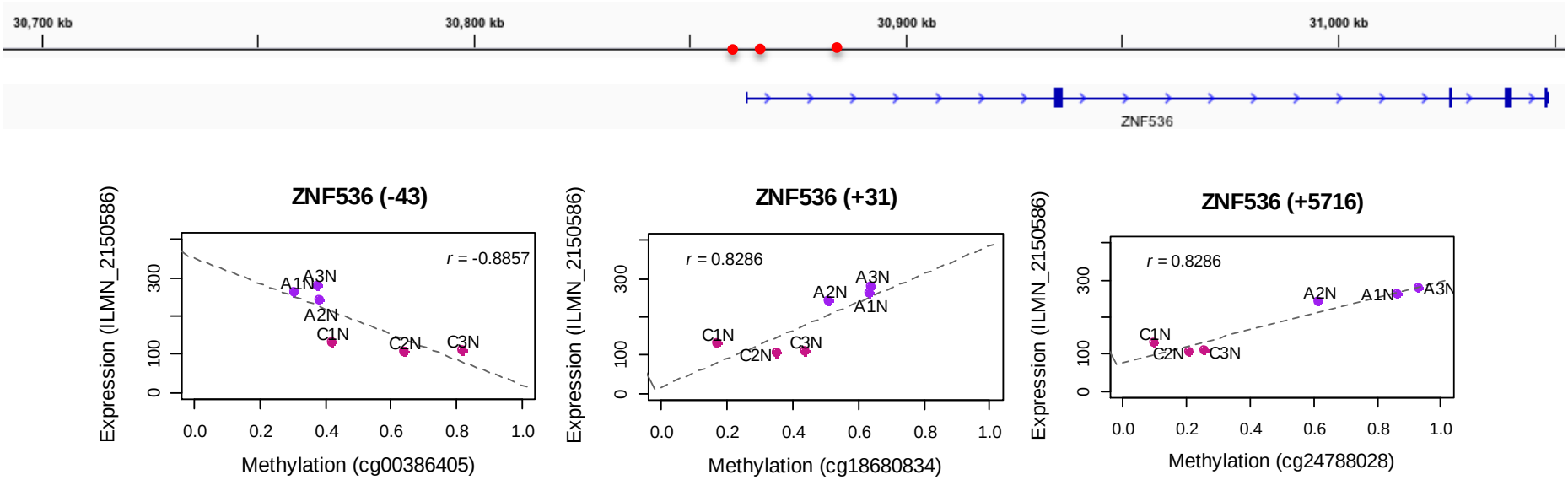


**Figure S5.6 Simultaneous negative correlation between promoter methylation and *ZNF536* gene expression and positive correlation between gene-body methylation and *ZNF536* gene expression levels in patient-matched CAM and ATM samples detected in genome–wide DNA methylation and gene expression studies using Illumina 450k and HT-12 arrays.** Individual sample methylation levels (β-value, *X-axis*) are plotted against individual sample gene expression levels (intensity, *Y-axis*) for 3 representative CpG loci (red dots) associated with *ZNF536* regulation; *magenta – CAMs, purple – ATMs*; numbers in brackets indicate the distance to transcription start site of *ZNF536*; *r - Spearman's rank correlation rho*.

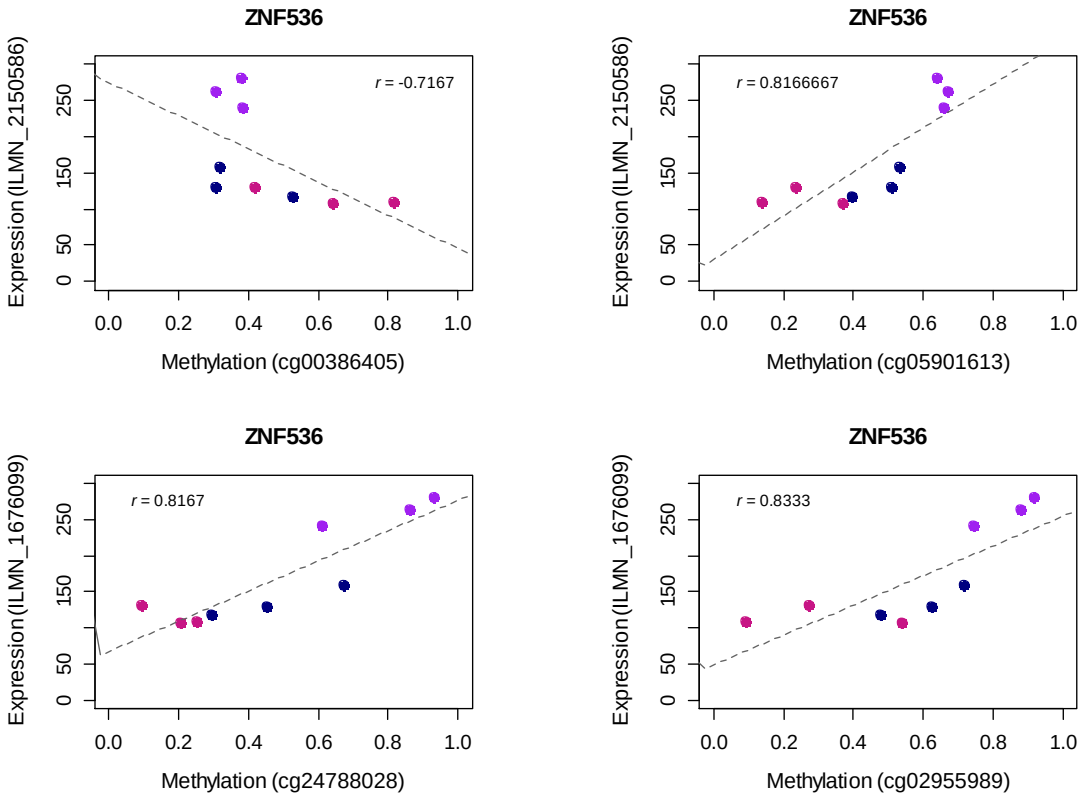


**Figure S5.7 Simultaneous negative correlation between promoter methylation and *ZNF536* gene expression and positive correlation between gene-body methylation and *ZNF536* gene expression levels in patient-matched CAMs and ATMs in comparison to NTMs, as detected in genome–wide DNA methylation and gene expression studies performed using Illumina 450k and HT-12 arrays.** Individual sample methylation levels (β-value, *X-axis*) are plotted against individual sample gene expression levels (intensity, *Y-axis*) for 4 representative CpG loci associated with *ZNF536* regulation; *magenta – CAMs, purple – ATMs, navy - NTMs; r - Spearman's rank correlation rho*.

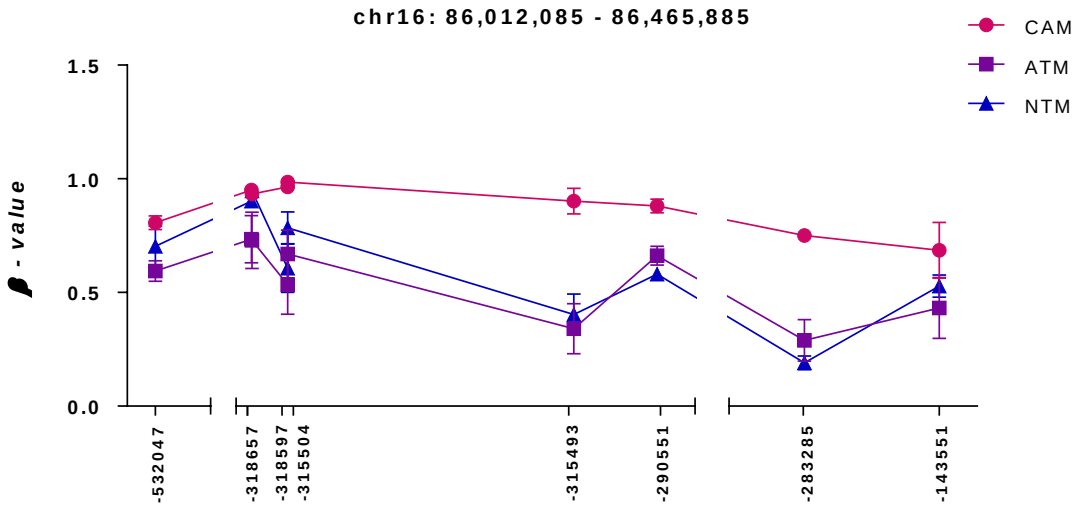


**Figure S5.8 Comparative DNA methylation profiles within a genomic region on chromosome 16 spanning 526,340bp identified by Illumina 450k array in gastric stromal myofibroblasts.** Mean β-values (n=3) for probes identified as differentially methylated in CAM vs ATM and CAM vs NTM comparisons are plotted. The *X*-axis indicates distance of Illumina 450k probes to the *FOXF1* transcription start site. Error bars represent SEM.

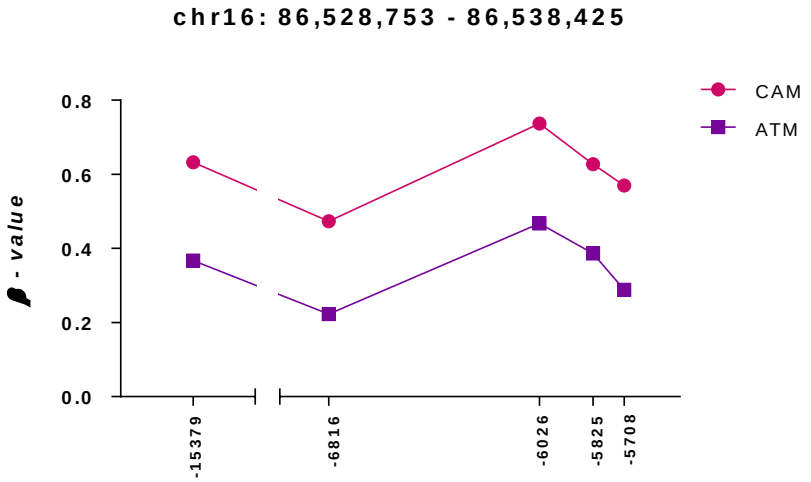


**Figure S5.9 Comparative DNA methylation profiles within a genomic region on chromosome 16 spanning 9,673bp identified by Illumina 450k array in oesophageal patient-matched CAM and ATM samples.** Mean β-values (n=3) for probes identified as differentially methylated in oesophageal CAM vs ATM comparison are plotted. The *X*-axis indicates distance of Illumina 450k probes to the *FOXF1* transcription start site.

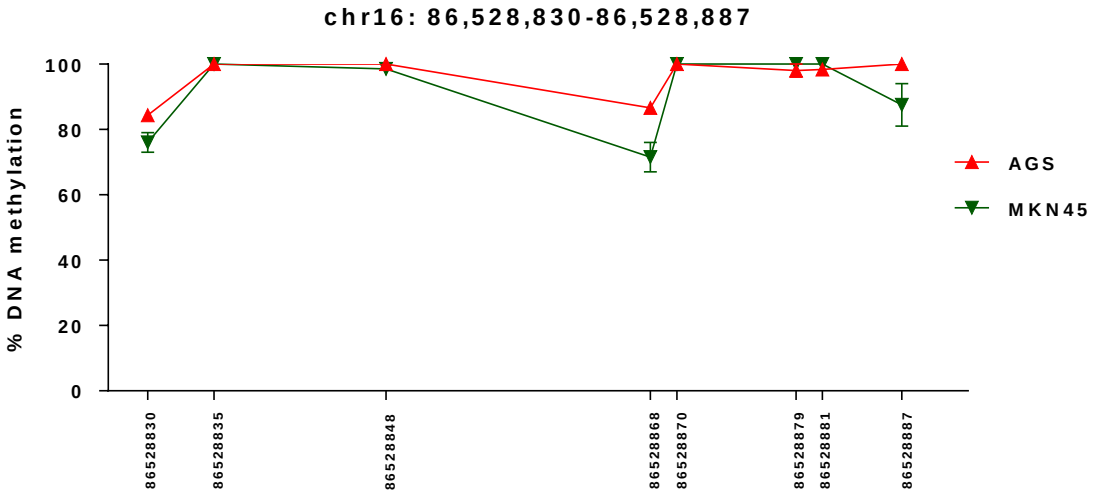


**Figure S5.10 Pyrosequencing analysis of the *FOXF1* promoter region in gastric cancer cell lines.** Methylation means for 8 individual CpG sites in interrogated promoter region are plotted for AGS cells (n=3) and MKN45 cells (n=2). The *X*-axis indicates chromosomal position of examined CpG sites. Error bars represent SEM.

