University of Nottingham

UK | CHINA | MALAYSIA

Faculty of Medicine and Health Sciences

School of Medicine

Department of Translational Medical Sciences

The role of Proline-Rich Homeodomain protein PRH/HHEX in cholangiocarcinoma

Nahlah Mahmoud A. Ghouth

MBBS, MSc.

Thesis submitted to the University of Nottingham

For the Degree of Doctor of Philosophy

January 2024

Declaration

Except where acknowledged in the thesis, I declare that this thesis “The role of Proline-Rich Homeodomain protein PRH/HHEX in cholangiocarcinoma” is my own work that was generated during the registered period at the University of Nottingham. No part of this thesis has been submitted for any other academic degree at any other institution.

Nahlah Ghouth, January 2024

Acknowledgements

I would like to express my gratitude to Allah Almighty for bestowing upon me the blessings of knowledge and strength to complete this thesis.

The Messenger of Allah said "Whoever is not grateful to the people, he is not grateful to Allah"

I would like to express my deepest thanks and appreciation to my principal supervisor, Professor Kevin Gaston, for his unwavering guidance, assistance, and support from the first day at the university of Bristol until the final hours before submitting this thesis to the University of Nottingham. Your boundless patience is inspiring. I would also like to extend my gratitude to Professor Sheela Jayaraman for her incredible and valuable inputs to my PhD work. I must likewise acknowledge Professor David Bates and Dr Isioma for their appreciated advice.

I also extend my thanks to Dr. Chris Toseland, Dr. Alia Santos, and Dr. Robert Markus for their help with PALM dSTORM work. Additionally, I am grateful for the aid with FCS work I received from Dr. Joelle Goulding.

I would like to acknowledge Dr. Eudmar Marcolino and Dr. Danielle Clark for their invaluable assistance during my early stages of my PhD. I am indeed fortunate to have Dr. Marcolino as a friend and a senior.

I cannot go without thanking my former desk mate Vicky Norris. Your shoulder was a source of solace for many of TVBG students during our moments of distress, and your advice will always be deeply appreciated.

The originals of TVBG QMC group; Roa'a, Nas, Ria, Eduardo, Nick, Aida, Santi, Cristina, Eric, Mireia thank you for welcoming me into your group. You made C floor lab a home.

My travel companions and top Whatsapp contacts, Amany, Mussarat, Khawla, and Ghaneemah thank you for all the enjoyable trips and late-night outings we embarked. You were the remedy for a broken heart after failed experiments.

The survivors in BDI3 (Asma, Mashael, Rinad, Rania, Abrar, Reem, Nairouz, Nada, Amna, Wafa’a, Khadijah, Annie, Bushra, Shatha, Nabeelah, Dewi, Grace and Chris) we shared laughter, tears, meals, bench talks and joyful company. Thank you for making this PhD journey a memorable and rewarding experience.

I sincerely would like to express my thanks to Wendy and Sedrick. Your warm greetings always made it worthy to start lab work at 6 a.m.

Last but not least, there are no words to express my deepest gratitude to my family; my father, my siblings Maram, Mashael, Nada, Ala'a, Emad, and Hattan. This extends to ChiChi who witnessed all the horrors of nearly losing my thesis drafts. Thank you all for believing in me and keeping me in your prayers. I owe you all that I have accomplished thus far.

Dedication

This thesis is dedicated to the memory of my late mother. She battled breast cancer with courage and grace. Her infinite love continues to strengthen my journey.

Abstract

Cholangiocarcinoma (CCA) is a malignancy of the bile ducts that arises from their lining epithelial cells. CCA has increasing incidence and mortality. Proline Rich Homeodomain / Haematopoietically Expressed Homeobox Protein (PRH/HHEX) is a transcription factor that acts as an oncoprotein in CCA. Here, using cumulative growth and EdU incorporation assays, I show that PRH expression in immortalised cholangiocytes (AKN1 cells) increased cell growth and proliferation. Additionally, I show using immunofluorescence imaging and Western blotting that PRH expression in AKN1 cells induces cellular phenotypes that resembles partial epithelial to mesenchymal transition (EMT), and that this morphological transformation is accompanied by an increase in cell migration and cell invasion. Recently, PRH expression has been reported to upregulate the expression of the mesenchymal marker protein Vimentin in CCA cells. To date most published research has focused on Vimentin expression in mesenchymal cells and its roles in the cytoplasm. Vimentin is known to participate in several biological processes such as cell proliferation, EMT, cell adhesion, cell migration, and cell invasion which have been reported to be upregulated in cancer cells with the consequence of increased tumour growth and disease progress. Knowledge on the role or roles of Vimentin in the nucleus is limited. However, nuclear Vimentin was reported to exhibit gene regulatory activity in Neuroblastoma cells. To expand knowledge on the importance of Vimentin in CCA, I used a combination of high and super-resolution imaging techniques to show that PRH may spatially exist near Vimentin in the nuclear subcellular regions of CCA cells. In addition, I used

siRNA to deplete Vimentin in CCA cells and to reveal that Vimentin loss inhibits cell proliferation, induces an epithelial-like morphology, and reduces cell migration. Gene ontology (GO) and Gene Set Enrichment Analysis (GSEA) of RNAseq data generated using these cells revealed that Vimentin, in conjunction with PRH, regulates multiple signalling pathways in line with the phenotype of Vimentin-depleted cells. This project provides new insights into the importance of PRH and Vimentin in CCA. The novel PRH and Vimentin spatial organisation and the identified target genes co-regulated by PRH and Vimentin serve as new potential targets for future interventional studies.

COVID-19 IMPACT

The lockdown occurred shortly after we moved our lab from Queens Medical Centre to Biodiscovery Institute 3. This mean I was unable to perform experiments between November 2019 and January 2020. Resuming lab work in the new building required multiple inductions. So, even after I was granted lab access, the beginning was slow. The abrupt closure of the university during COVID-19 pandemics forced me to stop lab work again from mid-March 2020 until early August 2020. This delayed several experiments originally planned to take place during that period specially microscopy work. Just after I was back in the lab, 3 new PhD students joined our group. Going through training sessions, although joyful, it impacted my already restricted time in the lab during that period. There were also consequences from the ordering and delivery system in BDI3 which delayed receiving lab reagents.

Furthermore, I suffer a physical malformation that suddenly worsened starting late 2020 throughout 2021 and afterwards. Accessing NHS healthcare during that period was difficult. This forced me to sometimes terminate experiments due to the pain.

Table of Contents

Declaration *1*

Acknowledgements *2*

Dedication *3*

Abstract *4*

COVID-19 IMPACT *6*

Table of Contents *7*

Table of Figures *14*

List of Tables *24*

Abbreviations *26*

1 Chapter 1: Introduction *1*

1.1 Cancer **1**

1.2 Hallmarks of cancer **2**

1.2.1 Genetic instability 3

1.2.2 Resisting cell death 5

1.2.3 Sustained proliferation 7

1.2.4 Cell cycle key regulators 8

1.2.5 Evading the inhibition of proliferation 10

1.2.6 Metastasis 10

1.3 Gene expression and regulation in cancer **23**

1.3.1 The central dogma of molecular biology and gene expression 23

1.3.2 Chromatin organisation 24

1.3.3 Chromatin remodelling 25

1.3.4 Transcription complex components 27

1.3.5 Transcriptional complex formation and gene transcription 30

1.3.6 Gene regulation in cancer 31

1.4 Biliary tract and cholangiocarcinoma 33

1.4.1 Biliary System 33

1.4.2 Cholangiocarcinoma 36

1.5 Proline Rich Homeodomain / Haematopoietically Expressed Homeobox Protein (PRH/HHEX) 42

1.5.1 PRH background and discovery 42

1.5.2 Regulation of PRH gene expression 43

1.5.3 PRH expression during embryonic development and in adults 44

1.5.4 PRH protein structure 45

1.5.5 PRH oligomerisation and high-order assembly 48

1.5.6 PRH post translational modification by CK2 phosphorylation 49

1.5.7 PRH subcellular distribution 50

1.5.8 The role of PRH in the regulation of target gene expression 53

1.5.9 PRH in Cancer 53

1.5.10 PRH in CCA 55

1.6 PRH and Vimentin in CCA 57

1.7 Vimentin 58

1.7.1 Background 58

1.7.2 Regulation of Vimentin gene expression 59

1.7.3 Vimentin expression 61

1.7.4 Vimentin structure and organisation 62

1.7.5 Post-translational modification of Vimentin (PTM) 64

1.7.6 Vimentin forms and subcellular localisation 65

1.7.7 Vimentin subcellular localisation on immunofluorescence imaging 71

1.7.8 Vimentin functions / roles 71

1.7.9 Vimentin in cancer 78

1.8 Aims of the project 79

2 Chapter 2: Materials and methods 81

2.1 Cell culture 81

2.1.1 Cell lines 81

2.1.2 Cell growth and maintenance 83

2.1.3 Cell counting 83

2.1.4 Cell pelleting 84

2.1.5 Cell thawing 84

2.1.6 Cell freezing 85

2.2 Preparation of cell lines 85

2.2.1 Stable AKN1 PRH overexpressing cell line 85

2.2.2 Vimentin knockdown cells 96

2.3 Microscopy 98

2.3.1 Fluorescent microscopy 98

2.3.2 Immunofluorescence staining and imaging 99

2.3.3 Photo-activated localisation microscopy – direct stochastic optical reconstruction microscopy (PALM dSTORM) immunofluorescence staining 102

2.3.4 Fluorescent correlation spectrometry 111

2.4 Western blotting 112

2.4.1 Whole cell protein extraction and quantification 112

2.4.2 Protein quantification 112

2.4.3 Loading sample preparation 113

2.4.4 Sodium dodecyl sulphate - poly acrylamide gel electrophoresis 113

2.5 Nucleic acid experiments 118

2.5.1 RT-qPCR primer design, preparation, and storage 118

2.5.2 Primer optimisation 119

2.5.3 RNA extraction 123

2.5.4 cDNA reverse transcription 124

2.5.5 RT-qPCR 124

2.5.6 Analysis 126

2.6 Proliferation assays 127

2.6.1 Cumulative growth curve 127

2.6.2 EdU incorporation assay 128

2.7 Transwell migration assay (chemotaxis) 130

2.7.1 Seeding for transwell migration assay 130

2.7.2 CalceinAM standard curve 131

2.8 Transwell invasion assay 132

2.8.1 Inserts coating with ECM gel layer 133

2.8.2 Seeding, labelling and detection 134

2.9 Gene expression profiling interactive analysis (GEPIA2) 134

2.10 RNAseq 135

2.10.1 Sample preparation 135

2.10.2 Novogene library preparation and sequencing 136

2.10.3 Data mining, retrieval, and upload 138

2.10.4 Galaxy web platform analysis for RNAseq data with biological repeats 138

2.10.5 Preparing for downstream analysis 139

2.10.6 Molecular signature database (MSigDB) 140

2.10.7 Enrichr for Gene Ontology 141

2.10.8 Pre-ranked GSEA 141

2.10.9 Overlapping with different RNAseq data 141

2.11 ChIPseq analysis 142

2.11.1 Data mining and retrieval 142

2.11.2 Galaxy platform analysis (flowchart) 142

2.11.3 Hypergeometric Optimization of Motif EnRichment (HOMER) 144

2.11.4 HOMER annotation 144

2.11.5 Downstream analysis 144

2.12 Statistical analysis 145

3 Chapter 3: GFP-tagged PRH expression in immortalised cholangiocytes and its effects on cell proliferation, cell morphology, cell migration, and cell invasion 146

3.1 Introduction 146

3.2 Results 148

3.2.1 Generation of a stable AKN1 GFP-PRH-Myc cell line 148

3.2.2 GFP-tagged PRH expression in AKN1 cells increases cell growth and proliferation ... 162

3.2.3 Morphological characterisation of AKN1 GFP-PRH-Myc cells 165

3.2.4 AKN1 cells expressing GFP-tagged PRH increases cell migration and cell invasion ... 173

3.2.5 Issues between early and late passages of AKN1 GFP and AKN1 GFP-PRH-Myc clone D 176

3.3 Discussion 179

4 Chapter 4: The investigation of PRH and Vimentin localisation in CCA cells 187

4.1 Introduction 187

4.2 Results 192

4.2.1 PRH is expressed in the nuclei and cytoplasm of CCLP1 cells 192

4.2.2 PRH and Vimentin spatial localisation in the nuclear and cytoplasmic compartments of CCLP1 cells 197

4.2.3 Investigating PRH-Vimentin nuclear localisation in CCLP1 cells using photoactivated localisation microscopy – direct stochastic optical reconstruction microscopy (PALM dSTORM) 210

4.2.4 PRH and PRH-induced Vimentin spatial localisation in AKN1 cells 229

4.2.5 The use of fluorescence correlation spectroscopy (FCS) to study the dynamics of GFP-PRH fused protein in AKN1 cells for long term GFP-PRH expression 233

4.3 Discussion 244

5 Chapter 5: The investigation of gene expression co-regulation by Vimentin and PRH in CCA cells 259

5.1 Introduction 259

5.2 Results 261

5.2.1 *VIM* and *HHEX* genes are highly expressed in CCA 261

5.2.2 Characterising Vimentin depletion in CCA cell lines 265

5.2.3 Vimentin Knockdown RNAseq and enrichment analysis in CCA cell lines 278

5.2.4 RNAseq overlap analysis to investigate co-regulation of gene expression by Vimentin and PRH in CCA 301

5.3 Discussion 310

6 Chapter 6: Conclusion 325

6.1 Overview 325

6.2 Conclusion 331

7 Appendix 333

8 References 338

Table of Figures

Figure 1-1 Hallmarks of Cancer. 3

Figure 1-2 Cell Cycle. 4

Figure 1-3 Metastasis. 11

Figure 1-4 Epithelial Cells and Junctions. 15

Figure 1-5 Epithelial to Mesenchymal Transition (EMT) (67, 68). 18

Figure 1-6 Organisation of DNA in chromatin (9, 120, 121). 24

Figure 1-7 Transcription complex at the promoter region with some of the regulatory elements (137). 28

Figure 1-8 Biliary System. 33

Figure 1-9 CCA Anatomical Classification. 36

Figure 1-10 Geographic distribution of CCA. 37

Figure 1-11 PRH $^{-/-}$ mice failed to develop their liver as indicated by its absence in the abdominal cavity (219). 45

Figure 1-12: PRH monomer. 46

Figure 1-13 PRH subcellular distribution according to COMPARTMENT. 50

Figure 1-14 PRH subcellular localisation. 52

Figure 1-15 A schematic illustration of intermediate filaments assembly (274). 58

Figure 1-16 Vimentin expression in normal human tissues as reported by the human protein atlas (296). 61

Figure 1-17 A monomer of Vimentin and its domains (307, 313, 314). 63

Figure 1-18 Vimentin assembly from monomers into cytoskeleton (317-320) 64

Figure 1-19 Vimentin subcellular distribution according to COMPARTMENT. 66

Figure 1-20 Vimentin approved filamentous expression in the cytoplasm. .. 71

Figure 2-1 si*VIM5* transfection protocol. 97

Figure 2-2 Summary of PALM dSTORM imaging and analysis (412). 108

Figure 2-3 RT-qPCR primer standard curve. 122

Figure 2-4 RT-qPCR reaction mix preparation. 125

Figure 2-5 EdU incorporation assay experimental steps. 129

Figure 2-6 Transwell migration (Chemotaxis) assay experimental steps. .. 130

Figure 2-7 Transwell invasion assay experimental steps. 133

Figure 2-8 Overall RNA processing by Novogene 136

Figure 2-9 Novogene cDNA library preparation. 137

Figure 2-10 Galaxy RNAseq pipeline 138

Figure 2-11 Galaxy ChIPseq pipeline (423). 142

Figure 3-1 Schematic presentation of pEGFP-C1 and pEGFP-PRH-Myc plasmids. 149

Figure 3-2 Restriction enzyme digest analysis of pEGFP-PRH-Myc plasmid. 150

Figure 3-3 Transfection efficiency of pEGFP-C1 and pEGFP-PRH-Myc plasmids in AKN1 cells. 153

Figure 3-4 AKN1 GFP and AKN1 GFP-PRH-Myc cells selection with G418. 154

Figure 3-5 FACS single-sorting of transfected AKN1 cells expressing GFP or GFP-tagged PRH. 157

Figure 3-6 AKN1 GFP and AKN1 GFP-PRH-Myc cells after FACS single-cell sorting and growth with maintenance doses of G418 (D clone). 160

Figure 3-7 Validation of PRH mRNA and GFP-tagged PRH protein expression in AKN1 GFP and AKN1 GFP-PRH-Myc cells (D clone cell line). 161

Figure 3-8 GFP-tagged PRH expression in AKN1 cells appears to significantly induce cell growth and shorten cell doubling time. 164

Figure 3-9 GFP-tagged PRH expression in AKN1 cells induces cell proliferation. 165

Figure 3-10 GFP-tagged PRH expression induces mesenchymal-like morphology in AKN1 cells. 167

Figure 3-11 Morphological analysis of AKN1 GFP and AKN1 GFP-PRH-Myc cells using immunofluorescence imaging. 171

Figure 3-12 GFP-tagged PRH expression induces E-Cadherin expression in AKN1 cells. 173

Figure 3-13 GFP-tagged PRH expression significantly increases AKN1 transwell migration (chemotaxis)..... 174

Figure 3-14 GFP-tagged PRH expression induces a significant increase in AKN1 transwell invasion. 175

Figure 3-15 Late passages of clone D AKN1 GFP-PRH-Myc cells show no change in PRH mRNA levels and various GFP-tagged PRH protein levels. 177

Figure 3-16 The reduction in cell proliferation of AKN1 GFP-PRH-Myc cells in later passages..... 178

Figure 4-1 ChIPseq track of CCLP Myc-PRH visualised on UCSC genome browser. 188

Figure 4-2 Vimentin is identified as a protein partner for PRH in CCLP Myc-PRH mass spectrometry. 189

Figure 4-3 PRH interacts with Vimentin in the cytoplasmic and nuclear fractions of CCLP Myc-PRH cells. 189

Figure 4-4 PRH expression in CCLP1 cells. 194

Figure 4-5 Nuclear PRH aggregates away from DAPI dark-stained regions (heterochromatin)..... 197

Figure 4-6 PRH and Vimentin show partial co-localisation in CCLP1 cells.200

Figure 4-7 PRH and Vimentin overlap in the cytoplasmic and to a lesser extent in the nuclear subcellular compartments of CCLP1 cells. 205

Figure 4-8 RGB profile analysis of PRH and Vimentin in the subcellular compartment of CCLP1 cells. 209

Figure 4-9 PRH is clustered in the nucleus. 213

Figure 4-10 PRH is segregated from condensed heterochromatin and clustered in the nucleus. 214

Figure 4-11 Spatial organisation of PRH and Vimentin in CCLP1 cells in PALM dSTORM images. 216

Figure 4-12 PRH and Vimentin single molecules are precisely localised using PALM dSTORM. 217

Figure 4-13 Example STORM-rendered images of nuclear PRH (red) and Vimentin (green) in CCLP1 cells showing various sizes of cluster formations at different magnifications. 220

Figure 4-14 Cluster analysis of nuclear RPH and Vimentin. 221

Figure 4-15 The window for ClusDoC tool showing the parameters of degree of co-localisation and DBSCAN. 223

Figure 4-16 Cluster maps for PRH and Vimentin based on the selected regions for cluster analysis. 225

Figure 4-17 Statistical analysis of nuclear PRH and Vimentin cluster organisations in CCLP1 cells. 226

Figure 4-18 Degree of co-localisation analysis of nuclear PRH and Vimentin clusters in CCLP1 cells in seven selected ROIs. 228

Figure 4-19 PRH nuclear clusters show a significant Degree of co-localisation with nuclear Vimentin (*p*= 0.0042). 229

Figure 4-20 Weak PRH and PRH-induced Vimentin overlap in AKN1 cells expressing GFP-tagged PRH..... 231

Figure 4-21 Western blotting of late passaged AKN1 GFP and AKN1 GFP-PRH-Myc cells. 234

Figure 4-22 Confocal imaging of AKN1 GFP and AKN1 GFP-PRH-Myc showing the three observed patterns of GFP expression. 237

Figure 4-23 The recorded dynamic of the first pattern of GFP expression (uniformed and very bright) in AKN1 GFP cells. 238

Figure 4-24 The recorded dynamic of the second pattern of GFP dynamics (diffused and very dim) in AKN1 GFP-PRH-Myc cells. 240

Figure 4-25 The recorded dynamic of the third pattern of GFP expression (diffused and very dim in foci and clusters) in AKN1 GFP-PRH-Myc cells. 242

Figure 5-1 Dot plot gene expression profiling of *VIM* and *HHEX* displayed as transcripts per million in different cancer (red) and normal tissues (green). 262

Figure 5-2 Box plot indicating the high expression of *VIM* and *HHEX* in 36 cancer tissues compared to 9 normal samples. 263

Figure 5-3 *VIM* expression correlates with *HHEX* in CCA..... 264

Figure 5-4 Vimentin mRNA fold change using RT-qPCR after si*VIM5* transfection. 265

Figure 5-5 Determination of Vimentin relative fold change using Western blotting(48 hours after *siVIM* transfection). 266

Figure 5-6 PRH mRNA and protein expression after knocking down Vimentin. 268

Figure 5-7 Vimentin loss in CCA cells significantly decreases cell proliferation. 270

Figure 5-8 Brightfield images of the control and Vimentin knockdown cells. 271

Figure 5-9 Morphological analysis of CCLP siVimentin cells using immunofluorescence imaging. 273

Figure 5-10 Morphological analysis of CCSW siVimentin cells using immunofluorescent imaging. 275

Figure 5-11 Vimentin knockdown significantly decreases cell migration in CCA cells. 278

Figure 5-12 RNA samples integrity was assessed using gel electrophoresis. 280

Figure 5-13 Differentially expressed genes in CCLP siVimentin RNAseq.. 282

Figure 5-14 Gene ontology analysis on CCLP siVimentin RNA seq data. . 284

Figure 5-15 GSEA using MSigDB Hallmarks on CCLP siVimentin significant and regulated DEG. 285

Figure 5-16 Pre-ranked GSEA analysis using Hallmarks on CCLP siVimentin significant DEG. 287

Figure 5-17 Differentially expressed genes in CCSW siVimentin RNAseq. 289

Figure 5-18 Gene ontology analysis on CCSW siVimentin RNA seq data. 290

Figure 5-19 GSEA using MSigDB Hallmarks on CCSW siVimentin significant and regulated DEG. 291

Figure 5-20 Pre-ranked GSEA analysis using Hallmarks on CCSW siVimentin significant DEG. 292

Figure 5-21 Common DEG enriched in EMT hallmark in CCLP siVimentin RNAseq and CCSW siVimentin RNAseq. 294

Figure 5-22 Summary on DEG counts in CCLP siVimentin and CCSW siVImentin RNAseq. 296

Figure 5-23 Venn diagram showing 558 significant DEG were regulated by Vimentin in both RNAseq data sets. 297

Figure 5-24 GSEA using MSigDB Hallmarks on 118 significant and regulated DEG by Vimentin in the same direction. 298

Figure 5-25 The effect of Vimentin knockdown on the expression of genes with roles in the progression of the cell cycle. 300

Figure 5-26 Venn diagram showing that 131 significant DEG were co-regulated by Vimentin and PRH. 301

Figure 5-27 UpSet plot showing the number of extracted significant DEG co-regulated by Vimentin and PRH in the same direction. These DEG also have peaks for PRH on ChIPseq. 302

Figure 5-28 GSEA using MSigDB Hallmarks on 18 significant DEG co-regulated by Vimentin and PRH in the same direction. 303

Figure 5-29 Venn diagram showing that 599 significant DEG were co-regulated by Vimentin and PRH. 304

Figure 5-30 UpSet plot showing the number of extracted significant DEG co-regulated by Vimentin and PRH in opposite direction after applying FC cut-off on the common DEG. 305

Figure 5-31 GSEA using MSigDB Hallmarks on 26 significant DEG co-regulated by Vimentin and PRH in opposite directions. 306

Figure 5-32 Significant DEG co-regulated by Vimentin and PRH and the number of PRH peaks. 307

Figure 5-33 Validation of RNAseq findings using RT-qPCR. 308

Figure 5-34 Flowchart analysis model for the investigation of gene co-regulation by Vimentin and PRH in CCA 311

Figure 6-1 Schematic illustration of gene regulation by PRH and Vimentin. 331

Figure 7-1 GSN RT-qPCR primer optimisation. 333

Figure 7-2 STAT3 RT-qPCR primer optimisation. 334

Figure 7-3 WASF1 RT-qPCR primer optimisation 335

Figure 7-4 Plasmid sequencing repor.t 336

List of Tables

Table 2-1 Cell lines properties 82

Table 2-2 Selecting antibiotics 86

Table 2-3 QIAGEN plasmid Maxiprep Kit buffers 89

Table 2-4 Diagnostic restriction digest experiment preparation 90

Table 2-5 AKN1 PEI transfection 92

Table 2-6 si*VIM5* transfection 96

Table 2-7 List of immunofluorescence staining buffers 100

Table 2-8 List of immunofluorescence antibodies 101

Table 2-9 List of PALM dSTORM buffers 104

Table 2-10 List of PALM dSTORM antibodies 105

Table 2-11 List of PALM dSTORM imaging buffers 106

Table 2-12 List of Western blotting buffers 115

Table 2-13 List of Antibodies used in Western blotting 116

Table 2-14 RT-qPCR cycles 121

Table 2-15 Qiagen RNeasy buffers 123

Table 2-16 List of primers used in RT-qPCR 126

Table 3-1 Cell count input and output for fluorescent-activated cell sorting (FACS). 158

Table 5-1 RNAseq samples prepared for pre-sequencing quality check-up. 279

Abbreviations

5mC	5-methylcytosine
a.a.	Amino acids
AFP	Alpha-fetoprotein
AJCC/UICC	American Joint Cancer Committee/Union for International Cancer Control
ASE	Anti-silencer elements
Bak	Bcl-2 homologous antagonist killer
Bax	Bcl-2 Associated X-protein
Bcl-2	B-cell lymphoma 2
Bcl-XL	B-cell lymphoma-extra large
BECs	Biliary epithelial cells
Bim	Bcl-2-like protein 11
BMP	Bone morphogenetic protein
BSA	Bovine serum albumin
CA 19-9	Carbohydrate antigen 19-9
CBD	common bile duct
CCA	cholangiocarcinoma
CDH1	E-cadherin
CDK	Cyclin-dependent protein kinases
CDK1	Cyclin-dependent protein kinases 1
CDK4	Cyclin-dependent protein kinases 4
CDK6	Cyclin-dependent protein kinases 6
CDKI	Cyclin-dependent protein kinases inhibitor
CDKN1A	Cyclin-dependent protein kinases inhibitor 1 A
CDKN1B	Cyclin-dependent protein kinases inhibitor 1 B
CDKN2A	Cyclin-dependent protein kinases inhibitor 2 A
CDKN2B	Cyclin-dependent protein kinases inhibitor 2 B
cDNA	Complementary DNA
CEA	Carcinoembryonic antigen
CHRD	Chordin
CisGem	Cisplatin and Gemcitabine
CK2	Protein kinase
ClusDoC	combined cluster detection and colocalization
CM	complete media
CN	Cyclin
CNB/	Cyclin B
CND	Cyclin D
CND1	Cyclin D1
CNE	Cyclin E
CoIP	Co-immunoprecipitation assay
CpG	Cytosine phosphate Guanine

CK8	Cytokeratin 8
CK18	Cytokeratin 18
CK19	Cytokeratin 19
CPM	Count per molecule
CT	Computing tomography
CTCs	Circulating tumour cells
DAPI	4′,6-diamidino-2-phenylindole
DMEM	Dulbecco’s Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DR6	Death receptor 6
DRiPs	defective by-products of ribosomal proteins
DTCs	Disseminated tumour cells
E. coli	Escherichia coli
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
eIF4E	Eukaryotic translation initiation factor 4E
ELISA	Enzyme-linked immunosorbent assay
EM	Electron microscopy
EMT	Epithelial to mesenchymal transition
ENG	Endoglin
ERCP	Endoscopic retrograde cholangiopancreatography
ES	enrichment score
ESM1	Endothelial specific molecule 1
FBS	Fetal Bovine Serum
FC	Fold change
FCS	Fluorescent correlation spectrometry
FDR	False discovery rate
FGF	Fibroblast growth factor
FIJI	FIJI is just image J
FISH	Fluorescence in situ hybridization
FAK	Focal adhesion kinase
FRAP	Fluorescence Recovery After Photobleaching
FRET	Fluorescence Resonance Energy Transfer
G-ABP	G-actin binding proteins
GGT	Gamma-glutamyl transpeptidase
G1	Gap 1 phase
G2	Gap 2 phase
GEO	Gene Expression Omnibus
GEPIA2	Gene expression profiling interactive analysis
GFP	Green fluorescent protein

GJA1	Gap Junction Protein Alpha 1
GO	Gene ontology
GO-BP	Gene ontology biological processes
GO-CC	Gene ontology cellular component
GO-MF	Gene ontology molecular function
GSC	Goosecoid
GSEA	Gene set enrichment analysis
GSN	Gelsolin
GTEx	Genotype-tissue expression
H3K4me2	Histone di-methylation in histone 3 Lysine 4
H3K9me3	Histone 3 Lysine 9
HAT	Histone acetyl transferase
HDACs	Histone deacetylase enzymes
HNF3β	Hepatocyte Nuclear Factor 3β
HEA	Human epithelial antigen
HOMER	Hypergeometric Optimization of Motif EnRichment
IARC	International Agency for Research on Cancer
IGFBP3	Insulin Like Growth Factor Binding Protein 3
INK4A	Inhibitor of CDK4A
ITGB5	Integrin Subunit Beta 5
KAP7	Karyopherin/Importin 7
LB	Luria-Bertani
lncRNA	long ncRNA
M	Mitotic phase
MBD	Methyl-CpG binding domain
MET	Mesenchymal-to-epithelial transition
MF	Mouse fibroblasts
miRNAs or miR	micro RNAs
MMPs	Metalloproteinases
MRI	Magnetic resonance imaging
MRTF	Myocardin-related transcription factors
MSigDB	Molecular signature database
NCBI	National Centre for Biotechnology Information
ncRNA	non-coding RNA
NDR	Nucleosome depleted regions
NEAA	Non-essential amino acids
NF-kB	Nuclear factor kappa B
NIH	National Human Genome Research Institute Home
NTC	No template control
NTCP	Sodium/taurocholate co-transporting polypeptide
ORA	Over-representation analysis
p21	Cyclin-dependent protein kinases inhibitor p21
p27	Cyclin-dependent protein kinases inhibitor p27

P2Y2	Purinoceptor 2 protein receptors
p53	Protein 53
p57	Cyclin-dependent protein kinases inhibitor 057
PALM dSTORM	Photo-activated localisation microscopy – direct stochastic optical reconstruction microscopy
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEA3	Polyoma enhancer activator 3
PIC	Pre-initiator complex
PIGF	Placental growth factor
PML	Promyelocytic leukaemia
PPI	Protein-protein interactions
PSC	primary sclerosing cholangitis
PTM	Post-translational modification
PVDF	Polyvinylidene difluoride
QC	Quality control
RB1	Retinoblastoma tumour suppressor protein 1
RNA	Ribonucleic acid
RNAP	RNA polymerases
ROI	Regions of interests
rpm	Round per minute
RTK	Receptor tyrosine kinase
RT-qPCR	real time quantitative polymerase chain reaction
S	Synthetic phase
SDC1	Syndecan 1
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SERPINE1	Serpin Family E Member 1
SERPINE2	Serpin Family E Member 2
siRNA	Small interfering RNA
SMLM	Single molecule localisation microscopy
sncRNA	small ncRNA
STAT3	Signal transducer and activator of transcription 3
SWI/SNF	switching defective/sucrose non-fermenting
TAF	TBP associated factors
TBP	TATA box binding protein
TCGA	The Cancer Genome Atlas
TF	Transcription factor
TG	Thyroglobulin
TGFBR3	Transforming Growth Factor Beta Receptor 3
TGF-β	Transforming growth factor-beta
TLE	Transducin-like enhancer
TNM	Tumour-node-metastasis
TP53	Tumour protein 53
TPM	Transcripts per million

TPM4	Tropomyosin 4
TSG	Tumour suppressor gene
TSP1	Thrombospondin-1
TSS	Transcription starting site
TTF	Thyroid-specific transcription factor
UCSC	University of California Santa Cruz
ULF	Unit-length filament
VEGFs	Vascular endothelial growth factors
VWF	Von Willebrand factor
WASF1	WASP Family Member 1
ZBP-89	Zinc finger Binding Protein of 89
Zeb1/2	Zinc-finger E-box binding homeobox
α-SMA	α-smooth muscle actin

Chapter 1: Introduction

1.1 Cancer

Cancer is defined by the National Cancer Institute as a collection of related diseases characterised by a rapid and uncontrolled division of abnormal transformed cells anywhere in the body often caused by the accumulation of detrimental variations in the genome (1). The primary cause of cancer mortality is related to disseminated tumours as a result of migratory cells detached from the original growth, circulated far from their primary locations and established as secondary cancers, a process known as metastasis. As a consequence, the disruption in normal tissue homeostasis will contribute to a loss of normal physiology of the affected organs.

According to the International Agency for Research on Cancer (IARC), cancer is a leading cause of death; the most recent database from GLOBOCAN 2022 indicates almost 20 million incidents and 9.7 million deaths in the year of 2022 (2). Incident rate is expected to exceed 35 million cases in the year of 2050 (2). The impact of cancer on the world is disproportionate for developing countries with health inequality issues. In 2012, the American Cancer Society reported that the financial loss from cancer is a global economic burden with costs reaching $900 billion, and the expenditure on lowering the exposure to cancer risk factors is approximately $1.8 billion (3). Recently, the American medical association reported that the global burden will reach $25.2 trillion over the next 30 years between 2020 and 2050 (4)

Prevention, early diagnosis, and treatment are key elements in eradicating diseases. Understanding how normal cells transform into cancer cells and their consequences in the body can unveil targets for those elements.

1.2 Hallmarks of cancer

The development of cancer is described as a multistep process during which normal cells gain, through an evolutionary process, characteristics enabling them to become malignant (5). These malignant cells survive beyond the normal extent seen for normal cells.

Hanahan & Weinberg (2011) summarised the distinguishing features of cancer and referred to them as “The Hallmarks of Cancer” (6). These 10 hallmarks are: genetic instability, resisting cell death, evading the inhibition of proliferation, sustaining proliferation, replicative immortality, angiogenesis, reprogramming metabolism, inducing inflammation, escaping the immune system, and metastasis (figure 1-1). Recent reviewers have suggested four emerging hallmarks; non-mutational epigenetic re-programming, polymorphic microbiomes, senescent cells, and unblocking phenotypic plasticity (7).

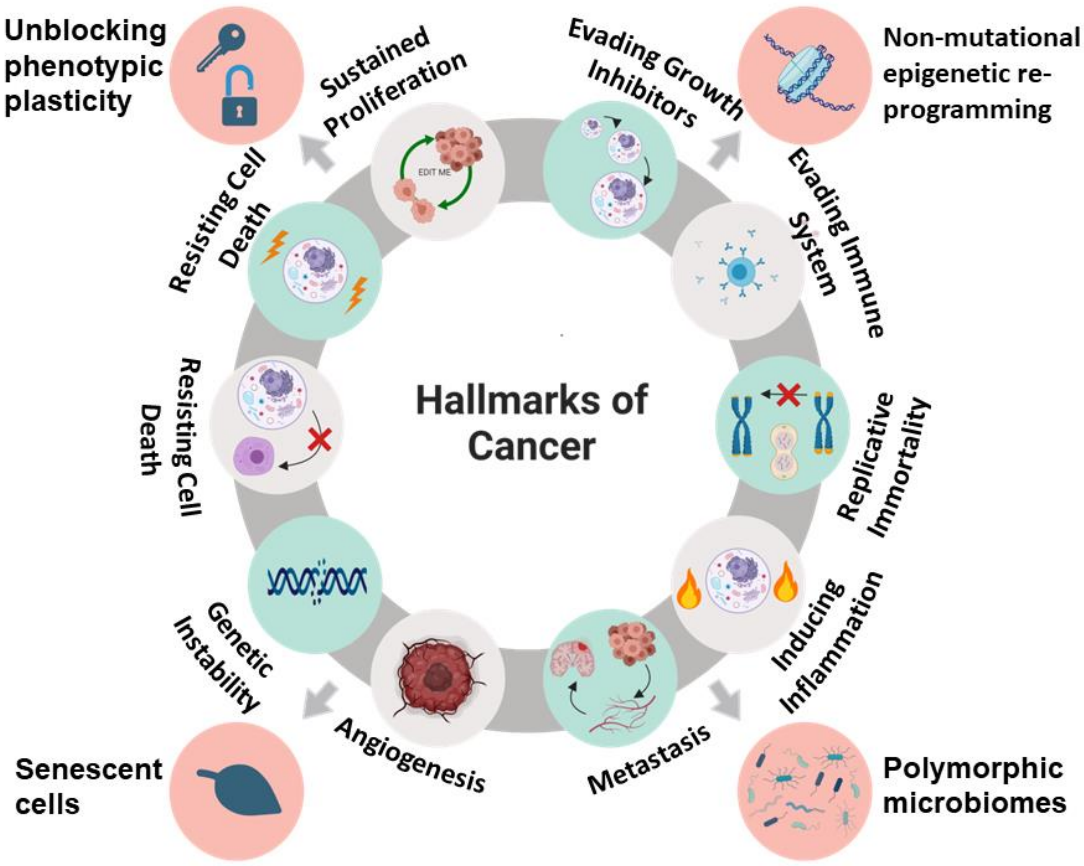


Figure 1-1 Hallmarks of Cancer.

There are ten original hallmarks mentioned in the literature that allow a normal cell to become cancerous. These hallmarks are genetic instability, resisting cell death, evading growth inhibitors (evading inhibition of proliferation), sustaining proliferation, replicative immortality, angiogenesis, reprogrammed metabolism, inducing inflammation, escaping the immune system, and metastasis(6). There are four emerging hallmarks: non-mutational epigenetic re-programming, polymorphic microbiomes, senescent cells, and unblocking phenotypic plasticity (7). Created with BioRender.com

1.2.1 Genetic instability

Genetic stability is the maintenance of faithful genetic material passaged from a single somatic cell to the following daughter cells (8). The stable hierarchical structure of deoxyribonucleic acid (DNA) in the nucleus must be monitored during cell division to ensure a complete copy of (DNA) is passed down to daughter cells. In a normal cell cycle (figure 1-2), Gap 1 phase (G1) is the window where cells either commit to divide and prepare themselves to start

DNA synthesis or exit the cycle to stay in a non-proliferative status, i.e. quiescence in (G0). DNA duplication takes place during the synthetic phase (S). This stage is followed by a check-up point Gap 2 phase (G2) where DNA is condensed, and chromosomes allocate themselves centrally in preparation for the next stage. The next mitotic (M) phase is when a parent cell undergoes a structural change of the cytoskeleton and narrows at some locations to split into two daughter cells with each one having half of the parental genetic material. If replication errors are detected, the cell has the ability to halt the cycle, repair the damage, and resume the cycle (9).

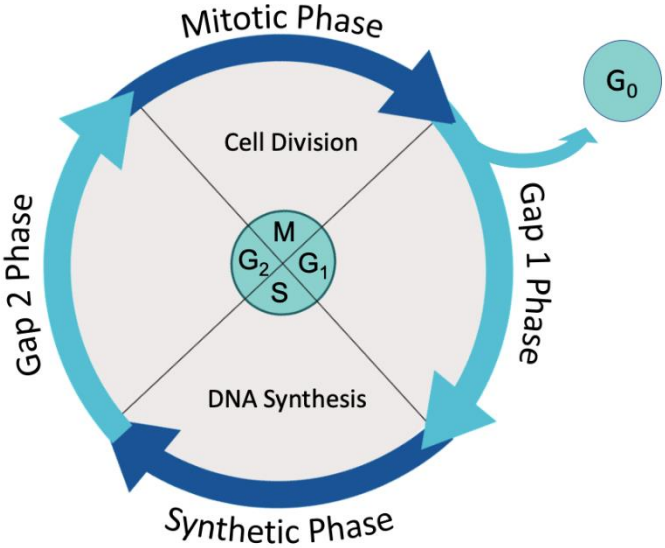


Figure 1-2 Cell Cycle.

The cell cycle is composed of four main phases: Gap 1 phase where there is a rapid increase in growth and metabolism in preparation for the following step, Synthetic phase during which DNA synthesis occurs, Gap 2 phase a checkpoint for DNA synthesis errors, and Mitotic phase where cell division happens. G0 is the stage of cell arrest.(10). Created with BioRender.com

Another protective measure is to eliminate the faulty cell through cell death. An example for this is cell apoptosis, which is a terminating programmed cell death that can take place when a cell undergoes DNA replication errors (11). Therefore, the overall mutation rate is maintained at very low levels. However,

defective cells can still escape the repair processes leading to increased numbers of cells with DNA damage continuing through mitosis, passing the damage to their daughter cells, and accumulating mutations beyond the normal rate (12). This phenomenon was hypothesised by Fearon and Vogelstein in 1990 as a pathogenesis underlying cancer development when they were working on a colorectal neoplasia model to study its genetic basis (13). They noticed that the presence of mutation is not sufficient for colorectal cancer, instead the amount or frequencies of mutations was a definite reason of neoplasia in their model (neoplasia is an abnormal growth of cells) (13).

Chromosome instability is an example genetic instability where some nucleotide deletions, insertions, or substitution leads to abnormal chromosomal count or structure (14). The repair mechanism of this abnormality may be accompanied by random extra chromosomal DNA replications that later get integrated in the genome (15). One such process was identified later as chromothripsis, which is a clustered rearrangement of DNA along with loss of DNA sequences resulting from a single somatic DNA alteration (16, 17). If this occurs in genes driving cell proliferation or inhibiting cell death, it could assist cancer development.

1.2.2 Resisting cell death

Normal tissue homeostasis is maintained by several mechanisms of cell death including apoptosis, necrosis, and autophagy.

Apoptosis, as mentioned earlier, is a programmed cell death through which healthy tissues dispose of abnormal cells. An important regulator in apoptosis

is tumour protein 53 (p53), (encoded by the *TP53 gene*), which is known as the guardian of the genome (18). *TP53* is a tumour suppressor gene (TSG) and p53 induces cell cycle arrest in G1/S phase and allows cells to undergo DNA damage repair which normally prevents the formation of abnormal cells (18). It can also start a cascade of inflammation-independent proteolysis and cellular fragmentation. Deactivation of TSGs interferes with the cellular surveillance systems. The most common example in cancer cells is *TP53* loss of function; making the DNA repair system less sensitive to DNA damage and giving rise to apoptosis resistant cells (18).

Other examples of genes regulating apoptosis can encode apoptotic proteins (e.g. Bcl-2 Associated X-protein (Bax),Bcl-2 homologous antagonist killer (Bak), and Bcl-2-like protein 11 (Bim)), and anti-apoptotic proteins (e.g. Bcl-2 and Bcl-XL) that allow cells to overcome stressful stimuli and go on through the cell cycle (19, 20). Their functions are dysregulated in cancer due to mutations in their encoding genes, most commonly *Bcl2* and *Bcl-XL*, leading to accumulation of anti-apoptotic proteins. As a result, cancer cells become insensitive to apoptotic signals and overgrow (21).

Necrosis is defined as a premature cell death where dying cells release their contents into the surrounding microenvironment (22). Unlike apoptosis, necrosis is triggered mostly through external stimuli and is associated with inflammatory responses and inflammatory cytokines release e,g, interleukin-1 (23). In the cancer microenvironment, necrosis could promote cancer via released stimulatory factors that enable adjacent cells to proliferate and thus contribute to cancer cells persistence (24). This occurs in response to tumour

cells rapid growth depriving their inner core population of blood supply and inducing hypoxia. The shedding of the necrotic core releases molecular fragments capable of inducing inflammation and recruiting inflammatory cells that in turn release pro-tumourigenic cytokines and growth factors. Therefore, necrosis in cancer can promote further tumour growth (25).

Autophagy is a catabolic process where lysosomes initiate cellular degradation in response to stress stimuli (26). This has been suggested to be one reason behind the regrowth of tumours following chemotherapy (27). Chemotherapeutic agents induce cell damage which upregulates autophagy to remove degradation products (28). Despite the protective effect of autophagy, it enables few cancer cells to survive therapy leaving behind residual cancer cells (28). These cells can resume cell proliferation and regrow upon cessation of chemotherapy.

1.2.3 Sustained proliferation

Cell proliferation in normal tissues is achieved by the progress of cells through the cell cycle. We described the cell cycle earlier under “Genetic Instability”. that the cell cycle includes two Gap phases (G1 and G2), DNA synthesis phase, and a mitosis phase. The transition through G1/S or G2/M checkpoints is regulated by ,multiple checkpoint signalling pathways, e.g. p53 pathway, cyclin D – CDK 4 / 6 pathway, TGF-β and PI3K/AKT pathways (9), while the control of cells passing through S phase is achieved by ATR.Chk1 pathway (9). Details on Cyclins and their associated kinases are in the following section.

Normal cells Normal cells maintain strict control over the cell cycle and respond appropriately to growth signals. However, cancer cells have the ability to sustain cell proliferation in several ways including changes in growth factors, growth factor signalling and growth factor receptors. Cancer cells can overproduce growth factors or their receptors to increase growth factor signalling. Growth receptors in cancer cells can also become hypersensitive or structurally altered to work either in a ligand-independent manner or to create a positive feedback loop (29).

1.2.4 Cell cycle key regulators

The key regulators of the cell cycle are different combinations of cyclin (CN) proteins and cyclin-dependent protein kinases (CDK). **CNs** are proteins that can be found at different concentration throughout the cell cycle. CNs undergo periodic synthesis and degradation to peak at certain phases of the cell cycle and serve their purposes when they are required. For instance, CND is important during G1 phase, but its levels fluctuate throughout the cell cycle. CNE is known as a transition CN at G1/S phases. CNA starts elevating at the initiation of S phase and drops toward the beginning of phase M. Whereas, CNB peaks during M phase (30). CNs convert the other key regulators of the cell cycle CDKs into their active forms. **CDKs** are serine-threonine kinases that need CN to be functional (31). CDKs are considered the driving forces of the cell cycle. However, CDKs on their own are inactive. CDKs along with their partner CNs form complexes that can phosphorylate their target proteins (32).

There are several CN/CDK complexes that are important for the cell cycle. The most important combinations are CND/CDK4 (33, 34), CND/CDK6 (33),

and CNB/CDK1 (35, 36). A prominent example of a CND/CDK4 target protein is retinoblastoma tumour suppressor protein (RB1). Non-phosphorylated RB1 is inactive and inhibits a transcription factor called E2F1 and other E2F family members. RB1 undergoes a phosphorylation during G1 phase by CND/CDK4 and CND/CDK6 and dissociates from E2F1. The latter becomes active and upregulates the expression of several target genes including CNE. CNE in turn activates CDK2 that further phosphorylates RB1. When RB1 is hyperphosphorylated, it can drive DNA replication and encourage the cells to proceed towards S phase. (34)

1.2.4.1 Aberrant CN and CDK

CN are degraded via ubiquitination-induced proteolysis (37). Their persistence at increased concentration throughout the cell cycle allows cells to divide uncontrollably and initiate abnormal growth. In addition, the levels of CN in the cell are kept under control via several signalling pathways. These pathways regulate the genes encoding CNs e.g. MAPK and Wnt signalling pathways (38). For example, in hepatic cancer stem cells, the gene encoding CND1 is over expressed (38). Another example of aberrant CN in cancer proliferation is activated by the binding of epidermal growth factor (EGF) to its receptor (EGFR) which is known to stimulate the *RAS* signalling pathway. *RAS* can trigger downstream effects to induce cancer proliferation through the upregulation of CND1 (39). **CDK** can also be inhibited by CDK inhibitors (CDKI) (40). When CDKI binds to an activated CN/CDK complex, CDK undergoes conformational changes that expels CN from the complex and blocks the ATP binding site (41). An important example of CDKI is a family of

Inhibitors called Inhibitor of CDK4A (INK4A) (42). Currently, there are 4 well known INK4A proteins known as p15, 16, 18, and 19. More examples of CDKI are p21, p27 and p57. CDKI can inactivate CDK and block the cell cycle at the G1/S checkpoint (43). Additionally, p21 and p27 are also capable of blocking the cell cycle at S phase. Mutations in any of these CDKI genes can lead to cancer development (43). Therefore, genes encoding CDKI are considered important TSGs e.g., *CDKN1A*, *CDKN2A*, *CDKN1B*, and *CDKN2B* (44).

1.2.5 Evading the inhibition of proliferation

Normal cellular mitosis is regulated by balanced pro and anti-proliferative signals. Cancer cells, however, can proliferate uncontrollably (12). One mechanism of cell growth inhibition occurs via cell-to-cell contact and is known as contact inhibition. When cells come into very close contact with each other as the population is getting denser, they stop proliferating. This feature is absent or reduced in cancer cells (45).Cancer cells also simultaneously dysregulate other previously discussed growth-related signalling pathways and regulators including: p53 (18), the Rb pathway (34), and CDKI, to escape proliferation inhibition (44).

1.2.6 Metastasis

Metastasis as defined earlier is the spread of tumour cells from their primary origin into distant sites where they form secondary growths. It is known as the real killer in cancer since it replaces normal functioning tissues with abnormal cells (46, 47). Metastasis is a complex and multistep process that relies first on the presence of a primary tumour showing increased cell survival and

proliferation as well as lack of differentiation ability. In addition to that, metastasis also involves other elements including angiogenesis, epithelial to mesenchymal transition (EMT), invasion, intravasation, circulation survival, extravasation and dormancy, and secondary tumour formation (48)(figure 1-3).

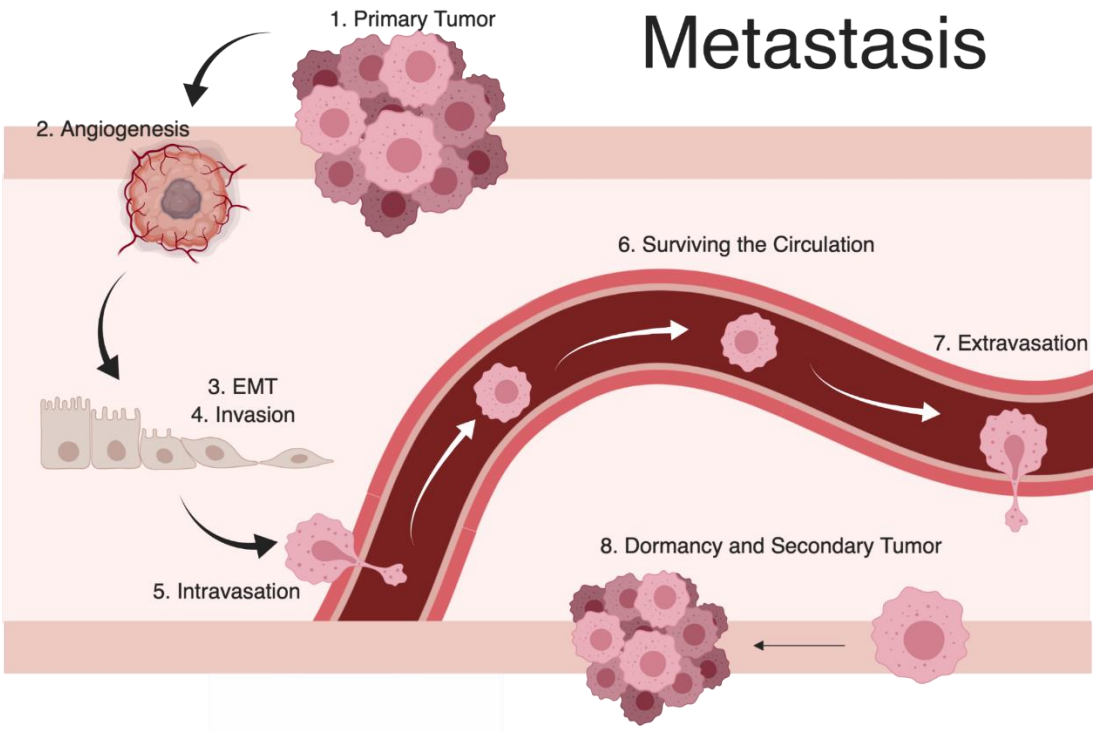


Figure 1-3 Metastasis.

There are eight main steps in the process of metastasis: presence of a primary tumour, angiogenesis, epithelial to mesenchymal transition (EMT), invasion, intravasation, surviving the circulation, extravasation and dormancy and secondary tumour formation. The outcome is having cancer cells in a new location distant from their original growth (49). Created with BioRender.com

1.2.6.1 Primary tumour

The presence of a primary tumour composed of genetically unstable cells containing mutations which make them death-resistant, capable of evading growth inhibitory signals, and continuously proliferating with immortal

replication ability can make the primary tumour mass significantly large and the tumour aggressive. This mass can trigger inflammation and establish its own vasculature. Only a small fraction of the original tumour (around 0.01% of the primary growth) was found to metastasise (50, 51). The hallmark angiogenesis provides a gateway to the blood circulation into which the tumour can shed cells that can survive the immune system and travel to distant sites. The metastatic cells utilise their newly acquired microenvironment to create a metastatic niche through intercellular interaction and cross-talking, and undergo mesenchymal-to-epithelial transition (MET) to partially regain their epithelial characteristics, thereby facilitating their colonisation (49, 52). Eventually, these seeded cells can make a secondary tumour (53).

1.2.6.2 Neoangiogenesis in metastasis

Avascular tumours lack the potential of growth and metastasis. An important add-on angiogenic characteristic for cancer cells is to acquire an altered balance between factors that induce and inhibit angiogenesis, i.e. an angiogenesis switch (54). Angiogenic inducers, such as vascular endothelial growth factors (VEGFs), are potent growth factors that are produced by cells to stimulate the formation of blood vessels and other normal physiological functions (55, 56). The family of VEGF includes VEGF-A, B, C, D and placental growth factor (PIGF) (56). VEGFs turn the switch on if their levels exceeded the inhibitors, and the opposite happens when the level of an inhibitor such as Thrombospondin-1 (TSP-1) increases (57-59). Overall, neoangiogenesis is an important hallmark for cancer progression and metastasis.

1.2.6.3 Epithelial to mesenchymal transition

1.2.6.3.1 Epithelial cells

The epithelial cells in normal tissues are stationary and held together by multiple interactions between cell-surface molecules (figure 1-4). They form a strong barrier contributing to the body's defence against harmful substances and pathogens as they line the inner and outer surfaces of organs and cavities. They tend to have cuboidal morphology with apical to basal polarity. They also strongly interact with their extracellular matrix (ECM) from their basal side and that provides support for epithelial cells, tissues and organs (60). This interaction is achieved via several cell-basement membrane junctions: Hemidesmosomes, focal adhesion, and basal adherens junctions (figure 1-4) (59). **Hemidesmosomes** are specialised structures composed of dense protein structures that interlink intracellular intermediate filaments (Keratins) with laminin and collagen IV in ECM and convey signalling (61). The main components of Hemidesmosomes junctions are Plectin, bulbous pemphigoid antigen 1 (BP230), Integrin α, and Integrin αβ (61). **Focal adhesion junctions** are dynamic complexes composed of several protein structures, e.g. Integrin α, Integrin αβ, focal adhesion kinase (FAK), Vinculin, and Talin (62). They provide the link between intracellular Actin filaments and ECM (62). Basal adherens junctions contain Cadherin, Catenin, and Actin as main structural components. They provide intercellular connection as well as cell-basement membrane anchoring to epithelial cells (62).

Epithelial cells form tight cellular interactions with the adjacent cells; apical tight junctions, that protect from paracellular space infiltration, adherens

junctions that link neighbouring cells via their actin cytoskeletons and mainly represented by cadherins, gap junctions that allow direct horizontal communication, and finally desmosomes (63).

Losing the contact between epithelial cells and ECM can trigger cell migration, EMT and cell survival. Pathway signalling through components of Hemidesmosomes and focal adhesion junctions direct cells to migrate through various ECM rigidities, particularly FAK, Integrins and Vinculin structures (64). In addition, FAK signalling is reported to play integral roles in cancer cells survival where FAK upregulation has been associated with anoikis resistance, which is cell death induced by detachment. This occurs through FAK activating pathways like PI3K-AKT signalling leading to enhancement of cell survival upon detachment (65). Furthermore, TGF-β induces EMT via downregulating E-Cadherin which is reported to require FAK phosphorylation and activation (66)

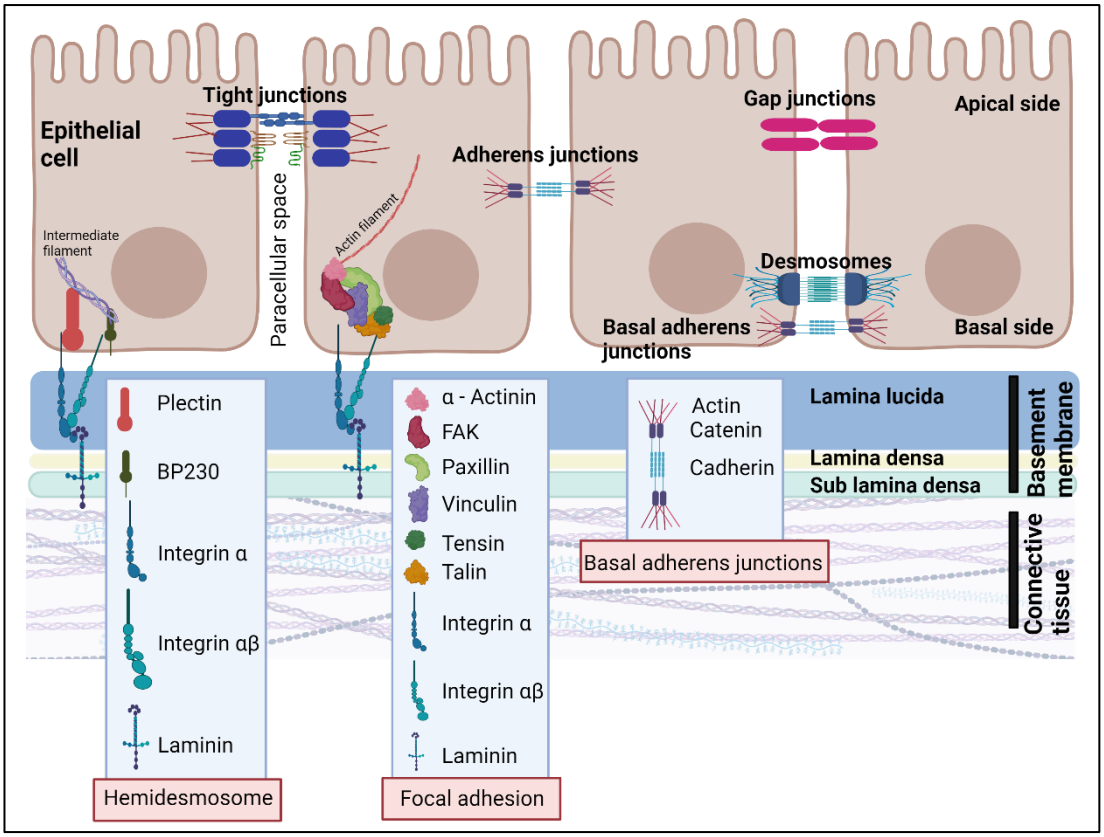


Figure 1-4 Epithelial Cells and Junctions.

Close and tight arrangements of epithelial cells is maintained in part by different types of junctional protein; tight junctions, adherens junctions and desmosomes (67, 68). Anchoring of epithelial cells on their basement membrane is maintained by basal junctions: hemidesmosomes, focal adhesion junctions, and basal adherens junctions (59). Created with BioRender.com

1.2.6.3.2 E-Cadherin

E-Cadherin is a calcium-dependent cell adhesion protein that facilitates the close and tight arrangements of epithelial cells over their basement membrane (69). Therefore, it is usually a strong antagonist for cell migration and invasion. When the gene encoding E-Cadherin (*CDH1*) is mutated, E-Cadherin levels, activity and regulation are affected. *CDH1* mutation is associated with several cancers, e.g. breast and gastric cancer (70). Loss of E-Cadherin allows cells to break free from their neighbours and to invade the basement membrane.

1.2.6.3.3 EMT

EMT is a process that was first described by Elizabeth Hay in her study of primitive streak formation (71). It is a reversible physiological process in which differentiated epithelial cells, that are usually laterally contacted and basally fixed to their basement membrane, lose both their cell-to-cell and cell-to-ECM contacts (72). There are three main types of EMT based on the underlying context. Type 1 EMT is the physiological form that is important during early embryonic development (73). Type 2 EMT is the outcome of tissue injuries, growth and fibrosis (73). Whereas type 3 EMT is involved in cancer progression and metastasis (74).

1.2.6.3.3.1 EMT in normal physiology

In normal physiology, **EMT type 1** is primarily involved in embryogenesis and tissue regeneration (75). During embryonic development, EMT is essential for gastrulation, which is the reorganisation of embryonic tissues into the three primary germ layers: ectoderm, endoderm, and mesoderm (75). EMT enables these layers to further develop and specialise into specific tissues with distinct functions. For example, the neuroepithelial cells of the neuro crest tube undergo EMT and migrate rostrally contributing to most of the head structure (76). **EMT type 2** is associated with inflammation and ceases when inflammation subsides (75). It is significantly important for wound healing where it stimulates cells motility and proliferation at the edge of wounds and contributes to tissue repair (75). In chronic inflammation, EMT is involved in the abnormal transformation of myofibroblasts into permanent fibrosis that further progresses to organ failure (77).

1.2.6.3.3.2 EMT in cancer

EMT type 3 is associated with neoblastic cells in cancer. They undergo EMT and gain features allowing them to invade and migrate into distant locations leading to cancer progression (75). The characteristics of EMT in cancer are non-coordinated in comparison to physiological EMT (78).

EMT in cancer is controlled by a complex network of several regulators including transforming growth factor-beta (TGF-β), Wnt/β-catenin, NOTCH, hedgehog, signal transducer and activator of transcription 3 (STAT3), bone morphogenetic protein (BMP), receptor tyrosine kinase (RTK), (ECM)-induced pathways, and hypoxia-induced pathways (79-81). EMT is also regulated by epigenetic modification which also influences the expression of the previous list of EMT regulators particularly TGF-β and by hypoxia (82).

EMT in cancer results from an alteration in the proteins holding cells together (mainly E-Cadherin but including other proteins such as N-Cadherin, Claudin-1 Zona Ocludin-1, B-Catenin, Snail, Slug, and Vimentin) (83). EMT involves the downregulation of several important epithelial markers e.g. E-cadherin (figure 1-5), claudin, occlusion and keratin, whereas mesenchymal markers such as Vimentin, N-cadherin and α-SMA undergo an upregulation of expression (84). This alteration in marker expression occurs gradually over time and is induced by specific EMT transcription factors (TFs) that are often aberrantly expressed in cancer cells e.g., Snail, Slug and zinc-finger E-box binding homeobox (Zeb1/2), and Twist (83, 85-87).

1.2.6.3.3.3 Cellular and tissue characterisation of EMT

Along with the change in expression of the EMT-TFs, morphological changes of the cells take place to become more elongated, lacking polarity, expressing different marker proteins, capable of synthesizing ECM components and also motile and migratory (72). Loss of cell-cell contact is one of the mechanisms induced by the alteration in epithelial and mesenchymal markers. For example, loss of E-Cadherin induces stress on the structure of adherens junctions linking the cells laterally. In response, β-Catenin translocates into the nucleus to further progress EMT events (88). Additionally, Actin filaments undergo changes in locations where they are in contact with focal adhesion junctions leading to cells detaching from the basement membrane (89). Tumour cells, therefore, become invasive and migratory. EMT also enable tumour cells to confer chemotherapy and radiotherapy resistance via enhancing efflux and DNA repair mechanism (90, 91).

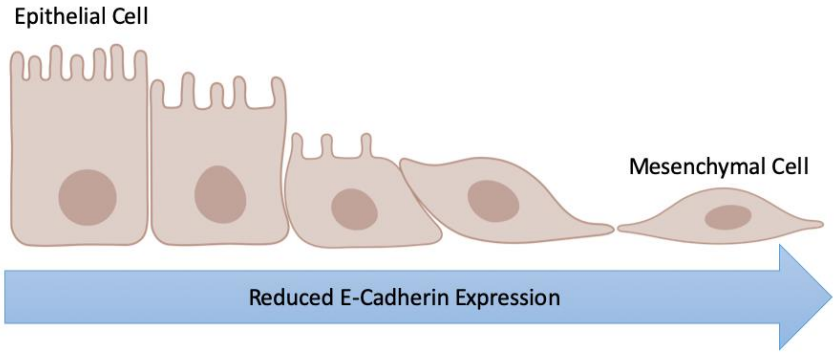


Figure 1-5 Epithelial to Mesenchymal Transition (EMT) (67, 68).
Created with BioRender.com

These EMT features can be characterised *in vitro* by investigating the changes in expression of epithelial and mesenchymal markers (e.g. RT-qPCR, Western

blotting, flowcytometry, and transcriptomic sequencing), assessing cell morphology (e.g. immunohistochemistry and immunofluorescence imaging), or performing functional assays (e.g. cell adhesion, cell invasion, cell migration, and wound healing) (75).

1.2.6.3.3.4 Partial EMT introduction

The reverse mechanism to EMT is called MET which has been reported to promote colonisation and metastatic niche formation (49). When cells are embracing both features as hybrid EMT-MET, it is referred to as partial EMT (92). The presence of a mixture of EMT-cells and MET- cells arising from epithelial-mesenchymal plasticity is seen more often associated with metastasis (93), immune evasion (94), and resistance to treatment (95)

1.2.6.3.3.4.1 Partial EMT phenotype

The co-expression of epithelial and mesenchymal markers are key elements in partial EMT. Early hybrid EMT states are characterised by expressed epithelial markers (E-Cadherin and /or Keratins) and mesenchymal markers (Vimentin) with the cells being more rounded as reported in several cancer cell lines (e.g. breast (96), pancreatic (97), and colorectal cancers (98)) (68) As the hybrid EMT advances towards the late stages, Vimentin marker increases accentuating the elongated morphology of the cells (68). (Further impacts of Vimentin expression in cancer and EMT is discussed in later sections). Additionally, cells in a state of partial EMT are reported to be more invasive (99), migratory (100), and with enhanced survival in suspension (101). In vitro

experiments to identify EMT markers and phenotypes can be applied in partial EMT as well.

1.2.6.4 Invasion

After EMT, cancer cells can invade the ECM and the surrounding tissues until they reach blood vessels. The two main components of ECM are the basement membrane to which epithelial cells are anchored, and the interstitial matrix responsible for providing elasticity to the tissues (9). One way a tumour cell can invade and migrate through the ECM is by pseudopod formation (102). Cancer mesenchymal cells release proteases (e.g. metalloproteinases (MMPs)) to free a space for their formed pseudopod. This pseudopod, then, will direct itself towards the created space to interact with ECM to which it adheres. Next, the cell undergoes cytoskeletal remodelling and contracts itself towards the leading pseudopod. Finally, the cell becomes free from adhesion with the ECM. This action is repeated as cells migrate either individually, or in groups (103). Another form of cancer cell motility for invasion is amoeboid movement; an EMT-independent process that lacks ECM degradation (104). Ameboid movement is associated with nuclear compression and distortion. It can also occur with cancer cells either individually (105), or in groups where the leading cells move depending on their pseudopods while the rest of the group relies on their interlinking connection with leading cells (106). Both pseudopod and amoeboid movements rely on having a leading edge of the cell that is responsible of initiating the motility and directing the movement towards chemoattractant, i.e. cell polarisation (107).

1.2.6.5 Intravasation

Intravasation is the process via which cancer cells gain access to the bloodstream through or between endothelial cells (108). To cross through the endothelial cell barrier, cancer cells follow a process similar to their invasion motility through the ECM. Intravasation can also happen through the lymphatic system and eventually into the blood circulation. The final drainage of lymph is into the venous system of the blood circulation.

1.2.6.6 Surviving the circulation

Once tumour cells are in the bloodstream, they are called circulating tumour cells (CTCs) (48). The path CTCs must travel depends on the anatomical part of the vascular system to which they intravasated. The site where CTCs enter the circulation can also determine the possible destination for their metastasis if they survived within the circulation. It is important to bear in mind that these CTCs are amenable to the body immune defence as well as removal by the lungs as tumour emboli (50). Additionally, CTCs have to avoid death from the stress forces generated by the turbulent blood flow and from the shearing pressure developed as blood is being pumped in the circulation (109). CTCs can be used for diagnosis and prognosis purposes (110).

1.2.6.7 Extravasation

Extravasation is defined as the re-entry of CTCs from a vessel interior into an organ parenchyma i.e., their potential secondary location. Due to the ongoing hurdles CTCs face while in the bloodstream, their rapid extravasation is important for their permanent survival. Extravasation is a multistep

mechanism. First, because CTCs tend to aggregate and form clusters while in the circulation, CTCs must break their intercellular adhesion and unfold themselves into single-file chain allowing their transportation across the blood vessels wall (111). Another mechanism through which CTCs can extravasate is by forming a CTC-platelet complexes. CTC-platelet interactions facilitates their adhesion to endothelial cells via purinoceptor 2 protein receptors (P2Y2) that in turn make endothelial cells more permeable (112). In addition, CTC-platelet interaction can also induce the release of growth factors and cytokines to assist with extravasation (113). CTCs were recently found to induce necrosis in endothelial cells via death receptor 6 (DR6) (114). Once CTCs successfully leave the circulation, they are known as disseminated tumour cells (DTCs).

1.2.6.8 Dormancy and secondary growth

The secondary location for CTCs is not random. For example, breast cancer often metastasises to certain organs in different patients, the other side of the breast, the liver, the brain and the bones. The earliest attempt to explain this was by Paget et al. in the hypothesis of seed and soil where he explained that CTCs can only be attracted and grown in tissues where they can thrive (115). In the secondary site, DTCs can become dormant cells, i.e. they will survive but not necessarily proliferate (116). However, they have the potential to reactivate at any time and proliferate forming a secondary tumour.

1.3 Gene expression and regulation in cancer

1.3.1 The central dogma of molecular biology and gene expression

The central dogma of molecular biology was first introduced in 1958 by Francis Crick as an explanation of the molecular interactions leading to protein synthesis (117). It simply stated that the genetic information stored in the DNA is conveyed through gene transcription into ribonucleic acid (RNA), and later this RNA is translated into proteins (117). It is fundamental for understanding molecular biology. It is also important to fathom the current understanding of gene expression and regulation.

Gene expression is defined by the national institute of human genome research (NIH) as “The process by which the information encoded in a gene is turned into a function.” (118). The flow of information from DNA into RNA is called gene transcription, and it occurs by the action of RNA polymerases (RNAP). The RNA that is produced in transcription can either lead to protein production via a translation process using RNA transcripts in the ribosomes or it can give rise to untranslated non-coding RNA (ncRNA) that serve other direct functions in the cell (9).

To understand how gene expression is regulated in cancer, the following sections will include a brief explanation of chromatin organisation and regulation, the components of a gene transcription complex, the process of gene transcription, and gene regulation in cancer.

1.3.2 Chromatin organisation

DNA was first reported to consist of a double-helix polymer strand by Watson and Crick in 1953 (119). The DNA is condensed as chromosomes at different levels of compactness (figure 1-6).

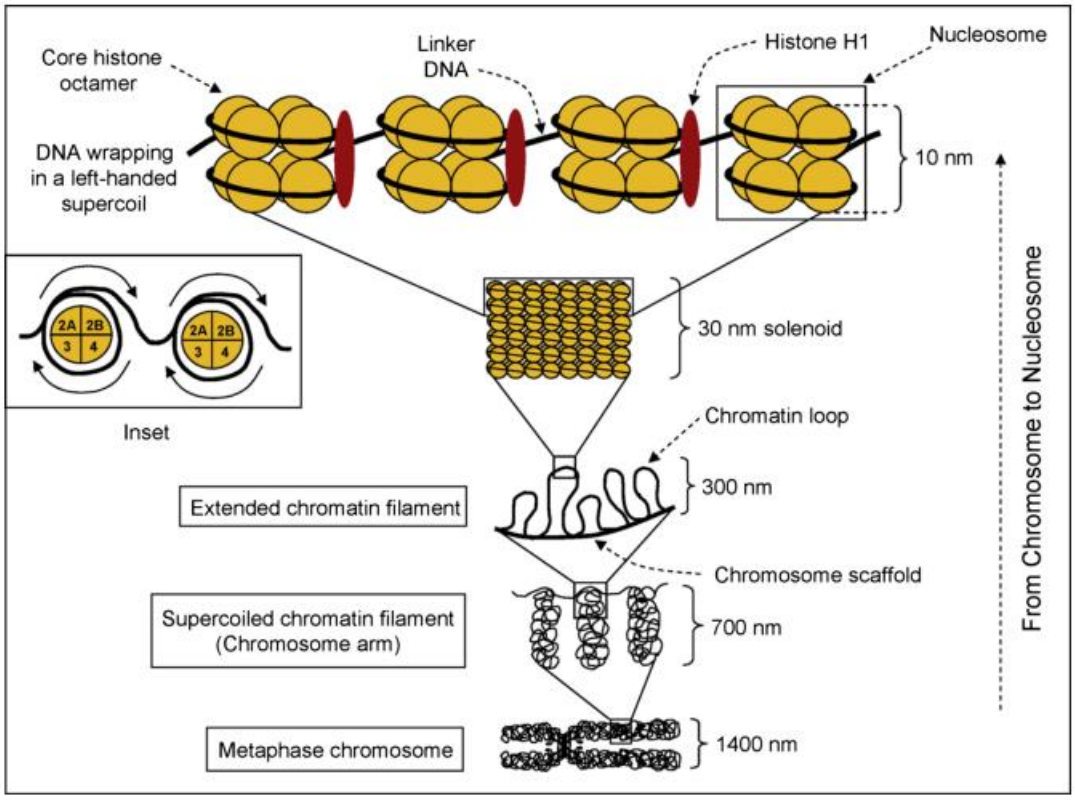


Figure 1-6 Organisation of DNA in chromatin (9, 120, 121).
Adopted from Fundamentals of genes and genome (122)

Chromosomes are composed of DNA double-helix strands coiling as chromatin. Chromatin consists of DNA helical fibres that are wrapped around proteins called histones and other proteins. Histones form octamers with four subunits proteins known as histone H2A, H2B, H3 and H4. There are two amino acids (a,a), lysine and arginine, on the histone proteins that are positively charged while the DNA has negatively charged phosphate groups.

Therefore, an attraction force brings DNA closer to wrap around histones. When DNA is wrapped twice around a histone tetramer, such an organisation gives the nucleosome (9, 120, 121). Nucleosomes are packaged into two forms of DNA chromatin. The first one is a highly condensed chromatin, or heterochromatin, which is also referred to as closed chromatin. H1 linkers are small proteins that link nucleosomes together. They have positive charge which helps in condensing chromatin. The second form chromatin, or euchromatin, is less tightly packaged and is also known as the open chromatin (9). This open form can be accessible to various transcription factors and regulatory proteins as will be explained later.