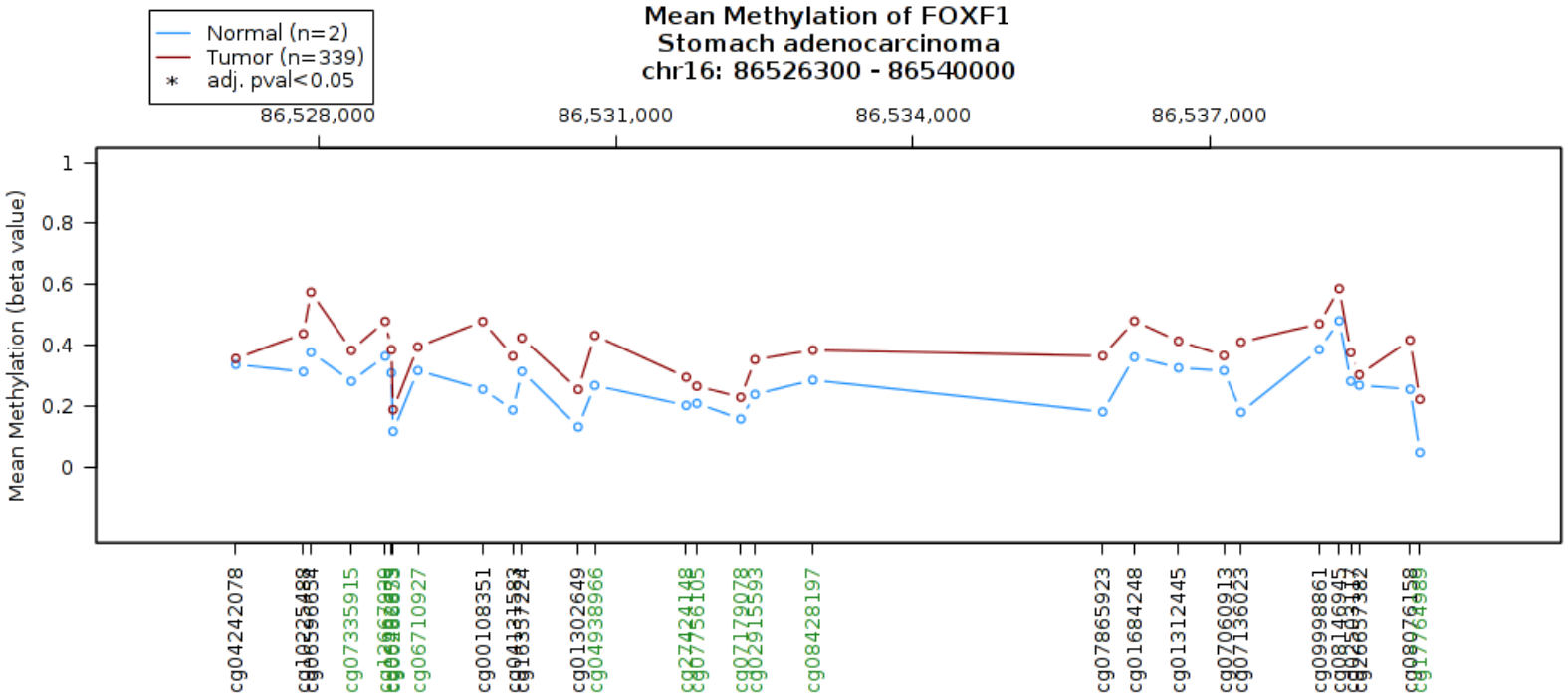


**Figure S5.11 Differential DNA methylation data for stomach adenocarcinoma from the Cancer Genome Atlas (TCGA).** The presented region corresponds to the differentially methylated region identified in gastric CAMs compared to either ATMs or NTMs as shown in Figure 5.6A. Mean β values for stomach tumour (n=339; red line) and normal tissue (n=2; blue line) are plotted. The *X-axis* represents the Illumina 450k probes. TCGA data was obtained using Wanderer (Diez-Villanueva, Mallona et al. 2015).

# *Chapter VI*

## *Effects of Hypoxia on Gene Expression and Secretion of Pro-tumorigenic Factors in Gastric Stromal Myofibroblasts*

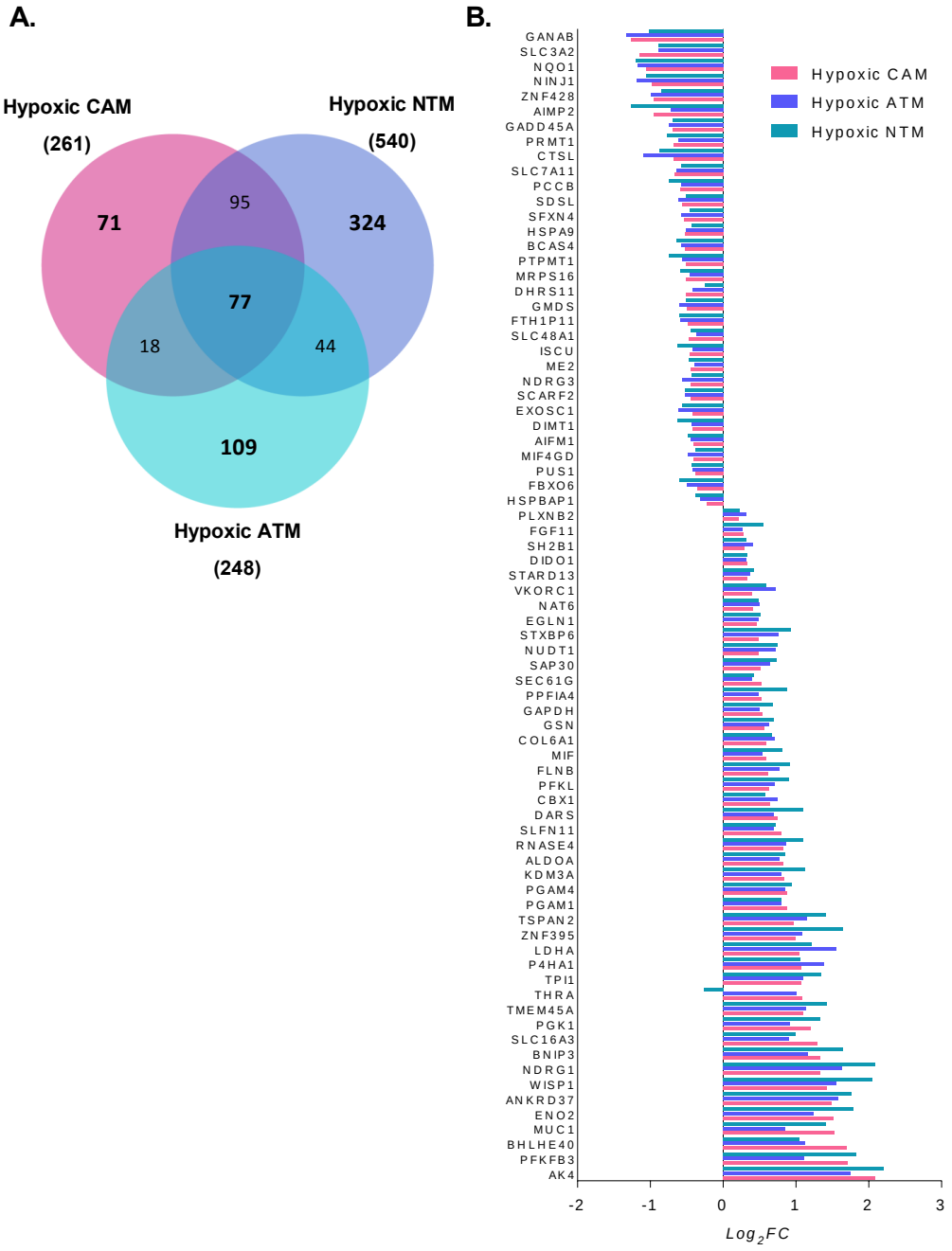


**Figure S6.1 Universal gene expression changes induced under hypoxia in gastric CAMs, ATMs and NTMs. A.** Comparison of hypoxia-induced gene expression signatures identified in gastric CAMs, ATMs and NTMs; *FDR p-value < 0.05*. Overlapping intersections of 77 genes is shown in B.**B.** Genes universally changed in the same direction under hypoxia in gastric CAMs, ATMs and NTMs.

## S6.1 Gene Set Enrichment Analysis (GSEA)

Unique CAM, ATM and NTM hypoxia-induced gene profiles were subjected to gene set enrichment analysis (GSEA) using the complete Molecular Signatures Database (MSigDBv5.0). The analysis identified: (i) total of 114 gene sets (all for hypoxia phenotype) in unique CAM hypoxia-induced gene profile; (ii) total of 102 gene sets (including 48 in hypoxia phenotype and 54 in normoxia phenotype) in unique NTM hypoxia-induced gene profile; and (iii) total of 104 gene sets (all for hypoxia phenotype) in unique ATM hypoxia-induced gene profile with *FDR q-value <0.05*. Overlap of identified CAM, ATM and NTM significantly enriched gene sets is shown in Figure S6.2. Although the gene sets were identified based on unique CAMs, ATMs and NTMs hypoxia-changed genes there are some overlapping gene sets indicating that some of the uniquely changed genes are involved in regulation of the same processes.

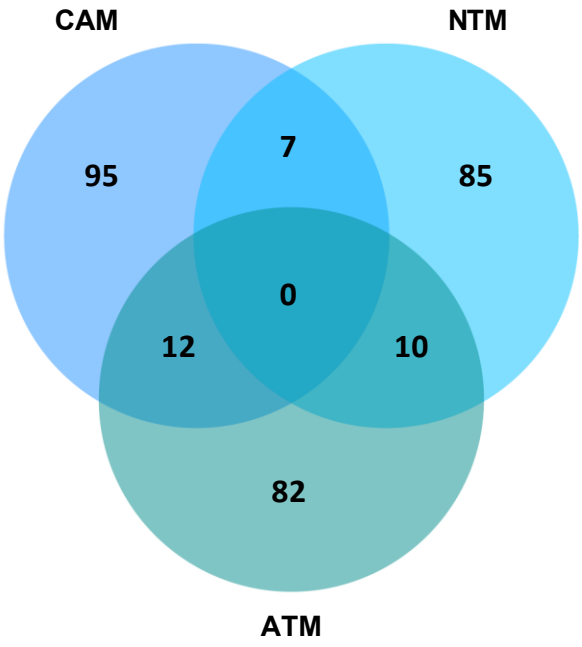


**Figure S6.2 Comparison of enriched MSigDB gene sets identified for unique CAM, ATM and NTM hypoxia-induced gene expression profiles;** *FDR q-value <0.05.*

## S6.2 Leading Edge Analysis

Enriched gene sets with *FDR q-value <0.05* identified in respective unique CAM, ATM or NTM hypoxia-induced gene expression profiles were subjected to leading edge analysis (LEA) in order to examine genes that are in the leading-edge subsets of the enriched gene sets. These genes may have a key role in each of the three analysed phenotypes. A summary of the respective LEA results for CAMs, ATMs and NTMs enriched gene sets is shown in Table S6.1.

**Table S6.1 Summary of leading edge analysis (LEA) of enriched MSigDB gene sets identified for unique CAM, ATM and NTM hypoxia-induced gene expression profiles.** Top 10 genes identified in respective leading edge analyses and number of leading-edge subsets in which the given gene appears is shown; *pink* – genes upregulated in hypoxia; *blue*- genes downregulated in hypoxia.

| **CAM** | | **ATM** | | **NTM** | |
|---|---|---|---|---|---|
| Gene symbol | No of subsets | Gene symbol | No of subsets | Gene symbol | No of subsets |
| **IDI1** | 45 | **AURKA** | 42 | **CYB5A** | 13 |
| **SQLE** | 40 | **BIRC5** | 41 | **SOX4** | 13 |
| **HMGCS1** | 37 | **CCNA2** | 39 | **EGR1** | 13 |
| **MSMO1** | 36 | **TPX2** | 38 | **JUN** | 12 |
| **HMGCR** | 35 | **MAD2L1** | 36 | **NME1** | 12 |
| **SCD** | 32 | **AURKB** | 36 | **ENG** | 11 |
| **FDFT1** | 31 | **CDKN3** | 35 | **APOE** | 11 |
| **LDLR** | 31 | **MELK** | 34 | **KLHL24** | 11 |
| **FDPS** | 31 | **MYL9** | 28 | **MBNL1** | 10 |
| **DHCR7** | 29 | **CDCA3** | 27 | **CREB5** | 9 |

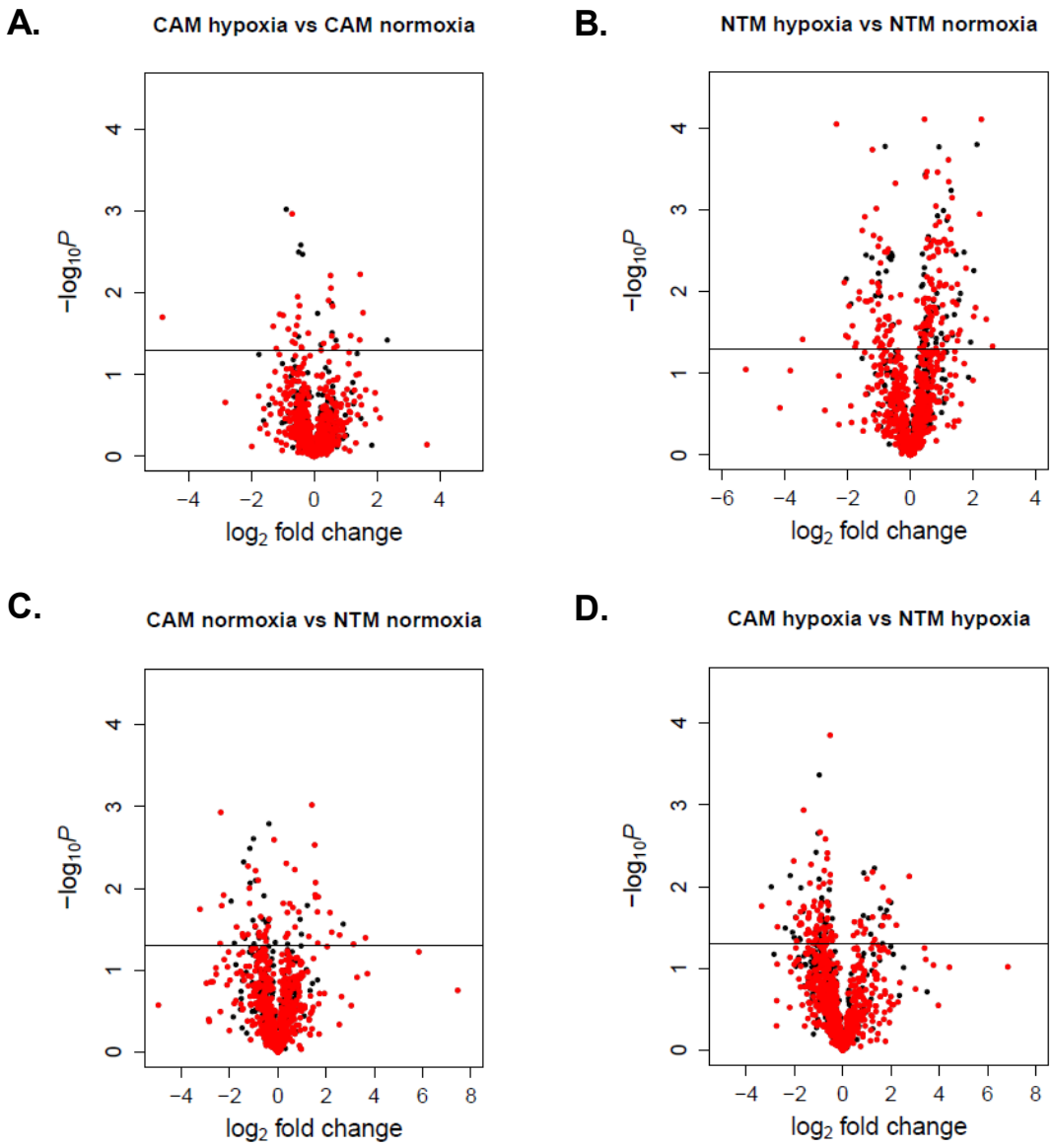


**Figure S6.3 Volcano plots of differentially secreted proteins** in the following comparisons: **A.** CAM-hypoxic-CM vs CAM-ctrl-CM **B.** NTM-hypoxic-CM vs NTM-ctrl-CM **C.** CAM-ctrl-CM vs NTM-ctrl-CM **D.** CAM-hypoxic-CM vs NTM-hypoxic-CM. Proteins highlighted in red were annotated as secreted/extracellular or identified in exosomes based on MetazSecKB, Matrisome and ExoCarta searches (as described in section 6.3.4.1). Proteins above the line are considered to be differentially secreted between the given conditions.