1.3.3 Chromatin remodelling

The dynamics of chromatin wrapping and unwrapping is necessary for gene expression. Unwrapping or loosening reduces the tightness of the chromatin, making spaces that allow the entry of various factors. This is also called chromatin remodelling. This remodelling can occur via histone modification, nucleosome modification, or protein binding.

There are several so-called ***epigenetic*** mechanisms that regulate histone modification including acetylation, methylation, ubiquitination, and phosphorylation (123). The most well-known being histone acetylation and methylation. Histone acetylation can relax chromatin condensation. This occurs upon the addition of an acetyl group by histone acetyl transferase (HAT) enzymes on the lysine or arginine residues in histones (124). This acetylation neutralises the positive charge of histones, causing the negatively charged DNA to repel from histones, and leads to chromatin de-condensation.

Acetylated histones can be recognised by a group of proteins known as Ac-Histone binding proteins including readers, writers, and erasers. **Readers** have motifs that recognise and bind Ac-Histones (125), **writers** are HAT enzymes that add acetyl groups to histones (124), and **erasers** reverse acetylation for example histone deacetylase enzymes (HDACs) which condense the chromatin again and inhibit transcription (126).

Histone methylation is a process where a methyl group is added to lysine and arginine residues on histones by writer enzymes such as lysine/arginine-specific-methyltransferase, and it is associated with both heterochromatin and euchromatin depending on how a residue is methylated (124, 127). For instance, histone di-methylation in histone 3 Lysine 4 (H3K4me2) can result in chromatin unwrapping and active gene expression (128). However, if multiple methyl groups are added e.g., Histone 3 Lysine 9 (H3K9me3) which includes a tri-methylation, the chromatin forms heterochromatin and gene expression is repressed (129). Histone remodelling can also be achieved via ***protein complex regulators*** such as the switching defective/sucrose non-fermenting (SWI/SNF) family of ATP-dependent modifying enzymes. SWI/SNF enzymes are able to apply twist and pull forces on the DNA. This combination of actions slides the DNA along nucleosomes, destroys contacts between DNA and histones, and repositions the nucleosome (130). Histone remodelling is also transcriptionally regulated by ***pioneer TFs***. These factors can scan the DNA and discover their binding motifs (short DNA target sites) even when they are in heterochromatin. Hepatonuclear factors are well characterised examples of pioneer TFs (131, 132). **Nucleosome** remodelling can also be brought about

via histone subunit alteration, position sliding or through nucleosome loss (133).

Overall, for gene expression to take place, heterochromatin needs to undergo chromatin remodelling to be accessible by the different components that are required to form what is called a transcription complex to initiate gene transcription.

1.3.4 Transcription complex components

Before going through the details of gene transcription, this section will briefly introduce the main components of the transcription complex. They are RNAP, gene promoters and other regulatory elements, transcription factors, and other regulatory proteins.

1.3.4.1 RNA Polymerases

They can be classified into RNAP I, II, and III. RNAPII can recognise and bind to a diverse collection of regulatory elements on gene promoters, while RNAPI and RNAPIII bind to less elements to produce non-coding RNA. RNAPII cannot bind to DNA directly, and it require the help of basal TFs to bind to promoter DNA (134, 135).

1.3.4.2 Gene promoters and their regulatory elements

Promoters are regions on the DNA, consisting of 100 – 1000 base pairs, that define where gene transcription begins. They can be located upstream of a gene transcription starting site (TSS), or downstream of a TSS, or overlapping with a gene coding region on the DNA (136). The scheme in (figure 1-7) shows a scheme of a typical eukaryotic gene promoter with the different regulatory

elements that can be found within the promoter and have sequences for the binding of TFs (137).

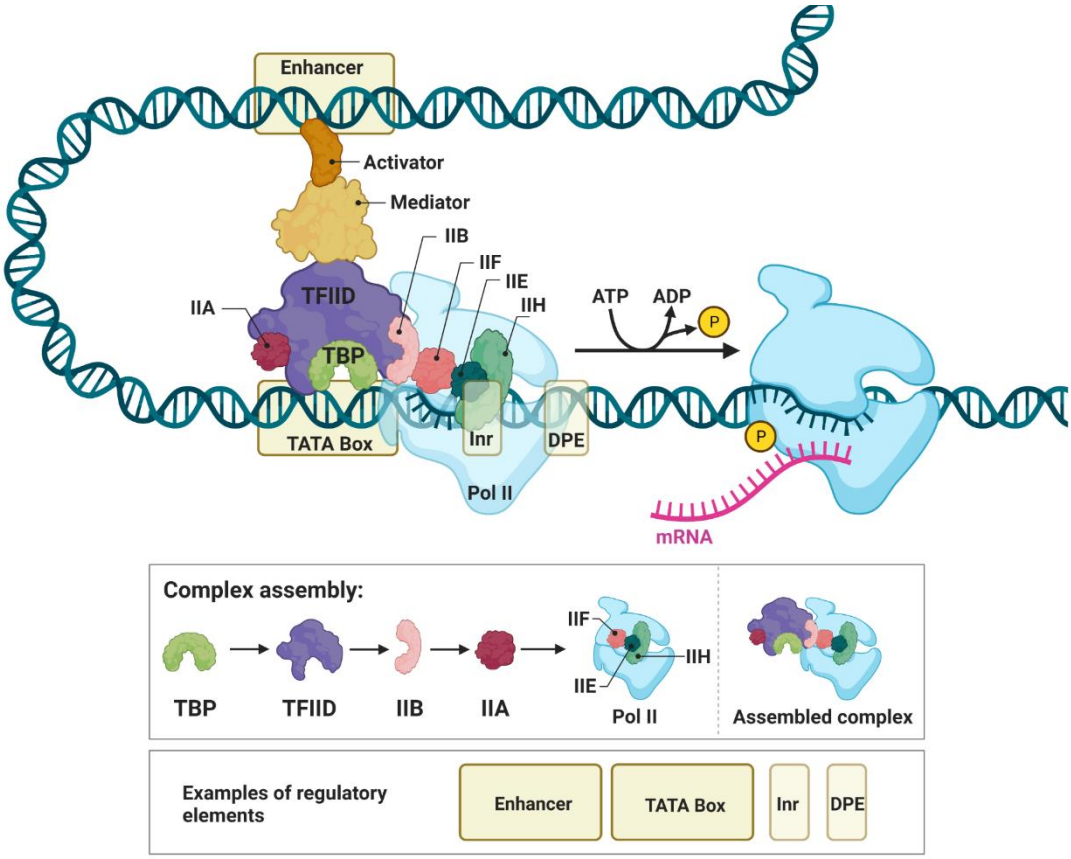


Figure 1-7 Transcription complex at the promoter region with some of the regulatory elements (137).

Created with BioRender.com

The most common promoter element is the TATA-box. It is characterised by repeated T and A nucleotides that have 2 hydrogen bonds in between making them easier to be pulled apart for transcription. The TATA box is a binding sequence for TFIID, specifically the TATA box binding protein (TBP) subunit of TFIID, that in turn binds to RNAPII (138, 139). Other regulatory DNA elements can bind TFs to regulate gene expression in certain tissues, thereby called tissue-specific regulatory elements. Although these elements are often

located in promoter regions, they are also found in enhancer regions (140). If a promoter element is overlapping with the TSS and the gene coding area, it is referred to as an initiator (Inr) motif. Inr are usually bound to TFIID (141).

In addition to the promoter regulatory elements, there are also other distal regulatory regions on the DNA, found farther upstream or downstream from the TSS called enhancers. These regions can also bind to TFs . Upon binding, these TFs are brought closer to the promoter via DNA looping. When it loops closer to the promoter, the enhancer can bind to other TF or mediator proteins on the promoter forming a transcription initiation complex. This complex in turn facilitates the binding of RNAP to the promoter to start gene transcription and enhance gene expression (137).

1.3.4.3 Transcription factors

Tissue-specific TF can be simply defined as proteins with a sequence-specific DNA-binding ability that can regulate gene transcription. Their DNA-binding domains can recognise certain regulatory elements on the promoter region or enhancers of the target gene (142). When TFs bind to DNA, they can regulate multiple processes in the cell. In addition, TF can often act in conjunction with non-DNA-binding co-repressors and co-activators, via their other domains, to bring about changes in gene expression. As mentioned earlier, some TF can reach their motifs on heterochromatin DNA. Such TF are called pioneer TF (131). In contrast basal TF recruit RNAPII to DNA to produce mRNA at all genes transcribed by this RNA polymerase and are classified, according to the nuclear extract fraction from which they were eluted upon discovery, into TFIIA, TFIIB, TFIIE, TFIIF, TFIIH, and TFIID (143).

1.3.4.4 Regulatory proteins

Regulatory proteins include the various co-activators and co-repressors that can interact with DNA-bound TFs to regulate gene expression. Generally, the role of the co-regulatory proteins is to recruit chromatin remodelling enzymes or chromatin-binding proteins to modify the structural configurations of chromatin (144).

1.3.5 Transcriptional complex formation and gene transcription

To regulate gene expression, it is required to expose gene promoters through chromatin opening. This is facilitated by the action of three components: specific DNA regions e.g., CpG islands and nucleosome depleted regions (NDR), pioneer TFs, and enhancers. CpG islands are regions on the DNA characterised by repeated and dispersed CpG dinucleotides and the lack of TATA box elements (145). DNA regions rich in CpG and NDR attracts HATs and demethylases which results in loosening of the chromatin (145, 146). (142). Enhancers, which are DNA sequences usually located within 10 – 100 kb from the promoters, provide an indirect mechanism to open the chromatin. They can recruit large amounts of TFs, and pioneer TFs, and histone modifiers. Enhancers are brought closer to the promoter by the action of various DNA looping proteins, thus increasing the concentration of TF at the promoter regions (137). Gene transcription by basal TF is achieved by forming pre-initiator complex formation (PIC). PIC is assembled by the recruitment of general TF and RNAPII at the promoter site (147). The first approaching TF are TFIID, which are composed of TBP and TBP associated factors (TAF). This is followed by the accumulation of other types of TFs and RNAPII. RNAPII

starts the process of DNA transcription aided by other factors like capping enzymes and elongation factors. Together, they help stabilise the newly transcribed RNA for elongation to commence (147). Transcription termination is RNAPII dependent and can be achieved by three pathways: cleavage of the mRNA, polyadenylation of mRNA at 3' end, and disassembly of RNAPII (148).

1.3.6 Gene regulation in cancer

Gene transcription and expression regulation is tightly maintained through a harmonious work of various players including epigenetic modifiers, TFs, regulatory elements, and RNAPII. Profound gene dysregulation contributes to the onset and progression of various diseases including cancer (149). Significant alteration in chromatin organisation and remodelling has a great impact on the dynamics of gene expression. Particularly, the aberrant functioning of histone modifiers may result in unwanted chromatin condensation preventing TFs and other regulatory protein from reaching their target motifs (gene silencing) or loosening other DNA regions making them accessible to regulators (gene activation) (124, 127). Gene dysregulation or mutation might affect oncogenes or TSG that directly disturb gene expression in cancer. They can also influence the expression of genes crucial for DNA repair, cell cycle progression, or apoptosis. This may influence the cellular homeostasis and lead to uncontrolled proliferation and persistence of cells carrying DNA mutations (149). Additionally, cancer may also cause alteration in signalling pathways that can impact the activity or expression of various key TFs linked to gene expression (150). Moreover, several super-enhancers in

cancer are connected to the activation of genes associated with tumorigenesis, e.g., the *Myc* oncogene (151).

1.4 Biliary tract and cholangiocarcinoma

1.4.1 Biliary System

1.4.1.1 Anatomy of the biliary system

The biliary system is part of the gastrointestinal tract located in close relation to the liver. It consists of two main components: the gall bladder and the bile ducts. The major functions carried out by this system are related to bile synthesis, storage and delivery into the duodenum where bile plays an important role in the digestion of dietary fats (60).

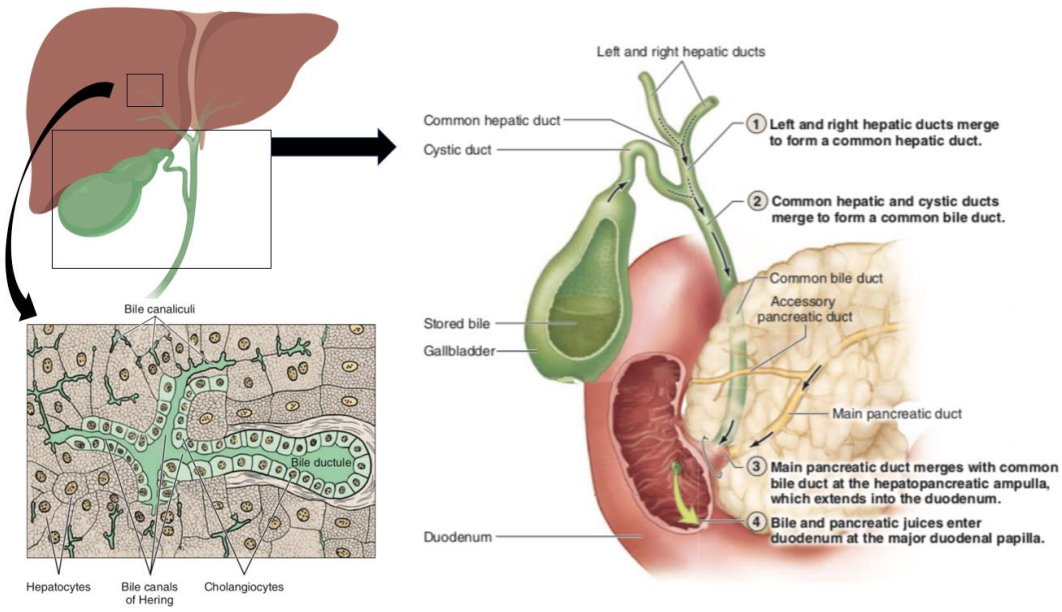


Figure 1-8 Biliary System.

The biliary system is composed of the gall bladder and bile ducts. The bile duct tree starts within the parenchyma of the liver and directs bile towards the gall bladder or the duodenum. Along the way, bile ducts keep converging until, finally, a single common bile duct drains into the duodenum. Source: Junqueira's Basic Histology Text and Atlas (60).

As seen in (figure 1-8), **the gall bladder** is a pear-shaped hollow organ that is located just below the liver. It serves as a storage site for diluted bile produced by hepatocytes, and it also concentrates the bile via water reabsorption. **The**

bile duct tree of the biliary system is a branching tree distributed inside and outside the liver parenchyma (the functional tissues in an organ that carry out specific tasks). Tracing the bile ducts as they deliver the bile and other secretions into the duodenum will also show the size-wise classification. The smallest branches are thin tubes known as bile canaliculi located between hepatocyte lobules. These canaliculi collect the bile synthesised by hepatocytes and drain it into the short canals of Hering. From there, bile drains into the intrahepatic bile ductules. These ductules in turn direct the bile into bigger interlobular ducts. Interlobular ducts from both lobes of the liver merge to form right and left hepatic ducts. These two ducts merge outside the liver hilum (point of entry and exit of liver vessels, ducts, and nerves) to make the main hepatic duct. Then, this hepatic duct is joined by the cystic duct of the gall bladder and together they form the common bile duct (CBD). This point of connection is bidirectional, towards the gall bladder for bile storage and towards the intestine for secretion and excretion. Finally, CBD will eventually open into the duodenum through the ampulla of Vater delivering bile to its site of action.

1.4.1.2 Development of the biliary system

During the early development of the gastrointestinal tract, the primitive gut forms a blind-ended tube divided into three parts: the foregut, mid gut, and hind gut. From the foregut, a small endodermal outgrowth is developed which is called a bud. This bud consists of rapidly proliferating cells and is later separated into liver and biliary buds. The liver bud cells have the ability to differentiate into hepatocytes or ductal cells, while the biliary bud cells

differentiate into the bile duct, gall bladder and cystic duct. This embryologic finding explains the heterogeneity in cellular makeup of liver and bile duct cancers (152-154). Bile duct cancer can originate from either hepatic progenitor cells or from biliary tree cells (155). The presence of cancer stem cells, capable of differentiating into multiple cell types within the tumour growth, contributes to the diversity of cholangiocarcinoma (CCA) cell populations.

1.4.1.3 Normal cholangiocytes

The wall of the hepatic ducts along with the cystic and CBD is made of four layers: mucosa, lamina propria, submucosa and muscularis layer. The lining mucosa is made up of simple epithelial cells called cholangiocytes (60). These cholangiocytes range in morphology and size from small cuboidal to tall columnar as the size of the ducts increases. Under normal conditions, these cells perform a range of important roles including the synthesis of several components of bile juices. Bile juice is a yellowish-green solution consisting majorly of water. Other constituents in this solution are bile salts, bile acids, bilirubin, cholesterol, fatty acids, vitamins, and salts. The bile acid portion is synthesised by the lining of the biliary ducts. In addition to their secretory function, cholangiocytes are also capable of reabsorbing water. This diversity in morphology and properties, under stressful conditions, produces several diseases in the biliary system including the malignant CCA (156, 157).

1.4.2 Cholangiocarcinoma

1.4.2.1 Background and classification

CCA is a heterogenous group of malignancies that can arise anywhere in the biliary tree (158). CCA is classified anatomically based on its location in relation to the liver (figure 1-9). When the tumour is developed within the liver parenchyma, it is called intrahepatic CCA. If it arises in either of the main hepatic duct adjacent to the hilum of the liver, it is called perihilar CCA (159, 160). A clinical term for perihilar CCA in the converging point of the two ducts is Klatskin tumours (161). When it is found in CBD, CCA is classified as extrahepatic(159, 160). Around 50% of CCA is perihilar, while extrahepatic tumours account for approximately 40% of CCA, and finally 10% of CCA are intrahepatic tumours (162).

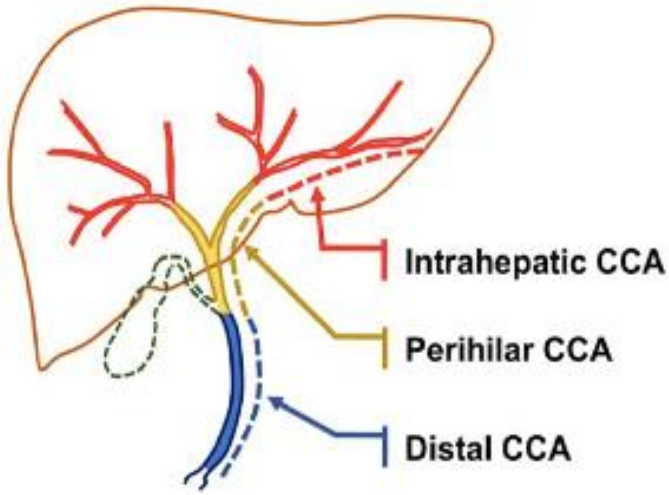


Figure 1-9 CCA Anatomical Classification.

CCA is classified for surgical purposes based on its location in the biliary tree. Intrahepatic if it buried in the liver tissues. Perihilar when it is in the main right and left hepatic ducts. Extrahepatic if develops in the CBD. Source: Kendall et al., 2019 (163).

1.4.2.2 CCA Epidemiology, aetiologies and risk factors

CCA is relatively rare. In the UK, 2.17 cases were registered per 100,000 of the population at risk. Whereas, in the northeast regions of Thailand 85 cases per 100,000 of the population at risk were reported and this is considered the highest prevalence in the world (158) (figure 1-10).

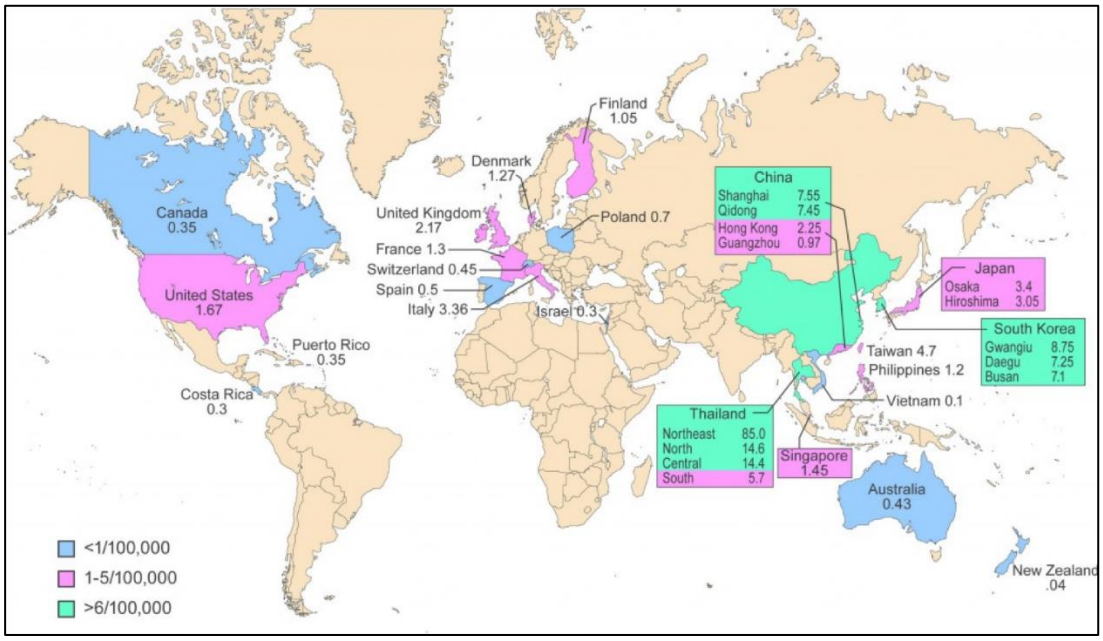


Figure 1-10 Geographic distribution of CCA.

The map shows the number of reported cases per 100,000 of the population. Source: RARECARE project.

The geographical difference in CCA is also denoted by its risk factors. The causative aetiologies appear to vary among populations. Infestation from liver flukes e.g. *Opistorchis viverrini* and *Clonorchis sinensis* tops the list of the risk factors in Thailand and the rest of Southeast Asia (164-166). In 1994 both parasites were declared carcinogenic for human by IARC (167).

Liver flukes are acquired from ingesting raw koi pla fish carrying the adult form of the parasites which initiate and sustain a chronic type of inflammation in the

biliary tract epithelium (168). However, in Western countries liver flukes are very uncommon and 6% to 36% of the diagnosed CCA patients are found to have primary sclerosing cholangitis (PSC), which is now considered the main factor underlying the pathogenesis of CCA in the West (169, 170). PSC is an autoimmune disease that affects the biliary lining epithelium. PSC produces chronic inflammation of cholangiocytes that can progress to cause a narrowing in the pathway of bile flow leading to a decrease in bile flow that progresses to stagnation, i.e. cholestasis (171).

The list of other hepatobiliary diseases that could potentially lead to CCA include primary hepatobiliary diseases e.g. cholelithiasis (172), hepatolithiasis and cholecystitis (173), chronic viral and non-viral liver diseases (164, 174-176), congenital fibro polycystic liver diseases (177, 178), and genetic disorders e.g. cystic fibrosis (179). Other risk factors for CCA include metabolic syndrome (180), hyperglycaemia, and obesity (181, 182).

1.4.2.3 Pathogenesis

Chronic inflammation has been found to induce the molecular aberrations behind the development of CCA in animal models (183). In human CCA, chronic inflammatory injuries induced by chronic cholestasis are believed to lead to the development of precancerous lesions. Inflammation subjects the cells to stress conditions that trigger epigenetic changes resulting in altered DNA methylation. This results in changes the expression of genes potentially activating cell survival signalling (184).

1.4.2.4 Clinical presentation and diagnostic work up

CCA is a clinical challenge as it does not produce a specific clinical picture. The earliest signs, symptoms, and physical findings in CCA patients are so common that the differential diagnoses would not have CCA at the top of the list. In most of the cases, CCA is asymptomatic until the appearance of jaundice, that is a yellow discoloration of the skin and mucous membrane. Jaundice occurs when the flow of bile within the bile ducts is obstructed often late in the disease (185, 186).

No specific laboratory tests are available to investigate CCA. Instead, patients presenting with the previously mentioned clinical picture would go through routine blood tests, biochemistry, liver enzymes and liver function tests. Biomarkers including Carbohydrate antigen 19-9 (CA 19-9), carcinoembryonic antigen (CEA) and Alpha-fetoprotein (AFP) can be elevated however these are also nonspecific. Contrast based imaging techniques (e.g., ultrasound, computing tomography (CT) scans, magnetic resonance imaging (MRI) and endoscopic retrograde cholangiopancreatography (ERCP) are used to confirm CCA diagnosis and to acquire biopsies for histopathological studies. Imaging, histopathology and molecular profiling are considered the best options for CCA diagnosis (187).

Overall, the prevalence of common presentation accompanied by a cycle of non-determining routine laboratory investigations usually trigger the need for complexed imaging and invasive biopsy. This is why CCA is often discovered late and incidentally in clinics (188, 189). There are active efforts to introduce non-invasive diagnostic markers promising early detection, but these are

awaiting validation. Recently, biliary next-generation sequencing were recommended to be taken forward for further research (190).

1.4.2.5 CCA TNM staging

Tumour-node-metastasis (TNM) staging system for CCA is approved by the American Joint Cancer Committee/Union for International Cancer Control (AJCC/UICC). TNM staging is made separately for each type of CCA (191, 192).

1.4.2.6 CCA treatment

On a global scale, complete surgical resection with negative margins is the best option for resectable tumours, but even with surgery there is a high recurrence rate of around 50% one-year post operation (193, 194). Unfortunately, since the presentation for CCA is late, most of the diagnosed cases are inoperable. CCA shows resistance to both chemotherapies and radiotherapies.

The most recent updates on CCA diagnosis and management in the UK were commissioned in September 2023 by the British society of Gastroenterology Liver Section (190). Complete surgical resection remains the most curative option for resectable CCA in operable patients (190). Liver transplantation is evaluated based on tumour size and status of extrahepatic diseases (190). Palliative jaundice drainage is the recommended management for inoperable CCA with endoscopic ultrasound/ERCP, or ERCP alone (190). A combination of Cisplatin and Gemcitabine (CisGem) is the first line chemotherapy with/without immunotherapy (when available and approved) (190). In

November 2023, Durvalumab (Imfinzi) immunotherapy has been added as a first option along CisGim. Several targeted therapies are under clinical trials (195).

1.5 Proline Rich Homeodomain / Haematopoietically Expressed Homeobox Protein (PRH/HHEX)

1.5.1 PRH background and discovery

Proline rich homeodomain protein or haematopoietically expressed homeobox protein (PRH/HHEX) is a TF with an estimated molecular mass of 30 kDa (196-198). The protein was first identified by Crompton et al. in a study conducted to characterise the biochemical functions of homeodomain proteins in the haematopoietic system of chicken (199). The same study revealed that 20% of the a.a. in the PRH N-terminal domain consisted of proline residues in addition to the presence of two conserved and consecutive proline residues in the homeodomain, hence the name PRH (199). PRH is a highly conserved protein and the PRH homeodomain sequence shows a high percentage of a.a. similarity to those of chicken (75%), murine (94%) (199, 200).

The PRH protein is encoded by the gene *HHEX* (199-201). This gene was referred to as an orphan gene because it is found on chromosome 10, and not within the chromosomal clusters containing many of the other human homeobox genes (201). In mice, however, *HEX* gene is located on chromosome 19 (202). Morgutti et al. characterised in detail the genomic organisation of human *HHEX* gene by comparing it to the mouse *hex* gene *(203).* They mapped human *HHEX specifically* via FISH analysis to chromosome 10q24. In addition, they reported that *HHEX* spans around 5.7 kb of genomic DNA over 4 exons (203). Morgutti et al. also indicated a high

sequence homology in regions around the TSS including TATA box, 5' UTR and 3' UTR regions required for full HHEX promoter activity (203).

Upon the initial discovery of PRH, Crompton et al. predicted the potential gene transcription activity of PRH because they identified it as a homeodomain protein encoded by a homeobox gene. PRH binds to DNA via its homeodomain to regulate genes important for vertebrate cell differentiation and development (199, 201).

1.5.2 Regulation of PRH gene expression

1.5.2.1 Signalling pathway regulation

Early during embryonic development several signalling pathways modulate the activity of TFs that bind and regulate PRH promoter. Wnt/β-catenin signalling pathway regulates the expression of PRH which is essential to mediate liver and pancreas formation (204, 205). Wnt/β-catenin and TGF-β pathways coordinate PRH expression which is required for the early specification of the anterior endomesoderm (204). Several pathways such as TGF-β / through BMP (206-208), and BMP/ fibroblast growth factor (FGF) (207, 208) are also reported to regulate PRH expression and to play an important role in organ development (209).

1.5.2.2 Transcription factor regulation

Several TFs can regulate the expression of PRH in a cell-type specific manner by recognising and binding specific motifs in the intronic regions or 5' flanking regions of the promoter. In hematopoietic cells, TFs such as GATA-1, GATA-2, and cMyb bind to enhancer elements located in the first intron of *HHEX*

gene (210). Other TFs belonging to Sp family of TFs, namely Sp1 and Sp3, bind to GC boxes located in the 5' flanking region of *HHEX* promoter in MH_1C_1 and K562 cells (211). Thyroid-specific transcription factor (TTF-1) (212), and Pax8 (213) can bind and activate the *HHEX* promoter in thyroid follicular cells. In addition, PRH can exert a positive feedback expression on its own promoter (212, 214). PRH also forms a positive feedback expression loop with Notch3 (215). Other TFs capable of activating *HHEX* promoter include HNF3β and GATA-4 in HepG2 liver cells (216).

1.5.3 PRH expression during embryonic development and in adults

PRH, like other members of homeodomain protein family, has a spatial and temporal expression pattern that is crucial for the expression of key genes during embryonic development (217, 218). PRH is expressed and involved in the differentiation of various cells into vital organs during embryogenesis such as the brain, heart, pancreas, and liver (219-222). Disruption of PRH expression and function during embryogenesis is lethal. PRH knockout was found to be life incompatible as mice under study were suffering defective organogenesis in the brain, blood, and circulatory systems (219, 223). The liver was particularly absent in the abdominal cavity of $PRH^{-/-}$ mice (219) (figure 1-11). In addition, absence of PRH in the hepatic tissues was associated with loss of gall bladder and extrahepatic bile ducts indicating the importance of PRH for hepatic post developmental differentiation (224).

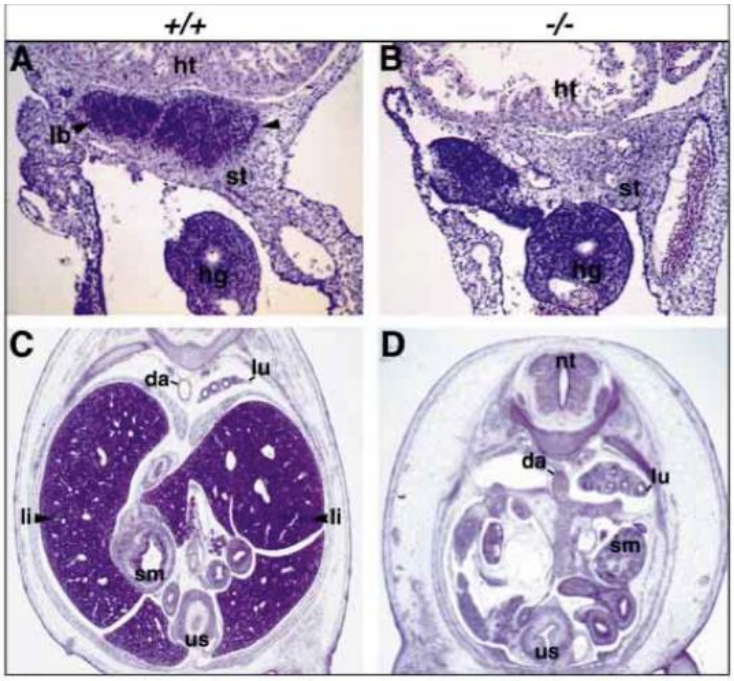


Figure 1-11 PRH $^{-/-}$ mice failed to develop their liver as indicated by its absence in the abdominal cavity (219)**.**

In adults, RNA and protein analysis revealed that PRH is widely expressed in organs such as the thyroid, liver and other tissues (219, 223) and, it is essential for the differentiation of hematopoietic cells (225, 226).

1.5.4 PRH protein structure

A monomer of human PRH is composed of 270 a.a. with a molecular mass of 30 kDa. However, it appears to have a molecular mass of 37 kDa when run on sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (227). As shown in (figure 1-12-A), the PRH monomer has three functional domains by which it affects different target genes: an N-terminal domain, a central homeodomain, and a C-terminal domain (228-230).

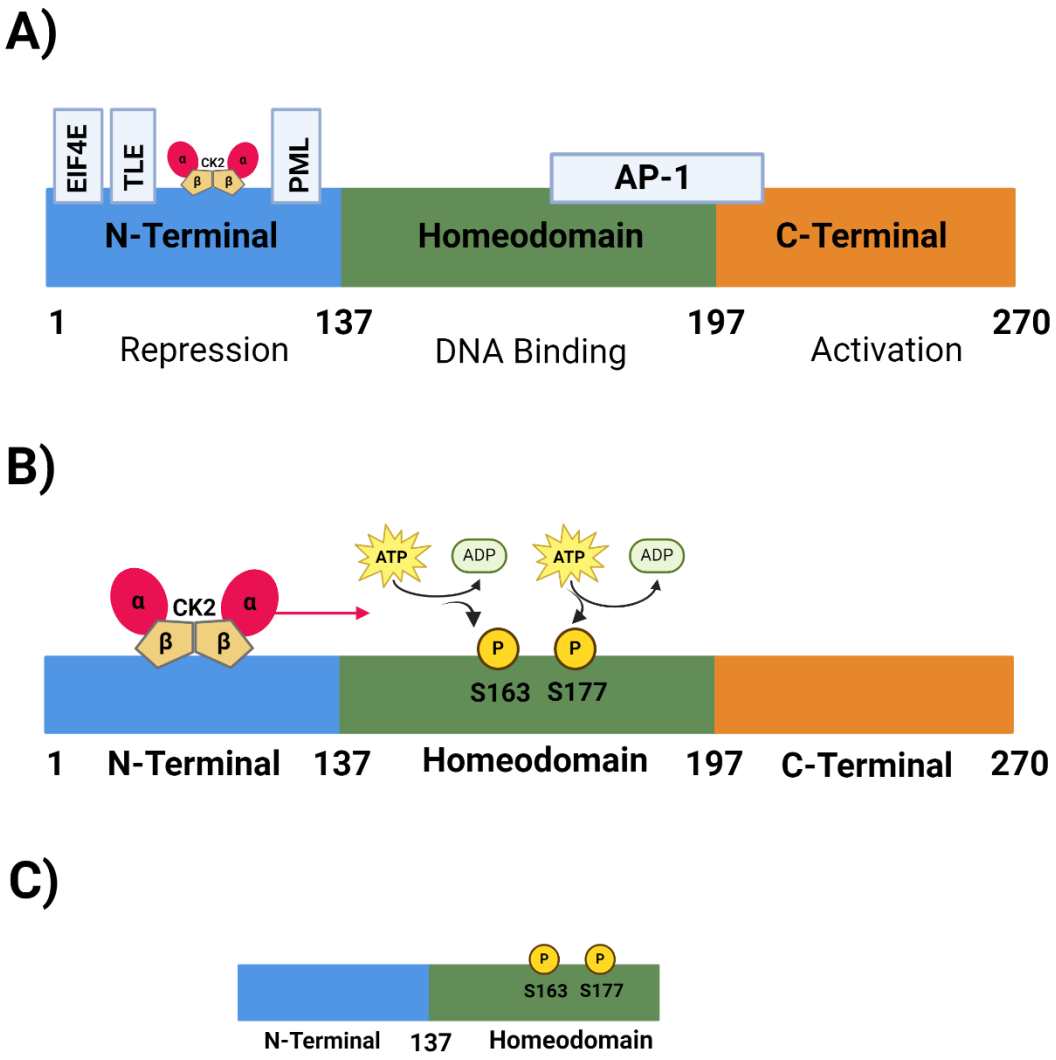


Figure 1-12: PRH monomer.

(A)The three domains of PRH: N-terminal repressing domain, DNA binding domain and C-terminal activating domain. (B) PMT of PRH by CK2 phosphorylation. (C) Truncated forms of PRH – phosphorylated homeodomains. Source: Adopted and modified from Gaston et al., Cell Biosci (2016)**(231)**.

1.5.4.1 The N-terminal domain

PRH N-terminal domain is also known as the proline-rich domain since it consists of 136 a.a. [1-136] where 20% of those residues are proline (199). Functionally, this region is referred to as the repression domain because via several protein-protein interactions (PPI) it results in transcriptional repression

(199, 232). Examples of those interacting proteins include eukaryotic translation initiation factor 4E (eIF4E) to inhibit cell proliferation (233), members of the groucho/transducin-like enhancer (TLE) co-repressor family (234), the ring domain of promyelocytic leukaemia (PML) protein (227, 233), and the β subunit of protein kinase CK2 (228). However, the PRH N-terminal region also interacts with HC8 proteasome subunit (235). The absence of the N-terminal domain in PRH produces a truncated form of PRH that can still bind to DNA, but fails to repress transcription (236). Moreover, this domain helps with forming higher-order arrangements of PRH proteins, a process known as PRH oligomerisation (196).