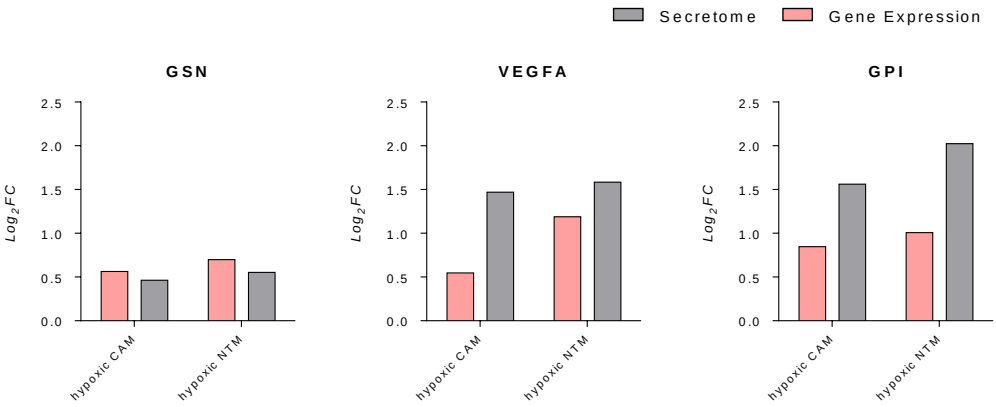


**Figure S6.4 Differential expression and secretion** of gelsolin (GSN), vascular endothelial growth factor A (VEGFA) and glucose-6-phosphate isomerase (GPI) in hypoxic CAMs (n=3) and hypoxic NTMs (n=3) compared to respective control-normoxic myofibroblasts; *p-value <0.05.*

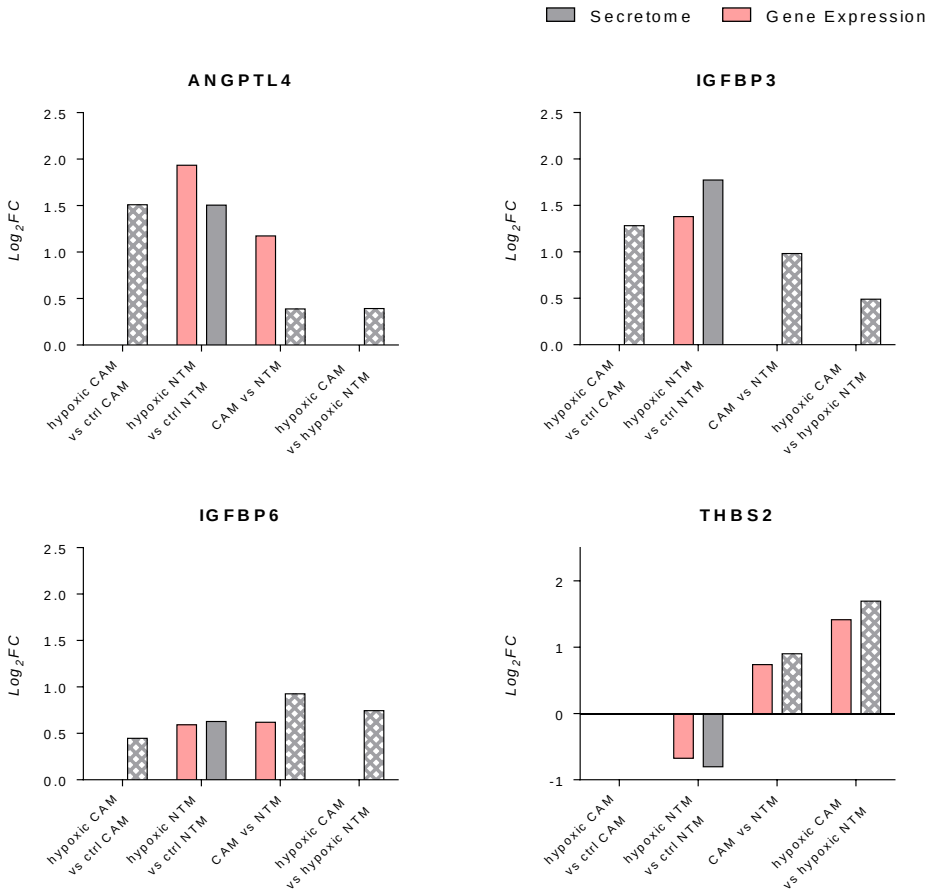


**Figure S6.5 Differential expression and secretion** of angiopoietin-related protein 4 (ANGPTL4), insulin-like growth factor-binding protein 3 (IGFBP3), insulin-like growth factor-binding protein 6 (IGFBP6) and thrombospondin-2 (THBS2) in gastric CAMs (n=3) and NTMs (n=3) under hypoxia and normoxia; *checked* – proteins detected in respective comparisons but not identified as differentially secreted (*p-value>0.05)*.

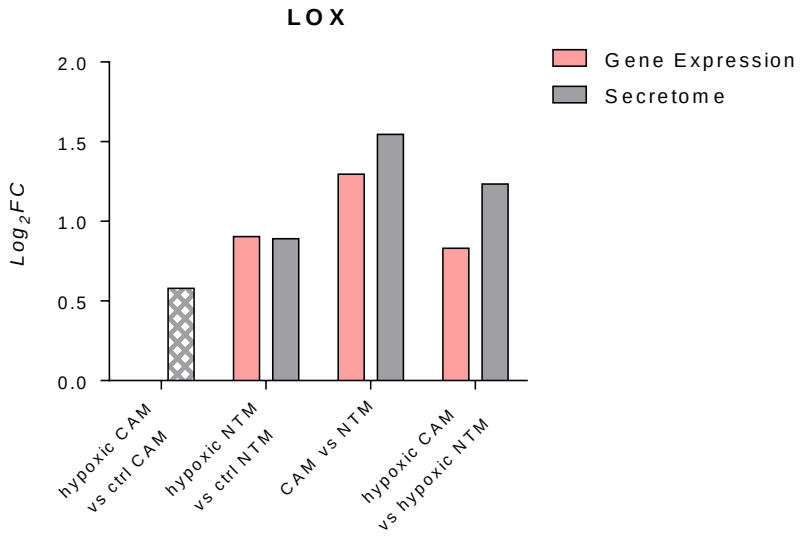


**Figure S6.6 Differential expression and secretion of enzyme lysyl oxidase (LOX) in gastric CAMs (n=3) and NTMs (n=3) under hypoxia and normoxia.** LOX mRNA transcript was not identified as differentially expressed in hypoxic CAMs compared to control-normoxic CAMs whereas LOX protein (*checked*) was detected in CAM-hypoxic-CM and CAM-ctrl-CM however it was not identified as significant in differential secretome analysis (*p-value>0.05)*.

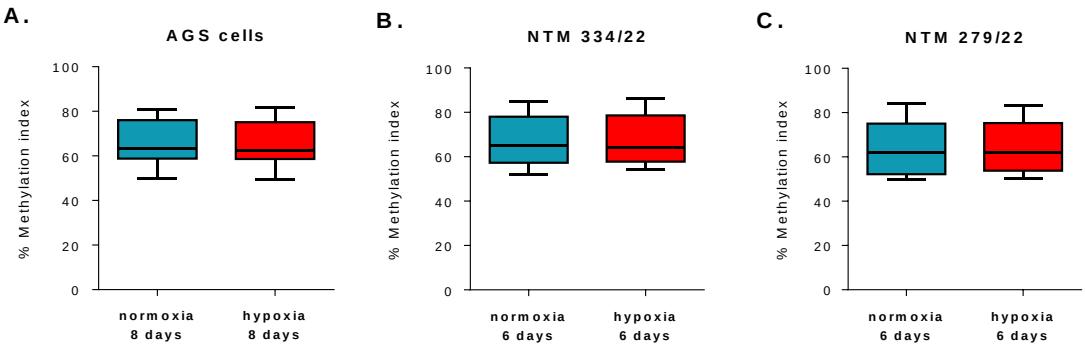


**Figure S6.7 Global DNA methylation after long-term hypoxia treatment assessed by LINE-1 pyrosequencing assays. A.** AGS gastric cancer cell line (n=3) was cultured under hypoxia and normoxia for 8 days **B, C.** Two normal tissue myofibroblast (NTM) cell lines 334/22 (n=3) and 279/22 (n=3) were cultured under hypoxia and normoxia for 6 days. Boxplots represent methylation distribution and mean for 6 CpG sites examined inside the CpG-rich 5' internal LINE-1 promoter.

# References

# References

Abbaszadegan, M. R., Moaven, O., Sima, H. R., Ghafarzadegan, K., A'Rabi, A., Forghani, M. N., Raziee, H. R., Mashhadinejad, A., Jafarzadeh, M., Esmaili-Shandiz, E., Dadkhah, E. (2008). p16 promoter hypermethylation: a useful serum marker for early detection of gastric cancer. *World Journal of Gastroenterology* **14**(13): 2055-2060.

Afanas'ev, I. (2015). Mechanisms of superoxide signaling in epigenetic processes: relation to aging and cancer. *Aging and Disease* **6**(3): 216-227.

Afratis, N., Gialeli, C., Nikitovic, D., Tsegenidis, T., Karousou, E., Theocharis, A. D., Pavao, M. S., Tzanakakis, G. N., Karamanos, N. K. (2012). Glycosaminoglycans: key players in cancer cell biology and treatment. *FEBS Journal* **279**(7): 1177-1197.

Al-Rakan, M. A., Colak, D., Hendrayani, S. F., Al-Bakheet, A., Al-Mohanna, F. H., Kaya, N., Al-Malik, O., Aboussekhra A. (2013). Breast stromal fibroblasts from histologically normal surgical margins are pro-carcinogenic. *Journal of Pathology* **231**(4): 457-465.

Allen, M., Jones, J. L. (2011). Jekyll and Hyde: the role of the microenvironment on the progression of cancer. *Journal of Pathology* **223**(2): 162-176.

American Cancer Society, *www.cancer.org*

Ao, Z., Shah, S. H., Machlin, L. M., Parajuli, R., Miller, P. C., Rawal, S., Williams, A. J., Cote, R. J., Lippman, M. E., Datar, R. H., El-Ashry D. (2015). Identification of cancer-associated fibroblasts in circulating blood from patients with metastatic breast cancer. *Cancer Research* **75**(22): 4681-4687.

Ashcroft, G. S., Yang, X., Glick, A. B., Weinstein, M., Letterio, J. J., Mizel, D. E., Anzano, M., Greenwell-Wild, T., Wahl, S. M., Deng, C. X., Roberts A. B. (1999). Mice lacking Smad3 show accelerated wound healing and an impaired local inflammatory response. *Nature Cell Biology* **1**(5): 260-266.

Assenov, Y., Muller, F., Lutsik, P., Walter, J., Lengauer, T., Bock, C. (2014). Comprehensive analysis of DNA methylation data with RnBeads. *Nature Methods* **11**(11): 1138-1140.

Bae, J. H., Kang, M. J., Yang, K. M., Kim, T. O., Yi, J. M. (2015). Epigenetically silenced microRNAs in gastric cancer: Functional analysis and identification of their target genes. *Oncology Reports* **34**(2): 1017-1026.

Balabanova, S., Holmberg, C., Steele, I., Ebrahimi, B., Rainbow, L., Burdyga, T., McCaig, C., Tiszlavicz, L., Lertkowit, N., Giger, O. T., Oliver, S., Prior, I., Dimaline, R., Simpson, D., Beynon, R., Hegyi, P., Wang, T. C., Dockray, G. J., Varro, A. (2014). The neuroendocrine phenotype of gastric myofibroblasts and its loss with cancer progression. *Carcinogenesis* **35**(8): 1798-1806.

Bauer, M., Su, G., Casper, C., He, R., Rehrauer, W., Friedl, A. (2010). Heterogeneity of gene expression in stromal fibroblasts of human breast carcinomas and normal breast. *Oncogene* **29**(12): 1732-1740.

Bedard, P. L., Hansen, A. R., Ratain, M. J., Siu, L. L. (2013). Tumour heterogeneity in the clinic. *Nature* **501**(7467): 355-364.

Benetti, R., Gonzalo, S., Jaco, I., Munoz, P., Gonzalez, S., Schoeftner, S., Murchison, E., Andl, T., Chen, T., Klatt, P., Li, E., Serrano, M., Millar, S., Hannon, G., Blasco, M. A. (2008). A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. *Nature Structural & Molecular Biology* **15**(3): 268-279.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**(1): 289-300.

Bertino, E. M., Otterson, G. A. (2011). Romidepsin: a novel histone deacetylase inhibitor for cancer. *Expert Opinion on Investigational Drugs* **20**(8): 1151-1158.

Bertout, J. A., Patel, S. A., Simon, M. C. (2008). The impact of $O_2$ availability on human cancer. *Nature Reviews Cancer* **8**(12): 967-975.

Bertuccio, P., Chatenoud, L., Levi, F., Praud, D., Ferlay, J., Negri, E., Malvezzi, M., La Vecchia, C. (2009). Recent patterns in gastric cancer: A global overview." *International Journal of Cancer* **125**(3): 666-673.

Bhowmick, N. A., Chytil, A., Plieth, D., Gorska, A. E., Dumont, N., Shappell, S., Washington, M. K., Neilson, E. G., Moses, H. L. (2004). TGF-beta signaling in fibroblasts modulates the oncogenic potential of adjacent epithelia. *Science* **303**(5659): 848-851.

Bhowmick, N. A., Neilson, E. G., Moses, H. L. (2004). Stromal fibroblasts in cancer initiation and progression. *Nature* **432**(7015): 332-337.

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., Delano, D., Zhang, L., Schroth, G. P., Gunderson, K. L., Fan, J. B., Shen, R. (2011). High density DNA methylation array with single CpG site resolution. *Genomics* **98**(4): 288-295.

Bird, A. (2007). Perceptions of epigenetics. *Nature* **447**(7143): 396-398.

Bird, A. (1986). CpG-rich islands and the function of DNA methylation. *Nature* **321**(6067): 209-213.

Bock, C. (2012). Analysing and interpreting DNA methylation data. *Nature Reviews Genetics* **13**(10): 705-719.

Brenner, H., Rothenbacher, D., Arndt, V. (2009). Epidemiology of Stomach Cancer. *Methods in Molecular Biology*. M. Verma. **472**: 467-477.

Brettingham-Moore, K. H., Duong, C. P., Heriot, A. G., Thomas, R. J. S., Phillips, W. A. (2011). Using gene expression profiling to predict response and prognosis in gastrointestinal cancers-the promise and the perils. *Annals of Surgical Oncology* **18**(5): 1484-1491.

Calon, A., Tauriello, D. V. F., Batlle, E. (2014). TGF-beta in CAF-mediated tumor growth and metastasis. *Seminars in Cancer Biology* **25**: 15-22.

Campbell, I., Polyak, K., Haviv, I. (2009). Clonal mutations in the cancer-associated fibroblasts: the case against genetic coevolution. *Cancer Research* **69**(17): 6765-6768.

Cancer Research UK, *www.cancerresearchuk.org*

Carmeliet, P., Dor, Y., Herbert, J. M., Fukumura, D., Brusselmans, K., Dewerchin, M., Neeman, M., Bono, F., Abramovitch, R., Maxwell, P., Koch, C. J., Ratcliffe, P., Moons, L., Jain, R. K., Collen, D., Keshet, E. (1998). Role of HIF-1 alpha or in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis. *Nature* **394**(6692): 485-490.

Casazza, A., Di Conza, G., Wenes, M., Finisguerra, V., Deschoemaeker, S., Mazzone, M. (2014). Tumor stroma: a complexity dictated by the hypoxic tumor microenvironment. *Oncogene* **33**(14): 1743-1754.

Chaffer, C. L., Weinberg, R. A. (2011). A perspective on cancer cell metastasis. *Science* **331**(6024): 1559-1564.

Chan, A. O. O., Broaddus, R. R., Houlihan, P. S., Issa, J. P. J., Hamilton, S. R., Rashid, A. (2002). CpG island methylation in aberrant crypt foci of the colorectum. *American Journal of Pathology* **160**(5): 1823-1830.

Cedar, H., Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. *Nature Reviews Genetics* **10**(5): 295-304.

Chen, D., Fu, L. Y., Zhang, Z., Li, G., Zhang, H., Jiang, L., Harrison, A. P., Shanahan, H. P., Klukas, C., Zhang, H. Y., Ruan, Y., Chen, L. L., Chen, M. (2014). Dissecting the chromatin interactome of microRNA genes. *Nucleic Acids Research* **42**(5): 3028-3043.

Chen, Y. A., Lemire, M., Choufani, S., Butcher, D. T., Grafodatskaya, D., Zanke, B. W., Gallinger, S., Hudson, T. J., Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **8**(2): 203-209.

Cho, J. Y., Lim, J. Y., Cheong, J. H., Park, Y. Y., Yoon, S. L., Kim, S. M., Kim, S. B., Kim, H., Hong, S. W., Park, Y. N., Noh, S. H., Park, E. S., Chu, I. S., Hong, W. K., Ajani, J. A., Lee, J. S. (2011). Gene expression signature-based prognostic risk score in gastric cancer. *Clinical Cancer Research* **17**(7): 1850-1857.

Claus, R., Lucas, D. M., Stilgenbauer, S., Ruppert, A. S., Yu, L., Zucknick, M., Mertens, D., Bühler, A., Oakes, C. C., Larson, R. A. (2012). Quantitative DNA methylation

analysis identifies a single CpG dinucleotide important for ZAP-70 expression and predictive of prognosis in chronic lymphocytic leukemia. *Journal of Clinical Oncology* **30**(20): 2483-2491.

Comito, G., Giannoni, E., Di Gennaro, P., Segura, C. P., Gerlini, G., Chiarugi, P. (2012). Stromal fibroblasts synergize with hypoxic oxidative stress to enhance melanoma aggressiveness. *Cancer Letters* **324**(1): 31-41.

Correa, P., Haenszel, W., Cuello, C., Tannenbaum, S., Archer, M. (1975). A model for gastric cancer epidemiology. *Lancet* **2**(7924): 58-60.

Costea, D. E., Hills, A., Osman, A. H., Thurlow, J., Kalna, G., Huang, X., Murillo, C. P., Parajuli, H., Suliman, S., Kulasekara, K. K., Johannessen, A. C., Partridge M. (2013). Identification of two distinct carcinoma-associated fibroblast subtypes with differential tumor-promoting abilities in oral squamous cell carcinoma. *Cancer Research* **73**(13): 3888-3901.

Cox, J., Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **26**(12): 1367-1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., Mann, M. (2011). Andromeda: a peptide search engine integrated into the maxquant environment. *Journal of Proteome Research* **10**(4): 1794-1805.

Cox, T. R., Erler, J. T. (2011). Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. *Disease Models & Mechanisms* **4**(2): 165-178.

Crawford, Y., Kasman, I., Yu, L., Zhong, C., Wu, X., Modrusan, Z., Kaminker, J., Ferrara, N. (2009). PDGF-C mediates the angiogenic and tumorigenic properties of fibroblasts associated with tumors refractory to anti-VEGF treatment. *Cancer Cell* **15**(1): 21-34.

Crew, K. D., Neugut, A. I. (2006). Epidemiology of gastric cancer. *World Journal of Gastroenterology* **12**(3): 354-362.

De Vlieghere, E., Verset, L., Demetter, P., Bracke, M., De Wever, O. (2015). Cancer-associated fibroblasts as target and tool in cancer therapeutics and diagnostics. *Virchows Archiv* **467**(4): 367-382.

Dedeurwaerder, S., Desmedt, C., Calonne, E., Singhal, S. K., Haibe-Kains, B., Defrance, M., Michiels, S., Volkmar, M., Deplus, R., Luciani, J., Lallemand, F., Larsimont, D., Toussaint, J., Haussy, S., Rothe, F., Rouas, G., Metzger, O., Majjaj, S., Saini, K., Putmans, P., Hames, G., van Baren, N., Coulie, P. G., Piccart, M., Sotiriou, C., Fuks, F. (2011). DNA methylation profiling reveals a predominant immune component in breast cancers. *EMBO Molecular Medicine* **3**(12): 726-741.

Desmouliere, A., Redard, M., Darby, I., Gabbiani, G. (1995). Apoptosis mediates the decrease in cellularity during the transition between granulation tissue and scar. *American Journal of Pathology* **146**(1): 56-66.

Diez-Villanueva, A., Mallona, I., Peinado, M. A. (2015). Wanderer, an interactive viewer to explore DNA methylation and gene expression data in human cancer. *Epigenetics & Chromatin* **8**: 22.

Do, T. V., Xiao, F., Bickel, L. E., Klein-Szanto, A. J., Pathak, H. B., Hua, X., Howe, C., O'Brien, S. W., Maglaty, M., Ecsedy, J. A., Litwin, S., Golemis, E. A., Schilder, R. J., Godwin, A. K., Connolly, D. C. (2014). Aurora kinase A mediates epithelial ovarian cancer cell migration and adhesion. *Oncogene* **33**(5): 539-549.

Doi, A., Park, I. H., Wen, B., Murakami, P., Aryee, M. J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S., Miller, J., Schlaeger, T., Daley, G. Q., Feinberg, A. P. (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nature Genetics* **41**(12): 1350-U1123.

Du, P., Kibbe, W. A., Lin, S. M. (2008). lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**(13): 1547-1548.

Duraes, C., Almeida, G. M., Seruca, R., Oliveira, C., Carneiro, F. (2014). Biomarkers for gastric cancer: prognostic, predictive or targets of therapy? *Virchows Archiv* **464**(3): 367-378.

Eden, E., Lipson, D., Yogev, S., Yakhini, Z. (2007). Discovering motifs in ranked lists of DNA sequences. *PLoS Computational Biology* **3**(3): e39.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**: 48.

Egger, G., Liang, G. N., Aparicio, A., Jones, P. A. (2004). Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **429**(6990): 457-463.

Ehrlich, M., Wang, R. Y. H. (1981). 5-Methylcytosine in eukaryotic DNA. *Science* **212**(4501): 1350-1357.

Emoto, S., Ishigami, H., Yamashita, H., Yamaguchi, H., Kaisaki, S., Kitayama, J. (2012). Clinical significance of CA125 and CA72-4 in gastric cancer with peritoneal dissemination. *Gastric Cancer* **15**(2): 154-161.

Enkelmann, A., Heinzelmann, J., von Eggeling, F., Walter, M., Berndt, A., Wunderlich, H., Junker, K. (2011). Specific protein and miRNA patterns characterise tumour-associated fibroblasts in bladder cancer. *Journal of Cancer Research and Clinical Oncology* **137**(5): 751-759.

Erler, J. T., Bennewith, K. L., Cox, T. R., Lang, G., Bird, D., Koong, A. Le, Q. T., Giaccia, A. J. (2009). Hypoxia-induced lysyl oxidase is a critical mediator of bone marrow cell recruitment to form the premetastatic niche. *Cancer Cell* **15**(1): 35-44.

Esteller, M. (2007). Cancer epigenomics: DNA methylomes and histone-modification maps. *Nature Reviews Genetics* **8**(4): 286-298.

Esteller, M. (2008). Molecular origins of cancer: Epigenetics in cancer. *New England Journal of Medicine* **358**(11): 1148-1159.

Eyden, B. (2008). The myofibroblast: phenotypic characterization as a prerequisite to understanding its functions in translational medicine. *Journal of Cellular and Molecular Medicine* **12**(1): 22-37.

Fang, X. Y., Pan, H. F., Leng, R. X., Ye, D. Q. (2015). Long noncoding RNAs: novel insights into gastric cancer. *Cancer Letters* **356**(2): 357-366.

Farmer, P., Bonnefoi, H., Anderle, P., Cameron, D., Wirapati, P., Becette, V., Andre, S., Piccart, M., Campone, M., Brain, E., MacGrogan, G., Petit, T., Jassem, J., Bibeau, F., Blot, E., Bogaerts, J., Aguet, M., Bergh, J., Iggo, R., Delorenzi, M. (2009). A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. *Nature Medicine* **15**(1): 68-74.

Feinberg, A. P. (2007). Phenotypic plasticity and the epigenetics of human disease. *Nature* **447**(7143): 433-440.

Feinberg, A. P., Tycko, B. (2004). Timeline - The history of cancer epigenetics. *Nature Reviews Cancer* **4**(2): 143-153.

Feinberg, A. P., Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301**(5895): 89-92.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* **136**(5): E359-E386.

Fernandez, C., Lobo, M., Gomez-Coronado, D., Lasuncion, M. A. (2004). Cholesterol is essential for mitosis progression and its deficiency induces polyploid cell formation. *Experimental Cell Research* **300**(1): 109-120.

Fernandez, C., Martin, M., Gomez-Coronado, D., Lasuncion, M. A. (2005). Effects of distal cholesterol biosynthesis inhibitors on cell proliferation and cell cycle progression. *Journal of Lipid Research* **46**(5): 920-929.

Fiegl, H., Millinger, S., Goebel, G., Muller-Holzner, E., Marth, C., Laird, P. W., Widschwendter, M. (2006). Breast cancer DNA methylation profiles in cancer cells and tumor stroma: association with HER-2/neu status in primary breast cancer. *Cancer Research* **66**(1): 29-33.

Finak, G., Bertos, N., Pepin, F., Sadekova, S., Souleimanova, M., Zhao, H., Chen, H., Omeroglu, G., Meterissian, S., Omeroglu, A., Hallett, M., Park, M. (2008). Stromal gene expression predicts clinical outcome in breast cancer. *Nature Medicine* **14**(5): 518-527.

Gabbiani, G., Ryan, G. B., Majno, G. (1971). Presence of modified fibroblasts in granulation tissue and their possible role in wound contraction. *Experientia* **27**(5): 549-550.

Gal-Yam, E. N., Saito, Y., Egger, G., Jones, P. A. (2008). Cancer epigenetics: modifications, screening, and therapy. *Annual Review of Medicine* **59:** 267-280.

Geeleher, P., Hartnett, L., Egan, L. J., Golden, A., Raja Ali, R. A., Seoighe, C. (2013). Gene-set analysis is severely biased when applied to genome-wide methylation data. *Bioinformatics* **29**(15): 1851-1857.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y., Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**(10): R80.