1.5.4.2 The homeodomain

This domain consists of 60 a.a. [137-196] and characterised by the presence of a helix-loop-helix motif (229). It recognises and binds to specific DNA sequences (5'-CAAG- 3' and 5'-ATTAA-3') on its target genes, hence it is also referred to as the DNA-binding domain (197, 237). This homeodomain-DNA binding is sometimes coupled with N-terminal domain PPI to regulate target genes (196, 197). Pellizari et al reported that this sort of mechanism is required for PRH to bind Thyroglobulin promoter, regulate its expression, and affect thyroid differentiation (237). Other examples of PRH interaction with other proteins via this domain include JUN (238), and HNF1α (239),

1.5.4.3 The C-terminal (activation) domain

The C-terminal region of PRH is made up of 74 a.a. [197-270] is an acidic domain. It is the transcriptional activating part of the PRH protein regardless

of cell type (230). It can bind to other co-activator and TFs e.g. serum response factor (240). In addition, both PRH acidic domain and homeodomain are required to activate sodium-dependent bile acid transporter (NTCP) gene in hepatocytes (230), and hepatocyte nuclear factor 1α (HNF1α) (239). Without this activating domain, NTCP is negatively impacted (230).

1.5.5 PRH oligomerisation and high-order assembly

Soufi et al. were first to demonstrate PRH oligomerisation via its N-terminal domain *in vitro* as well as confirming this via cross-linking *in vivo* (196). In their work, they saw PRH elute as a 250 kDa form from a gel filtration column. They also revealed via analytical ultracentrifugation that PRH can exist at 280 kDa and larger complexes. With cross-linking in cells, they reported that PRH is oligomeric (196). They proposed that PRH N-terminal domain has a novel dimerising motif [1-46 a.a.] that enables PRH intermolecular interactions to appear in different assemblies ranging from dimers to double octamers (196).

Soufi et al. proposed a model to explain PRH octamer formation. They suggested considering the full length PRH to comprise three parts; P, R, and H. “P” which represent PRH dimerisation region can interact with another “P” forming a dimer. In addition, the other two parts are “R” which is the repression region, and “H” the homeodomain and C-terminal together. Parts “R” and “H” within a certain dimer can interact with other “R” and “H” located on another dimer, therefore, forming a tetramer. The free “R” and “H” in this tetramer can in turn interact with other “R” and “H” on another similarly formed tetramer resulting in an octamer PRH (196). Furthermore, the authors used linear dichroism and electron microscopy (EM) to indicate that PRH octamers appear

as spheres that self-associate to form larger particles, or hexadecameric spheres (198). In fact, octameric and hexadecameric PRH were shown by dynamic light scattering to be the dominant species *in vitro* (198). These structures are very stable capable of surviving heat-induced and chemical denaturation (241).

Nuclear PRH complexes can self-associate and accumulate more PRH domains leading to DNA compaction. Such configurations allow PRH to bind tandem arrays of PRH-binding sites on target genes, therefore, providing high affinity bindings to their target genes (196, 198). For example, PRH oligomers can spread along multiple DNA-binding sites on *Goosecoid* (*GSC*)promoter to modulate chromatin structures and repress gene expression (196, 197).

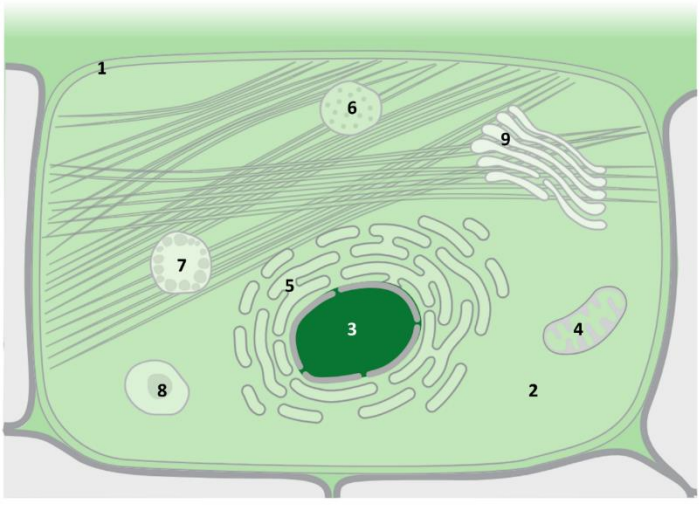
1.5.6 PRH post translational modification by CK2 phosphorylation

Protein Kinase CK2 (Casein Kinase 2 or CK2) is an enzyme comprised of four subunits: two catalytic α subunits and two regulatory β subunits. CK2 is known to affect the survival of cells through controlling progression of the cell cycle (242). CK2, by its β subunit, can interact with PRH while CK2 α subunit phosphorylate S163 and S177 of PRH homeodomain to deactivate its transcriptional activity and send it for cleavage by the proteasome (figure 1-12-B and C). This phosphorylation and processing lead to the formation of a truncated form of PRH known as PRHΔC. PRHΔC can suppress the repression activity of other full-length PRH (PRH octamers) by sequestering the co-repressor TLE1 and most likely via other mechanisms. (197, 228, 234).

Phosphorylated PRH interacts with TEAD4 in colorectal cancer to modulate YAP/TEAD and promotes tumourigenesis (243).

1.5.7 PRH subcellular distribution

The database COMPARTMENT is developed by Binder et al. which integrates findings from curated literature that used high throughput microscopy, findings from protein localisation proteomic databases, and text mining to provide information on protein subcellular localisation (244). According to the database COMPARTMENT, PRH is mainly a nuclear protein with lesser abundance in the cytosol and in various cytosolic organelles such as the mitochondria and the lysosomes (figure 1-13) (244). Nuclear PRH (dark green) has the strongest evidence since knowledge from curated literature were referenced. Several microscopic imaging techniques helped revealing this distribution including immunofluorescence (196), immunohistochemistry (245), and subcellular fractionation experiments (246).



No.	Location	Data Mining
1	Plasma membrane	Text
2	Cytosol	Text
3	Nucleus	Knowledge – Text – PSORT - YLoc
4	Mitochondrion	Text
5	Endoplasmic reticulum	Text
6	Lysosome	Text
7	Endosome	Text
8	Peroxisome	Text
9	Golgi apparatus	Text

Figure 1-13 PRH subcellular distribution according to COMPARTMENT.

Adopted from COMPARTMENT (244)

PRH spatial distribution has been reviewed in the literature. It is context dependent, and the redistribution of PRH localisation has been found to be causally linked to changes in its function, specifically in relation to cancer development (231, 247).

1.5.7.1 Nuclear PRH

PRH in the nucleoplasm, as mentioned earlier, is a TF that interacts via its homeodomain with DNA motifs to regulate gene expression and control tissue differentiation (196, 241). Ectopic PRH within nuclear bodies was reported to be partially co-localising with eIF4E and PML bodies to inhibit eIF4E transport in U937 and K562 myeloid cells to repress cell proliferation (233).

1.5.7.2 Cytoplasmic PRH

Normal breast (214), and thyroid (245) tissues exhibit nucleocytoplasmic distribution of PRH on immunohistochemistry. Ghosh et al. conducted the first experiment of PRH subcellular localisation in COS and MLE15 cell lines and mouse embryo (248). They reported PRH being primarily cytoplasmic in the definitive endoderm cells of mice embryo (7.5 days) using immunoblotting and immunofluorescence staining (248). Cytoplasmic compartmentalisation of PRH is a mechanism to inhibit its nuclear functions on target genes, but it may have other additional functions (245, 249). Several reports showed that a shift from nuclear to cytoplasmic localisation is implicated in the neoplastic transformation of increased cell proliferation and migration in breast (214, 249), thyroid (245), and haematopoietic cancers (250).

1.5.7.3 Nucleocytoplasmic trafficking of PRH

One way to illustrate the shuttling mechanism of PRH in cells was carried out by Ploski et al. (251). They used digitonin permeabilised cells to shown that PRH is one of many import cargos for Karyopherin/Importin 7 (KAP7) which is a structural protein in the cell. KAP7 mediates the transport of PRH into the nuclear compartment through direct protein binding. KAP7 depletion prevented the accumulation of PRH in the nucleus of HEK293 cells (251).

1.5.7.4 PRH subcellular localisation on immunofluorescence imaging

Immunofluorescence imaging is one of many methods applied to show PRH spatial distribution in cells. The Human Atlas provides reference images of PRH subcellular localisation in three different cell lines: HEL (an erythroleukemia cells). K562 (lymphoblast cells from myelogenous leukaemia), and U2OS (osteosarcoma cells) (252). (figure 1-14)

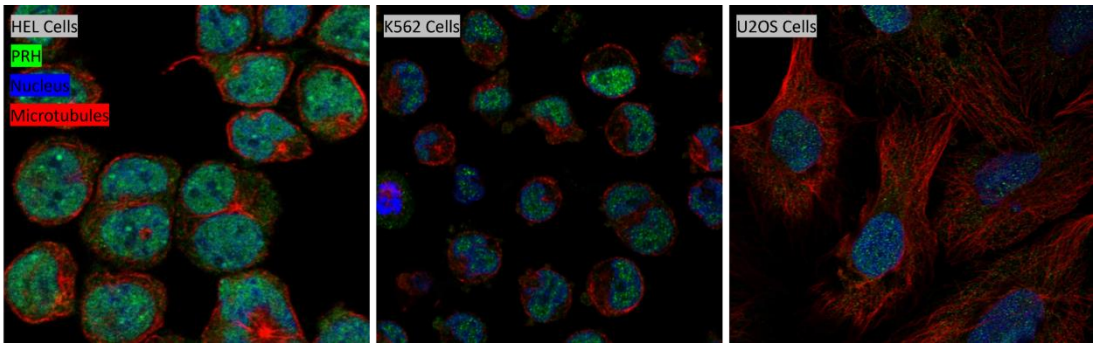


Figure 1-14 PRH subcellular localisation.
Adopted from The Human Atlas (252)

These images are approved for PRH in the nucleoplasm, nuclear bodies, and in the cytoplasm. They all show high PRH nuclear to cytoplasmic ratio. Moreover, they all show PRH as discrete foci in both compartments (196, 252).

1.5.8 The role of PRH in the regulation of target gene expression

PRH exhibits either a repressing and activating effect via direct promoter binding, or by interacting with other TFs to ultimately regulate the transcription of its target genes (8, 9). The PRH homeodomain can directly bind to DNA to downregulate its target genes e.g. endothelial specific molecule (*ESM-1*) (253), thyroglobulin (*TG*) (237), chordin (*CHRD*), *GSC* (197, 254) and *VEGFA* (255). PRH can act via recruitment of co-repressors proteins from the Groucho/TLE family and by competing with TATA box protein (256, 257). PRH activates the transcription of other target genes e.g. Endoglin (*ENG*) and sodium/taurocholate co-transporting polypeptide (*NTCP*) again by direct contact to their promoter DNA (230, 258). PRH, via its N-terminal domain, suppresses eIF4E-mediated transport of certain TF mRNA across the nuclear membrane. Therefore, it interferes with the translation of those trapped mRNA. For example, PRH can interfere with the cell cycle by preventing the translation of cyclin D1 in leukaemia cells (233).

1.5.9 PRH in Cancer

Any perturbations in the status of PRH, expression and/or subcellular localisation, can result in target genes dysregulation which is implicated in the development of cancer in liver, thyroid, breast, prostate, bile duct, colon, and haematopoietic cells (214, 215, 243, 245, 259, 260).

PRH can act as a TSG to inhibit angiogenesis as well as cell proliferation and migration. PRH overexpression can inhibit the action of VEGF and EIF-4E to reduce angiogenesis and cell proliferation in leukaemia cells, respectively

(233, 255, 261). In hepatocellular carcinoma, a decline in PRH expression was observed as the cancer progresses from grade II to grade III. In addition, PRH overexpression in a hepatoma cell line resulted in increased expression of T*P53* and *RB1*, and decreased expression of *Bcl2* proto-oncogenes. When these cells were injected in nude mice, they prevented tumour formation in the liver (259).

ENG has an important role in angiogenesis and proliferation. Endoglin is a co-receptor for TGF- β type I and II. It function as a growth inhibitor by regulating cellular proliferation and survival i.e. prevent tumour formation (262, 263). PRH inhibits tumour cell migration and invasion by direct alteration of *ENG* transcription in prostate and breast cancers (260).

PRH also represses the EMT inducer *GSC* (a homeobox gene that promotes tumourigenesis and increase migration) in breast cancer and therefore downregulates EMT (264). In contrast, PRH knockdown works in favour of tumourigenesis. In breast and thyroid cancer cells, PRH knockdown and cytoplasmic mis-localisation de-repress *VEGF* expression and induce angiogenesis and cell survival (255). PRH also directly represses *ESM-1* expression which was found to be associated with lung and kidney cancers (265). Exogenous PRH also downregulates *MMP10* which is important for breast tumour cells invasion (261).

PRH has been also reported as an oncoprotein in several cancers. For example, PRH is upregulated by *LMO2* oncogene to promote acute T cell lymphoblastic lymphoma (266). In colorectal cancer (CRC), **PRH interacts**

with the complex YAP/TEAD4, and promotes tumorigenesis. This interaction is associated with poor prognosis in CRC patients (243).

PRH mis-localisation is linked to certain cancers. In breast cancer, PRH is predominantly located in the cytoplasm compared to normal breast tissues where PRH is both nuclear and cytoplasmic (267). A similar observation regarding PRH cytoplasmic localisation was reported in thyroid cancer (245). Loss of nuclear localisation was also found to associate with human myeloid leukaemia (250, 268)

1.5.10 PRH in CCA

Recently, aberrant PRH was reported to induce molecular and cellular profiles suggesting that PRH has an oncoprotein role in CCA (260). PRH expression is very low in primary biliary epithelial cells (BECs) and normal immortalised bile duct cells (AKN1 and MMNK1) but PRH expression at protein level is high in several CCA cell lines (CCLP1, CCSW1, rmCCA1, HuCCA1). In addition, PRH at mRNA level showed various levels of positive fold changes in CCA patients compared to normal subjects(215). These observations were supported by *in vitro* works where PRH depletion in CCA cells resulted in a reduction in tumour cell proliferation, colony formation, migration, and invasion. Loss of PRH in CCA cells is associated with induced E-Cadherin and decreased Vimentin expression which correlates with the phenotypes seen. Moreover, *in vivo* implantation of PRH-depleted CCA cells in xenograft mouse model generated significantly smaller and less tumours compared to control cells 21 days after injection. In contrast, PRH overexpression in CCA cell lines and in normal immortalised or primary bile duct cells promoted cell

proliferation, migration, and invasion. Gene set enrichment analysis (GSEA) and gene ontology (GO) of PRH depletion and overexpression showed enrichment of genes associated with MET, EMT which is consistent with the *in vitro* findings (215).

PRH was reported to regulate two important signalling pathways in CCA: Notch3 and Wnt. PRH is an indirect regulator of Notch3 in CCA where it forms a positive feedback loop with Notch3 contributing to the upregulation of genes driving CCA tumourigenesis (215). In addition, PRH expression in CCA cells directly upregulates the expression of genes that are involved in Wnt pathway (*WNT11*, *WNT16*, *TCF7L1*, and *DKK1*) (215). Wnt activation in CCA by PRH was also mediated through Notch3. GSEA and GO revealed that PRH and Notch3 co-regulate genes in EMT that were found to promote CCA tumourigenesis (215).

1.6 PRH and Vimentin in CCA

Kitchen et al. indicated that PRH depletion was accompanied by a reduction in Vimentin expression at the mRNA and protein level (215). This is also reflected in cell morphology and phenotype as they became more epithelial and less proliferative and migratory (215). In contrast, PRH overexpression in AKN1 and BECs cells is associated with more mesenchymal cells that showed increased Vimentin expression along with increased cell proliferation, migration, and Invasion (215). The regulation of Vimentin expression by PRH in CCA was found to be Notch3 dependent (215).

Yet, unpublished data from the Jayaraman laboratory shows that the PRH and Vimentin proteins interact in CCA cells. Co-immunoprecipitation assay (CoIP) revealed Vimentin as a protein binding partner for PRH in CCLP Myc-PRH and CCLP AdEmpty cellular fractions. This was confirmed by mass spectrometry in CCLP1 CCA cells (Ka Ying Lee and Jayaraman) (269). Investigating the interaction and interplay between PRH and Vimentin in CCA is important to reveal whether Vimentin plays a role in the transcriptional activity of PRH and vice versa, or if both proteins are needed together as a complex to regulate common biological functions e.g. cell proliferation, or protein degradation.

1.7 Vimentin

1.7.1 Background

Vimentin is a 57 kDa protein encoded by the gene *VIM* located on chromosome 10 (270, 271). It was given the name Vimentin in 1978 by Franke et al. to differentiate it from other rod-like proteins (272). Vimentin is a major cytoskeletal protein that belongs to class III intermediate filaments family (273). These intermediate filaments are one of three major types; the others being microfilaments and microtubules (10). Altogether, they generally provide a mechanical support that maintain cellular shape as well as playing roles in cell division and motility (10). These intermediate filaments consist of laterally connected groups of 8-rod strands that collectively give the final formation of an intermediate filament (figure 1-15) (274).

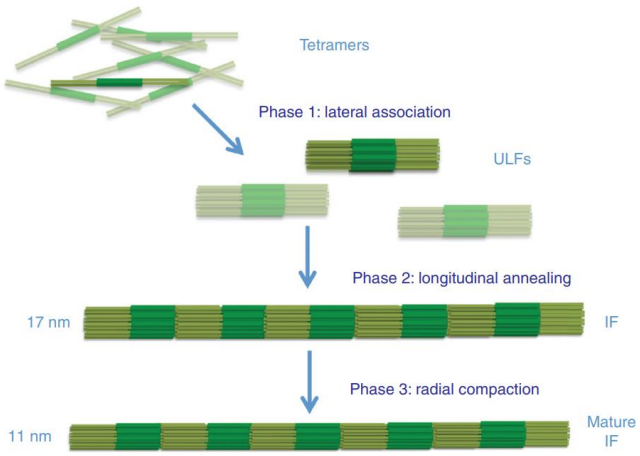


Figure 1-15 A schematic illustration of intermediate filaments assembly (274).

Intermediate filaments include six subtypes and each has a well-known representative protein: type 1 Keratin, type II Keratin, type III Vimentin, type IV Neurofilaments, type V nuclear Lamins, and type VI Nestin (9). Vimentin is

mainly found in mesenchymal cells; it is a well-known mesenchymal marker with a high importance during development and in disease states (273).

1.7.2 Regulation of Vimentin gene expression

1.7.2.1 Transcriptional regulation

The Vimentin gene promotor has been functionally analysed and reported to comprise three parts; a negative regulatory region flanked by positive regulatory regions (275). Several TFs binds directly or indirectly to the *VIM* promoter, including Sp1 TF which recognises GC boxes on the promoter and regulate Vimentin expression, and thereby control cell proliferation (276-278). *VIM* is also regulated by AP-1, which is a dimeric TF composed of two subunits (Jun and Fos)(279). In fact, AP-1 has a tandem binding sequences on the *VIM* promoter, and AP-1 increases Vimentin expression in breast cancer cells (276), and in colorectal cancer cells (280). Other TFs binding to enhancer regions on *VIM* promoter include NF-kB (281, 282), PEA3 (283), and β-catenin (284). STAT3 also activates Vimentin expression by binding to an anti-silencer elements (ASE) only when ZBP-89 is bound to a silencer element on the promoter (285).

Indirect transcriptional promoter regulation can also occur. For instance, c-Jun interacts with Sp1 when the latter is bound to GC-boxes (286). In addition, TGFβ1 also signals through AP-1, Sp1 and Smads in order to differentiate skeletal myogenic cells via controlling Vimentin expression (287).

1.7.2.2 Epigenetics

DNA methylation is a chemical modification involves the addition of a methyl group on a cytosine to form a 5-methylcytosine (5mC). This modification silences gene expression by recruiting repressing proteins, such as methyl-CpG binding domain (MBD) proteins. As a result, MBD proteins attract other chromatin modifiers e.g. HDAC1. HDAC1 in turn causes **DNA deacetylation**, condensate the chromatin, and interfere with TF binding (288). A similar observation of promoter methylation and Vimentin expression level has been linked to cervical cancer (289). In contrast, treating colon cancer cells with an anti-methylation has seen to activate Vimentin expression (290). Moreover,

ZBP-89 binds to a repressor element on Vimentin promoter, recruits HDAC1 to silence Vimentin expression in HeLa cells. Reversing deacetylation by using a HDAC1 inhibitor seemed to revoke that observation (291).

1.7.2.3 Non-coding RNAs

ncRNA are single stranded RNA that are not translated or encoding a protein, but they are still functional (292). When these strands are longer than 200 nucleotides, they are classified as **long ncRNA (lncRNA)**. Whereas **small ncRNA (sncRNA)** are shorter than 200 nucleotides. One example of this class are **micro RNAs** (**miRNAs or miR)** with a length ranging from 15 to 20 nucleotides and an important role in gene expression (292). miRNAs can bind to target mRNA and cause a post-transcriptional mRNA silencing (293). They regulate several EMT related genes including Vimentin. In 2020, Lin et al. discovered a novel long ncRNA called RNA-01488 that was found to reduce

Vimentin expression via miR-124-3p/miR-138-5p which inhibited metastasis (294).

1.7.3 Vimentin expression

Vimentin protein is widely distributed in normal human tissues as indicated by the Human Protein Atlas project (295). Vimentin is expressed in cells of major organs e.g., the brain, lungs, liver, immune system cells, blood vessels, skin and adipose tissues as reviewed by Danielsson et al. (296) (figure 1-16).

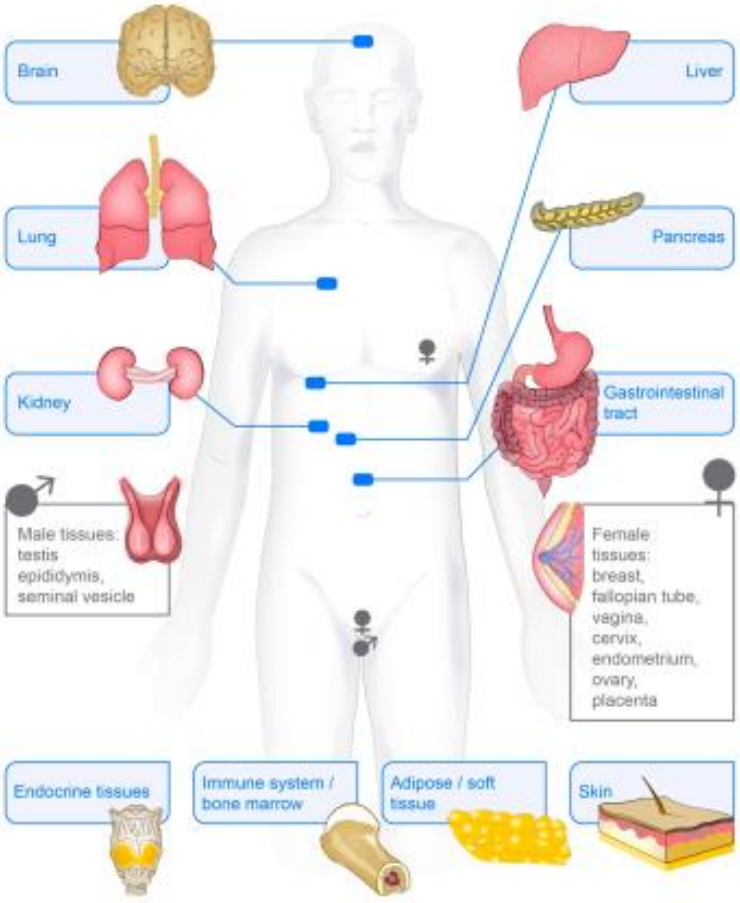


Figure 1-16 Vimentin expression in normal human tissues as reported by the human protein atlas (296).

Vimentin expression occurs early during embryogenic development in mice, specifically starting on day 8.5 during the formation and migration of primary mesenchymal cells of the primitive streak (297), a process that represents physiological EMT-MET. In adult mice, Vimentin is expressed in connective tissues, muscle and brain tissues (298). Vimentin null mice (*VIM* $^{-/-}$) grow suffering multiple developmental defects including neuronal growth anomalies (299), impaired wound healing (300), and fragile vasculature (301).

1.7.4 Vimentin structure and organisation

Vimentin was predicted to have filaments with a 10 nm diameter when it was investigated in various cell lines from mice, chickens, and humans using various techniques including EM (302-306). A Vimentin monomer is comprised of 466 a.a. residues (figure 1-17). A monomer of Vimentin and other intermediate filaments includes a central α-helical domain flanked by two non-α-helical domains. The central part is also known as the “rod” domain. The N-terminal region is also called the “head” domain. The C-terminal part includes 61 a.a. residues and called the “tail” domain (307). The head domain comprises approximately 96 a.a through which Vimentin can identify and bind G-rich highly repetitive region on DNA (308). To investigate the functions of genes harbouring G-rich highly repetitive region, the sequences were retrieved for gene ontology. Functional pathway analysis revealed that Vimentin binds to G-Quadruplex region on DNA at genes associated with cell-cell communication, signal transduction, and neurogenesis (309). This head domain has been investigated extensively, and it is a site for phosphorylation and glycosylation and it interacts with other Vimentin heads or tails in Vimentin

monomers (310). Truncated Vimentin (1 – 411 a.a.), or “tailless Vimentin”, was thought to have little impact on mature filaments formation (310, 311). However, it was reported later that tailless Vimentin disrupts mitosis or leads to asymmetrical division (312).

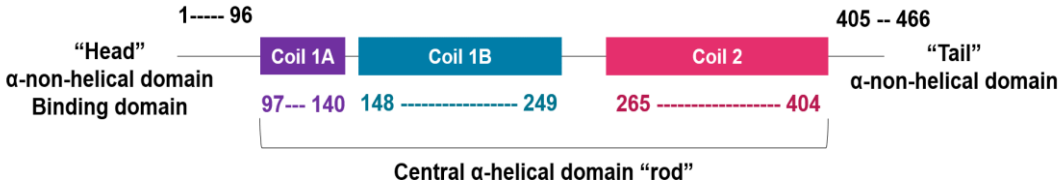


Figure 1-17 A monomer of Vimentin and its domains (307, 313, 314).

The “rod” central part in intermediate filaments is characterised by having coiled subsegments 1A, 1B, 2A, 2B1 and 2B2 with connecting linkers in between (315). These coiled-coiled-α-helical subsegments attribute to Vimentin interacting with other proteins under class III (e.g. Desmin) or class IV (Neurofilament NF-L). This interaction enables Vimentin monomers to form stable polymers ranging from homodimers to tetramers and octamers (316). This is indicated through the stepwise organisations which Vimentin can be seen forming in live cell imaging and EM (317) (figure 1-18). Meire et al reported that a group of 8-linked tetramers form a unit-length filament (ULF) (318). This ULF phenomena explained an older suggestion that there are smaller non-filamentous Vimentins serving as building blocks via aggregating to form short intermediate filaments, also referred to as squiggles, followed by further lateral linking to finally give the mature Vimentin intracellular network (317). Plectin is an example of a linker protein that crosslinks Vimentin with the rest of the cytoskeleton (319).

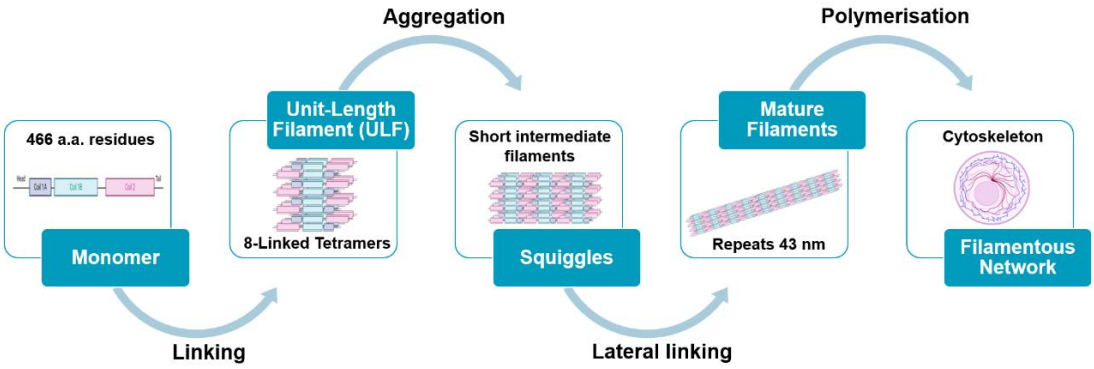


Figure 1-18 Vimentin assembly from monomers into cytoskeleton (317-320)
Created with BioRender.com

Vimentin polymerisation is controlled by the phosphorylation status of the proteins (321). Depolymerised Vimentin can repolymerise *in vitro* indicating distinct biochemical features for Vimentin, such as filaments dynamism and elasticity, as compared to the rest of the cytoskeleton (322). Further to Vimentin structural polymerisation which forms its intracellular networks, assembly and disassembly through precursor exchange between polymerised and soluble Vimentins allows the cells to regulate this network formation (323). This remodelling is required for cellular motility such as cell migration. Another mechanism influencing Vimentin network construction is enzymatic proteolysis via caspases or viral proteases. These proteases cleave Vimentin at various a.a. locations dismantling Vimentin filaments into smaller aggregates and ultimately leading to cell dysmorphology and apoptosis (324, 325).

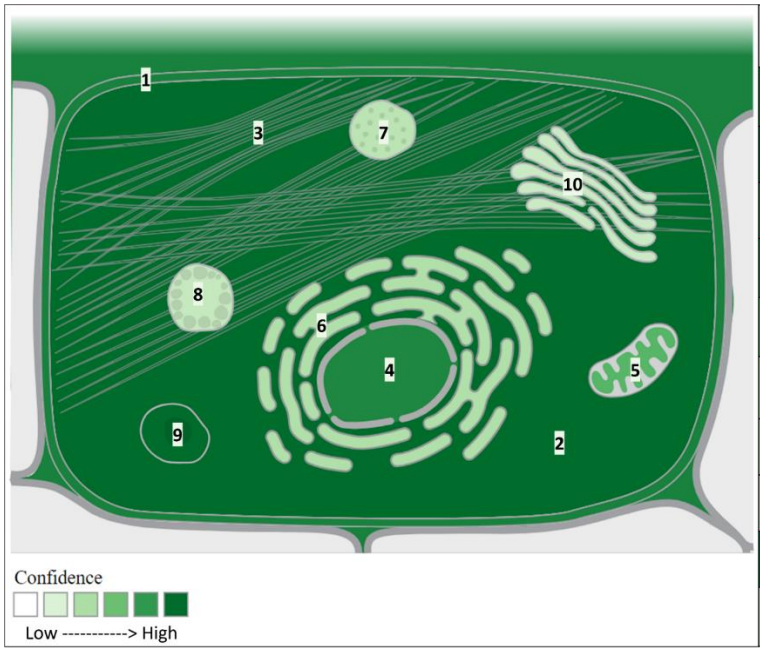
1.7.5 Post-translational modification of Vimentin (PTM)

Proteomic analysis identified several PMT mechanisms regulating Vimentin activity including phosphorylation, proteolysis, ubiquitination, citrullination,

Acytylation, ADP ribosylation, s-nitrosylation, s-glutathylnation, glycosylation, and SUMOylation (326). PMT regulates the spatial organization of the Vimentin network in the cell which impacts on the cell behavior and interactions with the surrounding ECM (327). For instance, Vimentin filaments undergo phosphorylation to link with other cytoskeletal filaments to form the cytoskeleton. Phosphorylation of Vimentin was found to re-organise the filaments to help with cell motility (328).

1.7.6 Vimentin forms and subcellular localisation

Vimentin was first proposed as an intracellular and cytoplasmic filamentous protein, via microscopic imaging (272, 329). However, several studies later reported the presence of Vimentin on the cell surface (330), as well as a secretory form by Golgi apparatus (331), and on the surface of circulating cells (332). Moreover, it was proposed that Vimentin could be nuclear (333), interacting with Actin during cell mitosis (312). According to the database COMPARTMENT, Vimentin is mainly an intracellular protein with abundant cytoplasmic distribution in addition to its nuclear existence (figure 1-19) (244). This different localisation of Vimentin influences its biological roles.



No.	Location	Data Mining
1	Plasma membrane	Knowledge – Text
2	Cytosol	Knowledge – Text – PSORT - YLoc
3	Cytoskeleton	Knowledge – Text – HPA
4	Nucleus	Knowledge – Text – PSORT
5	Mitochondrion	Text – PSORT
6	Endoplasmic reticulum	Text
7	Lysosome	Text
8	Endosome	Text
9	Peroxisome	Knowledge – Text
10	Golgi apparatus	Text

Figure 1-19 Vimentin subcellular distribution according to COMPARTMENT.
Adopted from COMPARTMENT (244).

Cytoplasmic filamentous Vimentin is the most common and well-investigated form of the protein. Cells exploit the polymerisation and phosphorylation of Vimentin, as mentioned earlier, to form structures ranging from small squiggling proteins to a large network in the cytosol (317, 323). This network serves a plethora of functions that have been reviewed extensively in the literature and found to be required physiologically to sustain normal cellular performance, or pathologically for the cells to initiate and survive in diseases including cancer (296). For instance, the expression of Vimentin is important as cells migrate or invade the environment (334, 335). To confirm the need for cytoplasmic Vimentin filaments in biological cell processes, Berr et al. conducted scratch wound healing assay and transwell invasion assay after they knocked out Vimentin in (KPV $^{-/-}$) mice model cell line of non-small cell lung carcinoma where they reported significant reduction in cell motility and

invasion compared to control cells.(336) Similar findings were observed using 3D (KPV $^{-/-}$) spheroids. They concluded that mature Vimentin filaments are required for the *in vivo* progression of non-small cell lung carcinoma (336).

Cell surface Vimentin interacts via its N-terminus with phospholipids in the cell membrane (337). Vimentin was also detected on the cell surface of natural killer cells (338). One proposed mechanism to explain the presence of Vimentin on the surface of cells is the potential interaction between charged residues in Vimentin and polar cell membranes, since Vimentin is not known to have a secretory signal sequence (339). More recent studies suggested that Vimentin can act as a receptor or ligand of a receptor owing to its cell surface expression. This was proposed upon an observation of Vimentin interacting with von Willebrand factor (VWF) in ELISA experiments (330). VWF is needed for platelet aggregation and adhesion which is an initial reaction to stop bleeding or start an atheroma (340). This Vimentin-VWF interaction was further confirmed later through a bleeding time measurement in Vimentin knockout mice tails, where bleeding was found longer compared to control mice group (330, 341). These findings indicate that Vimentin PPI on cell surface is important for disease development, e.g. stroke (330, 341). Recently, a study was conducted to investigate the significance of cell surface Vimentin on human articular chondrocytes (342). The authors reported that disturbing cytoplasmic Vimentin network increased the levels of Vimentin on the chondrocyte surface. As a result, surface Vimentin expression enhanced senescence-associated phenotypes of the chondrocyte cells which are typically observed in osteoarthritis (342). The same study indicated that high

levels of surface Vimentin increased the expression of genes encoding Integrin proteins of adherens junctions, which was seen *in vitro* as delayed detachment of cells upon trypsinisation (342). Overall, the authors concluded that cytoplasmic Vimentin network is essential for maintaining a normal differentiation of chondrocytes, while cell surface Vimentin is an important marker of dysfunctional chondrocytes and is also important for cell adhesion (342).

Secreted Vimentin is related to the previous extracellular form since Vimentin was observed to be released to the outer surface of the cells in response to inflammatory signals (331). Vimentin was also detected on HMEC-1 endothelial cells as an antigen for PAL-E antibodies through cell-specific PMT (343). Moreover, secreted Vimentin was also identified by using antibodies specific for secreted form on Sezary cells (atypical malignant T cells). Therefore, Vimentin has been proposed as a marker in Sezary disease, which is a rare form of cutaneous T-cell lymphoma (344). In addition, Vimentin was reported to exist in high abundance within activated human macrophages (331). Interestingly, Vimentin was secreted extracellularly in TNF-α-activated macrophages forming a pool within the vicinity of the cells. This Vimentin was found to exist in a non-filamentous form that is eventually released into the surrounding cell media. Subsequent stimulation of non-activated macrophages using the conditioned media significantly increased the migratory speed and phagocytosis activity of the cells (345). Altogether, these findings highlighted the roles of secreted Vimentin as marker in diseases as

well as a facilitator of macrophages which could be link to their importance for bacterial killing.

Nuclear Vimentin has been reported recently in multiple cell types. *In vitro* observation of Vimentin tail and nuclear envelop Lamin B interaction through different experiments including immunoprecipitation was one of the earliest suggestions of Vimentin extending between the cytosolic skeleton and the protein framework inside the nucleus (346). More recently, Vimentin has been reported to regulate gene expression (347).

Mergui et al. investigated the regulation of $p21^{Waf1}$ expression by Vimentin and they confirmed nuclear localisation (348). The authors reported that downregulation of Vimentin was accompanied by a similar change in $p21^{Waf1}$ mRNA. Similarly, upregulating Vimentin expression led to activation of $P21^{Waf1}$ activity and nuclear localisation of Vimentin (348). This finding along other reports of nuclear Vimentin reviewed in the literature raised the hypothesis that Vimentin plays a role in transcriptional regulation (347).

1.7.6.1 Cytoplasmic to nuclear trafficking of Vimentin

In 1998, Hertig et al. suggested that Vimentin could gain nuclear access depending on the transport of DNA through nuclear pores (349). The authors relied on a previous experiment where non-karyophilic substances were introduced into the nucleus of mammalian cells via oligodeoxyribonucleotide-mediated nuclear import (350). The authors injected FITC-labelled Vimentin into Human GMO 0970C fibroblasts and followed with confocal laser scanning of the cells to monitor the integration of injected Vimentin with endogenous

Vimentin filaments. Later, they injected Oligo (dG)25 into the cells where they reported that Oligo (dG)25 managed to extract considerable amounts of FITC-labelled Vimentin from the cytoplasm and channelled into the nucleus through formation of nuclear pore complexes (349). Therefore, they suggested that Vimentin can be transported to the nucleus together with DNA fragments through nuclear pore complexes (349).