Giaccia, A. J., Schipani, E. (2010). Role of carcinoma-associated fibroblasts and hypoxia in tumor progression. *Diverse Effects of Hypoxia on Tumor Progression*. M. C. Simon. **345:** 31-45.

Giannoni, E., Bianchini, F., Masieri, L., Serni, S., Torre, E., Calorini, L., Chiarugi, P. (2010). Reciprocal activation of prostate cancer cells and cancer-associated fibroblasts stimulates epithelial-mesenchymal transition and cancer stemness." *Cancer Research* **70**(17): 6945-6956.

Gigek, C. O., Chen, E. S., Calcagno, D. Q., Wisnieski, F., Burbano, R. R., Cardoso Smith, M. A. (2012). Epigenetic mechanisms in gastric cancer. *Epigenomics* **4**(3): 279-294.

Gilkes, D. M., Semenza, G. L., Wirtz, D. (2014). Hypoxia and the extracellular matrix: drivers of tumour metastasis. *Nature Reviews Cancer* **14**(6): 430-439.

Gonda, T. A., Kim, Y. I., Salas, M. C., Gamble, M. V., Shibata, W., Muthupalani, S., Sohn, K. J., Abrams, J. A., Fox, J. G., Wang, T. C., Tycko, B. (2012). Folic acid increases global DNA methylation and reduces inflammation to prevent *Helicobacter*-associated gastric cancer in mice. *Gastroenterology* **142**(4): 824-833e7.

Gonda, T. A., Varro, A., Wang, T. C., Tycko, B. (2010). Molecular biology of cancer-associated fibroblasts: Can these cells be targeted in anti-cancer therapy?" *Seminars in Cell & Developmental Biology* **21**(1): 2-10.

Grady, W. M., Markowitz, S. D. (2002). Genetic and epigenetic alterations in colon cancer. *Annual Review of Genomics and Human Genetics* **3**: 101-128.

Grant, S., Easley, C., Kirkpatrick, P. (2007). Vorinostat. *Nature Reviews Drug Discovery* **6**(1): 21-22.

Griffiths, E. A., Pritchard, S. A., Valentine, H. R., Whitchelo, N., Bishop, P. W., Ebert, M. P., Price, P. M., Welch, I. M., West, C. M. L. (2007). Hypoxia-inducible factor-1 alpha expression in the gastric carcinogenesis sequence and its prognostic role in gastric and gastro-oesophageal adenocarcinomas. *British Journal of Cancer* **96**(1): 95-103.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Waehrisch, S., Beisaw, A., Macura, K., Blaess, G., Kellis, M., Werber, M., Herrmann, B. G. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse." *Developmental Cell* **24**(2): 206-214.

Gu, J., Qian, H., Shen, L., Zhang, X., Zhu, W., Huang, L., Yan, Y., Mao, F., Zhao, C., Shi, Y., Xu, W. (2012). Gastric cancer exosomes trigger differentiation of umbilical cord derived mesenchymal stem cells to carcinoma-associated fibroblasts through TGF-beta/Smad pathway. *Plos One* **7**(12): e52465.

Gu, Y., Chen, T., Li, G., Yu, X., Lu, Y., Wang, H., Teng, L. (2015). LncRNAs: emerging biomarkers in gastric cancer. *Future Oncology* **11**(17): 2427-2441.

Guo, J., Wang, B., Fu, Z., Wei, J., Lu, W. (2015). Hypoxic microenvironment induces EMT and upgrades stem-like properties of gastric cancer cells." *Technology in Cancer Research & Treatment*: 1-9.

Ha, S. Y., Yeo, S. Y., Xuan, Y. H., Kim, S. H. (2014). The prognostic significance of cancer-associated fibroblasts in esophageal squamous cell carcinoma. *Plos One* **9**(6):e99955.

Hager, M. H., Solomon, K. R., Freeman, M. R. (2006). The role of cholesterol in prostate cancer. *Current Opinion in Clinical Nutrition and Metabolic Care* **9**(4): 379-385.

Hale, M. D., Hayden, J. D., Grabsch, H. I. (2013). Tumour-microenvironment interactions: role of tumour stroma and proteins produced by cancer-associated fibroblasts in chemotherapy response. *Cellular Oncology* **36**(2): 95-112.

Han, L., Witmer, P. D., Casey, E., Valle, D., Sukumar, S. (2007). DNA methylation regulates microRNA expression. *Cancer Biology & Therapy* **6**(8): 1284-1288.

Hanahan, D., Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* **144**(5): 646-674.

Hanson, J. A., Gillespie, J. W., Grover, A., Tangrea, M. A., Chuaqui, R. F., Emmert-Buck, M. R., Tangrea, J. A., Libutti, S. K., Linehan, W. M., Woodson, K. G. (2006). Gene promoter methylation in prostate tumor-associated stromal cells. *Journal of the National Cancer Institute* **98**(4): 255-261.

Harper, K. N., Peters, B. A., Gamble, M. V. (2013). Batch effects and pathway analysis: two potential perils in cancer studies involving DNA methylation array analysis. *Cancer Epidemiology, Biomarkers & Prevention* **22**(6): 1052-1060.

Harris, A. L. (2002). Hypoxia - A key regulatory factor in tumour growth. *Nature Reviews Cancer* **2**(1): 38-47.

Harris, L. G., Samant, R. S., Shevde, L. A. (2011). Hedgehog signaling: networking to nurture a promalignant tumor microenvironment. *Molecular Cancer Research* **9**(9): 1165-1174.

Hecker, L., Jagirdar, R., Jin, T., Thannickal, V. J. (2011). Reversible differentiation of myofibroblasts by MyoD. *Experimental Cell Research* **317**(13): 1914-1921.

Heldin, C. H., Moustakas, A. (2012). Role of Smads in TGF beta signaling. *Cell and Tissue Research* **347**(1): 21-36.

Heller, G., Schmidt, W. M., Ziegler, B., Holzer, S., Muellauer, L., Bilban, M., Zielinski, C. C., Drach, J., Zoechbauer-Mueller, S. (2008). Genome-wide transcriptional response to 5-aza-2 '-deoxycytidine and trichostatin a in multiple myeloma cells. *Cancer Research* **68**(1): 44-54.

Hellman, A., Chess, A. (2007). Gene body-specific methylation on the active X chromosome. *Science* **315**(5815): 1141-1143.

Herman, J. G., Baylin, S. B. (2003). Mechanisms of disease: gene silencing in cancer in association with promoter hypermethylation. *New England Journal of Medicine* **349**(21): 2042-2054.

Herrera, M., Islam, A. B. M. M. K., Herrera, A., Martin, P., Garcia, V., Silva, J., Garcia, J. M., Salas, C., Casal, I., Garcia de Herreros, A., Bonilla, F., Pena, C. (2013). Functional heterogeneity of cancer-associated fibroblasts from human colon tumors shows specific prognostic gene expression signature. *Clinical Cancer Research* **19**(21): 5914-5926.

Hetz, C., Chevet, E., Harding, H. P. (2013). Targeting the unfolded protein response in disease. *Nature Reviews Drug Discovery* **12**(9): 703-719.

Hinz, B. (2007). Formation and function of the myofibroblast during tissue repair. *Journal of Investigative Dermatology* **127**(3): 526-537.

Holmberg, C., Quante, M., Steele, I., Kumar, J. D., Balabanova, S., Duval, C., Czepan, M., Rakonczay, Z., Jr., Tiszlavicz, L., Nemeth, I., Lazar, G., Simonka, Z., Jenkins, R., Hegyi, P., Wang, T. C., Dockray, G. J., Varro, A. (2012). Release of TGFbetaig-h3 by gastric myofibroblasts slows tumor growth and is decreased with cancer progression. *Carcinogenesis* **33**(8): 1553-1562.

Hu, C., Wang, Z., Zhai, L., Yang, M., Shan, L., Chai, C., Liu, M., Wang, L. (2013). Effects of cancer-associated fibroblasts on the migration and invasion abilities of SGC-7901 gastric cancer cells. *Oncology Letters* **5**(2): 609-612.

Hu, M., Yao, J., Cai, L., Bachman, K. E., van den Brule, F., Velculescu, V., Polyak, K. (2005). Distinct epigenetic changes in the stromal cells of breast cancers. *Nature Genetics* **37**(8): 899-905.

Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oleś, A. K., Pagès, H., Reyes, A., Shannon, P., Smyth, G. K., Tenenbaum, D., Waldron, L., Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods* **12**(2): 115-121.

Hurwitz, H., Fehrenbacher, L., Novotny, W., Cartwright, T., Hainsworth, J., Heim, W., Berlin, J., Baron, A., Griffing, S., Holmgren, E., Ferrara, N., Fyfe, G., Rogers, B., Ross, R., Kabbinavar, F. (2004). Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *New England Journal of Medicine* **350**(23): 2335-2342.

Ikushima, H., Miyazono, K. (2010). TGF beta signalling: a complex web in cancer progression. *Nature Reviews Cancer* **10**(6): 415-424.

Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J. B., Sabunciyan, S., Feinberg, A. P. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics* **41**(2): 178-186.

Janowska-Wieczorek, A., Wysoczynski, M., Kijowski, J., Marquez-Curtis, L., Machalinski, B., Ratajczak, J., Ratajczak, M. Z. (2005). Microvesicles derived from activated platelets induce metastasis and angiogenesis in lung cancer. *International Journal of Cancer* **113**(5): 752-760.

Jeschke, J., Collignon, E., Fuks, F. (2015). DNA methylome profiling beyond promoters - taking an epigenetic snapshot of the breast tumor microenvironment. *FEBS Journal* **282**(9): 1801-1814.

Jiang, L., Gonda, T. A., Gamble, M. V., Salas, M., Seshan, V., Tu, S., Twaddell, W. S., Hegyi, P., Lazar, G., Steele, I., Varro, A., Wang, T. C., Tycko, B. (2008). Global hypomethylation of genomic DNA in cancer-associated myofibroblasts. *Cancer Research* **68**(23): 9900-9908.

Johnson, A. B., Denko, N., Barton, M. C. (2008). Hypoxia induces a novel signature of chromatin modifications and global repression of transcription. *Mutation Research* **640**(1-2): 174-179.

Jones, P. A. (1999). The DNA methylation paradox. *Trends in Genetics* **15**(1): 34-37.

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics* **13**(7): 484-492.

Jones, P. A., Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. *Nature Reviews Genetics* **3**(6): 415-428.

Jones, P. A., Baylin, S. B. (2007). The epigenomics of cancer. *Cell* **128**(4): 683-692.

Kaelin, W. G., Jr., Ratcliffe, P. J. (2008). Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. *Molecular Cell* **30**(4): 393-402.

Kagan, H. M., Li, W. D. (2003). Lysyl oxidase: properties, specificity, and biological roles inside and outside of the cell. *Journal of Cellular Biochemistry* **88**(4): 660-672.

Kamburov, A., Pentchev, K., Galicka, H., Wierling, C., Lehrach, H., Herwig, R. (2011). ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Research* **39**(Database issue): D712-D717.

Kaminskas, E., Farrell, A. T., Wang, Y. C., Sridhara, R., Pazdur, R. (2005). FDA drug approval summary: Azacitidine (5-azacytidine, Vidaza((TM))) for injectable suspension. *Oncologist* **10**(3): 176-182.

Kanda, M., Fujii, T., Takami, H., Suenaga, M., Inokawa, Y., Yamada, S., Nakayama, G., Sugimoto, H., Koike, M., Nomoto, S., Kodera, Y. (2014). The combination of the serum carbohydrate antigen 19-9 and carcinoembryonic antigen is a simple and accurate predictor of mortality in pancreatic cancer patients. *Surgery Today* **44**(9): 1692-1701.

Kanda, M., Kodera, Y. (2015). Recent advances in the molecular diagnostics of gastric cancer. *World Journal of Gastroenterology* **21**(34): 9838-9852.

Kang, C., Song, J. J., Lee, J., Kim, M. Y. (2014). Epigenetics: an emerging player in gastric cancer. *World Journal of Gastroenterology* **20**(21): 6433-6447.

Kanzleiter, T., Jaehnert, M., Schulze, G., Selbig, J., Hallahan, N., Schwenk, R. W., Schuermann, A. (2015). Exercise training alters DNA methylation patterns in genes related to muscle growth and differentiation in mice. *American Journal of Physiology-Endocrinology and Metabolism* **308**(10): E912-E920.

Kharaishvili, G., Simkova, D., Bouchalova, K., Gachechiladze, M., Narsia, N., Bouchal, J. (2014). The role of cancer-associated fibroblasts, solid stress and other microenvironmental factors in tumor progression and therapy resistance. *Cancer Cell International* **14:** 41.

Khatri, P., Drăghici, S. (2005). Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics* **21**(18): 3587-3595.

Kidd, S., Spaeth, E., Watson, K., Burks, J., Lu, H., Klopp, A., Andreeff, M., Marini, F. C. (2012). Origins of the tumor microenvironment: quantitative assessment of adipose-derived and bone marrow-derived stroma. *Plos One* **7**(2): e30563.

Kim, Y. I., Fawaz, K., Knox, T., Lee, Y. M., Norton, R., Arora, S., Paiva, L., Mason, J. B. (1998). Colonic mucosal concentrations of folate correlate well with blood measurements of folate status in persons with colorectal polyps. *American Journal of Clinical Nutrition* **68**(4): 866-872.

King, H. W., Michael, M. Z., Gleadle, J. M. (2012). Hypoxic enhancement of exosome release by breast cancer cells. *BMC Cancer* **12**: 421.

Klein, A., Fluegel, D., Kietzmann, T. (2008). Transcriptional regulation of serine/threonine kinase-15 (STK15) expression by hypoxia and HIF-1. *Molecular Biology of the Cell* **19**(9): 3667-3675.

Klopp, A. H., Zhang, Y., Solley, T., Amaya-Manzanares, F., Marini, F., Andreeff, M., Debeb, B., Woodward, W., Schmandt, R., Broaddus, R., Lu, K., Kolonin, M. G. (2012).