Another mechanism via which Vimentin gain nuclear access and nuclear localisation is by PTM. Vimentin interacts with SUMO1 and SUMO2/3 (small ubiquitin-like modifiers) in a process called SUMOylation (351). SUMOylation regulates nuclear localising signals that in turn facilitate the recognition of proteins by the nuclear receptors that mediate nuclear import (352). Vimentin trafficking is upregulated by controlling the levels of SUMO1 and SUMO2/3 expression through P62 protein in polyploid giant cancer cells. Once Vimentin is in the nucleus, it acts as a TF to promote cell migration and invasion. Subsequent Vimentin ChIPseq and gene ontology analysis in these cells confirmed the implication of Vimentin in pathways associated with cell invasion and migration for cancer metastasis (351). Particularly, RT-qPCR validation of gene ontology confirmed that Vimentin knockdown in these cells downregulated CD42 which was associated with a reduction in transwell cell migration. This suggested that Vimentin trafficking into the nucleus via SUMOylation is required for the transcription of CD42 gene to modulate its function in polyploid giant cancer cells (351)

1.7.7 Vimentin subcellular localisation on immunofluorescence imaging

The Human Atlas provides reference images of Vimentin subcellular localisation in three different cell lines: ASC52telo (hTERT immortalised adipose derived Mesenchymal stem cell line). U-251MG (malignant glioblastoma tumour), and U2OS (osteosarcoma cells) (252). (figure 1-20)

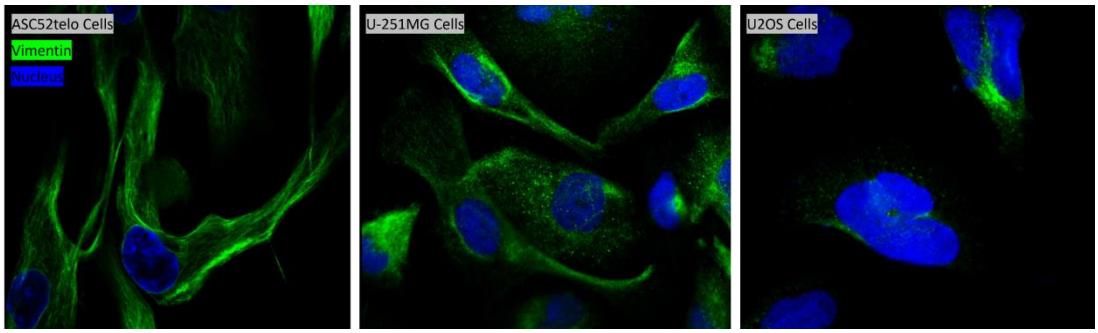


Figure 1-20 Vimentin approved filamentous expression in the cytoplasm.
Adopted from The Human Atlas (252).

These images are approved for Vimentin immediate filaments morphology in the cytoplasm. In addition, Vimentin in U-251MG cells can also be seen as small foci. Overall, all the cells show high Vimentin cytoplasmic to nuclear ratio.

1.7.8 Vimentin functions / roles

Cytosolic Vimentin forms an elastic network which can withstand mechanical stresses by allowing a range of deformations and recoiling without affecting the integrity of the cytoplasm; therefore, contributing to cell **stiffness and shape preservation** (334, 353-355). This feature was seen during transwell migration by (*VIM* $^{+/+}$) mouse fibroblasts (MF) compared to (*VIM* $^{-/-}$), as offering

protection to the nuclei by the Vimentin nuclear cage (356). The consequence of Vimentin loss on stiffness was reported in (*VIM* $^{-/-}$) MF where they exhibited more surface spread in response to mechanical indentation compared to control cells indicating the need for Vimentin to maintain elasticity (357). Lack of Vimentin was found to increase the stiffness of (*VIM* $^{-/-}$) mice carotid arteries due to sub-endothelial regulation of smooth muscles differentiation and membrane matrix interaction (358).

Another major role of Vimentin is in **EMT regulation**, for which Vimentin is a well-known canonical marker (359). EMT is an essential step in many events including cell plasticity, adhesion, migration, invasion, metastasis, and wound healing. These processes are directly or indirectly dependent on Vimentin-linked cytoskeletal remodelling, and they are tightly regulated by different signalling pathways where Vimentin is either an upstream or downstream player (359).

Cell **plasticity** is defined as the ability of the cells to reversibly alter their phenotype and adopt a new identity that renders adaptation to their surrounding environment (360, 361). Plasticity can happen normally e.g. during stem cell differentiation (361), or pathologically conditions such as pre-cancerous states like oesophageal metaplasia or trans-differentiation (362). An inverse relationship between Vimentin expression and cell differentiation has been reported by Chen et al. where they reported increased neuronal cell differentiation in phosphorylated-Vimentin deficient mice (363).

Single cell **motility** is achieved via cellular contractile forces generated by the interaction of different types of cytoskeleton filaments including Vimentin and

actomyosin with the surrounding extracellular matrix (364). One explanation behind this motility is that it is initiated by EMT where cells gain the mesenchymal phenotype and become amoeboid after breaking free from their neighbouring cells by downregulating the linking junctional proteins, losing their apical to basal polarity, becoming more undifferentiated, and therefore, more motile (365). For instance in gene expression profiling of epithelial-mesenchymal-amoeboid transition in breast cancer samples there was *VIM* upregulation and *CDH1* downregulation (365).

Cell **adhesion** is a process induced by several adhesion molecules, protein complexes or junctions via which cells maintain their lateral cell-cell contacts or their anchoring on a basement membrane (9). After the cells gain an amoeboid morphology, the cells lose their adhesiveness with their surroundings setting them free to move (68). In endothelial cells, Vimentin was found to maintain cell adhesion via interaction with β3-Integrins and Plectin after which Vimentin is recruited to the cell surface (366). B3-Integrin is a major adhesion-regulating cell receptor (367). (368). Additionally, Vimentin was found to bind via its Ser38 residue to another Integrin protein α5β1-Integrin facilitating cell adhesion (369). Plectin is a cytoskeletal linker protein that can undergo interaction with Vimentin to facilitate Vimentin polymerisation (370, 371).

Cell **migration** is a multistep process initiated by EMT in a similar fashion to cell motility (372, 373). It differs from motility in the fact that it is considered more of an organised outcome to signalling chemicals rather than a random event. In addition, migration can occur to either single cells or for a population

of close cells, with the latter needing an attraction signal from a leading migrating cell plus cytoskeletal crosslinking and ECM interaction (374). One suggested mechanism explaining how directed migration is regulated by Vimentin is the increase in Actin filament flow rates when they encounter Vimentin compared to areas deficient of Vimentin (375). Vimentin also stabilise microtubules through direct interaction via its tail domain and thereby regulates cell migration, adhesion, and cell division (376, 377). Furthermore, cell migration was one of the reported outcome of (*VIM* $^{-/-}$) MF in comparison to their control where they experienced not only a reduction in cell stability and cell motility, but also a reduction in directed cell migration (378). Vimentin over expression in gastric cancer GES-1 cells led to more migratory cells in transwell migration assay (379). Vimentin-regulated EMT also has a major impact on **wounds healing** process. It regulates TGFβ-slug signalling pathway in EMT which activates keratinocytes to differentiate (380). An evidence of Vimentin expression importance for healing is demonstrated by (*VIM* $^{-/-}$) MF that were showing scarring and impaired wound closure (300). The reasoning for this compromised wound healing was the impaired keratinisation (380) and decreased TGFβ-stimulated fibroblast proliferation (380).

Cell **invasion** also involves EMT (359). However, cell invasion is characterised by having the cells degrading ECM with the help of matrix metalloproteinases giving them access to neighbouring tissues (381). In MCF7 breast cancer cells, Vimentin overexpression correlates with poor prognosis and activates various EMT-related TF such as Slug which in turn induces EMT

and tumorigenesis (335). In cancer, upregulated Vimentin was found to regulate tumour cell invasion in different tissues including lung, prostate and melanoma (347).

Cell **metastasis** is a possible consequence following cell migration and invasion. Vimentin affects metastasis by controlling EMT first, and then regulates both of cell migration and invasion. For example, (*VIM* $^{-/-}$) MF of lung adenocarcinoma have reduced metastasis (382).

Cell proliferation is a biological process by which cells grow in number via cell division. Vimentin expression is induced in response to wound injury, via the activation of TGFβ pathway, to promote cell proliferation and encourage wound healing. Cell proliferation correlates with cell plasticity since several oncogenes that control Vimentin stiffness and morphology were found to induce cell proliferations e.g c-Myc. *C-Myc* oncogene over expression induced Vimentin re-organisation into thicker filaments around the nuclei of Rat1 cells affecting their stiffness (383). Although Vimentin expression has been mostly known to induce cell proliferation, Vimentin loss was recently reported to also promote cell proliferation in HepG2 liver cancer cells via Rictor/AKT/β-catenin signalling pathway (384).

Vimentin also acts in cell **mitosis** by interacting with Actin filaments and microtubules (312). Actin filaments belongs to microfilaments family of the cytoskeleton (9). Towards the end of a mitosis, actin filaments go through remodelling to allow cell division. Vimentin filaments play a role in normal cell division by interacting with cortical actin (312). Aberrant bundling of Vimentin localisation was found to produce mitosis catastrophe signs, while Vimentin

wildtype was more peripheral and intertwining with actomyosin cortex on super-resolution microscopy (312). Further 3D images of Vimentin and Actin co-staining and distribution in image stacks revealed that Vimentin affects the distribution of cortical Actin during mitosis (312). Vimentin is also involved in cell **apoptosis** during which caspases have been reported to disassemble Vimentin filaments, forming pro-apoptotic fragments that activate apoptosis, and this later leads to nuclear fragmentation (324). Caspase-induced Vimentin collapse was antagonised by using a caspase inhibitor which confirmed the observation (324).

Vimentin is an essential player in Notch-Jagged signalling pathway. This pathway is involved in vascular remodelling and angiogenesis (301). **Vascular remodelling** is a response mechanism to vascular stress. To explain how Vimentin is exactly affecting vascular remodelling, Langlois et al. investigated the response of the arteries in (*VIM $^{-/-}$)* mice to the vasomotor reagents: phenylephrine, acetylcholine, and sodium nitroprusside. The authors reported a reduction in arterial relaxation associated with a stiffness in the arterial walls reflecting the underlying defective endothelial differentiation (358). Later, it was suggested that defective vascular remodelling and differentiation in (*VIM $^{-/-}$)* mice were connected to Notch3-Jagged signalling dysregulation (301, 385). Vimentin phosphorylation at serine 38 interacts with Jagged1; this interaction is required for an effective propagation of Notch3-Jagged1 activation across the arterial wall (301). **Angiogenesis** is the emergence of new blood vessels from existing vasculature (386). To assess the importance of Vimentin in angiogenesis, Antofolk et al. compared the placental vascular development

pattern between (*VIM* $^{-/-}$) mice and *VIM* wildtype mice. On day 11.5, (*VIM* $^{-/-}$) mice showed an abnormal placenta vascular pattern which later became less branching and less complex compared to the control (387). This reduction of angiogenesis and its developmental attenuation was seen to be rescued by adding recombinant Jagged1 which indicates the importance of Vimentin in Jagged1-Notch signalling for angiogenesis (387).

There is abundant evidence on the need of mature Vimentin filaments for pathway signalling. Some studies stated that Vimentin organisation into filaments was required for the activation of certain signalling pathways, e.g. β3-adrenergic receptor-activated ERK/MAP signalling pathway in adipocytes (388). When β3-adrenergic receptors are activated, they associate with Vimentin filaments in a Src kinase-dependent mechanism. This leads to ERK/MAP activation and lipolysis stimulation. There was a significant decrease in lipolysis when Vimentin filaments were disturbed (388). In addition, phosphorylation of Vimentin filaments initiated the interaction between Vimentin and Jagged 1 of Notch signalling which resulted in the remodelling of arterial walls (388). Although phosphorylation of Vimentin affects the assembly of Vimentin filaments and generates active short Vimentin filaments (squiggles) which can modulate cell migration (389), there are no reports on whether that phosphorylation breaks down Vimentin into active monomers, or monomers of Vimentin are driving pathway signalling in the literature.

Additional biological functions, where Vimentin is taking part in their occurrences, include lipid metabolism and homeostasis (390), response to viral infections (391) and bacterial inflammatory responses (392).

In addition to Vimentin and other subtypes of intermediate filaments, there are other two major types of skeletal proteins: microfilaments (Actin) and microtubules (Tubulin) (10). Recent interests to study cytoskeletal protein-TF interaction and gene regulation have emerged. The interaction between cytoskeletal Actin microfilaments and the nuclear components has been reviewed to occur via intermediate protein monomers e.g. G-actin binding proteins (G-ABP) (393). Myocardin-related transcription factors (MRTFs) are G-ABP that can translocate into the nucleus and modulate the activity of SRF transcription factors on muscle-contractile genes (394). Dysregulated Actin-induced MRTF-SRF circuit was associated with Infant acute megakaryoblastic leukaemia (395). This adds on the complexity of functions played by the cell cytoskeleton and highlight a new target for interventional studies.

1.7.9 Vimentin in cancer

Increased Vimentin expression is linked to the tumorigenesis and aggressiveness of several epithelial cancers e.g. cervical (396), and renal cancers (397), prostate (398), breast (399), lung (400), oral (401), and gastrointestinal cancers (379). It was even suggested to serve a good marker for either diagnosis in oral cancer (401), or prognosis in gastric cancer (402), colorectal cancer (403, 404), and in hepatocellular carcinoma where it was seen secreted by cancer cells (405).

1.8 Aims of the project

PRH has been investigated extensively as a TF regulating target genes essential to embryonic development and tumourigenesis. Research regarding PRH in diseases development, cancer specifically, are ongoing. However, our understanding of the specific mechanisms via which PRH regulates gene expression and interaction with other proteins to promote the development of CCA is limited and relatively new. Vimentin is highly expressed in CCA cells. However, the roles of Vimentin other than as a cytoskeletal protein and mesenchymal cell marker are also poorly understood.

Hypothesis: PRH expression in CCA is associated with alterations in cellular phenotype and drives PRH-Vimentin nuclear localisation to co-regulate gene expression that contributes to CCA tumourigenesis.

The general aim of this project is to shed light on the potential collaborative regulation of gene expression in CCA by PRH and Vimentin.

The specific aims for this project are:

1. To express PRH tagged with green fluorescent protein (GFP) in AKN1 cells and to examine the change in phenotype of these immortalised cholangiocytes using cell proliferation, cell migration, and cell invasion assays.

2. To investigate nuclear spatial existence of PRH and Vimentin in CCA by examining protein subcellular localisation in CCA cell lines using confocal and the super-resolution PALM dSTORM imaging techniques.

3. To explore the biological changes brought about by Vimentin loss in CCA and examine the effect on gene expression co-regulation by PRH and Vimentin in CCA cells using siRNA-mediated protein knockdown and genome-wide mRNA sequencing.

Chapter 2: Materials and methods

2.1 Cell culture

2.1.1 Cell lines

Three cell lines were used throughout the project: AKN1, CCLP1, and CCSW1. A summary of all the used cell lines is provided in table 2-1.

AKN1 cells are human immortalised cholangiocytes isolated from a donated normal liver without a recipient in 1998 (406). Sections of histologically normal liver cells underwent serial dilutional cloning to separate hepatocytes and non-parenchymal cells (406). The cells were grown in a long-term cell culture where they went through histological, morphological, microscopic, cytogenetic, karyotypic characterisation. AKN1 cells were reported as hepatocyte-like cells with biliary epithelium characteristics expressing cytokeratin 8 (CK8), cytokeratin 18 (CK18), cytokeratin 19 (CK19), human epithelial antigen (HEA), and gamma-glutamyl transpeptidase (GGT). However, the cells do not express Albumin that is a hepatocyte marker. Genomic DNA testing showed a significant nuclear accumulation of wildtype p53 protein without mutation of *p53* gene (406). Routine morphologic inspection revealed that the cells acquired neoplastic transformation associated with abnormal karyotypic changes of chromosomes 2 and 8 which led to cells immortalisation. AKN1 cells injection in nude mice generated low frequency small tumours within 14 days (406). There are no reports in the literature regarding neoR gene in AKN1 cells. No other mutations were stated in AKN1 cells.

CCLP1 and CCSW1 are CCA cell lines isolated from resected tumours; moderate to poorly differentiated intrahepatic CCA (CCLP1) and moderately differentiated intrahepatic CCA (CCSW1)(407). The cells of both cell lines were isolated from dissociated small tumour cuts followed by mesh filtration and centrifugation separation of tumour cells. The cell lines were established in a long-term culture growth undergoing characterisation and removal of contaminating fibroblasts. They both strongly expressed cytokeratin AEI, but not CEA-19 or Albumin. Both CCLP1 and CCSW1 were tumourigenic in nude mice generating tumours that were similar to the original tumours (407).

All the cell lines were tested regularly for mycoplasma infection.

Table 2-1 Cell lines properties

Cell Line	AKN1	CCLP1	CCSW1
Origin	Normal donated liver with no recipient. Liver bile ducts	Intrahepatic CCA tumour	Intrahepatic CCA tumour
Differentiation	Differentiated	Moderate to poor	Moderate
Morphology in 2D cell culture	Epithelial	Mesenchymal	Mesenchymal
Growth Properties	Adherent	Adherent	Adherent
Isolation	Serial dilutional clonal of donated normal human liver	Enzymatic dissociation + filtration + centrifugation	Enzymatic dissociation + filtration + centrifugation
Mutation profile	Accumulation of p53 proteins without mutation in p53 gene. No further information available regarding specific mutations.	No TP53, KRAS, NRAS, HRAS, PIK3CA, PTEN, or MYC mutations. (408).	

2.1.2 Cell growth and maintenance

The cells were grown in various sizes of flasks with vent caps (25 cm^2 - 225 cm^2 Rectangular Canted Neck Cell Culture Flask with Vent Cap, Corning). The growth media, or complete media (CM), for all the above cell lines consisted of Dulbecco's Modified Eagle Medium – high glucose (DMEM, Sigma Aldrich, D5796) supplemented with 10% (v/v) foetal bovine serum (FBS, GIBCO®, 10270106) and 1% (v/v) non-essential amino acids (NEAA, Sigma-Aldrich, M7145).

The cells were incubated in a humidified environment with 5% CO_2 at 37°C. Passaging was done when the cells were around 80% confluent. Briefly, old media was aspirated and discarded. The cells were washed 2 -3 times with warm Phosphate Buffered Saline (PBS, Sigma-Aldrich®, D8537). Then, cells were trypsinised with 3 ml Trypsin-EDTA solution (Sigma-Aldrich®, 15400054) diluted (1:10 v/v) in PBS. The flask was incubated for 5 min in a humidified environment with 5% CO_2 at 37°C until the cells were detached. Trypsin was neutralised with an equal volume of CM. The suspension was centrifuged at 300xg for 5 minutes. The cells were then resuspended in fresh CM and subcultured into a new flask at the required dilution ratio (1:3 to 1:10).

2.1.3 Cell counting

For counting purposes, cells were trypsinised as mentioned in the previous section. A sample of the cell suspension was diluted 1:1 (v/v) in 0.4% trypan blue in distilled water (Trypan Blue, GIBCO®, 15250061). From this mix, 10 µl

of the suspension was loaded under the coverslip of a haemocytometer. Blue-stained (dead) and non-blue (live) cells were counted in all 4 corners and the counts were averaged. The final cell number as (cell/ml) and percentage of live cells were calculated considering the dilution factor from the addition of trypan blue. Alternatively, an automated cell counter (CellDrop™ BF – brightfield, DeNovix, bioscience) was used to obtain the same information. Finally, viable cells were re-seeded at the desired density for different experimental protocols.

2.1.4 Cell pelleting

When cell pellets were needed for either RNA extraction or protein analysis, the whole cell suspension was centrifuged at 300xg for 5 minutes. Supernatant was discarded and the pellet were resuspended in PBS to wash away traces of trypsin and media. From there, the cells were centrifuged in an Eppendorf microcentrifuge 5415 – (1000 round per minute (rpm)) for 1- 3 minutes to collect the final pellet. Pellets were stored in -80°C if not used immediately.

2.1.5 Cell thawing

To revive the cells from long-term storage, they were brought on ice and submerged quickly in a 37°C water bath to defrost. Afterwards, cells were transferred into a sterile falcon tube containing 5 ml of fresh CM and centrifuged at 300xg for 5 minutes. The supernatant was discarded, and cells were resuspended in an appropriate volume of fresh CM, transferred to a suitable sterile flask, and incubated in a humidified environment with 5% CO_2 at 37°C. The next day, the old media was replaced with a fresh media to

remove any dead cells. Cells were left to reach the desired confluency for passaging. Cells went through at least one passage before being used for experiments.

2.1.6 Cell freezing

For cryo-storage, at least $1X10^6$ cells were pelleted as described in (2.1.4Cell pelleting) and resuspended in 1 ml of freezing medium consisted of 80% (v/v) CM with 10% (v/v) FBS and 10% (v/v) Dimethyl sulfoxide (DMSO, Sigma-Aldrich®, 10213810). The cells were transferred into labelled cryo-vials and placed in a Mr Frosty container containing isopropanol to gradually freeze the cells at a cooling rate of -1°C per minute. The Mr Frosty was kept at -80°C overnight before transferring the cryo-vials to -150°C (or liquid nitrogen) for prolonged storage.

2.2 Preparation of cell lines

2.2.1 Stable AKN1 PRH overexpressing cell line

2.2.1.1 Plasmids

To generate AKN1 cells stably expressing enhanced GFP, we used an empty EGFP C1 plasmid (pEGFP-C1 from Addgene – 45769). These cells were used as control cell line in experiments. To generate AKN1 cells stably expressing PRH fused to EGFP and a Myc tag, we used a pEGFP-PRH-Myc plasmid that has the human PRH cDNA cloned from the vector pBluescript-PRH in-frame with N-terminal GFP and a C-terminal Myc tag followed by a translation stop codon. Both plasmids were provided by (Professor Sheela Jayaraman).

2.2.1.2 Plasmid amplification

2.2.1.2.1 Bacterial strain

Escherichia coli (E. coli) XL1 blue. Genotype: recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F′ proAB laclq ZΔM15 Tn10 (Tetr)] (Agilent, 200249).

2.2.1.2.2 LB broth and LB agar plates

To prepare one litre of Luria-Bertani broth (LB) media, 25 g of LB powder (LB, Sigma Aldrich, L3022) was dissolved in double-distilled water (ddH_2O) and mixed to homogenise. LB broth was autoclaved for 15 minutes at 121°C for sterilisation and stored at 4°C.

To prepare LB agar for agar plates, 2 g of LB powder and 2 g of bacterial agar (Sigma Aldrich, A5306) were added to 100 ml ddH_2O and mixed to complete dissolving. The solution was sterilised by autoclaving at 121°C for 15 minutes and stored at 22°C if not used immediately. To prepare the agar plates, 15 – 20 ml was poured into 10 cm Petri dishes close to Bunsen flame. The plates were left on the bench at 22°C to solidify for 15-20 minutes.

Before using the LB broth or pouring the agar, an appropriate dose of a specific antibiotic was added as suggested in table 2-2.

Table 2-2 Selecting antibiotics

Antibiotic	Recommended Concentration
Ampicillin	100 µg/mL
Kanamycin	50 µg/mL

2.2.1.2.3 Preparation of competent cells

E. coli XL1 blue were resuspended in 700 µl of LB broth. The tube was placed in a shaking incubator for 45 – 90 minutes at 37°C. Next, the tube was centrifuged in an Eppendorf microcentrifuge 5415 – (1000 rpm) for 1- 3 minutes to make a concentrated vial. Most of the supernatant was discarded by simple inversion, while the left over was used to resuspend the pelleted bacteria and the cells were spread on a 50 µg/ml Tetracycline-supplemented agar plate. The inverted plate was incubated overnight at 37°C. The following day, a single colony was picked and inoculated into 5 ml of LB broth media supplemented with 50 µg/ml Tetracycline. The tube was placed in a shaking incubator where it was left to grow overnight at 37°C. The following day, the whole culture was transferred to a 500 ml LB broth with 50 µg/ml Tetracycline. The bottle was placed in a shaking incubator overnight at 37°C. The next day, 50 ml of the culture was transferred into a pre-chilled falcon tube and incubated on ice for 10 minutes. After that, the tube was centrifuged at 3000 xg for 10 minutes at 4°C. The supernatant was discarded, and the pellet was resuspended in 5 ml ice-cold 0.1 M $CaCl_2$ ($CaCl_2$, Sigma-Aldrich®, 449709) and again placed on ice to chill for 10 minutes. The suspension was centrifuged again at 3000 xg for 10 minutes at 4°C. Again, the supernatant was discarded, and the pellet was resuspended in 1 ml ice cold 0.1 M $CaCl_2$ and incubated on ice for 30-40 minutes (up to 1 hour). 1 ml of sterile 100% Glycerol (Fisher Scientific, 10021083) was added into the suspension, mixed and aliquoted as 100 µl into prechilled Eppendorf tubes. The bacteria were snap frozen in liquid nitrogen before storage at -80°C.

2.2.1.2.4 Bacterial transformation

An aliquot of the competent cells was placed on ice to thaw. DNA plasmids were added to the competent cells (0.5 – 1 µg of DNA/100 µl of cells) and left on ice for 30 minutes. After incubation, the mixture was placed for 45 seconds (up to 2 minutes) in a heating plate for a heat shock at 42°C then immediately placed on ice for 2 minutes (up to 5 minutes) to allow DNA incorporation. 700 µl of LB broth was added to the cells. The tube was incubated in a shaker for 1 hour at 37°C. Later, the tube was centrifuged in an Eppendorf microcentrifuge 5415 – (5000 rpm) for 1 minute. Most of the supernatant was discarded by simple inversion and the formed pellet was resuspended in the leftover LB broth. The mixture was spread on an agar plate supplemented with 50mg/ml of Ampicillin (Rosch, 10835242001) or 10mg/ml Tetracycline as required, and left in a 37°C incubator overnight.

The following day, A single colony was inoculated into 5 ml LB broth with antibiotic and placed in a 37°C shaking incubator overnight. After that, the whole suspension was transferred into 250 ml of antibiotic-supplemented LB broth. The subculture was incubated again overnight before it was ready for plasmid purification.

2.2.1.2.5 Plasmid purification

To purify the plasmids, a Maxi prep kit (Qiagen®, 12163) was used following the protocol provided from the supplier. Briefly, the bacterial culture collected from the previous transformation was pelleted by centrifugation at 6000 xg for 15 min at 4°C. The formed pallet was then resuspended in 10 ml of a lysis

buffer P1. After that, an equal volume of buffer P2 was added, and the solution was mixed by inversion. The mix was incubated at 22°C for 5 minutes. After the incubation, 10 ml of buffer P3 was added to the mix and left to incubate on ice for 20 minutes. This was followed by two consecutive 4°C centrifugations: first at 14,000–18,000 xg for 10 minutes, then the supernatant from the first at ≥20,000 xg for 15 minutes. A Qiagen column was equilibrated by using a 10 ml of QBT buffer. The supernatant from the centrifugation above was transferred to the column to flow by gravity into the column's resin. The column was washed twice by passing 30 ml of buffer QC. DNA was eluted using a 15 ml of buffer QF and precipitated by adding 10 ml of isopropanol. The mix was then centrifuged at 4°C at ≥15,000 xg for 30 minutes. The pellet was washed with 5 ml of 70% ethanol and centrifuged at ≥15,000 xg for 10 min. The DNA pellet was air dried for 10 minutes and finally dissolved in an appropriate volume 10 mM Tris·Cl, pH 8.5 buffer. All buffers were supplied in the kit.

Table 2-3 QIAGEN plasmid Maxiprep Kit buffers

Buffer	**Purpose**
P1	Re-suspension/RNase A buffer
P2	Lysis buffer
P3	Neutralisation buffer
QBT	Equilibration buffer
QC	Washing buffer
QF	Elution buffer

2.2.1.2.6 DNA quantification

To quantify the purified plasmid DNA, first, the absorbance at 260 nm was obtained using a nanodrop UV spectrophotometer (NanoDrop™ 2000 UV – Thermo Scientific). The concentration was measured by applying the formula (A260 of 1.0 = 50 μg/ml pure DNA plasmid). To check purity, the ratio of A260 / A280 was obtained; aiming for a value between 1.9 and 2.1 which indicated a highly pure plasmid DNA sample.

2.2.1.3 Plasmid verification

2.2.1.3.1 Diagnostic restriction digest

To perform plasmid restriction digestion, four sterile PCR tubes were prepared as in the example table 2-4.

Table 2-4 Diagnostic restriction digest experiment preparation

Component	Uncut plasmid	EcoRI	KpnI	Cut with both enzymes
DNA (μg/μL)	1	1	1	1
10X Buffer (μL)	2	2	2	2
Restriction enzyme 1 (μL)	--	1	--	1
Restriction enzyme 2 (μL)	--	--	1	1
Distilled water (μL)	17	16	16	15

The tubes were incubated for 1 hour at 37°C. The reaction was stopped by heat inactivation at 65°C for 15 minutes. All the restriction enzymes and buffers were purchased from BioLabs.

2.2.1.3.2 2% Agarose gel for DNA electrophoresis

A 2% agarose gel was prepared by dissolving agarose (Sigma Aldrich, A9539) in Tris-acetate EDTA (TAE) buffer composed of 40mM Tris-acetate (Sigma Aldrich, T1258) and 1mM EDTA (Sigma Aldrich E1644). The 2% agarose solution was microwaved until it became clear and left to cool down on the bench. 0.5 µg/ml Ethidium Bromide (Promega, H5041) was added to the solution and mixed by swirling. Agarose solution was poured into a gel tray with a well-comb in place and left to solidify at 22°C.

2.2.1.3.3 DNA gel electrophoresis

The comb was removed from the gel and the tray was placed in an electrophoresis tank containing TAE buffer. DNA gel loading dye was added to each restriction reaction mix. A 5 µl of DNA hyper ladder (Bioline, BIO-33056) was added to the first lane in the gel. Uncut plasmid was loaded into the second lane. Lanes 3 and 4 were used for plasmids cut with a single restriction enzyme. Finally, the reaction representing plasmid with both restriction enzymes was loaded in lane 5. The voltage was adjusted to 100 V and left to run until loaded samples were 80% away from the wells. The gel was imaged using a Gel Doc™ EZ Gel Documentation System from Biorad. The bands were analysed using the software Image Lab™. Bands were examined for backbone and insert sizes.

2.2.1.4 pEGFP-PRH-Myc transfection and selection

2.2.1.4.1 Transfection

To produce an AKN1 cells expressing GFP-PRH-Myc, $3X10^5$ cells/well were seeded in a 6-well plate and left to adhere overnight in a humidified environment with 5% CO_2 at 37°C. When cells reached 60-70% confluence, they were considered ready for transfection. Branched polyethylenimine (PEI, Sigma Aldrich, 408727) was used at 6:1 ratio (PEI: DNA) for transfection. A working PEI solution was prepared by making a stock of 1 mg/ml. The pH was adjusted to 7.5. The solution was sterilised by filtration. Aliquots were stored at -20°C. PEI transfection protocol is summarised in table 2-5.

Table 2-5 AKN1 PEI transfection

	pEGFP	**pEGFP-PRH-Myc**
OptiMEM media (µl)	100	100
DNA GFP (µg)	2	--
DNA GFP-PRH-Myc (µl)	--	2
25 mg/ml PEI (µl)	12	12
Incubation (minutes)	10	10
Complete media (µl)	600	600
Transfection mix on cells (µl)	714	714
Incubation with transfection mix (hours)	2-3	2-3
Complete media (ml)	2	2

First, a mix of serum-free media (Opti-MEM, 31985062, Gibco™), DNA and PEI was prepared and immediately pulse-vortexed for 15 second. The mix was incubated for 10 minutes at 22°C. After incubation, 600 µl of complete media was added to both tubes and mixed thoroughly. This transfection mix was used to replace the old media for AKN1 cells prepared earlier. 2-3 hours later, 2 ml of fresh CM was added to all wells. The plate was incubated in a humidified environment with 5% CO_2 at 37°C. 24-hour post-transfection the cells were inspected for GFP expression using Leica DMI3000 B fluorescence microscope at 10X magnification.

2.2.1.4.2 Geneticin (G418) selection

5 X 10^4 AKN1 cells were seeded in a 24-well plate and left to adhere overnight in an incubator with 5% CO_2 at 37°C. A range of G418 doses (G418, Sigma-Aldrich®, A1720) was added to the cells after reaching 70% confluence, 2 wells / dose. A pair of wells was grown free of G418 as a control. The plate was inspected visually every day for signs of cell death. Media with or without G418 was replaced every 3 days. The optimal dose of G418 for selection was determined as the lowest dose required to kill all the cells over 5-7 days.

Transfected AKN1 cells were selected with 750 ug/ml Geneticin antibiotic. This eliminated the cells that did not express GFP and maintained the growth of cells expressing *neoR* for antibiotic resistance. The same growth and maintenance protocols were applied after transfection with the addition of a maintenance dose of G418 to the conditioning medium.

2.2.1.5 Fluorescent-Activated Cell Sorting (FACS) for monoclonal/polyclonal stable cell line

Flow cytometry is a technology that enables studying cell type and status by analysing the cell's response to fluorescent illumination. A single-cell suspension is passed through a fluorescent light beam that the cells scatter differently depending on various factors including fluorescent labelling, or fluorescent protein expression. Therefore, the cells can be counted and sorted into different groups (409).

Wild type AKN1, AKN1 GFP and AKN1 GFP-PRH-Myc cells were trypsinised and quantified to make suspensions of $2X10^7$ cells/ml and 5 $X10^6$ cells/ml, respectively. New labelled Eppendorf tubes or 96-well plates were prepared with fresh CM to receive the single-sorted cells. All vessels were transferred on ice to the flow cytometry facility. Wildtype AKN1 cells were used as a reference for cells without GFP using MoFlo™ XDP cell sorter. Scatter-gating steps were applied to isolate single GFP-expressing cells. GFP protein was excited using 488 nm laser, and the fluorescent signals were detected using a 528/29 filter. Single-sorted cells were collected in 96-well plates containing fresh media, while polyclonal-sorted cells were collected in Eppendorf tubes. The cells were immediately taken back to the cell culture hood where polyclonal-sorted cells were suspended in an appropriate-sized flask. The cells were incubated in a humidified environment with 5% CO_2 at 37°C for growth expansion. Cell maintenance and GFP expression imaging were carried out when needed. Data were analysed using Kaluza software and represented in histograms and density plots.

2.2.1.6 Limiting serial dilution

Limiting dilution of cells is a cost-effective and convenient method to isolate single cells (410). It enables the separation of individual cells based on various factors such as GFP expression. It relies on preparing high dilutions of cell suspensions to allow the transfer of a single cell per aliquot with simple pipetting. Due to the high chance of having multiple cells per aliquot, this technique might need to be repeated to achieve a true single-cell / well (411).

Adherent cells were trypsinised and neutralised with equal volume of CM. The cells were pelleted and resuspended in fresh CM at a concentration of 0.5-0.8 cell / 100 µl. 100 µl of the suspension was transferred into each well of 96-well plate. This would increase the chances of having a single cell with each 100 µl aliquot. The cells were incubated in a humidified environment with 5% CO_2 at 37°C for growth expansion.

2.2.2 Vimentin knockdown cells

AllStar negative control siRNA (1027418, Qiagen) and *VIM5* siRNA (Hs_VIM_5 FlexiTube siRNA, NM_003380, Qiagen) were used to transfect CCLP1 and CCSW1 cells. RNAiMAX lipofectamine was used to produce CCLP1 and CCSW1 Vimentin knockdown cells. The protocol for transfection is summarised in table 2-6 and (figure 2-1).

Table 2-6 si*VIM5* transfection

	siControl (AllStar negative)	**siVIM5**
Opti-MEM (ml)	2	2
50 nM siControl (µl)	2	--
50 nM siVIM5 (µl)	--	2
RNAiMAX (µl)	25	25

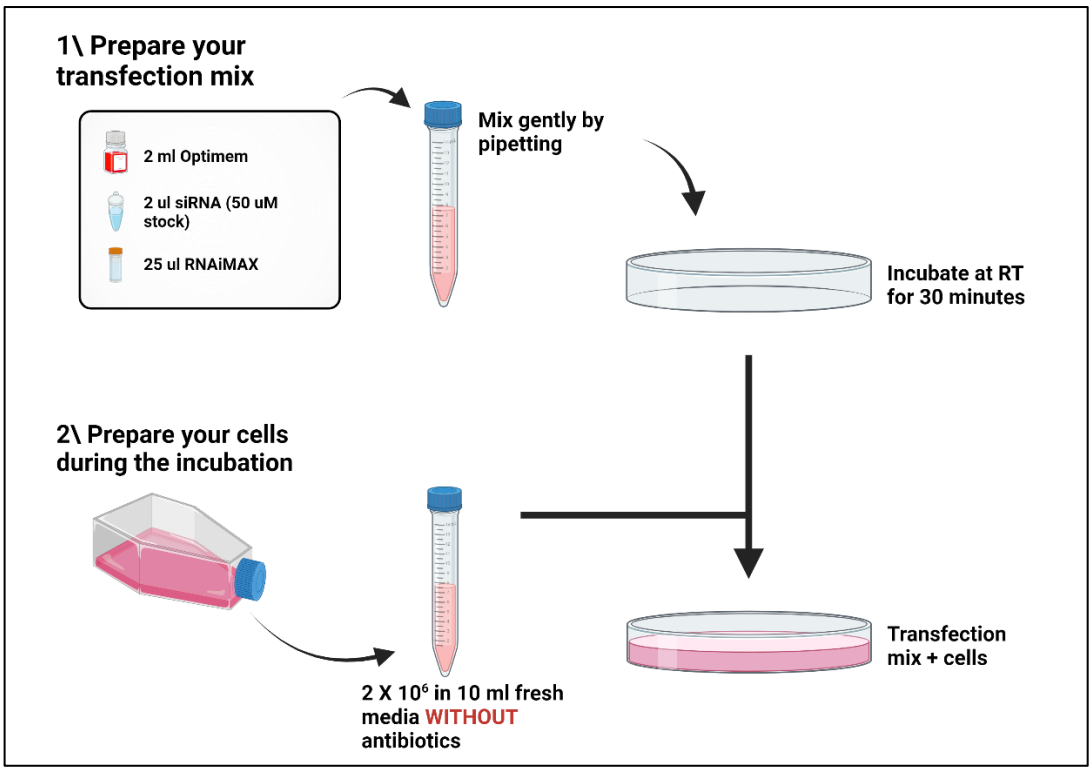


Figure 2-1 si*VIM5* transfection protocol.

Created with BioRender.com

In brief, two 15 ml falcon tubes were labelled as siControl and si*VIM5*. The transfection mix for each tube consisted of Opti-MEM media, siRNA and RNAiMAX lipofectamine (ThermoFischer, 13778150). The contents were mixed gently, and immediately transferred to their respective labelled vessels. Both vessels were incubated for 30 minutes at 22°C. During incubation, cells were trypsinised and quantified to make a suspension of 2 X 10^6 cells in 10 ml of growing media without antibiotics for each condition. After the 30 minutes incubation of the transfection mix, cells were added to their respective plates and incubated in a humidified environment with 5% CO_2 at 37°C for 48 hours. After 48 hours incubation, cells were trypsinised and pelleted for RNA

extraction, cDNA preparation and real time quantitative polymerase chain reaction (RT-qPCR). Cells were also used to seed for experiments.

2.3 Microscopy

2.3.1 Fluorescent microscopy

2.3.1.1 Live cell imaging

Cells grown in flasks or plates were examined for GFP expression using a Leica DMI3000 B fluorescence microscope choosing filters for brightfield and GFP. After obtaining the images, cells were immediately placed back in the incubator. The software “FIJI is just image J” (FIJI) was used to process and analyse the images.

2.3.1.2 Fixed cell imaging

Cells were grown on 19 mm diameter coverslips (Fisherbrand™, 12323138) and incubated overnight in a humidified environment with 5% CO_2 at 37°C to adhere. The following day cells were gently washed with PBS three times for 5 minutes. For fixation 4% Formaldehyde (FA, Thermo Scientific™, 28908) was prepared (FA/PBS, v/v), and was used to fix the cells for 10 minutes. After fixation, cells were washed with PBS three times for 5 minutes. For nuclear staining, 4′,6-diamidino-2-phenylindole (DAPI) (Sigma®, D9542) was prepared as DAPI:PBS (1:1000, v/v) using a 1mg/ml DAPI stock. The solution was added on the cells for 15 minutes. Cells were washed twice in PBS for 5 minutes. Coverslips were mounted with immu-mount medium (Fischer Scientific, 10662815) containing bisbenzimide (Sigma-Aldrich®, 33258) on glass slides (Fischer Scientific, 12372098). Slides were left to dry overnight in

the fridge. Images were obtained with a Leica DMI3000 B fluorescence microscope choosing the appropriate filters, using the same exposure time, shuttle opening and gain for all pictures. After taking the plate out of the incubator, all steps were performed at 22°C.

2.3.2 Immunofluorescence staining and imaging

2.3.2.1 Cell fixation, staining and imaging

Coverslips were sterilised in 100% ethanol for a minimum of 20 minutes and left to dry inside cell culture hood. The coverslips were placed in the wells of a 12-well plate. $50X10^3$ – $80X10^3$ cells were seeded on coverslips. The plate was incubated in a humidified environment with 5% CO_2 at 37°C for the cells to adhere overnight. The next day, old media was removed, and the cells were washed gently with PBS. Cells were fixed with 4% FA for 10-50 minutes. When FA was removed, cells were washed in PBS three times for 5 minutes. If cells were not to be stained immediately, they were stored with PBS supplemented with 0.02% sodium azide (NaN_3, Sigma Aldrich, 71289) in 4°C. After washing with PBS, cells were permeabilised for 15 minutes with 0.01-0.02% Triton® X-100 (Sigma Aldrich, T8787) prepared in PBS (v/v). To remove all traces of Triton® X-100, cells were washed four times with PBS for 5 minutes. After that, cells were blocked with 3% Bovine Serum Albumin (BSA, Sigma-Aldrich®, A9418) (BSA/PBS, w/v) for 30 minutes. Primary antibodies were diluted in 3% BSA as shown in table 2-8. Cells were incubated with primary antibody overnight in 4°C (or, for 2 hours at 22). The following day, cells were washed four times with PBS for 5 minutes. Secondary antibodies were prepared in the

blocking solution as shown in table 2-8. Cells were incubated with secondary antibodies for 50 minutes in room temperature. After the addition of secondary antibodies, the plate was covered with tin foil to protect from light exposure. Cells were washed four times with PBS for 5 minutes. For nuclear DAPI stain, cells were incubated in DAPI/PBS (1:1000, v/v) for 15 minutes in 22°C. DAPI/PBS solution was discarded, and cells were washed three times with PBS for 5 minutes. Coverslips were mounted with immu-mount medium containing bisbenzimide. Slides were left to dry overnight in 4°C. The following day, coverslips were sealed with fast dry transparent nail polish. Images were obtained with either a Leica DMI3000 B fluorescence microscope or a confocal Leica DMI4000B microscope.

Table 2-7 List of immunofluorescence staining buffers

Purpose	Buffer	Preparation
Fixation	4% FA	1:4 dilution of 16% FA stock in PBS
Washing	1X PBS	1X PBS
Blocking	3% BSA	1.5 g of BSA 50 ml of PBS
Permeabilisation	1X PBS+ 0.05% Triton	50 ml PBS 25 µl triton
Preservation	1X PBS + 0.02% NaN3	PBS supplemented with 0.02% NaN3 (1/500 dilution of 10% stock)

Table 2-8 List of immunofluorescence antibodies

Antibody	**Dilution (v/v)**	**Supplier**
Monoclonal Anti-HHEX antibody produced in mouse	1:500	Sigma Aldrich, SAB1403914
M3	1:5000	Dr. Sheela Jayaraman
Mouse monoclonal [OTI3E6] to Hex	1:100	Abcam, Ab117864
Rabbit polyclonal to Hex	1:500	Abcam, Ab34222
HHEX Antibody (2018B) [Unconjugated]	1:500	Novus Biologicals, 2018B
HHEX Antibody (906705)	1:500	Novus Biologicals, 906705
Vimentin (D21H3) XP® Rabbit mAb	1:500	Cell Signalling, 5741
Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody TRITC	1:1000	Thermo Fischer, A16040
Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	1:1000	Thermo Fischer, A-21206
Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 555	1:1000	Thermo Fischer, A-31570
Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor Plus 488	1:1000	Thermo Fischer, A32766
Goat anti-Rabbit IgG-H+L Cy5™ Conjugated	1:1000	Thermo Fischer, A10523

Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Cyanine5	1:1000	Thermo Fischer, A10524
Anti-Rabbit IgG (H+L), F(ab′)2 fragment, CF™680 antibody produced in goat	1:1000	Sigma Aldrich, SAB4600362
Anti-Mouse IgG (H+L), F(ab′)2 fragment, CF™680 antibody produced in goat	1:1000	Sigma Aldrich, SAB4600361
Texas Red™-X Phalloidin	1:40	(Thermo Fischer, T7471)

2.3.3 Photo-activated localisation microscopy – direct stochastic optical reconstruction microscopy (PALM dSTORM) immunofluorescence staining

2.3.3.1 Coverslip preparation

19 mm diameter round glass coverslips with high adherability surface (zeiss, 474030-9000-000) were cleaned prior to seeding cells for immunofluorescence PALM sSTORM. The entire process was carried out inside a fume-hood. A detergent solution consisted of 5:1:1 ration of H_2O:H_2O_2:NH_4OH was prepared and stored in glass bottles inside a flammables store cabinet. Coverslips were placed on corrosive-resistant rack (Mini rack, Thermo Fischer, C14784). The entire rack was submerged in a beaker filled with the detergent solution. This beaker was placed inside a bigger beaker containing water at a higher level than the coverslips.

Everything was covered with tin foil and placed on a hot plate prepared to keep the water at 70°C for 2-3 hours. During this incubation, the tin foil was slowly turning black. With a metal tong, the coverslips rack was removed and placed in a beaker filled with fresh water. The coverslips were washed 3 times by moving the rack up and down replacing the water in between. Another wash was done by submerging the rack in 100% ethanol. After that, coverslips were left to dry on clean tissue paper. Excess ethanol was removed by manually shaking every coverslip in air using metal tongs. To sterilise the coverslips for cell culture, they were quickly (less than a second) passed through a flame 3 times. Finally, coverslips were stored on the rack wrapped in tin foil to protected them from dust.

2.3.3.2 Cell preparation and staining

Cells were seeded on treated coverslips from the previous section and left overnight to adhere. On the following day, old media was discarded, and the cells were washed twice with PBS. Cells were fixed with 4% formaldehyde in PBS for 12 minutes. After fixation, cells were washed with PBS three times for 5 minutes to remove all the traces of the fixing reagent. To avoid autofluorescence from residual formaldehyde, 50 mM NH_4Cl (sigma Aldrich, A9434) was added. Cells were incubated for 15 minutes. Coverslips were washed three times with PBS for 5 minutes. For simultaneous permeabilization and blocking, cells were incubated for 30 minutes with 3% (w/v) BSA in PBS supplemented with 0.1% triton X-100. After that, the cells were immediately incubated with a primary antibody prepared in the same buffer used in the previous step. Cells were left to incubate overnight in the

cold room at 4C. The following day, coverslips were washed three times with 0.2% BSA, 0.05% triton X-100 for 10 minutes. Appropriate Alexafluor-conjugated secondary antibodies were prepared in the same buffer used in the previous step. Cells were incubated with secondary antibody for an hour at 22C. Cells were washed three times with 0.2% BSA supplemented with 0.05% Triton x-100 for 10 minutes. Another wash was performed using 1X Tris Buffer Saline (TBS, Sigma Aldrich, 6664) followed by 1X PBS for 10 minutes each. Cells were post-fixed with 4% formaldehyde in PBS for 10 minutes then washed with PBS three times for 10 minutes each wash. The last washing solution was replaced with PBS supplemented with 0.02% NAN_3 for prolonged storage of the coverslips. Cells were left at 4°C until imaging.

Table 2-9 List of PALM dSTORM buffers

Purpose	Buffer	Preparation
Fixation	4% FA	1:4 dilution of 16% FA stock in PBS
Quenching	50 mM NH4Cl	669 mg 250 ml water Filter
Blocking and permeabilization	3% BSA + 0.01% Triton	1.5 g of BSA 50 ml of PBS 50 µl triton
Washing	0.2% BSA + 0.05% Triton	100 mg BSA 50 ml PBS 25 µl triton
Washing	1X TBS	1X TBS
Washing	1X PBS	1X PBS
Preservation	1X PBS + 0.02% NaN_3	PBS supplemented with 0.02% NaN_3 (1/500 dilution of 10% stock)

Table 2-10 List of PALM dSTORM antibodies

Antibody	Dilution (v/v)	Supplier
Monoclonal Anti-HHEX antibody produced in mouse	1:500	Sigma Aldrich, SAB1403914
Vimentin (D21H3) XP® Rabbit mAb	1:500	Cell Signaling, 5741
Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	1:1000	Thermo Fischer, A-21206
Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Cyanine5	1:1000	Thermo Fischer, A10524

2.3.3.3 Coverslips preparation for imaging

Prior to imaging, cells on coverslips were incubated for 15 minutes at 22C with fiducial fluorescent beads prepared in PBS (TetraSpeck™ Microspheres, 0.1 µm, fluorescent beads, Invitrogen, T7279).

An imaging buffer was needed for PALM dSTORM acquisition. This buffer helps maintain the blinking ability of the fluorophore used to label the protein of interest. The fluorophore was excited, and a light was emitted. Then, the fluorophore goes into a dark state. This process is repeated sequentially until all the particles were imaged. Generally, three buffers were prepared. Buffer

A (10 mM Tris (pH8) + 50 mM NaCl), Buffer B (50 mM Tris (pH8) + 10 mM NaCl + 10% glucose (Sigma Aldrich, G2133-50KU)), and GLOX buffer (250 ul) composed of (10 mg Glucose oxidase (Sigma Aldrich, G2133-50KU))+ 50 µl Catalase (17 mg/ml) (Sigma Aldrich, C40-100MG) + 200 µl buffer A). On the day of imaging, a mix of buffer B, GLOX and BME was prepared on ice.

For each sample coverslip, a double-sided sticky spacer (Grace Bio-Labs SecureSeal™ imaging spacer, Sigma Aldrich, GBL654002) was placed on a glass slide. This spacer acted as a small reservoir to contain the buffer mentioned above. Sample coverslips were washed by submerging them quickly in a beaker full of distilled water. They were left to stand and dry on paper tissues for 30 seconds. The edges of sample coverslips were wiped with a cotton bud. 10 µl of the imaging buffer was added on the spacer. Then, coverslips were inverted on the spacer with cells facing down.

Table 2-11 List of PALM dSTORM imaging buffers

Buffer	Makeup	Preparation
Buffer A	10 mM Tris (pH8) + 50 mM NaCl	• 158 mg TrisHCl • 100 ml distilled water • Adjust pH 8 • 292 mg NaCl
Buffer B	50 mM Tris (pH8) + 10 mM NaCl + 10% glucose	• 788 mg TrisHCl • 100 ml distilled water • Adjust pH to 8 • 58 mg NaCl • 10 gm glucose

GLOX	Glucose + Catalase + Buffer A	• 10 mg Glucose oxidase • 50 µl Catalase (17 mg/ml) • 200 µl Buffer A • Vortex to dissolve glucose oxidase • Spin down 14000 rpm • Only use supernatant • Keep on ice • Store at 4 for 2 weeks • Everytime to use, spin down 14000 rpm

Imaging buffer	GLOX + Buffer B	• On ice
		  ○ 7 µl GLOX
		  ○ 7 µl BME
		  ○ 960 µl Buffer B
		Add enough on coverslips (10 µl)

2.3.3.4 Imaging

A Zeiss Elyra PS.1 system for super resolution imaging with ZENBlack software was used to image the slides. First, the view was adjusted while the definite focus was off. 488 laser excitation was used to screen cells and adjust the focus. Once the focus and the field were satisfying, the slide was fixed on the stage with a clip and the image was re-focused. The green laser was turned off to avoid bleaching the sample. The acquisition mode was changed to continuous, and the slide was put under definite focus and a snap image was captured. Next, the acquisition mode was changed to HiLO followed by imaging a wide-field view then a nuclear view. The microscope settings were

then adjusted to record a STROM image. 20,000 to 25,000 frames were taken with continuous red laser excitation first then green laser. Images were saved in CZI format.

2.3.3.5 PALM dSTORM image analysis

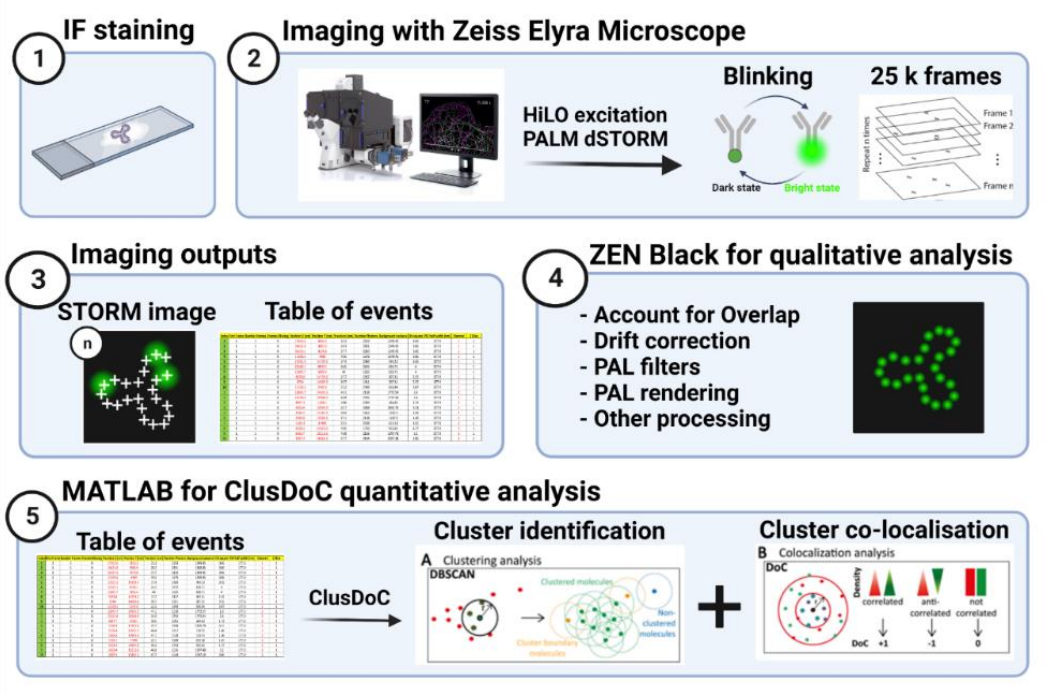


Figure 2-2 Summary of PALM dSTORM imaging and analysis (412)**.**
Created with BioRender.com

2.3.3.5.1 Qualitative analysis

ZENBlack software was used to produce rendered images for display and to generate data for quantitative analysis in Mathworks MATLAB. PLAM rendering parameters were adjusted. First, molecules settings were adjusted to account for molecules overlapping. In addition, signal to noise ratio was set at 6 to accommodate for background noise. Also, under localisation settings both localisation algorithms and Gaussian filter were selected. Frames were

scanned rapidly under the previous settings. Once satisfied, changes were applied, and a super resolution image was produced. This super resolution image was also rendered for the best visual presentation. ZENBlack has several filters that could be applied for that purpose. First, drift was corrected according to the fiducial bead if they were used, or automatically in the absence of beads. Under the statistics tab, image precision was checked while observing precision histogram. Poorly localised molecules must be removed to ensure high localisation accuracy. Particles with 5 to 75 nm precision were kept. Under PAL render, settings were applied appropriately to define the final image appearance. Both precision table and histogram were exported for quantitative analysis. Images were snapped at different magnifications to show molecules clustering. All the previous processing was done for the red and green channels separately.

2.3.3.5.2 Quantitative analysis

Precision tables were imported into MATLAB. The package for co-clustering analysis for single molecule localisation microscopy (SMLM) data (ClusDoC) was downloaded from (https://github.com/PRNicovich/ClusDoC) (412). Additional toolboxes were downloaded from MATLAB library include Image processing toolbox, Parallel computing toolbox and Statistics and machine learning. These tools ensured the optimum performance of the ClusDoC package.

To start quantitative analysis, ClusDoC package path was added and opened in MATLAB. The tool was initiated by inserting the script ClusDoC in the script window. Then, the localisation table was uploaded. Followed by choosing an

output folder for the results. Then using the tool selection, regions of interests (ROI) for analysis was traced (nuclear). This region could be taken as several parts rather than choosing the whole nuclear area all at once to prevent MATLAB crashing. Once the area was selected, the file of selection could be saved in case it was needed for repeating the analysis. *Ripley's K* was started to generate *Max R* for both channels. If the ROI was divided, then the average *MAX R* was used for DBSCAN. The parameters for *Ripley's K* were adjusted to start at 0 nm to 1000 nm and calculate *Ripley's K* every 10 nm steps. The value for maximum radius (Max r), which represented the radius with the highest probability of aggregations, was recorded for the following ClusDoC step.To scan PRH-Vimentin for clusters and characterise their clustering, we used the tool ClusDoC. This tool can perform two tasks simultaneously: cluster scanning with DBSCAN, and co-loclaisation with DoC. The parameters under the **DoC** were applied similarly on both channels. *L(r)-r* radius was calculated from the *Ripley's* function step. Co-localisation threshold was a standard value of 0.4 (413). The minimum number of co-localised clusters was chosen as 10. The remaining were left unchanged (412, 413). The **parameters for DBSCAN** were adjusted for each channel independently. Epsilon (nm) input was equal to the peak value of the localisation precision histogram for each fluorophore min (Pts) was adjusted as 5. Plot cutoff and parenthesis threat were kept as default, *L(r) -r* Radius (nm) value was taken from the *Ripley's* function step. Finally, the smooth radius was made at 7 nm. The outcome was exported as txt file for EXCEL analysis. In EXCEL, the average of all images was calculated and presented in violin plots.

2.3.4 Fluorescent correlation spectrometry

Fluorescent correlation spectrometry (FCS) is a tool to study molecular dynamics via analysing the intensity fluctuation of diffused fluorescent molecules through a focused light (414). It can provide details on the concentration, diffusion, and brightness properties of cellular fluorescent components.

Different cell conditions were seeded on a Nunc Lab-tek chambered cover glass 8-well chamber (155361, Thermo Scientific). Each condition was seeded in 4 different densities (5 – 10 – 20 – 30) $X10^3$ cells/well. The cells were incubated in a humidified environment with 5% CO_2 at 37°C to adhere overnight. The following day, the media was discarded, and the cells were washed gently with PBS. To image the cells for FCS, PBS was replaced with a 180 µl of HEPES-buffered saline solution (pH 7.45; Sodium pyruvate 2 mM, NaCl 145 mM, D-Glucose 10 mM, KCl 5 mM, $MgSO_4.7H_2O$ 1 mM, HEPES 10 mM, $CaCl_2$ 1.7 mM, $NaHCO_3$ 1.5 mM). The buffer was left on the cells for 10 minutes before imaging took place. Acquisition was done at 24°C to minimise cell movement and using a Zeiss LSM880 microscope on a Zeiss Axio Observer Z1 stand. Initial confocal images of the cells were acquired using a 488 nm laser for excitation. Images were processed using ZENBlack software (Carl Zeiss, Germany). GFP fluctuations were recorded following the method published by Goulding et al. (415). Correlation analysis from GFP dynamics data were analysed using ZENBlack software, and presented in Z-intensity scan, correlation 2x3D model.

2.4 Western blotting

2.4.1 Whole cell protein extraction and quantification

Cell pellets were resuspended in cold working solution lysis buffer consisted of 1X cell lysis buffer (table 2-12 for recipe), 1X cOmplete™ Protease Inhibitor Cocktail (Roche, 11697498001), 1XPhoStopTM (Roche, 4906845001), and 1X Phenylmethanesulfonyl fluoride (PMSF, Sigma-Aldrich®, P7626). The suspension was sonicated, while kept in iced water, using Ultrawave™ for 12 minutes. Alternatively, up and down pipetting with a fine tip or small gauge syringe needle was used. After that, the suspension was centrifuged at maximum speed for 30 minutes at 4°C. The supernatant was transferred to a new Eppendorf tube and stored at -80 °C.

2.4.2 Protein quantification

2 µl of the extract was used for protein quantification using a protein assay reagent (Bio-Rad®, 5000001) following the Bradford protocol (416). In brief, protein extract was diluted in distilled water. BSA stock solution was used to make serial dilutions stocks used to generate a BSA standard curve from which the concentration of proteins in the extract was interpolated. Bradford reagent was diluted 1:4 in distilled water (V/V). 10 µl of diluted lysate and BSA serial dilutions were transferred into separate wells of a pre-labelled 96-well plate in triplicate. 200 µl of working Bradford solution was added to all the wells. Extra blank wells were added to correct for the background later. The plate was read using a Tecan micro plate reader. The concentration of the unknown samples was interpolated from the BSA standard curve.

2.4.3 Loading sample preparation

For future sample loading, equal amounts of protein lysates were prepared with a 1X loading dye (table 2-12 for recipe) and distilled water. The samples were heated for 5 minutes at 95°C just before loading.

2.4.4 Sodium dodecyl sulphate - poly acrylamide gel electrophoresis

2.4.4.1 Protein loading

Equal amounts of proteins ranged between (5-100 µg) were loaded into washed wells of 10% Mini-PROTEAN® TGX™ precast protein gels (Bio-Rad®, 4561033). A 3 µl Precision-Plus Protein™ Dual Colour (Bio-Rad®, 1610374) was used as a reference protein ladder. The gels were run in SDS PAGE tank filed with a running buffer [0.25 M Trizma (Sigma Aldrich, T1503), 2 M Glycine (Severn biotech ltd®, 30-21-70), 10% SDS (SigmaAldrich, L3771)] at 100 V until the required level of protein separation was reached.

2.4.4.2 Transfer

Trans-Blot® Turbo™ RTA Mini PVDF Transfer Kit (Bio-Rad®, 1704272) was used to transfer the protein into Polyvinylidene difluoride (PVDF) membrane pre-activated in methanol (Sigma-Aldrich®, 179957). A commercial transfer buffer from Biorad, diluted in 100% ethanol and distilled water, was used for the transfer. A sandwich of stacks, membrane, gel, stacks was arranged. The transfer process was carried out using a Biorad Turbo transfer system with mixed molecular weight settings.

2.4.4.3 Blocking

The membrane was blocked in a 5% BSA prepared in 0.01% TBS-T [15 mM NaCl (Sigma-Aldrich®, 378860), 2 mM Tris-HCl (Sigma-Aldrich®, 10812846001), pH: 7.6, 0.01% Tween (Sigma-Aldrich®, P9416)] for either 1-2 hours at 22°C, or overnight at 4°C. This was performed to block non-specific protein binding.

2.4.4.4 Probing and washing

The next step was to probe the membrane with primary antibodies (table 2-13) diluted in the blocking solution. The incubation lasted for either (1-1.5 hour) at 22°C, or overnight at 4°C. To remove unbound antibody, the membrane was washed three times, for 10 minutes in TBS-T. Secondary fluorescent antibodies (table 2-13) were prepared in the blocking solution and the membrane was incubated in that solution for 1 hour at 22°C protected from light. Alternatively, HRP-conjugated secondary antibodies in 5% BSA were also used. The membrane was washed with TBS-T for 10 minutes three times. When HRP-conjugated antibodies were used, the membrane was further treated with a chemiluminescence blotting substrate (Rosch, 11520709001).

2.4.4.5 Imaging, densitometric measurement and statistical analysis

Membranes were imaged using Odyssey Fc imaging system (LI-COR Bioscience, UK). Densitometric quantification of the bands was performed using the Image Studio-Lite 5.2 software (LI-COR Biosciences Ltd. UK). Intensity values for protein of interest were normalised to the intensity of

housekeeping proteins. Results were presented as fold changes to show relative protein expression.

2.4.4.6 Membrane stripping

Whenever a membrane was used for a second round of probing, it was first stripped of the previous probing using a stripping buffer insisted of 0.4 M NaOH in distilled water. The membrane was incubated for 15 minutes in the stripping buffer at 22°C. That was followed by three washes with TBST, 5 minutes each. Then, the membrane was blocked as usual in 5% BSA before re-probing with a primary antibody.

Table 2-12 List of Western blotting buffers

Buffer	Purpose	Preparation
RIPA Lysis buffer (10X LB)	Whole cell lysis buffer	• 3 ml of 5 M Sodium chloride (150 mM final concentration) • 5 ml of 1 M, pH 8.0 Tris-HCl (50 mM final concentration) • 1 ml of 1% Nonidet P-40 • 5 ml of 10% Sodium deoxycholate (0.5% final concentration) • 1 ml of 10% SDS (0.1% final concentration)
cOmplete™ Protease Inhibitor Cocktail (25X)	Component of lysis buffer	• Dissolve 1 tab in 2 ml distilled water
PMSF (10837091001) 200X - 200 mM	Component of lysis buffer mix	• 1.39352 gm PMSF • 40 ml of isopropanol
Loading blue buffer (5X)	Loading dye	• 50% Glycerol (10 ml of 100% glycerol)

		• 300mM Tris-HCL pH: 6.8 (6 ml of 1 M solution) • 10% 2Mercaptoethanol (2 ml in final solution) • 5mg/ml bromophenol blue (100 mg (for total volume) in 2 ml methanol) • 10% SDS (2 gm for total volume)
TrisGlyc 10X	Component of run buffer	• 30.285 gm TrisBase • 150.012 gm Glycine • 1 L distilled water
10% SDS	Component of run buffer	• 50 gm SDS • 500 ml distilled water
TBS 10X	Component of washing buffer	• 24 gm Tris base • 88 gm NaCl • 600 ml dH2O • pH 7.6 with 12N HCl
Tris Glycine	Run buffer	• 100 ml TrisGlyc • 10 ml 10% SDS • 890 ml distilled water
Turbo transfer buffer (fridge)	Transfer buffer	• 200 ml Transfer buffer main stock • 200 ml Ethanol (or methanol) • 600 ml distilled water • Complete volume to 1 L with dH2O
5% BSA	Blocking buffer	• 2.5 g of BSA • 50 ml of TBST
TBST	Washing buffer	• 100 ml TBS 10X • 900 ml distilled water • 1 ml tween
0.4M NaOH	Stripping buffer	• 16 g NaOH • 1 L distilled water

Table 2-13 List of Antibodies used in Western blotting

Antibody	Dilution (v/v)	Supplier

Lamin A/C	1:1000	Cell Signaling technology®, 2032S
Lamin A/C	1:10000	Abcam, ab108922
β-Actin	1:5000	Cell Signaling technology®, 4967
β-Actin C4	1:10000	Santa Cruz, sc-47778
GAPDH	1:5000	Cell Signaling technology®, 2118
Monoclonal Anti-HHEX	1:1000	(Sigma Aldrich, SAB1403914)
HHEX Antibody (2018B) [Unconjugated]	1:500	Novus Biologicals, 2018B
M3	1:10000	Dr. Sheela Jayaraman
Myc-Tag (9B11) Mouse mAb	1:1000	Cell Signaling technology®, 2276
GFP Antibody #2555	1:1000	Cell Signaling technology®, 2555
Vimentin (D21H3) XP® Rabbit mAb	1:1000	Cell Signaling, 5741
E-Cadherin (24E10) Rabbit mAb	1:1000	Cell Signaling, 3195
Snail (C15D3) Rabbit mAb	1:1000	Cell Signaling, 3879
Slug (C19G7) Rabbit mAb	1:1000	Cell Signaling, 9585
Claudin-1 (D5H1D) XP® Rabbit mAb	1:1000	Cell Signaling, 13255
β-Catenin (D10A8) XP® Rabbit mAb	1:1000	Cell Signaling, 8480
ZO-1 (D7D12) Rabbit mAb	1:1000	Cell Signaling, 8193
Human/Mouse Cleaved Caspase-3 (Asp175) Antibody	1:1000	R&D system, MAB835
Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody	1:1000	(Thermo Fischer, A16040)
IRDye® 800CW Donkey anti-Mouse IgG Secondary Antibody	1:5000	Li-Cor, 926-32212

IRDye® 800CW Donkey anti-Rabbit IgG Secondary Antibody	1:5000	Li-Cor, 926-32213
IRDye® 680RD Donkey anti-Mouse IgG Secondary Antibody	1:5000	Li-Cor, 926-68072
IRDye® 680RD Donkey anti-Rabbit IgG Secondary Antibody	1:5000	Li-Cor, 926-68073
Anti-Rabbit IgG (H+L), F(ab′)2 fragment, CF™680 antibody produced in goat	1:5000	(Sigma Aldrich, SAB4600362)
Anti-Mouse IgG (H+L), F(ab′)2 fragment, CF™680 antibody produced in goat	1:5000	(Sigma Aldrich, SAB4600361)
Goat Anti-Rabbit IgG Antibody, Fc, HRP conjugate	1:5000	(Sigma Aldrich, AP156P)
Goat Anti-Mouse IgG Antibody, Fc, HRP conjugate	1:5000	(Sigma Aldrich, AP127P)

2.5 Nucleic acid experiments

2.5.1 RT-qPCR primer design, preparation, and storage

The DNA sequence of the target gene was acquired using the University of California Santa Cruz (UCSC) genome browser gateway available on (https://genome.ucsc.edu/cgi-bin/hgTracks) (417, 418). PCR primers were designed using the National Center for Biotechnology Information (NCBI) primer-blast. The selection of suitable primer pairs was done following criteria obtained from the literature to ensure primer efficiency and specificity (419). Primers were screened using the OligoAnalyzer tool from Integrated DNA Technologies available from (https://www.idtdna.com/pages/tools/oligoanalyzer?returnurl=%2Fcalc%2Fan

alyzer). Main (100 µM) and working (10 µM) stocks of the lyophilised primers were prepared in PCR grade water (Invitrogen™, 10977035) inside a PCR hood and stored in -20°C.

2.5.2 Primer optimisation

2.5.2.1 Gradient PCR

A PCR reaction mix was prepared to test a range of annealing temperature (Ta) for the primers. A single reaction mix was composed of diluted cDNA samples, PCR green master mix (Thermo Scientific™, K1081), PCR grade water, forward and reverse primers (final concentration 500 nM). PCR products were amplified using Applied Biosystems Veriti 96-Well Programmable Thermocycler.

2.5.2.2 DNA gel electrophoresis

2.5.2.2.1 Agarose gel

A 2% agarose gel was prepared as in section (2.2.1.2.2.1).

2.5.2.2.2 DNA electrophoresis

A 5 µl of DNA hyper ladder and samples from the above-mentioned gradient PCR run were loaded in the gel. The voltage was adjusted to 100 V and left to run until loaded samples were 80% away from the wells. The gel was imaged using a Gel Doc™ EZ Gel Documentation System from Biorad. The bands were analyzed using Image Lab™ software.

2.5.2.3 Primer standard curve

A 2-fold serial dilution of cDNA was used to prepare several reaction mixes for RT-qPCR run using Rotor-Gene SYBR® Green PCR Kit (Qiagen, 204074) following the provided protocol. Each reaction contained SYBR green, forward primer, reverse primer, PCR water and cDNA (replaced with PCR water for no template control = NTC). An example of sample preparation is illustrated in (figure 2-3). Ta for RT-qPCR cycles was adjusted according to DNA gel electrophoresis bands. The band with the best intensity represented the Ta for the tested primer pairs.

Amplification plots, melting curves and standard curves were analysed using Rotor-Gene Q software for Ct values, single peak production and primer efficiencies, respectively. The amplification plot is considered good if the plot was showing 1 Ct difference in between the samples (depending on the dilution factor used for serial dilution). The melting plot was considered good if peaks from all the samples were at the same melting temperature and NTC was flat. The standard curve was examined for efficiency; values within the range 90-110% indicated an amplification factor close to 2 representing good duplication of RT-qPCR products each cycle. If all three curves pass the inspection, primer pairs were considered good for RT-qPCR.

Table 2-14 RT-qPCR cycles

Cycle	Cycle Point
Hold	95°C, 10 minutes
Cycling (45 repeats)	Step 1: Hold at 95°C, 10 seconds
	Step 2: Hold at (Ta)°C, 15 seconds
	Step 3: Hold at 72°C, 10 seconds
Melt	Ramp from 72°C to 95°C 1 degree each step
	Wait 90 seconds (pre-melt conditioningfor first step)
	Wait 5 seconds for each step afterwards

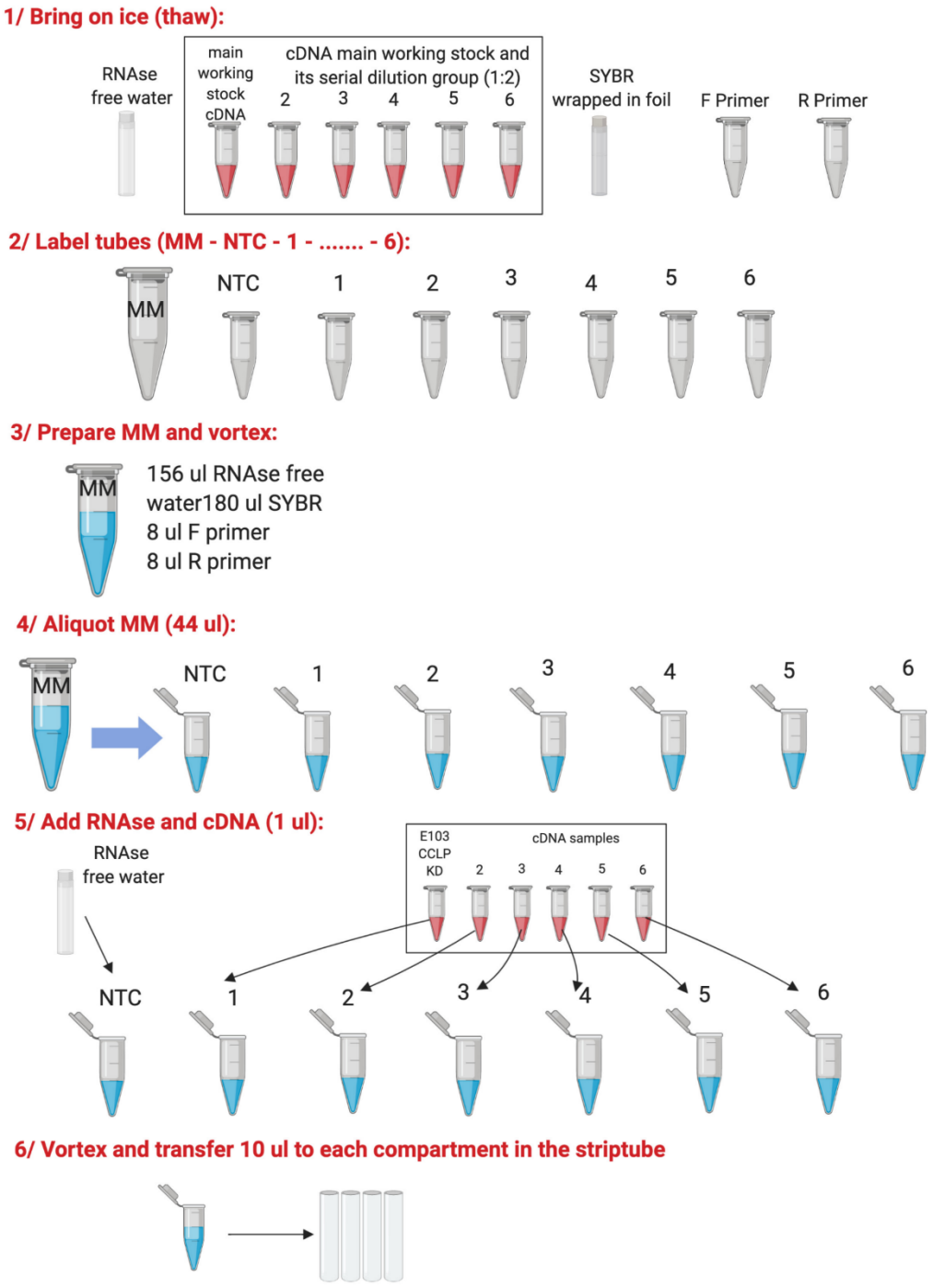


Figure 2-3 RT-qPCR primer standard curve.