Omental adipose tissue-derived stromal cells promote vascularization and growth of endometrial tumors. *Clinical Cancer Research* **18**(3): 771-782.

Klose, R. J., Bird, A. P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences* **31**(2): 89-97.

Kornberg, T. B. (2011). Barcoding hedgehog for intracellular transport. *Science Signaling* **4**(200): pe44.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* **128**(4): 693-705.

Kulshreshtha, R., Ferracin, M., Wojcik, S. E., Garzon, R., Alder, H., Agosto-Perez, F. J., Davuluri, R., Liu, C. G., Croce, C. M., Negrini, M., Calin, G. A., Ivan, M. (2007). A microRNA signature of hypoxia. *Molecular and Cellular Biology* **27**(5): 1859-1867.

Labiche, A., Heutte, N., Herlin, P., Chasle, J., Gauduchon, P., Elie, N. (2010). Stromal compartment as a survival prognostic factor in advanced ovarian carcinoma. *International Journal of Gynecological Cancer* **20**(1): 28-33.

Lauren, P. (1965). The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathologica Et Microbiologica Scandinavica* **64**(1): 31-49.

Lee, D. F., Su, J., Ang, Y. S., Carvajal-Vergara, X., Mulero-Navarro, S., Pereira, C. F., Gingold, J., Wang, H. L., Zhao, R., Sevilla, A., Darr, H., Williamson, A. J. K., Chang, B., Niu, X., Aguilo, F., Flores, E. R., Sher, Y. P., Hung, M. C., Whetton, A. D., Gelb, B. D., Moore, K. A., Snoeck, H. W., Ma'ayan, A., Schaniel, C., Lemischka, I. R. (2012). Regulation of embryonic and induced pluripotency by aurora kinase-p53 signaling. *Cell Stem Cell* **11**(2): 179-194.

Li, P., Shan, J. X., Chen, X. H., Zhang, D., Su, L. P., Huang, X. Y., Yu, B. Q., Zhi, Q. M., Li, C. L., Wang, Y. Q., Tomei, S., Cai, Q., Ji, J., Li, J. F., Chouchane, L., Yu, Y. Y., Sun, F. Z., Xu, Z. H., Liu, B. Y., Zhu, Z. G. (2015). Epigenetic silencing of microRNA-149 in cancer-associated fibroblasts mediates prostaglandin E2/interleukin-6 signaling in the tumor microenvironment. *Cell Research* **25**(5): 588-603.

Li, X. Y., Hu, S. Q., Xiao, L. (2015). The cancer-associated fibroblasts and drug resistance. *European Review for Medical and Pharmacological Sciences* **19**(11): 2112-2119.

Li, Y., Zhang, Y., Li, S., Lu, J., Chen, J., Wang, Y., Li, Y., Xu, J., Li, X. (2015). Genome-wide DNA methylome analysis reveals epigenetically dysregulated non-coding RNAs in human breast cancer. *Scientific Reports* **5**: 8790.

Lim, K. P., Cirillo, N., Hassona, Y., Wei, W., Thurlow, J. K., Cheong, S. C., Pitiyage, G., Parkinson, E. K., Prime, S. S. (2011). Fibroblast gene expression profile reflects the stage of tumour progression in oral squamous cell carcinoma. *Journal of Pathology* **223**(4): 459-469.

Lin, S. J., Gagnon-Bartsch, J. A., Tan, I. B., Earle, S., Ruff, L., Pettinger, K., Ylstra, B., van Grieken, N., Rha, S. Y., Chung, H. C., Lee, J. S., Cheong, J. H., Noh, S. H., Aoyama,

T., Miyagi, Y., Tsuburaya, A., Yoshikawa, T., Ajani, J. A., Boussioutas, A., Yeoh, K. G., Yong, W. P., So, J., Lee, J., Kang, W. K., Kim, S., Kameda, Y., Arai, T., zur Hausen, A., Speed, T. P., Grabsch, H. I., Tan, P. (2015). Signatures of tumour immunity distinguish Asian and non-Asian gastric adenocarcinomas. *Gut* **64**(11): 1721-1731.

Lin, S. M., Du, P., Huber, W., Kibbe, W. A. (2008). Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Research* **36**(2): e11.

Lin, X., Zhao, Y., Song, W. M., Zhang, B. (2015). Molecular classification and prediction in gastric cancer. *Computational And Structural Biotechnology Journal* **13**: 448-458.

Lindholm, M. E., Marabita, F., Gomez-Cabrero, D., Rundqvist, H., Ekstrom, T. J., Tegner, J., Sundberg, C. J. (2014). An integrative analysis reveals coordinated reprogramming of the epigenome and the transcriptome in human skeletal muscle after training. *Epigenetics* **9**(12): 1557-1569.

Liu, Q., Liu, L., Zhao, Y., Zhang, J., Wang, D., Chen, J., He, Y., Wu, J., Zhang, Z., Liu, Z. (2011). Hypoxia induces genomic DNA demethylation through the activation of HIF-1 alpha and transcriptional upregulation of MAT2A in hepatoma cells." *Molecular Cancer Therapeutics* **10**(6): 1113-1123.

Lo, P. K., Lee, J. S., Liang, X., Han, L., Mori, T., Fackler, M. J., Sadik, H., Argani, P., Pandita, T. K., Sukumar, S. (2010). Epigenetic inactivation of the potential tumor suppressor gene FOXF1 in breast cancer." *Cancer Research* **70**(14): 6047-6058.

Lopez-Serra, P., Esteller, M. (2012). DNA methylation-associated silencing of tumor-suppressor microRNAs in cancer. *Oncogene* **31**(13): 1609-1622.

Lorenzen, S., Lordick, F. (2011). How will human epidermal growth factor receptor 2-neu data impact clinical management of gastric cancer? *Current Opinion in Oncology* **23**(4): 396-402.

Luga, V., Zhang, L., Viloria-Petit, A. M., Ogunjimi, A. A., Inanlou, M. R., Chiu, E., Buchanan, M., Hosein, A. N., Basik, M., Wrana, J. L. (2012). Exosomes mediate stromal mobilization of autocrine Wnt-PCP signaling in breast cancer cell migration. *Cell* **151**(7): 1542-1556.

Lujambio, A., Portela, A., Liz, J., Melo, S. A., Rossi, S., Spizzo, R., Croce, C. M., Calin, G. A., Esteller, M. (2010). CpG island hypermethylation-associated silencing of non-coding RNAs transcribed from ultraconserved regions in human cancer. *Oncogene* **29**(48): 6390-6401.

Mahlapuu, M., Enerback, S., Carlsson, P. (2001). Haploinsufficiency of the forkhead gene Foxf1, a target far sonic hedgehog signaling, causes lung and foregut malformations. *Development* **128**(12): 2397-2406.

Mahlapuu, M., Ormestad, M., Enerback, S., Carlsson, P. (2001). The forkhead transcription factor Foxf1 is required for differentiation of extra-embryonic and lateral plate mesoderm. *Development* **128**(2): 155-166.

Maksimovic, J., Gordon, L., Oshlack, A. (2012). SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biology* **13**(6): R44.

Malik, P., Cashen, A. F. (2014). Decitabine in the treatment of acute myeloid leukemia in elderly patients. *Cancer Management and Research* **6**: 53-61.

Maor, G. L., Yearim, A., Ast, G. (2015). The alternative role of DNA methylation in splicing regulation. *Trends in Genetics* **31**(5): 274-280.

Martinez-Outschoorn, U., Sotgia, F., Lisanti, M. P. (2014). Tumor microenvironment and metabolic synergy in breast cancers: critical importance of mitochondrial fuels and function. *Seminars in Oncology* **41**(2): 195-216.

Martinez-Outschoorn, U. E., Balliet, R. M., Rivadeneira, D. B., Chiavarina, B., Pavlides, S., Wang, C., Whitaker-Menezes, D., Daumer, K. M., Lin, Z., Witkiewicz, A. K., Flomenberg, N., Howell, A., Pestell, R. G., Knudsen, E. S., Sotgia, F., Lisanti, M. P. (2010). Oxidative stress in cancer associated fibroblasts drives tumor-stroma co-evolution. A new paradigm for understanding tumor metabolism, the field effect and genomic instability in cancer cells. *Cell Cycle* **9**(16): 3256-3276.

Martinez-Outschoorn, U. E., Lisanti, M. P., Sotgia, F. (2014). Catabolic cancer-associated fibroblasts transfer energy and biomass to anabolic cancer cells, fueling tumor growth. *Seminars in Cancer Biology* **25**: 47-60.

Mataga, M. A., Rosenthal, S., Heerboth, S., Devalapalli, A., Kokolus, S., Evans, L. R., Longacre, M., Housman, G., Sarkar, S. (2012). Anti-breast cancer effects of histone deacetylase inhibitors and calpain inhibitor. *Anticancer Research* **32**(7): 2523-2529.

Mathivanan, S., Fahner, C. J., Reid, G. E., Simpson, R. J. (2012). ExoCarta 2012: database of exosomal proteins, RNA and lipids. *Nucleic Acids Research* **40**(D1): D1241-D1244.

Matsuoka, J., Yashiro, M., Doi, Y., Fuyuhiro, Y., Kato, Y., Shinto, O., Noda, S., Kashiwagi, S., Aomatsu, N., Hirakawa, T., Hasegawa, T., Shimizu, K., Shimizu, T., Miwa, A., Yamada, N., Sawada, T., Hirakawa, K. (2013). Hypoxia stimulates the EMT of gastric cancer cells through autocrine TGF beta signaling. *Plos One* **8**(5): e62310.

McCaig, C., Duval, C., Hemers, E., Steele, I., Pritchard, D. M., Przemeck, S., Dimaline, R., Ahmed, S., Bodger, K., Kerrigan, D. D., Wang, T. C., Dockray, G. J., Varro, A. (2006). The role of matrix metalloproteinase-7 in redefining the gastric microenvironment in response to Helicobacter pylori. *Gastroenterology* **130**(6): 1754-1763.

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M., Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology* **28**(5): 495-501.

Meinken, J., Walker, G., Cooper, C. R., Min, X. J. (2015). MetazSecKB: the human and animal secretome and subcellular proteome knowledgebase. *Database* **2015**.

Menendez, J. A., Lupu, R. (2007). Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nature Reviews Cancer* **7**(10): 763-777.

Mercer, T. R., Dinger, M. E., Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nature Reviews Genetics* **10**(3): 155-159.

Merry, C. R., Forrest, M. E., Sabers, J. N., Beard, L., Gao, X. H., Hatzoglou, M., Jackson, M. W., Wang, Z., Markowitz, S. D., Khalil, A. M. (2015). DNMT1-associated long non-coding RNAs regulate global gene expression and DNA methylation in colon cancer. *Human Molecular Genetics* **24**(21): 6240-6253.

Mi, H., Muruganujan, A., Thomas, P. D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research* **41**(D1): D377-D386.

Milne, A. N., Carneiro, F., O'Morain, C., Offerhaus, G. J. A. (2009). Nature meets nurture: molecular genetics of gastric cancer. *Human Genetics* **126**(5): 615-628.

Minciacchi, V. R., Freeman, M. R., Di Vizio, D. (2015). Extracellular vesicles in cancer: exosomes, microvesicles and the emerging role of large oncosomes. *Seminars in Cell & Developmental Biology* **40**: 41-51.

Mishra, P. J., Mishra, P. J., Humeniuk, R., Medina, D. J., Alexe, G., Mesirov, J. P., Ganesan, S., Glod, J. W., Banerjee, D. (2008). Carcinoma-associated fibroblast-like differentiation of human mesenchymal stem cells. *Cancer Research* **68**(11): 4331-4339.

Misra, S., Fu, A. A., Misra, K. D., Shergill, U. M., Leof, E. B., Mukhopadhyay, D. (2010). Hypoxia-induced phenotypic switch of fibroblasts to myofibroblasts through a matrix metalloproteinase 2/tissue inhibitor of metalloproteinase-mediated pathway: implications for venous neointimal hyperplasia in hemodialysis access. *Journal of Vascular and Interventional Radiology* **21**(6): 896-902.

Mitra, A. K., Zillhardt, M., Hua, Y., Tiwari, P., Murmann, A. E., Peter, M. E., Lengyel, E. (2012). MicroRNAs reprogram normal fibroblasts into cancer-associated fibroblasts in ovarian cancer. *Cancer Discovery* **2**(12): 1100-1108.

Mootha, V. K., Lindgren, C. M., Eriksson, K. F. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P, Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., Altshuler, D., Groop, L. C. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics* **34**(3): 267-273.

Mund, C., Lyko, F. (2010). Epigenetic cancer therapy: proof of concept and remaining challenges. *Bioessays* **32**(11): 949-957.

Naba, A., Clauser, K. R., Hoersch, S., Liu, H., Carr, S. A., Hynes, R. O. (2012). The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Molecular & Cellular Proteomics* **11**(4).

Navab, R., Strumpf, D., Bandarchi, B., Zhu, C. Q., Pintilie, M., Ramnarine, V. R., Ibrahimov, E., Radulovich, N., Leung, L., Barczyk, M., Panchal, D., To, C., Yun, J. J., Der, S., Shepherd, F. A., Jurisica, I., Tsao, M. S. (2011). Prognostic gene-expression signature of carcinoma-associated fibroblasts in non-small cell lung cancer. *Proceedings of the National Academy of Sciences of the United States of America* **108**(17): 7160-7165.

Nieman, K. M., Kenny, H. A., Penicka, C. V., Ladanyi, A., Buell-Gutbrod, R., Zillhardt, M. R., Romero, I. L., Carey, M. S., Mills, G. B., Hotamisligil, G. S., Yamada, S. D., Peter, M. E., Gwin, K., Lengyel, E. (2011). Adipocytes promote ovarian cancer metastasis and provide energy for rapid tumor growth. *Nature Medicine* **17**(11): 1498-U1207.

Noble, P. W., Barkauskas, C. E., Jiang, D. (2012). Pulmonary fibrosis: patterns and perpetrators. *Journal of Clinical Investigation* **122**(8): 2756-2762.