Created with BioRender.com

2.5.2.4 Primer concentrations

The range of primers concentration in PCR or RT-qPCR reactions was 250-500 nM. If the primers were still not efficient after trying a range of Ta, RT-qPCR for primer efficiency were repeated using a serial dilation of primers in the reaction mix. If primers were still inefficient (assuming $MgCl_2$ was adequate), designing new primer pairs was preferred over further optimisation.

2.5.3 RNA extraction

A cell pellet was used for the extraction of RNA following the protocol provided by RNeasy Kit (Qiagen, 74106). Briefly, cells were lysed in the presence of β-mercaptoethanol to minimise the effect of RNAse in the sample. RNA was precipitated with 70% Ethanol. The mix was passed through a silica gel column. The column was washed once with a stringent washing buffer (RW1), followed by two washes with a mild detergent buffer (RPE). RNA was eluted from the column using nuclease free water. The quality and concentration of the extracted RNA was checked using a Nanodrop. Samples free of protein and salt contamination were used for cDNA reverse transcription. All buffers were supplied in the kit.

Table 2-15 Qiagen RNeasy buffers

Buffer	Purpose
RLT	Lysis buffer
RW1	Stringent washing buffer
RPE	Mild detergent buffer

2.5.4 cDNA reverse transcription

Equal amounts of RNA from different conditions were reversely transcribed into cDNA following the protocol provided by QuantiNova for the reverse transcription kit (Qiagen, 205413). In summary, a mix of RNA, nuclease free water and genomic DNA removal mix was prepared and incubated at 45°C for 2 minutes. Then, a mix of reverse transcriptase enzyme and reverse transcription reaction mix was added. The reaction was incubated at several temperatures (annealing at 25°C for 3 minutes, reverse transcription at 45°C for 10 minutes, and inactivation at 85°C for 5 minutes) using Applied Biosystems Veriti 96-Well Programmable Thermocycler. cDNA was stored at -20°C until required.

2.5.5 RT-qPCR

Rotor-Gene SYBR® Green PCR kit was used to amplify and detect the expression of genes of interest and housekeeping genes. An example of RT-qPCR reactions preparation is illustrated in (figure 2-4). Briefly, the reaction mix composed of SYBR green, forward and reverse primers, PCR water, and cDNA (without cDNA in case of NTC). Primers we used are listed in table 2-16. The temperature profile for a standard RT-qPCR run was mentioned previously in (table 2-14) Results were exported as an EXCEL document for analysis.

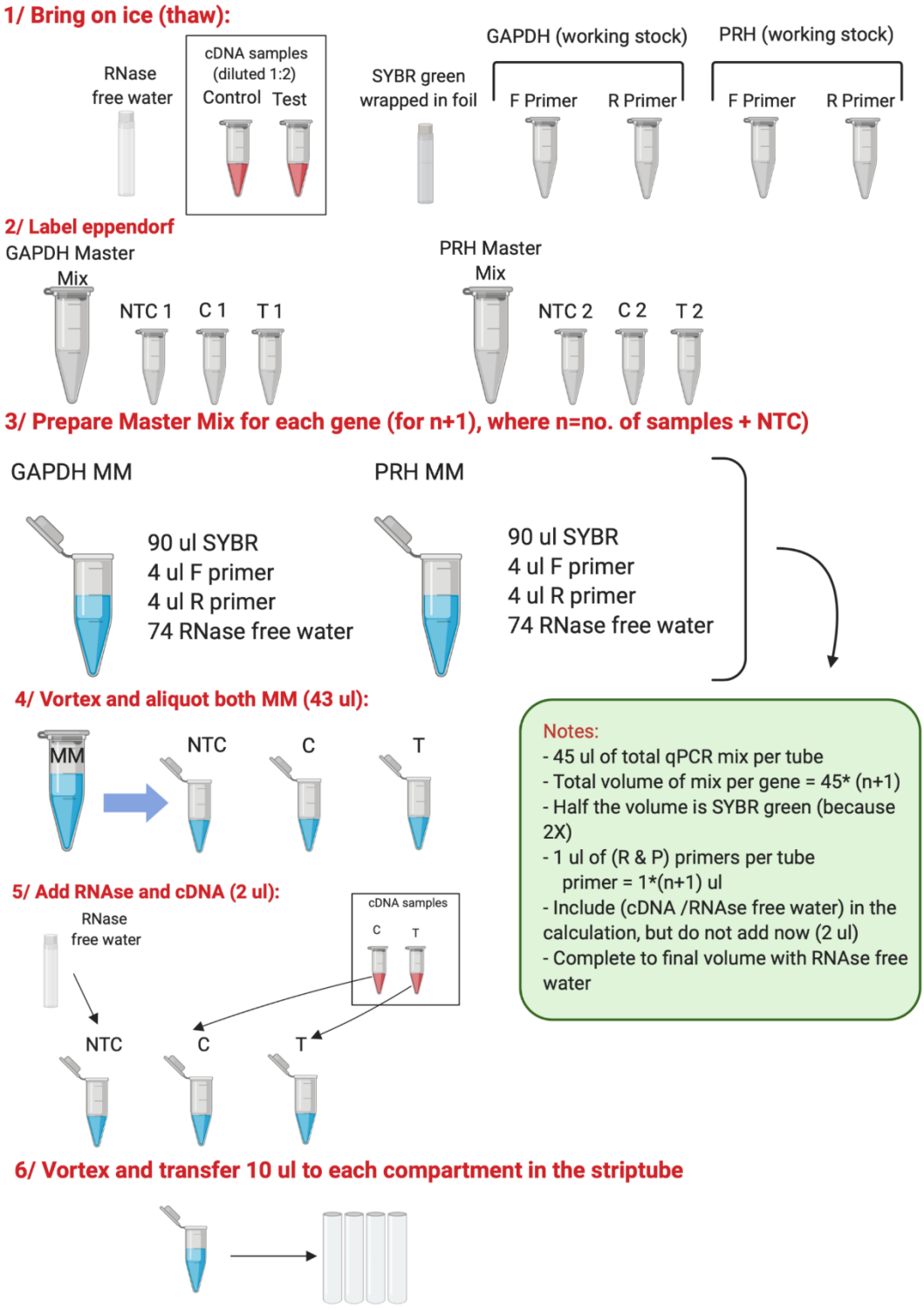


Figure 2-4 RT-qPCR reaction mix preparation.

Created with BioRender.com

2.5.6 Analysis

Ct values were analysed using the Pfaffle method which corrects gene expression by considering the values of primer efficiency (amplification factor) (420). First, Ct of samples were normalised to Ct of experimental controls. Then, the amplification factor of each primer was used to calculate the total amplification using the number of Ct for each sample or control. Finally, the fold change of GOI was calculated by normalising it to a housekeeping gene cDNA (HK). Mean average of multiple experiments was calculated to find the overall expression of gene of interest in samples and controls. Results were plotted as bar charts representing mRNA relative fold change.

Table 2-16 List of primers used in RT-qPCR

	Forward	Reverse
GAPDH	TCC TTG GAG GCC ATG TGG GCC AT	TGA TGA CAT CAA GAA GGT GGT GAA G
BACT	AAAGACCTGTACGCCAACAC	GTCATACTCCTGCTTGCTGAT
HPRT	AAATTCTTTGCTGACCTGCTG	TCCCCTGTTGACTGGTCATT
PRH	AAA CCT CTA CTC TGG AGC CC	GGT CTG GTC GTT GGA GAA TC
VIM	TGAGTACCGGAGACAGGTGCAG	TAGCAGCTTCAACGGCAAAGTTC
CDKN1A	ATCTGTGAAGCATGGTGGG	CCAGGCCAAGAAGAAGAGGA
CDKN1B	CTGCAACCGACGATTCTTTCTACT	GGGCGTCTGCTCCACAGA
CND2	GGACATCCAACCCTACATGC	CGCACTTCTGTTCCTCACAG
E-Cadherin	GTA ACG ACG TTG CAC CAA CC	AGC CAG CTT CTT GAA GCG AT

S1PR1	CCTGTGACATCCTCTTCAGAGC	CACTTGCAGCAGGACATGATCC
PLAC8	ACTCCAACTGGCAGACAGGCAT	GAGAGTCCTCATTGCGACGCTT
TNC	ATGTCCTCCTGACAGCCGAGAA	AGTCACGGTGAGGTTTTCCAGC
SLC7A11	TCCTGCTTTGGCTCCATGAACG	AGAGGAGTGTGCTTGCGGACAT
GSN	ATCTGCCATCCTGACTGCTCAG	CTTCCCACCAAACAGGCTCATG
WASF1	CCCTACCTGTAATCAGTGATGCC	GCTTCCTGTTCACGCTGCTCTT
STAT3	CTTTGAGACCGAGGTGTATCACC	GGTCAGCATGTTGTACCACAGG
MMP10	TCCAGGCTGTATGAAGGAGAGG	GGTAGGCATGAGCCAAACTGTG

2.6 Proliferation assays

2.6.1 Cumulative growth curve

2X10^4 cells were seeded in a 24-well plate as duplicates/condition. Total number of cells were counted (as descried earlier) every 48 hours over 12-day period. The cells were re-seeded at 1:5 dilution. The counting was corrected for both re-seeding dilution and suspension volume each time. Average of the duplicate counts was calculated for each condition at each time point. A cumulative growth curve was plotted using cell number against time.

To calculate the doubling time between the two conditions, the following formula was used:

$$Doubling\ time\ (hours) = \frac{t2 - t1}{3.32\ \times (Log\ n2 - Log\ n1)}$$

2.6.2 EdU incorporation assay

2.6.2.1 Seeding

To examine cell proliferation rate by detecting the number of cells in the S phase of cell cycle, a Click-iT™ EdU Microplate Assay (Invitrogen™, C10499) that can measure active DNA synthesis was performed. A summary of the experimental steps can be seen in (figure 2-5). $1X10^4$ cells were seeded in a 96-well plate in quadruplet per condition. The cells were incubated in a humidified environment with 5% CO_2 at 37°C overnight to adhere. The following day, the assay was performed following the protocol provided by the manufacturer. Briefly, the assay was initiated by adding EdU to the cells. Cells without EdU for background detection were kept in media only. After 2 hours of incubation in a humidified environment with 5% CO_2 at 37°C, the cells were fixed and incubated with EdU ClickIT cocktail at 22°C. After that, cells were blocked and washed in a BSA washing buffer. Finally, SIGMAFAST™ OPD (Sigma-Aldrich®, P9187) that is an HRP-substrate for Fluorescein/Oregon Green Polyclonal antibody (included in the kit) was added. The color intensity from adding OPD substrate was used to reflect the level of DNA synthesis. The absorbance was measured at 450 nm using a Tecan Infinite F50 microplate reader (Scientific laboratory supplies)

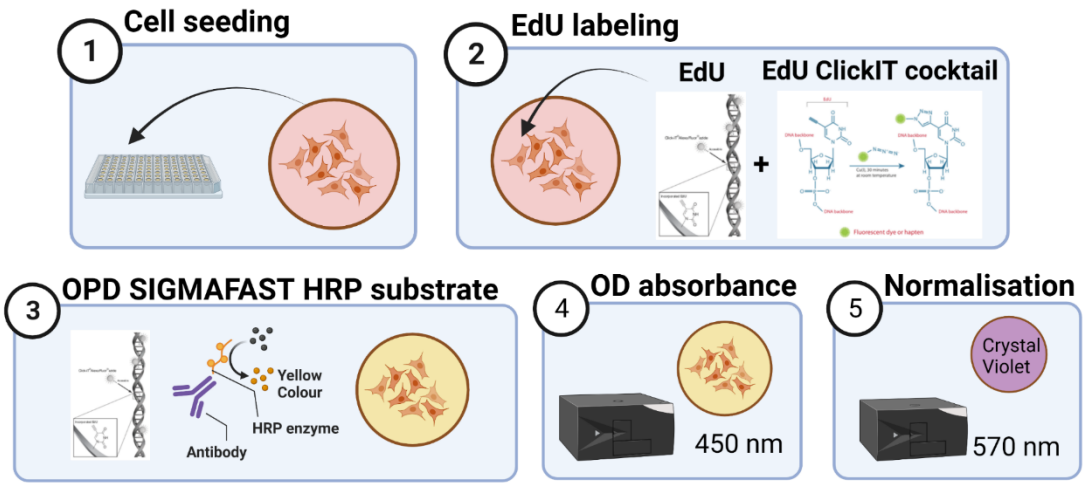


Figure 2-5 EdU incorporation assay experimental steps.
Created with BioRender.com

2.6.2.2 Crystal violet staining to correct for cell number

After the last step from the previous section, the substrate was removed, and the wells were washed once with tap water. The plate was dried, then 50 µl of 0.5% crystal violet was added to all the wells. The plate was left on a rotatory shaker for 30 minutes at 22°C. Crystal violet was removed and discarded appropriately. The plate was then washed by submerging in tap water until all the excess stain was removed. Then, 100 µl of 1% SDS was used to dissolve crystal violet over a period of 1 hour on a rotatory shaker at 22°C. The absorbance was afterwards detected using a Tecan Infinite F50 microplate reader at 570 nm.

2.6.2.3 Analysis

Readings obtained from OPD substrate measurement were normalised to crystal violet values. Non EdU values were averaged and subtracted from EdU

readings. Then, the average EdU readings were calculated for each condition and presented as fold changes.

2.7 Transwell migration assay (chemotaxis)

Transwell migration, also known as chemotaxis, is an *in vitro* method to assess the directed migratory potential of cells across a porous membrane in response to a chemical gradient (421). (figure 2-6 for experimental steps).

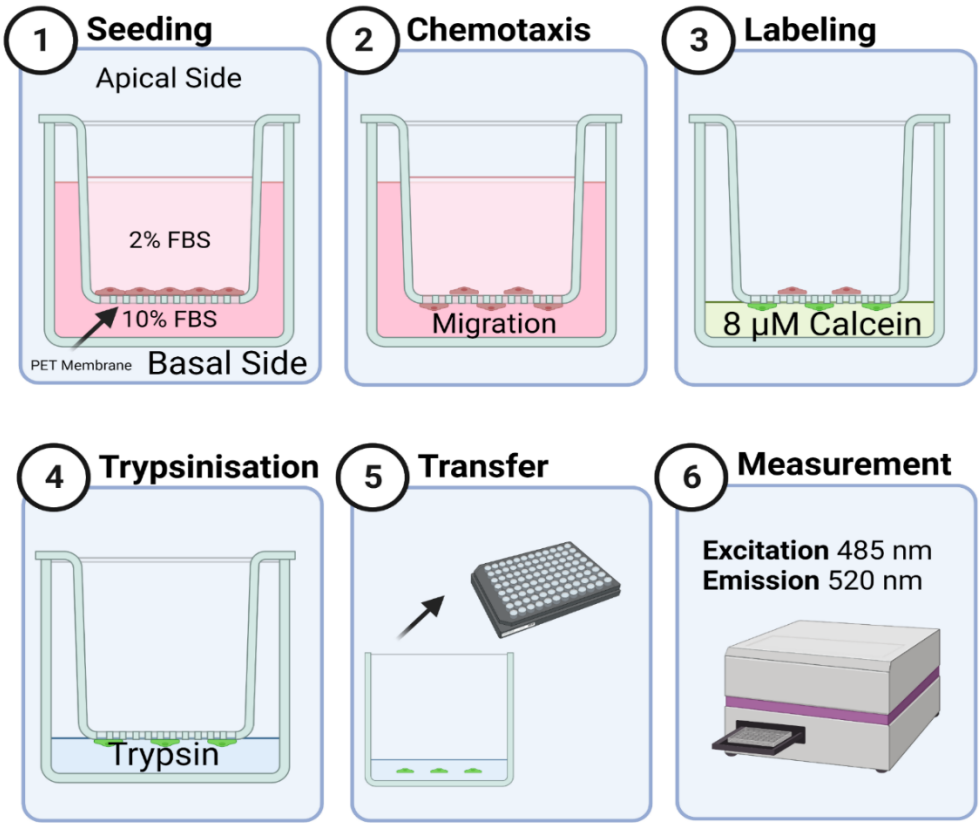


Figure 2-6 Transwell migration (Chemotaxis) assay experimental steps.
Created with BioRender.com

2.7.1 Seeding for transwell migration assay

Cells were starved overnight in 0.2% FBS media while incubated in a humidified environment with 5% CO_2 at 37°C. The next day, $8X10^4$ – $200\ X10^3$

cells were seeded in 24-well ThinCert™ inserts with pores size of 8 µM (Greiner Bio-one GmbH, 662638) as duplicates / condition. A chemoattracting gradient was made across the insert membranes by using a starving medium (2% FBS) in the inserts and (10% FBS) in the receiving wells. 1 mM of Hydroxyurea (Sigma Aldrich, H8627) was added on both sides of the membrane to prevent cell proliferation. The plate was incubated in a humidified environment with 5% CO_2 at 37°C. Hydroxyurea was topped up in both apical and basal compartment daily. After 48-72 hours, media was aspirated from both sides of all inserts. Cells on the basal side of the membrane (migrated cells) were labelled with 8 µM Calcein AM (Sigma Aldrich, 206700) in 450 µl of fresh serum free media for 45 minutes. Finally, migrated cells on the basal part of the insert were trypsinised with 500 µl of trypsin. 200 µl of each condition in duplicates was transferred into a black Clear® 96 well plate (Greiner Bio-one,655096) or a black 96-well plate (Greiner Bio-one, 655076). Additional wells containing trypsin only were used for background fluorescence. Fluorescence was detected at an excitation wavelength of 485 nm and emission wavelength of 520 nm using a FlexStation reader. The number of migrated cells was estimated by interpolating the fluorescence produced by the migrated cells against a standard curve of serially diluted cells labelled with Calcein AM. PC3 cells were used as positive controls, while inserts without chemoattraction were used as negative controls.

2.7.2 CalceinAM standard curve

Wild type cells were starved overnight in 0.2% FBS media. The following day, cells were trypsinised and normalised with equal volume of media. $2X10^6$ cell

were counted and pelleted to remove supernatant. The cells were resuspended in 5 ml free media containing 8 uM Calcein AM and incubated for 45 – 60 minutes in a humidified environment with 5% CO_2 at 37°C. The containing falcon tube was flicked every 10 minutes to resuspend the cells if they precipitated at the bottom due to gravity. When the incubation was complete, the cells were pelleted and resuspended in 2 ml of trypsin. A 2-fold serial dilution was performed in trypsin (1 – 1/16). 200 µl from each compartment was transferred into a black plate in triplicates, trypsin only wells were used to obtain background readings. Fluorescence was detected at an excitation wavelength of 485 nm and emission wavelength of 520 nm using a fluorescent reader flex station. Data were corrected by subtracting the background of trypsin only. Then, triplicates were averaged and plotted against cell number to make a Calcein standard curve. This standard curve was used to estimate the number of migrated cells. The number was corrected to represent the original volume of suspension containing the migrated cells. Replicates were averaged and percentage of migrated cells were calculated using the formula:

$$Migration\ \% = \frac{number\ of\ cells\ on\ basal\ side}{number\ of\ seeded\ cells} \times 100$$

Data were expressed as percentage of migration for all conditions.

2.8 Transwell invasion assay

Transwell invasion assay is an *in vitro* method to investigate the invasive potential of cells across a porous membrane coated with synthetic ECM in

response to chemical gradient (421). (figure 2-7 for experimental steps).

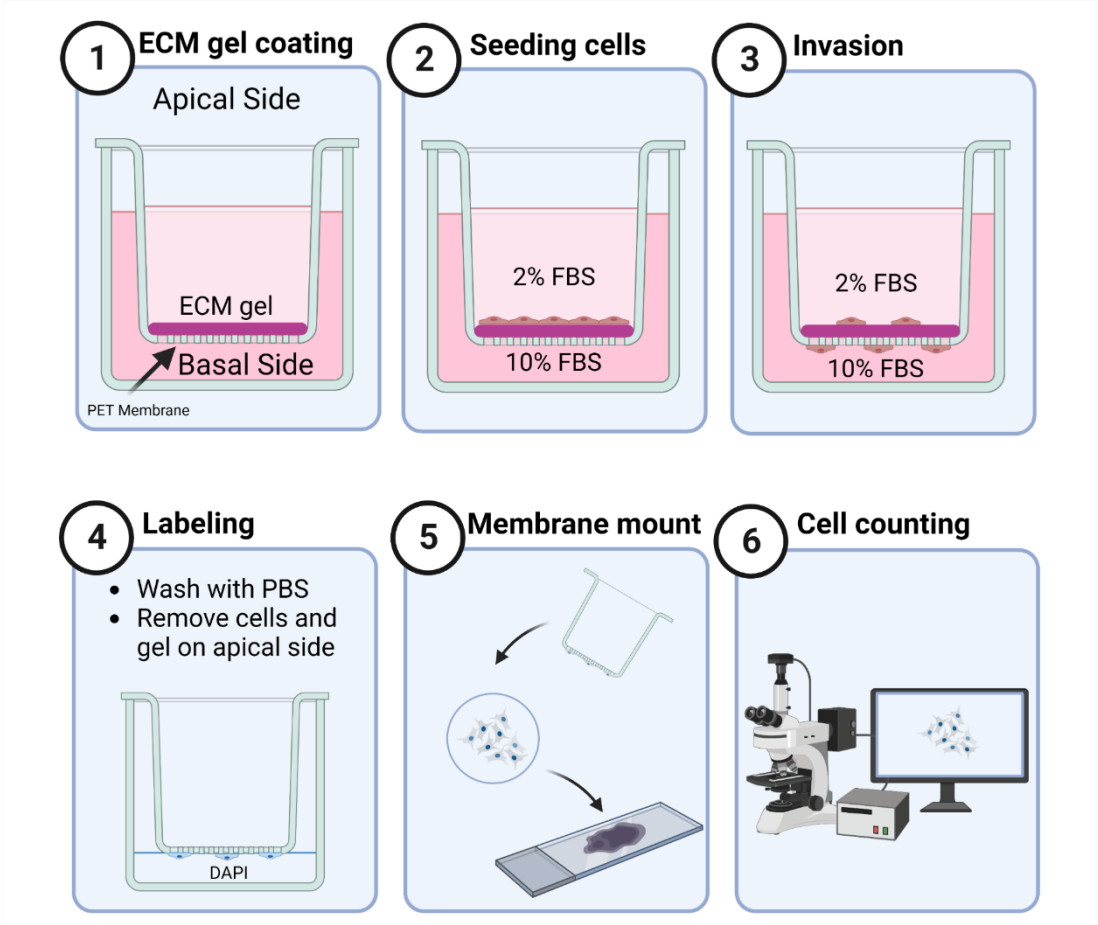


Figure 2-7 Transwell invasion assay experimental steps.
Created with BioRender.com

2.8.1 Inserts coating with ECM gel layer

ECM gel (Corning matrigel matrix, 356234) was thawed overnight at 4°C. Then, ECM was diluted to 8-12 mg/ml in serum free media while on ice. 200 – 300 µl was used to coat the apical side of the inserts. They were placed in their respective well plate and incubated for 2 hours or more to solidify. After the incubation, the apical side of the inserts was gently washed with PBS to remove the media leaving a thin layer of ECM at the bottom of the inserts. This

washing step was performed immediately before seeding to avoid drying the ECM gel.

2.8.2 Seeding, labelling and detection

Cells were starved overnight in 0.2% FBS media. Cell suspension was prepared to seed $5X10^4$ - 10 $X10^4$ cells in a 200 µl volume of 2% FBS media as duplicates per condition. 10% FBS was added in the receiving wells. 1 mM of Hydroxyurea was added on both sides of the membrane to prevent cellular proliferation. After 24- 48 hours, media was aspirated from both sides of all inserts. The gel was scraped using a Q-tip and the inserts were washed gently by submerging in PBS three times. Then, cells on the basal side of the membrane (invading cells) were fixed in 4% PFA for 15 minutes and stained with DAPI (1:1000) of 1 mg/ml stock. The membrane was cut out and mounted on glass slides for imaging. Finally, a Leica DMI3000 B fluorescence microscope was used to acquire images of different view fields on the membrane. The total number of DAPI stained nuclei was divided by the number of imaged fields using the equation:

$$Average\ Invasion = \frac{total\ number\ of\ cells\ on\ basal\ side}{number\ of\ fields}$$

2.9 Gene expression profiling interactive analysis (GEPIA2)

GEPIA2, available on (http://gepia2.cancer-pku.cn/#index), is a platform to analyse data from 9,736 tumours RNAseq and 8,587 normal samples from TCGA/GTEx project using a standard pipeline performed by a Python package from the Zhang lab (422). Gene expression profiling was done by choosing to

explore the expression of *HHEX* and *VIM* genes independently in normal and cancer patient samples. Gene expression data were represented in dot plots of transcripts per million (TPM) for each condition in a range of cancers. Correlation analysis of *HHEX* and *VIM* genes were performed in CCA cancer and normal tissues. HHEX and VIM correlation analysis were performed using Spearman and Pearson coefficients. Gene expression correlation data were represented in scatter plots.

2.10 RNAseq

2.10.1 Sample preparation

Total RNA was extracted using Qiagen RNAeasy columns as previously described with an additional step of an on-column DNAse digestion using the RNAse-free DNase set according to the supplier's protocol (Qiagen,79254). In brief, 1500 Kunitz units of lyophilised DNAse I was dissolved in 550 µL RNAse free water and stored as a stock at -20°C. 10 µl of DNAse I stock was mixed with 70 µl of the supplied RDD buffer and mixed gently. This mix was added to RNAeasy column after RW1 wash step. The column was left to incubate at 22°C for 15 minutes. This was followed up by another RW1 washing step. From here, the rest of the RNA extraction was resumed with the addition of RPE buffer according to the protocol.

RNA samples concentration and purity were tested using a nanodrop. If the sample concentration is high and does not show any protein or salt contamination, an aliquot was sent to Deep Seq Next Generation Sequencing facility at the university of Nottingham to acquire a second concentration

measurement as well as an RNA integrity Number (RIN) using the Agilent 4200 TapeStation system, DNA gel electrophoresis and Qubit. Only RIN above 9 was accepted to proceed with RNA sequencing. After generating TapeStation quality control (QC) report, RNA samples were shipped on dry ice to Novogene UK.

2.10.2 Novogene library preparation and sequencing

The overall flow of processing done by Novogene (excluding sample preparation) is illustrated in (figure 2-8).

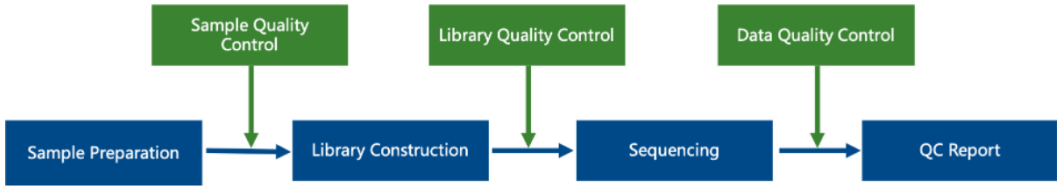


Figure 2-8 Overall RNA processing by Novogene
(Adopted form Novogene)

Novogene ran their own quality control screening on the samples before constructing the cDNA library and sequencing the RNA. RIN and concentration were checked again, and a sample QC report was generated based on the company's standards. Next, cDNA library was preparation performed as shown below (figure 2-9).

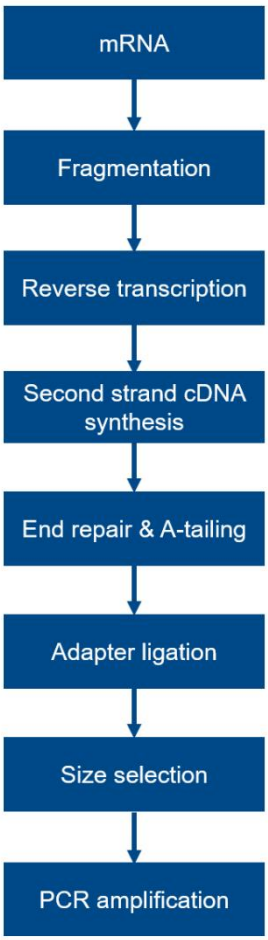


Figure 2-9 Novogene cDNA library preparation.
(Adopted form Novogene)

Briefly, mRNA was purified from the provided sample using magnetic beads. Then, mRNA was fragmented into shorter pieces. After fragmentation, the complementary strands were made using random hexamer primers. Further modification was applied on the produced cDNA including end repairs, A-tailing, adaptor ligation, size selection and PCR amplification. A second QC screening was done to check the generated library. The quantified library was then pooled and sequenced on an Illumina platform. Illumina used NovaSeq PE150 platform to build a poly A enriched RNA library and generate 5 G of

raw per sample. Sequences were of 150 bp length and paired-ended. Finally, a third QC screening was done on the reads for final data quality check-up.

2.10.3 Data mining, retrieval, and upload

RNAseq raw data were obtained from Novogene directly after sequencing, or from our group's previous sequencing, or from public resources e.g., NCBI Gene Expression Omnibus (GEO). The sequencing data (fastaq) were uploaded to the Galaxy web platform, and we used the public server available in (https://usegalaxy.eu/) to analyse the data (423).

2.10.4 Galaxy web platform analysis for RNAseq data with biological repeats

The pipeline of RNAseq raw data analysis we followed is summarised in (figure 2-10).

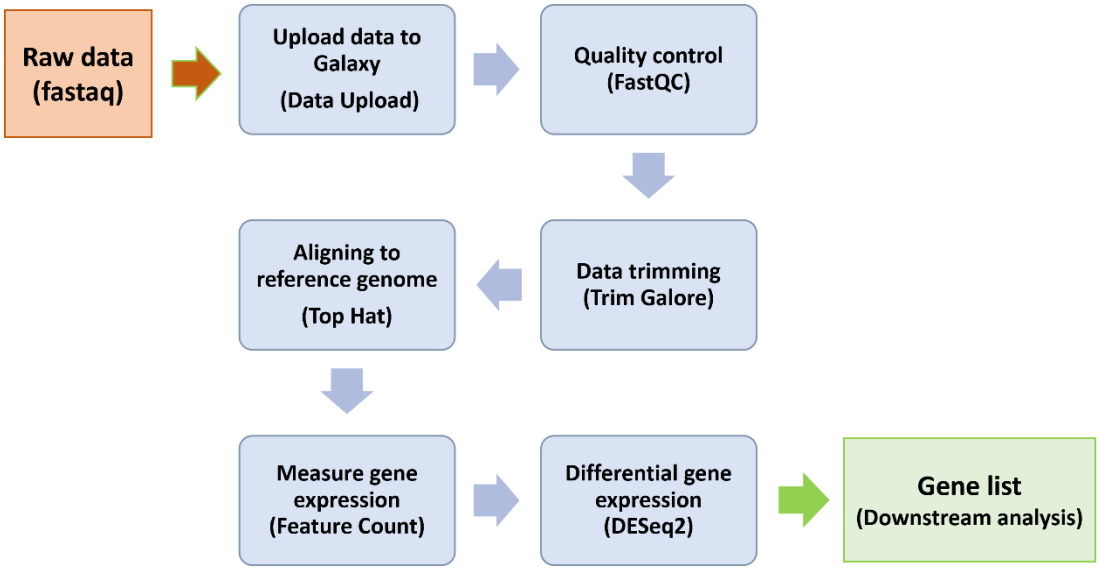


Figure 2-10 Galaxy RNAseq pipeline

The quality of the data was checked using FastQC tool (424). In general, this tool screen for sequence quality, contents, length, duplication and over-

representation (424). FastQC report was interpreted according to Galaxy website guidelines. Any red flags were checked for false flagging before the data deemed unqualified was used to proceed through the pipeline.

If the QC report was accepted, sequences were then trimmed using TrimGalore tool (425). This tool removes low quality sequences and adaptors from the reads. Sequences quality were identified according to phred scores; a phred score of >30 (i.e. 1/1000 error rate) indicated good sequences (426).

After that, reads were aligned to mammalian-sized genomes HG38 (or HG19) using the high-throughput short read aligner tool TopHat (427). To count the aligned features in SAM or BAM files, the tool FeatureCounts was used (428). This tool uses an annotated human genome HG38 (or HG19) obtained from the public data available through galaxy's internal library. To account for experimental biological repeats and calculate variance-mean dependence, DESeq2 tool was used to statistically test for differential expression (429). DESeq2 uses a negative binomial distribution model and compares the different conditions of the experiment to generate gene expression fold change (FC), nominal *p*-value (using Wald test), and correct the p-value by multiplying it by the number of tests producing an adjusted *p*-value (*q*-value) for significance purposes. The outcome files generated from DESeq2 tool (results and normalised counts) were downloaded and used for downstream analysis.

2.10.5 Preparing for downstream analysis

The first step performed was to convert ENSMBLE IDs to gene symbols for gene identification. This was done via the MART_export document obtained

from GitHub Bioinformatics forum (430). The second step was to filter out data with (#N/A) values. After that, further filtering was done based on adjusted *p*-values to keep significant genes which have values below *p*=0.05 (5%) only. Further sorting was done to classify the significant genes into upregulated and downregulated. A final list of statistically significant and regulated genes was produced. Data were represented in Venn diagrams using the tool FunRich3.1.3 available from (http://funrich.org/download) (431), volcano plots using graph pad prism, and heatmaps using the online tool ClustVis available on (https://biit.cs.ut.ee/clustvis/) (432).

2.10.6 Molecular signature database (MSigDB)

Molecular signature database (MSigDB) is an online database that has several collections of annotated gene sets (https://www.gsea-msigdb.org/gsea/index.jsp) (433, 434). This website can perform hypergeometric overlapping analysis, or over-representation analysis (ORA), GO and GSEA of gene sets with gene lists provided by the individual. Since the input is limited to 500 genes only, the DESeq2 result version used as an input underwent a stringent data filtering. The settings for the overlapping included choosing the reference annotated gene set, assigning a cutoff for *p*-value, and setting the number of overlapped sets. The outcome was presented in a horizontal bar chart of the gene sets and their false discovery rate (FDR) for *q*-values using Microsoft EXCEL (statistically significant terms were identified by an FDR <0.05.). The overlapped genes were plotted in heatmaps using the equivalent DESeq2 normalised counts from the different conditions of the experiment.

2.10.7 Enrichr for Gene Ontology

Enrichr is an online tool, developed by the Ma'ayan lab, that can perform several ORA analyses including GO (435, 436). It can be accessed via the link https://maayanlab.cloud/Enrichr/. Gene symbols of a certain gene list input were submitted into the online tool. Then GO of one or all three of biological processes, cellular component and molecular function was generated. Enrichment analysis using MSigDB library can also be done in Enrichr. Result outcomes were downloaded as images or PDF files.

2.10.8 Pre-ranked GSEA

Pre-ranked GSEA analysis is another tool to overlap the same annotated gene sets from previous MSigDB with the individual's gene list performed with GSEA software (433). This time, genes were ranked based on their Log_2 (FC) values from highest to lowest. This list was converted into a document with (rnk) extension and uploaded into GSEA software. The settings to run the analysis included choosing the annotated reference gene sets, assigning the rnk as the parameter file for collapsing the data, and selecting the minimum number of overlapped genes in a single set. The rest of the settings were left as default. The outcome was the enriched gene sets, and their enrichment score plots (ES). Overlapped gene sets and genes were presented as mentioned in the MSigDB section.

2.10.9 Overlapping with different RNAseq data

Different versions of DESeq2 results (i.e., after data filtration) were overlapped with gene lists from other RNAseq data. This helped to further shorten the list

of genes and make them more relevant to the project. Again, data were presented as described earlier.

2.11 ChIPseq analysis

2.11.1 Data mining and retrieval

PRH-related ChIPseq raw data were obtained from the group's previous sequencing projects. The sequencing data (fastaq) were uploaded to the Galaxy web platform, and we used the public server available in (https://usegalaxy.eu/) to analyse the data (423).

2.11.2 Galaxy platform analysis (flowchart)

The pipeline of ChIPseq raw data analysis we followed is summarised in (figure 2-11).