Ohlund, D., Elyada, E., Tuveson, D. (2014). Fibroblast heterogeneity in the cancer wound. *The Journal of Experimental Medicine* **211**(8): 1503-1523.

Okano, M., Bell, D. W., Haber, D. A., Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**(3): 247-257.

Orimo, A., Gupta, P. B., Sgroi, D. C., Arenzana-Seisdedos, F., Delaunay, T., Naeem, R., Carey, V. J., Richardson, A. L., Weinberg, R. A. (2005). Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell* **121**(3): 335-348.

Parsons, J. T., Horwitz, A. R., Schwartz, M. A. (2010). Cell adhesion: integrating cytoskeletal dynamics and cellular tension. *Nature Reviews Molecular Cell Biology* **11**(9): 633-643.

Paulsson, J., Micke, P. (2014). Prognostic relevance of cancer-associated fibroblasts in human cancer. *Seminars in Cancer Biology* **25**: 61-68.

Pavlides, S., Whitaker-Menezes, D., Castello-Cros, R., Flomenberg, N., Witkiewicz, A. K., Frank, P. G., Casimiro, M. C., Wang, C., Fortina, P., Addya, S., Pestell, R. G., Martinez-Outschoorn, U. E., Sotgia, F., Lisanti, M. P. (2009). The reverse Warburg effect aerobic glycolysis in cancer associated fibroblasts and the tumor stroma. *Cell Cycle* **8**(23): 3984-4001.

Peng, Q., Zhao, L., Hou, Y., Sun, Y., Wang, L., Luo, H., Peng, H., Liu, M. (2013). Biological characteristics and genetic heterogeneity between carcinoma-associated fibroblasts and their paired normal fibroblasts in human breast cancer. *Plos One* **8**(4).

Petersen, O. W., Nielsen, H. L., Gudjonsson, T., Villadsen, R., Rank, F., Niebuhr, E., Bissell, M. J., Ronnov-Jessen, L. (2003). Epithelial to mesenchymal transition in human breast cancer can provide a nonmalignant stroma. *American Journal of Pathology* **162**(2): 391-402.

Phipson, B., Oshlack, A. (2014). DiffVar: a new method for detecting differential variability with application to methylation in cancer and aging. *Genome biology* **15**(9): 465.

Pietras, K., Ostman, A. (2010). Hallmarks of cancer: Interactions with the tumor stroma. *Experimental Cell Research* **316**(8): 1324-1331.

Pinheiro, D. d. R., Silva Ferreira, W. A., Leao Barros, M. B., Araujo, M. D., Rodrigues-Antunes, S., Borges, B. d. N. (2014). Perspectives on new biomarkers in gastric cancer: diagnostic and prognostic applications. *World Journal of Gastroenterology* **20**(33): 11574-11585.

Polanska, U. M., Orimo, A. (2013). Carcinoma-associated fibroblasts: non-neoplastic tumour-promoting mesenchymal cells. *Journal of Cellular Physiology* **228**(8): 1651-1657.

Powell, D. W., Mifflin, R. C., Valentich, J. D., Crowe, S. E., Saada, J. I., West, A. B. (1999). Myofibroblasts. II. Intestinal subepithelial myofibroblasts. *American Journal of Physiology-Cell Physiology* **277**(2): C183-C201.

Provenzano, P. P., Cuevas, C., Chang, A. E., Goel, V. K., Von Hoff, D. D., Hingorani, S. R. (2012). Enzymatic targeting of the stroma ablates physical barriers to treatment of pancreatic ductal adenocarcinoma. *Cancer Cell* **21**(3): 418-429.

Pruitt, K. D., Brown, G. R., Hiatt, S. M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., Farrell, C. M., Hart, J., Landrum, M. J., McGarvey, K. M., Murphy, M. R., O'Leary, N. A., Pujar, S., Rajput, B., Rangwala, S. H., Riddick, L. D., Shkeda, A., Sun, H., Tamez, P., Tully, R. E., Wallin, C., Webb, D., Weber, J., Wu, W., DiCuccio, M., Kitts, P., Maglott, D. R., Murphy, T. D., Ostell, J. M. (2014). RefSeq: an update on mammalian reference sequences. *Nucleic Acids Research* **42**(D1): D756-D763.

Pufulete, M., Al-Ghnaniem, R., Khushal, A., Appleby, P., Harris, N., Gout, S., Emery, P. W., Sanders, T. A. B. (2005). Effect of folic acid supplementation on genomic DNA methylation in patients with colorectal adenoma. *Gut* **54**(5): 648-653.

Qin, Z., Ren, F., Xu, X., Ren, Y., Li, H., Wang, Y., Zhai, Y., Chang, Z. (2009). ZNF536, a novel zinc finger protein specifically expressed in the brain, negatively regulates neuron differentiation by repressing retinoic acid-induced gene transcription. *Molecular and Cellular Biology* **29**(13): 3633-3643.

Qu, Y., Dang, S., Hou, P. (2013). Gene methylation in gastric cancer. *Clinica Chimica Acta* **424**: 53-65.

Ramachandran, S., Ient, J., Gottgens, E. L., Krieg, A. J., Hammond, E. M. (2015). Epigenetic Therapy for solid tumors: highlighting the impact of tumor hypoxia. *Genes* **6**(4): 935-956.

Ramteke, A., Ting, H., Agarwal, C., Mateen, S., Somasagara, R., Hussain, A., Graner, M., Frederick, B., Agarwal, R., Deep, G. (2015). Exosomes secreted under hypoxia enhance invasiveness and stemness of prostate cancer cells by targeting adherens junction molecules. *Molecular Carcinogenesis* **54**(7): 554-565.

Record, M., Carayon, K., Poirot, M., Silvente-Poirot, S. (2014). Exosomes as new vesicular lipid transporters involved in cell-cell communication and various pathophysiologies. *Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids* **1841**(1): 108-120.

Rinn, J. L., Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annual Review of Biochemistry, Vol 81. R. D. Kornberg*. **81:** 145-166.

Robertson, K. D. (2005). DNA methylation and human disease. *Nature Reviews Genetics* **6**(8): 597-610.

Robinson, C. M., Neary, R., Levendale, A., Watson, C. J., Baugh, J. A. (2012). Hypoxia-induced DNA hypermethylation in human pulmonary fibroblasts is associated with Thy-1 promoter methylation and the development of a pro-fibrotic phenotype. *Respiratory Research* **13**:74.

Rodriguez-Canales, J., Hanson, J. C., Tangrea, M. A., Erickson, H. S., Albert, P. S., Wallis, B. S., Richardson, A. M., Pinto, P. A., Linehan, W. M., Gillespie, J. W., Merin, M. J., Libutti, S. K., Woodson, K. G., Emmert-Buck, M. R., Chuaqui, R. F. (2007). Identification of a unique epigenetic sub-microenvironment in prostate cancer. *Journal of Pathology* **211**(4): 410-419.

Ronnovjessen, L., Petersen, O. W. (1993). Induction of alpha-smooth muscle actin by transforming growth factor-beta 1 in quiescent human breast gland fibroblasts. Implications for myofibroblast generation in breast neoplasia. *Laboratory Investigation* **68**(6): 696-707.

Rutkowski, D. T., Hegde, R. S. (2010). Regulation of basal cellular physiology by the homeostatic unfolded protein response. *Journal of Cell Biology* **189**(5): 783-794.

Rysman, E., Brusselmans, K., Scheys, K., Timmermans, L., Derua, R., Munck, S., Van Veldhoven, P. P., Waltregny, D., Daniels, V. W., Machiels, J., Vanderhoydonc, F., Smans, K., Waelkens, E., Verhoeven, G., Swinnen, J. V. (2010). De novo lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation. *Cancer Research* **70**(20): 8117-8126.

Saito, R.-A., Micke, P., Paulsson, J., Augsten, M., Pena, C., Jonsson, P., Botling, J., Edlund, K., Johansson, L., Carlsson, P., Jirstrom, K., Miyazono, K., Ostman, A. (2010). Forkhead box F1 regulates tumor-promoting properties of cancer-associated fibroblasts in lung cancer. *Cancer Research* **70**(7): 2644-2654.

Salic, A., Mitchison, T. J. (2008). A chemical method for fast and sensitive detection of DNA synthesis in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **105**(7): 2415-2420.

Santi, D. V., Norment, A., Garrett, C. E. (1984). Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **81**(22): 6993-6997.

Sapari, N. S., Loh, M., Vaithilingam, A., Soong, R. (2012). Clinical potential of DNA methylation in gastric cancer: a meta-analysis. *Plos One* **7**(4).

Sarkar, S., Goldgar, S., Byler, S., Rosenthal, S., Heerboth, S. (2013). Demethylation and re-expression of epigenetically silenced tumor suppressor genes: sensitization of cancer cells by combination therapy. *Epigenomics* **5**(1): 87-94.

Satoyoshi, R., Kuriyama, S., Aiba, N., Yashiro, M., Tanaka, M. (2014). Asporin activates coordinated invasion of scirrhous gastric cancer and cancer-associated fibroblasts. *Oncogene* **34**(5): 650-60.

Schneider, B. G., Mera, R., Piazuelo, M. B., Bravo, J. C., Zabaleta, J., Delgado, A. G., Bravo, L. E., Wilson, K. T., El-Rifai, W., Peek, R. M., Jr., Correa, P. (2015). DNA methylation predicts progression of human gastric lesions. *Cancer Epidemiology Biomarkers & Prevention* **24**(10): 1607-1613.

Semenza, G. L. (1998). Hypoxia-inducible factor 1: master regulator of $O_2$ homeostasis. *Current Opinion in Genetics & Development* **8**(5): 588-594.

Semenza, G. L. (2003). Targeting HIF-1 for cancer therapy. *Nature Reviews Cancer* **3**(10): 721-732.

Shah, M. A., Khanin, R., Tang, L., Janjigian, Y. Y., Klimstra, D. S., Gerdes, H., Kelsen, D. P. (2011). Molecular classification of gastric cancer: a new paradigm. *Clinical Cancer Research* **17**(9): 2693-2701.

Shahrzad, S., Bertrand, K., Minhas, K., Coomber, B. L. (2007). Induction of DNA hypomethylation by tumor hypoxia. *Epigenetics* **2**(2): 119-125.

Shao, Z. M., Nguyen, M., Barsky, S. H. (2000). Human breast carcinoma desmoplasia is PDGF initiated. *Oncogene* **19**(38): 4337-4345.

Shiga, K., Hara, M., Nagasaki, T., Sato, T., Takahashi, H., Takeyama, H. (2015). Cancer-associated fibroblasts: their characteristics and their roles in tumor growth. *Cancers* **7**(4): 2443-2458.

Shimoda, M., Mellody, K. T., Orimo, A. (2010). Carcinoma-associated fibroblasts are a rate-limiting determinant for tumour progression. *Seminars in Cell & Developmental Biology* **21**(1): 19-25.

Shin, C. M., Kim, N., Jung, Y., Park, J. H., Kang, G. H., Park, W. Y., Kim, J. S., Jung, H. C., Song, I. S. (2011). Genome-wide DNA methylation profiles in noncancerous gastric mucosae with regard to Helicobacter pylori infection and the presence of gastric cancer. *Helicobacter* **16**(3): 179-188.

Simons, K., Gerl, M. J. (2010). Revitalizing membrane rafts: new tools and insights. *Nature Reviews Molecular Cell Biology* **11**(10): 688-699.

Singer, C. F., Gschwantler-Kaulich, D., Fink-Retter, A., Haas, C., Hudelist, G., Czerwenka, K., Kubista, E. (2008). Differential gene expression profile in breast cancer-derived stromal fibroblasts. *Breast Cancer Research and Treatment* **110**(2): 273-281.

Sinkkonen, L., Hugenschmidt, T., Berninger, P., Gaidatzis, D., Mohn, F., Artus-Revel, C. G., Zavolan, M., Svoboda, P., Filipowicz, W. (2008). MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nature Structural & Molecular Biology* **15**(3): 259-267.

Skalli, O., Schurch, W., Seemayer, T., Lagace, R., Montandon, D., Pittet, B., Gabbiani, G. (1989). Myofibroblasts from diverse pathologic settings are heterogeneous in their content of actin isoforms and intermediate filament proteins. *Laboratory Investigation* **60**(2): 275-285.

Smiraglia, D. J., Kulawiec, M., Bistulfi, G. L., Gupta, S. G., Singh, K. K. (2008). A novel role for mitochondria in regulating epigenetic modification in the nucleus. *Cancer Biology & Therapy* **7**(8): 1182-1190.

Smith, R. N., Aleksic, J., Butano, D., Carr, A., Contrino, S., Hu, F., Lyne, M., Lyne, R., Kalderimis, A., Rutherford, K., Stepan, R., Sullivan, J., Wakeling, M., Watkins, X., Micklem, G. (2012). InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. *Bioinformatics* **28**(23): 3163-3165.

Smith, Z. D., Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nature Reviews Genetics* **14**(3): 204-220.

Smyth, G. K. (2005). Limma: linear models for microarray data *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. C. V. Gentleman R, Huber W, Irizarry R, Dudoit S (eds.). New York, Springer**:** pp. 397-420.

Smyth, G. K., Michaud, J., Scott, H. S. (2005). Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21**(9): 2067-2075.

Sohn, B. H., Park, I. Y., Lee, J. J., Yang, S. J. J., Jang, Y. J., Park, K. C., Kim, D. J., Lee, D. C., Sohn, H. A., Kim, T. W., Yoo, H. S. S., Choi, J. Y., Bae, Y. S., Yeom, Y. I. (2010). Functional switching of TGF-beta1 signaling in liver cancer via epigenetic modulation of a single CpG site in TTP promoter. *Gastroenterology* **138**(5): 1898-1908.

Stoeltzing, O., McCarty, M. F., Wey, J. S., Fan, F., Liu, W., Belcheva, A., Bucana, C. D., Semenza, G. L., Ellis, L. M. (2004). Role of hypoxia-inducible factor 1α in gastric cancer cell growth, angiogenesis, and vessel maturation. *Journal of the National Cancer Institute* **96**(12).

Subarsky, P., Hill, R. P. (2003). The hypoxic tumour microenvironment and metastatic progression. *Clinical & Experimental Metastasis* **20**(3): 237-250.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**(43): 15545-15550.

Sumida, T., Kitadai, Y., Shinagawa, K., Tanaka, M., Kodama, M., Ohnishi, M., Ohara, E., Tanaka, S., Yasui, W., Chayama, K. (2011). Anti-stromal therapy with imatinib inhibits growth and metastasis of gastric carcinoma in an orthotopic nude mouse model. *International Journal of Cancer* **128**(9): 2050-2062.

Sun, C., Fukui, H., Hara, K., Zhang, X., Kitayama, Y., Eda, H., Tomita, T., Oshima, T., Kikuchi, S., Watari, J., Sasako, M., Miwa, H. (2015). FGF9 from cancer-associated fibroblasts is a possible mediator of invasion and anti-apoptosis of gastric cancer cells. *BMC Cancer* **15**.

Supek, F., Bošnjak, M., Škunca, N., Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS One* **6**(7).

Szyf, M., Bozovic, V., Tanigawa, G. (1991). Growth regulation of mouse DNA methyltransferase gene expression. *Journal of Biological Chemistry* **266**(16): 10027-10030.

Tahara, T., Arisawa, T. (2015). DNA methylation as a molecular biomarker in gastric cancer. *Epigenomics* **7**(3): 475-486.

Tan, I. B., Ivanova, T., Lim, K. H., Ong, C. W., Deng, N., Lee, J., Tan, S. H., Wu, J., Lee, M. H., Ooi, C. H., Rha, S. Y., Wong, W. K., Boussioutas, A., Yeoh, K. G., So, J., Yong, W. P., Tsuburaya, A., Grabsch, H., Toh, H. C., Rozen, S., Cheong, J. H., Noh, S. H., Wan, W. K., Ajani, J. A., Lee, J. S., Tellez, M. S., Tan, P. (2011). Intrinsic subtypes of gastric cancer, based on gene expression pattern, predict survival and respond differently to chemotherapy. *Gastroenterology* **141**(2): 476-U551.

Tang, D., Gao, J., Wang, S., Ye, N., Chong, Y., Huang, Y., Wang, J., Li, B., Yin, W., Wang D. (2015). Cancer-associated fibroblasts promote angiogenesis in gastric cancer through galectin-1 expression. *Tumour Biology* pp1-11.

Taverner, T., Karpievitch, Y. V., Polpitiya, A. D., Brown, J. N., Dabney, A. R., Anderson, G. A., Smith, R. D. (2012). DanteR: an extensible R-based tool for quantitative analysis of -omics data. *Bioinformatics* **28**(18): 2404-2406.

Teschendorff, A. E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450k DNA methylation data. *Bioinformatics* **29**(2): 189-196.

Ting-Song, Y., Xiao-Hu, Y., Xi, C., Xu-Dong, W., Jie, H., Dong-Lei, Z., Bo, Z., Zhen-Shun, S. (2014). MicroRNA-106b in cancer-associated fibroblasts from gastric cancer promotes cell migration and invasion by targeting PTEN. *FEBS Letters* **588**(13): 2162-9.

Tokuda, Y., Satoh, Y., Fujiyama, C., Toda, S., Sugihara, H., Masaki, Z. (2003). Prostate cancer cell growth is modulated by adipocyte-cancer cell interaction. *BJU International* **91**(7): 716-720.

Tomasek, J. J., Gabbiani, G., Hinz, B., Chaponnier, C., Brown, R. A. (2002). Myofibroblasts and mechano-regulation of connective tissue remodelling. *Nature Reviews Molecular Cell Biology* **3**(5): 349-363.

Tost, J., Gut, I. G. (2007). DNA methylation analysis by pyrosequencing. *Nature Protocols* **2**(9): 2265-2275.

Toullec, A., Gerald, D., Despouy, G., Bourachot, B., Cardon, M., Lefort, S., Richardson, M., Rigaill, G., Parrini, M. C., Lucchesi, C., Bellanger, D., Stern, M. H., Dubois, T., Sastre-Garau, X., Delattre, O., Vincent-Salomon, A., Mechta-Grigoriou F. (2010). Oxidative stress promotes myofibroblast differentiation and tumour spreading. *EMBO Molecular Medicine* **2**(6): 211-230.

Triche, T. J., Jr., Weisenberger, D. J., Van Den Berg, D., Laird, P. W., Siegmund, K. D. (2013). Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Research* **41**(7).

Tsuyoshi, H., Masakazu, Y., Takafumi, N., Junko, M., Yuhiko, F., Tamami, M., Tatsunari, F., Kiyoshi, S., Toshiyuki, S., Atsushi, M., Kosei, H. (2014). Cancer-associated fibroblasts might sustain the stemness of scirrhous gastric cancer cells via transforming growth factor-β signaling. *International Journal of Cancer* **134**(8): 1785-95.

Vaupel, P., Mayer, A. (2007). Hypoxia in cancer: significance and impact on clinical outcome. *Cancer and Metastasis Reviews* **26**(2): 225-239.

Vizoso, M., Puig, M., Carmona, F. J., Maqueda, M., Velasquez, A., Gomez, A., Labernadie, A., Lugo, R., Gabasa, M., Rigat-Brugarolas, L. G., Trepat, X., Ramirez, J., Moran, S., Vidal, E., Reguart, N., Perera, A., Esteller, M., Alcaraz, J. (2015). Aberrant DNA methylation in non-small cell lung cancer-associated fibroblasts. *Carcinogenesis* **36**(12): 1453-1463.

Vogelstein, B., Kinzler, K. W. (2004). Cancer genes and the pathways they control. *Nature Medicine* **10**(8): 789-799.

Wagner, A. D., Moehler, M. (2009). Development of targeted therapies in advanced gastric cancer: promising exploratory steps in a new era. *Current Opinion in Oncology* **21**(4): 381-385.

Wang, G. L., Jiang, B. H., Rue, E. A., Semenza, G. L. (1995). Hypoxia-inducible factor-1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular $O_2$ tension. *Proceedings of the National Academy of Sciences of the United States of America* **92**(12): 5510-5514.

Wang, K., Ma, W., Wang, J., Yu, L., Zhang, X., Wang, Z., Tan, B., Wang, N., Bai, B., Yang, S., Liu, H., Zhu, S., Cheng, Y. (2012). Tumor-stroma ratio is an independent predictor for survival in esophageal squamous cell carcinoma. *Journal of Thoracic Oncology* **7**(9): 1457-1461.

Wang, M., Kaufman, R. J. (2014). The impact of the endoplasmic reticulum protein-folding environment on cancer development. *Nature Reviews Cancer* **14**(9): 581-597.

Wang, S., Kaufman, R. J. (2012). The impact of the unfolded protein response on human disease. *Journal of Cell Biology* **197**(7): 857-867.

Wang, X., Zhou, Y. X., Qiao, W., Tominaga, Y., Ouchi, M., Ouchi, T., Deng, C. X. (2006). Overexpression of aurora kinase A in mouse mammary epithelium induces genetic instability preceding mammary tumor formation. *Oncogene* **25**(54): 7148-7158.

Wang, Y. (2009). Wnt/Planar cell polarity signaling: a new paradigm for cancer therapy. *Molecular Cancer Therapeutics* **8**(8): 2103-2109.

Watanabe, Y., Kim, H. S., Castoro, R. J., Chung, W., Estecio, M. R. H., Kondo, K., Guo, Y., Ahmed, S. S., Toyota, M., Itoh, F., Suk, K. T., Cho, M. Y., Shen, L., Jelinek, J., Issa J. P. J. (2009). Sensitive and specific detection of early gastric cancer with dna methylation analysis of gastric washes. *Gastroenterology* **136**(7): 2149-2158.

Watson, J. E. V., Doggett, N. A., Albertson, D. G., Andaya, A., Chinnaiyan, A., van Dekken, H., Ginzinger, D., Haqq, C., James, K., Kamkar, S., Kowbel, D., Pinkel, D., Schmitt, L., Simko, J. P., Volik, S., Weinberg, V. K., Paris, P. L., Collins, C. (2004). Integration of high-resolution array comparative genomic hybridization analysis of chromosome 16q with expression array data refines common regions of loss at16q23-qter and identifies underlying candidate tumor suppressor genes in prostate cancer. *Oncogene* **23**(19): 3487-3494.

Webber, J., Steadman, R., Mason, M. D., Tabi, Z., Clayton, A. (2010). Cancer exosomes trigger fibroblast to myofibroblast differentiation. *Cancer Research* **70**(23): 9621-9630.

Weber, K. T., Sun, Y., Bhattacharya, S. K., Ahokas, R. A., Gerling I. C. (2013). Myofibroblast-mediated mechanisms of pathological remodelling of the heart. *Nature Reviews Cardiology* **10**(1): 15-26.

Widschwendter, M., Jiang, G. C., Woods, C., Muller, H. M., Fiegl, H., Goebel, G., Marth, C., Muller-Holzner, E., Zeimet, A. G., Laird, P. W., Ehrlich, M. (2004). DNA hypomethylation and ovarian cancer biology. *Cancer Research* **64**(13): 4472-4480.

Wilson, V. L., Jones, P. A. (1983). DNA methylation decreases in aging but not in immortal cells. *Science (Washington D C)* **220**(4601): 1055-1057.

Wiseman, B. S., Werb, Z. (2002). Development - stromal effects on mammary gland development and breast cancer. *Science* **296**(5570): 1046-1049.

Wu, J., Zheng, H. Y., Zhang, Z. G., Wang, B (2013). Relationship and the clinical significance between the Helicobacter pylori infection and the expression of Hypoxia-inducible factor-1α in gastric carcinoma tissues. *Chinese Journal of Cancer Prevention and Treatment* **20**(6): 416-419.

Wu, M. C., Joubert, B. R., Kuan, P. F., Haberg, S. E., Nystad, W., Peddada, S. D., London, S. J. (2014). A systematic assessment of normalization approaches for the Infinium 450K methylation platform. *Epigenetics* **9**(2): 318-329.

Xu-Jun, H., Hou-Quan, T., Zhi-Ming, H., Ying-Yu, M., Ji, X., Hui-Ju, W., Ying-Jie, X., Li, L., Bao-Ying, F., Ya-Qing, L., Jian-Zhong, C. (2014). Expression of galectin-1 in carcinoma-associated fibroblasts promotes gastric cancer cell invasion through upregulation of integrin β1. *Cancer Science* **105**(11):1402-1410.

Xu, T. P., Huang, M. D., Xia, R., Liu, X. X., Sun, M., Yin, L., Chen, W.M., Han, L., Zhang, E. B., Kong, R., De, W., Shu Y. Q. (2014). Decreased expression of the long non-coding RNA FENDRR is associated with poor prognosis in gastric cancer and FENDRR regulates gastric cancer cell metastasis by affecting fibronectin1 expression. *Journal of Hematology & Oncology* **7**: 63.

Xu, W. S., Parmigiani, R. B., Marks, P. A. (2007). Histone deacetylase inhibitors: molecular mechanisms of action. *Oncogene* **26**(37): 5541-5552.

Yamanoi, K., Arai, E., Tian, Y., Takahashi, Y., Miyata, S., Sasaki, H., Chiwaki, F., Ichikawa, H., Sakamoto, H., Kushima, R., Katai, H., Yoshida, T., Sakamoto, M., Kanai Y. (2015). Epigenetic clustering of gastric carcinomas based on DNA methylation profiles at the precancerous stage: its correlation with tumor aggressiveness and patient outcome. *Carcinogenesis* **36**(5): 509-520.

Yamashita, M., Ogawa, T., Zhang, X., Hanamura, N., Kashikura, Y., Takamura, M., Yoneda, M., Shiraishi, T. (2012). Role of stromal myofibroblasts in invasive breast cancer: stromal expression of alpha-smooth muscle actin correlates with worse clinical outcome. *Breast Cancer* **19**(2): 170-176.

Yan, Y., Wang, L. F., Wang, R. F. (2015). Role of cancer-associated fibroblasts in invasion and metastasis of gastric cancer. *World Journal of Gastroenterology* **21**(33): 9717-9726.

Yang, T. S., Yang, X. H., Chen, X., Wang, X. D., Hua, J., Zhou, D. L., Zhou, B., Song, Z. S. (2014). MicroRNA-106b in cancer-associated fibroblasts from gastric cancer promotes cell migration and invasion by targeting PTEN. *FEBS Letters* **588**(13): 2162-2169.

Yoo, C. B., Jones, P. A. (2006). Epigenetic therapy of cancer: past, present and future. *Nature Reviews Drug Discovery* **5**(1): 37-50.

Zeisberg, E. M., Potenta, S., Xie, L., Zeisberg, M., Kalluri, R. (2007). Discovery of endothelial to mesenchymal transition as a source for carcinoma-associated fibroblasts. *Cancer Research* **67**(21): 10123-10128.

Zhao, L., Sun, Y., Hou, Y., Peng, Q., Wang, L., Luo, H., Tang, X., Zeng, Z., Liu, M. (2012). MiRNA expression analysis of cancer-associated fibroblasts and normal fibroblasts in breast cancer. *International Journal of Biochemistry & Cell Biology* **44**(11): 2051-2059.

Zhi, K., Shen, X., Zhang, H., Bi, J. (2010). Cancer-associated fibroblasts are positively correlated with metastatic potential of human gastric cancers. *Journal of Experimental & Clinical Cancer Research* **29**:66.

Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics* **39**(1): 61-69.

Ziller, M. J., Gu, H., Mueller, F., Donaghey, J., Tsai, L. T. Y., Kohlbacher, O., De Jager, P. L., Rosen, E. D., Bennett, D. A., Bernstein, B. E., Gnirke, A., Meissner, A. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**(7463): 477-481.