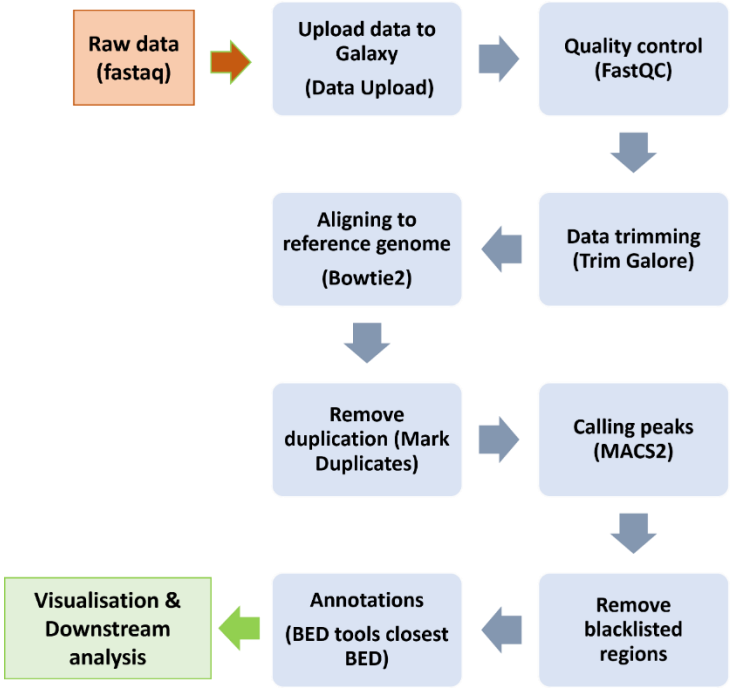


Figure 2-11 Galaxy ChIPseq pipeline (423).

First, raw data were uploaded into Galaxy in fastaq format. The quality of the data was checked using FastQC tool to screen for sequences quality, contents, length, duplication and over representation (424). QC report was interpreted as recommended by the Galaxy website. Like RNAseq QC report, any red flags were checked for false flagging before proceeding through the pipeline. The second step was using the tool TrimGalore to remove low quality sequences (< 30 phred score (426)) and adaptors from the reads (425). Next, Bowtie2 was used to map the sequences against a built-in reference genome HG38 (or HG19) (437, 438). Reads were then screened for duplicated molecules and filtered using the tool MarkDuplicate (439). After that, the tool MACS2 call peak was used to identify and evaluate the significance of the ChIP peaks (enriched binding sites) in the sequences (440, 441). The outcome BAM file can be converted into bigwig format and exported after this step to be visualised in any genome browser e.g., UCSC browser (418, 430, 442). Peak location and morphology were interpreted according to UCSC guidelines. To follow up after MACS2, a black-listed bed file of HG38 against which we mapped our reads earlier was downloaded from GitHub available from (https://github.com/Boyle-Lab/Blacklist/blob/master/lists/hg38-blacklist.v2.bed.gz). This file was uploaded into galaxy and used with the tool subtract to remove the black-listed regions of HG38 from our sequences. These regions are usually highly enriched areas such as centromeres. To annotate the peaks, first an annotated reference genome was retrieved from galaxy shared data (418). Our bed file was then annotated using the tool

BedTool closest bed (443). The outcome bed file was exported for further processing.

2.11.3 Hypergeometric Optimization of Motif EnRichment (HOMER)

Hypergeometric Optimization of Motif EnRichment (HOMER), which is provided by the Benner's Lab at UCSC, is a package for UNIX operating system containing several analysing tools for ChIPseq and other sequencing data (444, 445). The tool FindMotif from Homer was downloaded and run in Linux. Details of download and installation are available on (http://homer.ucsd.edu/homer/motif/index.html). The tool enables discovering novel motifs based on a differential motif discovery algorithm from the bed file generated in Galaxy. As a result, enriched specific regulatory elements can be identified. In addition, HOMER FindMotifs can assign a score along with binomial motifs enrichment (445). The result of FindMotifs tool is a webpage that shows both Homer (DeNovo) motifs and known motifs.

2.11.4 HOMER annotation

To annotate the bed file, i.e., associate the peaks with nearby genes, another tool from Homer "Annotate" was used for that purpose which is found in this link (http://homer.ucsd.edu/homer/ngs/annotation.html) (444, 445) . Again, this tool used run on Linux. The data was then saved as a text file to be further processed in EXCEL.

2.11.5 Downstream analysis

The annotated file generated by HOMER was cleaned before proceeding with downstream analysis. The spreadsheet was cleaned from empty cells, (#N/A)

entries and data entered as dates rather than genes. Although these entries could be referring to important peaks, unfortunately, they are useless with such format. Data were sorted based on peak distances from TSS into 100 kb, 10 kb and 2 kb. Further classification was done based on peaks epigenomic locations, e.g., TSS, promoter, and intronic. Moreover, the frequency of each nearby gene was calculated. These genes were also overlapped with differentially expressed genes from existing RNAseq data. Like RNAseq downstream analysis, MSigDB, Pre-ranked GSEA and GO terms were obtained for certain genes input. Bar charts, Venn diagrams, volcano plots and heatmaps were generated using the tools described under RNAseq analysis.

2.12 Statistical analysis

All statistical analyses were performed using Microsoft EXCEL and GraphPad Prism. Statistical significance was determined from N=3 individual biological repeats. *p*-value ≤ 0.05 was used as a cut-off for significance, where P ≤ 0.05 (*), P ≤ 0.01 (**), P ≤ 0.001 (***), P ≤ 0.0001 (****). Error bars were made based on SD. When statistical significance was determined from N=2, error bars represented standard error of the mean SEM. The type of statistical test used relied on the type of results obtained from the experiments.

Chapter 3: GFP-tagged PRH expression in immortalised cholangiocytes and its effects on cell proliferation, cell morphology, cell migration, and cell invasion

3.1 Introduction

Kitchen et al. reported for the first time that PRH acts as an oncoprotein in CCA (215). PRH is highly expressed in CCA cell lines and patient samples. PRH overexpression in cholangiocarcinoma cells lines (CCLP1, CCSW1, rmCCA1, HuCCA1), immortalised cholangiocytes (AKN1), and in biliary epithelial cells (BECs) induces EMT and promotes cell proliferation and migration (215). Depletion of PRH in CCA cell lines is associated with loss of mesenchymal morphology, decrease in cell proliferation and migration rates *in vitro,* and reduction of tumour growth in xenograft mouse models (215).

In this chapter, we aim to produce human immortalised cholangiocytes that express GFP-tagged PRH to enable further study of PRH. Cells expressing GFP-tagged PRH would enable experiments such as Fluorescence Recovery After Photobleaching (FRAP) to study protein dynamics (446), and Fluorescence Resonance Energy Transfer (FRET) to study molecular interaction (447). Establishing an *in vitro* immortalised system that constitutively expresses GFP-tagged PRH is a feasible, and beneficial way to explore PRH functions in cancer biology. In this chapter, we detail the process we followed to build this system. We used plasmid transfection to express

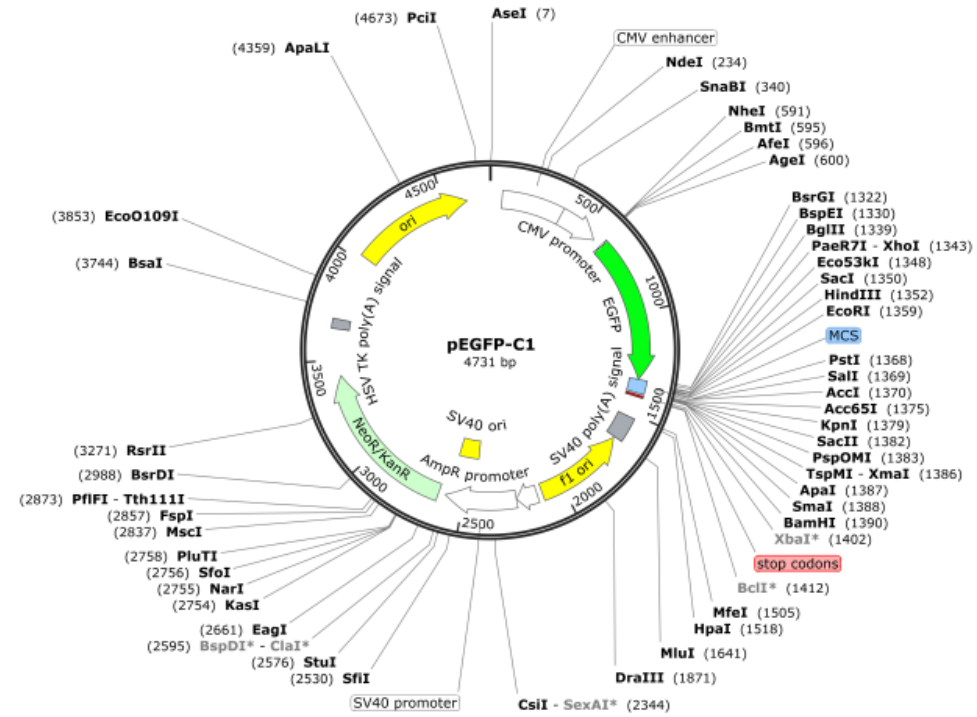
GFP-tagged PRH in AKN1 cells and generated a monoclonal cell line using FACS and G418 selection. AKN1 cells are human immortalised cholangiocytes isolated from a normal liver (406). They were first reported as hepatocyte-like cells with biliary epithelial markers (406). Having expressed GFP-tagged PRH in these cells we examined the relationship between PRH levels and changes in cell proliferation and cell migration. To examine effects on cell growth and proliferation, we used cumulative growth assays and EdU incorporation assays, respectively. We also used fluorescent imaging and morphological analysis to examine the change in cell morphology. In addition, to test whether the cells show altered migratory or invasive characteristics, we used transwell migration (chemotaxis) and transwell invasion assays.

3.2 Results

3.2.1 Generation of a stable AKN1 GFP-PRH-Myc cell line

To address the biological effects of PRH and its contribution to the tumorigenic phenotype and to facilitate further experiments, we generated a stable cell line expressing GFP-tagged PRH. The two plasmids we used, pEGFP-PRH-Myc and its control plasmid pEGFP-C1 (Addgene plasmid #45769), were provided by Professor Sheela Jayaraman. In (figure 3-1) we present the maps of both plasmids illustrating the genes for GFP, GFP-PRH-Myc, Kanamycin bacterial selection and Neomycin selection. pEGFP-PRH-Myc plasmid encodes the human PRH cDNA cloned from the vector pBluescript-PRH in-frame with GFP. After the PRH coding sequence, there is an in-frame Myc tag that is followed by a translation stop codon. Thus, pEGFP-PRH-Myc expresses a PRH fusion protein that has GFP at its N-terminus and a Myc tag at its C-terminus.

A) pEGFP-C1



B) pEGFP-PRH-Myc

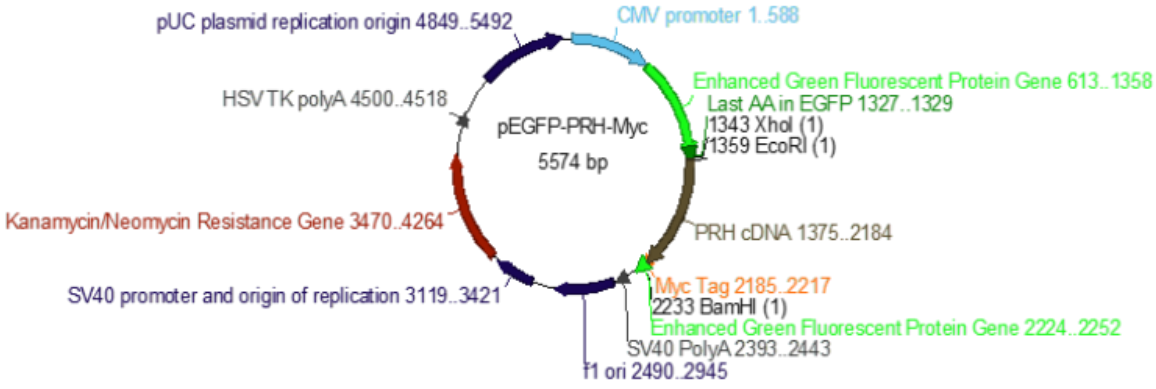


Figure 3-1 Schematic presentation of pEGFP-C1 and pEGFP-PRH-Myc plasmids.

(A) The map of pEGFP-C1 used to make a control cell line expressing GFP in AKN1 cells. (B) The map of pEGFP-PRH-Myc plasmid with human PRH cDNA between GFP and MycTag sequences which was used to generate AKN1 cells expressing GFP-tagged PRH.

To validate pEGFP-PRH-Myc plasmid prior to transfection, we used a restriction digest assay with EcoR1 (1359) and Kpn1 (2228) as described in the method chapter.

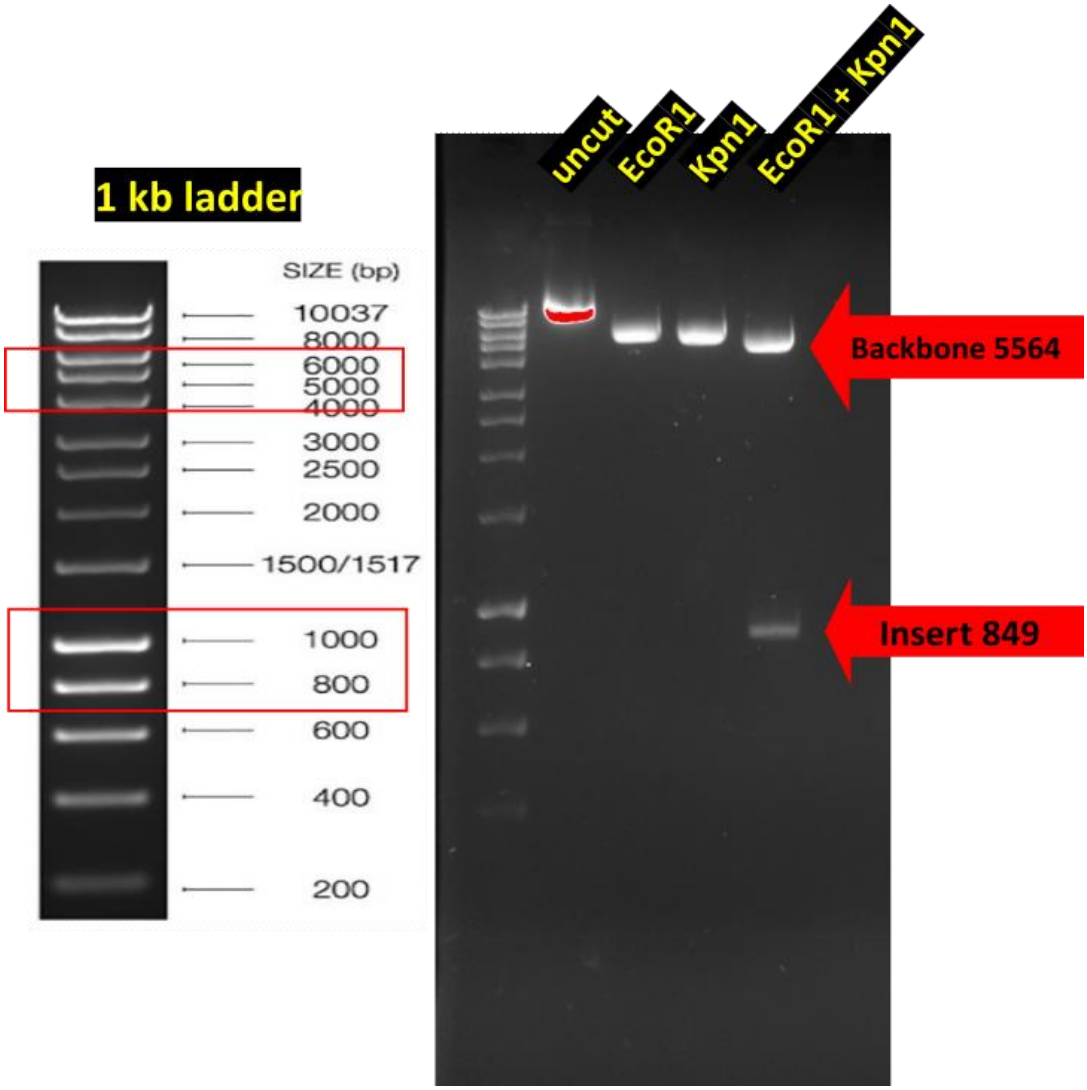


Figure 3-2 Restriction enzyme digest analysis of pEGFP-PRH-Myc plasmid.

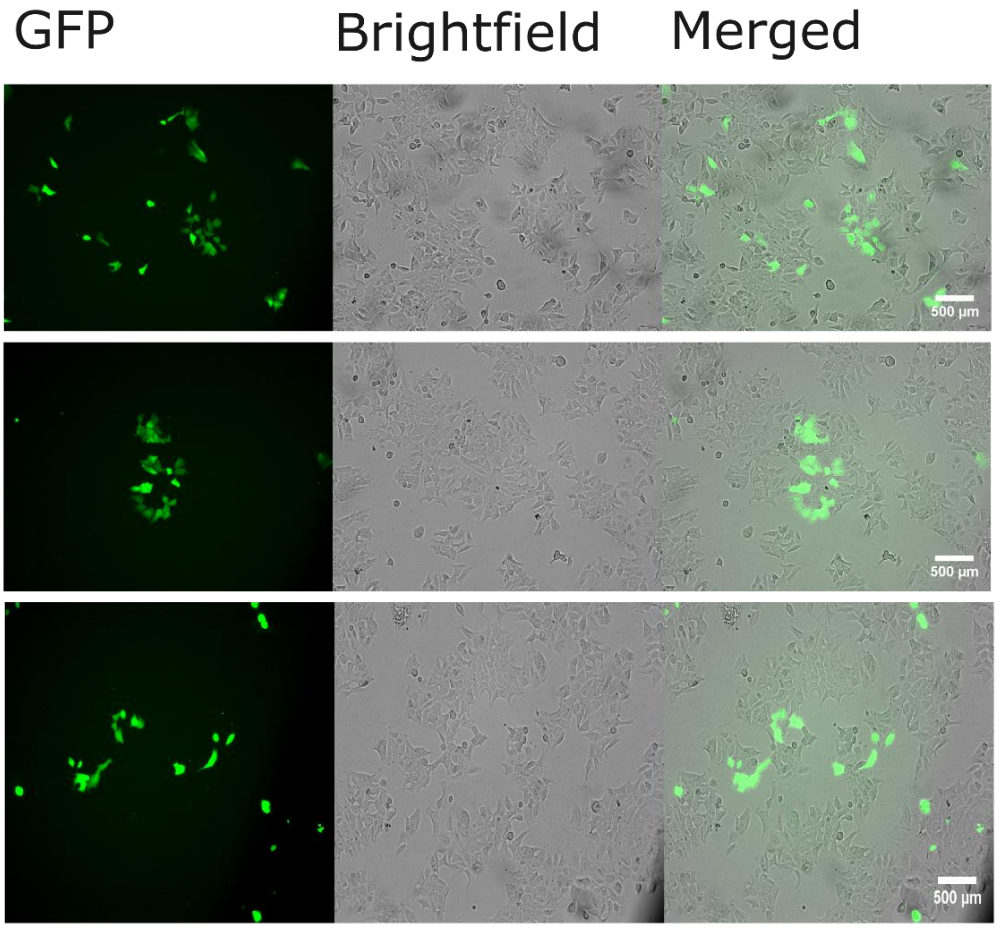
1 µl of 1 µg/µl plasmid was incubated with EcoR1 and Kpn1 restriction enzymes before subjecting them for 2% Agarose DNA gel electrophoresis along with plasmid samples incubated with each enzyme on its own, as well as undigested plasmid DNA. The image was produced using Image Lab™ software. The result shows that incubating the plasmid with both restriction enzymes released a single fragment with a size around 849 bp.

The gel image in (figure 3-2) of pEGFP-PRH-Myc diagnostic restriction digest using EcoR1 and Kpn1 enzymes has four lanes after the DNA (1kb) marker ladder. The uncut supercoiled DNA appeared as a band that did not travel down the gel, while DNAs cut with a single restriction enzyme showed single bands indicating linearised DNA. Double restriction digestion produced two

bands; the higher one represented the plasmid backbone at 5564, and the lower band indicated a DNA insert fragment of 849 corresponding to the size of DNA fragment calculated from the plasmid map.

To transfect AKN1 cells, we delivered the plasmids to $3X10^5$ cells seeded in a 6-well plate using a mix of OptiMEM, PEI and DNA (pEGFP or pEGFP-PRH-Myc). After 24 hours, we inspected the cells for GFP expression using a Leica DMI3000 B fluorescence microscope. In (figure 3-3) we show that approximately 30% of the cells were successfully transfected. AKN1 cells expressing GFP alone were green throughout the cells, while AKN1 cells expressing GFP-tagged PRH showed mainly green nuclei. We called the cells **AKN1 GFP** and **AKN1 GFP-PRH-Myc**, respectively.

A) AKN1 GFP



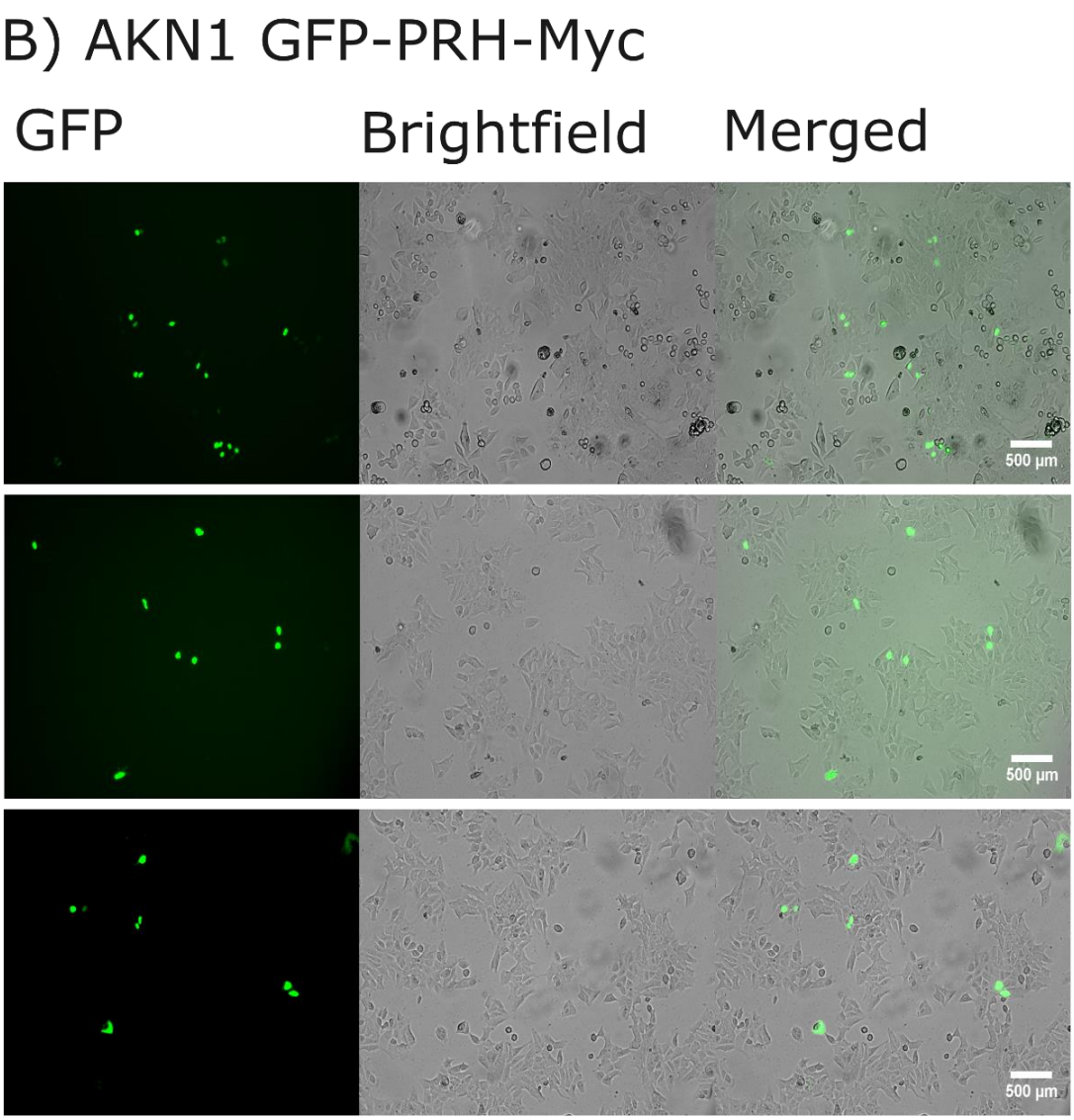


Figure 3-3 Transfection efficiency of pEGFP-C1 and pEGFP-PRH-Myc plasmids in AKN1 cells.

Live AKN1 GFP and AKN1 GFP-PRH-Myc images at 10X magnification acquired with a Leica DMI3000 B fluorescence microscope 24 hours after transfection showing approximately 30% of GFP expression (green cells and/or green nuclei). Scale bar 500 µm.

To remove non-transfected (non-green/non-GFP) cells, we started the selection of cells with 750 µg/ml (up to 1 mg/ml) G418 (Geneticin) 24 hours after the transfection. Every three days after that whenever we replaced the media, we added G418. Figure 3-4 shows that G418 killed almost 80% of non-GFP cells.

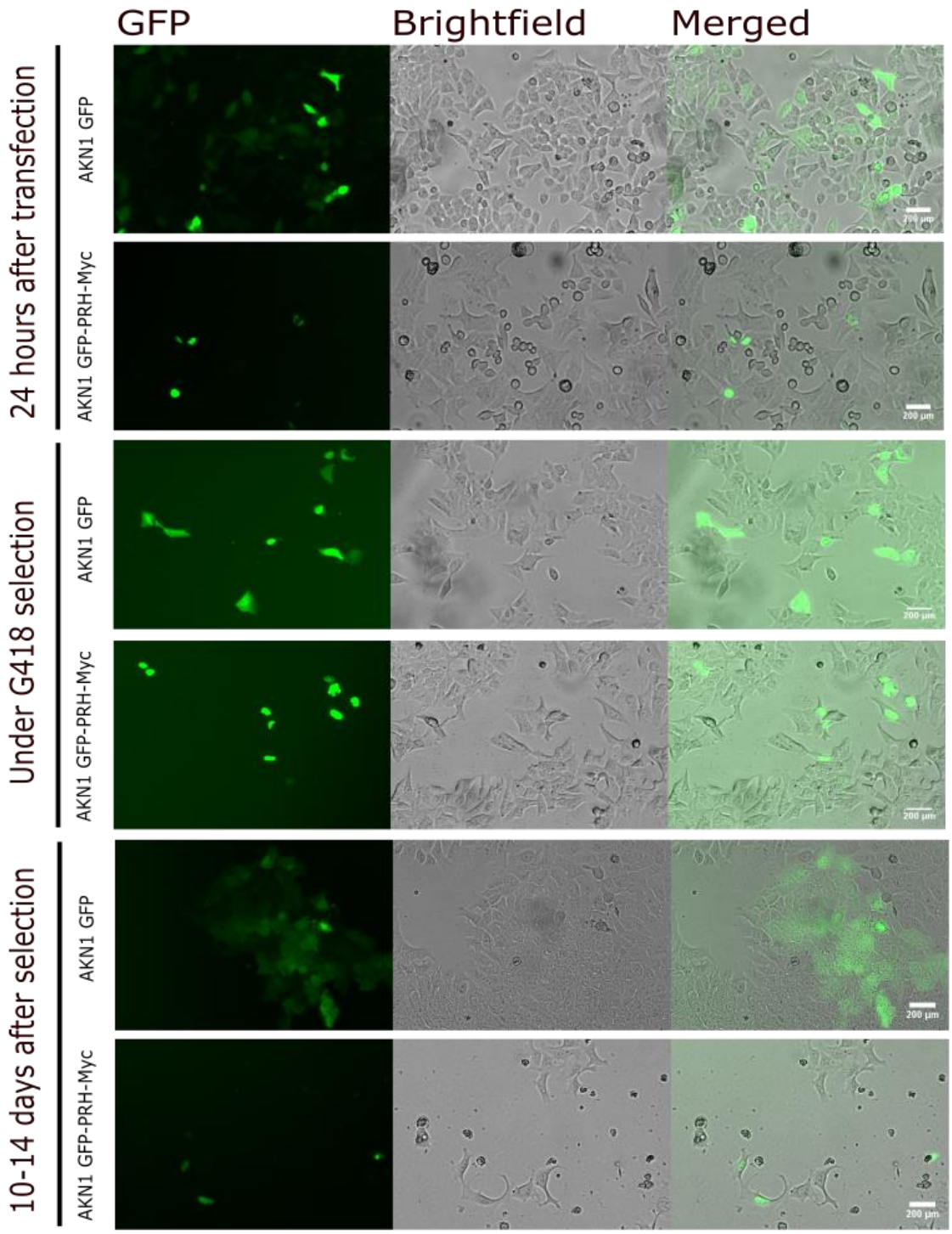


Figure 3-4 AKN1 GFP and AKN1 GFP-PRH-Myc cells selection with G418.

Cropped images of live AKN1 GFP and AKN1 GFP-PRH-Myc cells at a magnification of [1.5X (top) – 2.5X (middle and bottom)] acquired with a Leica DMI3000 B fluorescence microscope. (Top) 24 hours after transfection, (Middle) during G418 selection, and (Bottom) 10-14 days after G418 selection. It shows the reduction in non-GFP expressing cells with 0.750 – 1 mg/ml G418 selection over 10-14 days. Scale bars 200 µm.

Even though we added a maintenance dose of G418 to the conditioning media while we were growing the cells, non-GFP expressing cells were resistant to the antibiotic selection. The percentage of GFP expressing cells decreased while non-GFP expressing cells were dominating the growth vessel which gave rise to a mixed population cell line.

To overcome the issue of mixed population cell line and keep as many green cells as possible, we decided to single-sort the cells using fluorescent activated cell sorting (FACS). GFP single cell sorting by FACS is a rapid method to concurrently analyse and separate live GFP-expressing cells from a mixed population. With help from Dr David Onion, we sorted our cells twice with a duration of around five months in between. To prepare the cells for sorting, we prepared $2X10^7$ AKN1 GFP and $5X10^6$ AKN1 GFP-PRH-Myc cells in fresh complete media before transferring them to the Flowcytometry facility. The cells were single-sorted using (MoFlo™ XDP cell sorter) into a 96-well plate. The histogram and density plots in (figure 3-5) show the gating strategy we used to isolate AKN1 GFP and AKN1 GFP-PRH-Myc cells. Gating is the process of cell subpopulation selection for further investigation (448). We only needed to gate for GFP-expressing cells. We used AKN1 wildtype cells as a reference. We analysed forward scatter (FSC) against side scatter (SSC) to identify the population of interest. We followed with FSC width against FSC height to identify single cells in the previous population. The last gating was SSC height against fluorescence to isolate GFP-expressing cells. It is important to mention that gating was quite narrow initially, but we had to widen the gate for AKN1 GFP-PRH-Myc cells to include as many green cells as

possible since their signals were weaker compared to AKN1 cells expressing GFP alone.

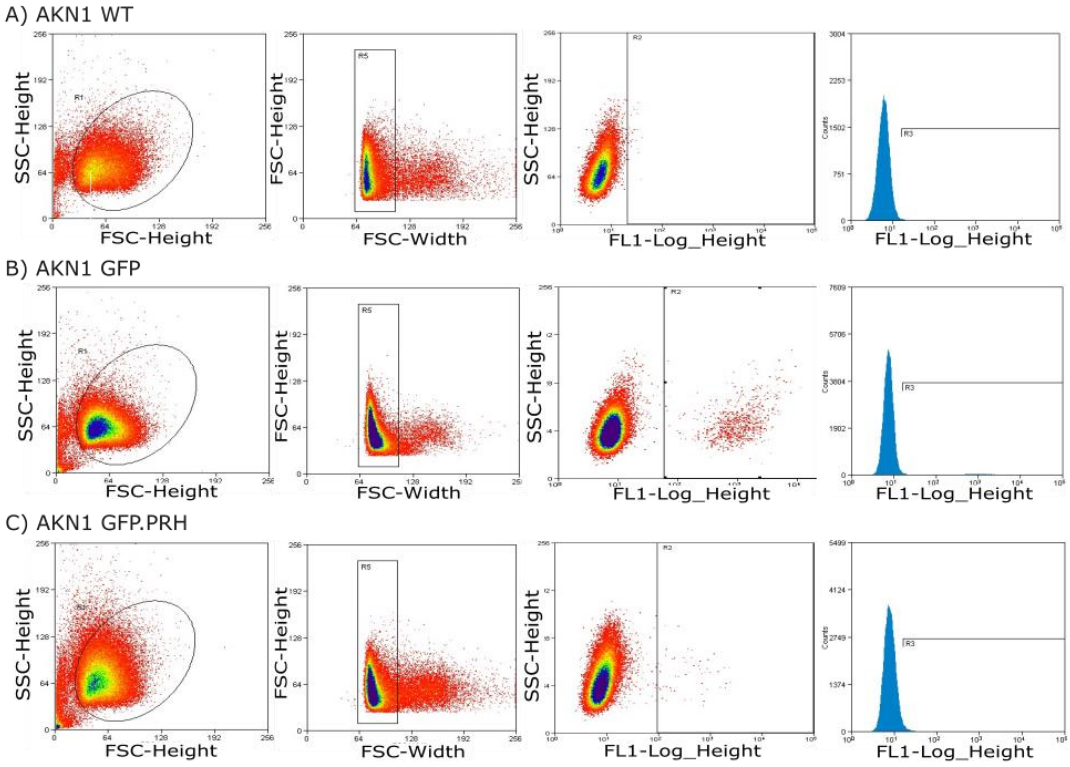


Figure 3-5 FACS single-sorting of transfected AKN1 cells expressing GFP or GFP-tagged PRH.

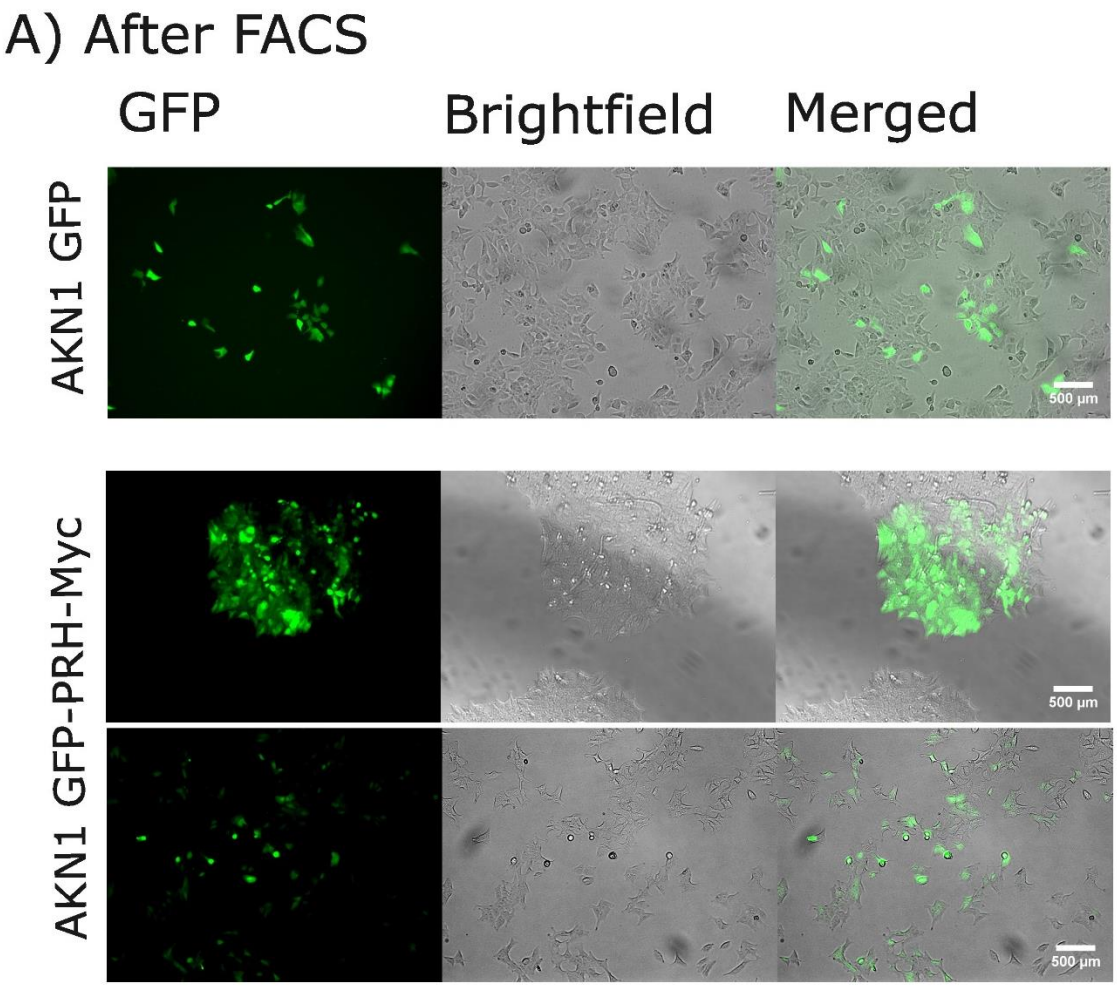
The AKN1 GFP and GFP-PRH-Myc cells were trypsinised and resuspended in fresh media at a specific concentration for sorting. AKN1 wild type was used as a reference. The left three columns of density plots depict the cell population selection process, with the left column showing FSC versus SSC (population of interest), the middle column showing FSC height versus FSC width (single cells), and the right column showing SSC height versus GFP fluorescence (GFP-expressing cells). The histogram column panel displays the selected cells for each cell line. GFP protein was excited using 488 nm laser, and the fluorescent signals were detected using a 528/29 filter.

Table 3-1 indicates that out of the total number of single-sorted cells, we had around $1X10^6$ AKN1 GFP cells and around $2x10^5$ of AKN1 GFP-PRH-Myc cells. Sorting purity for both conditions were over 90%.

Table 3-1 Cell count input and output for fluorescent-activated cell sorting (FACS).

	AKN1 GFP	**AKN1 GFP-PRH-Myc**
Input cell number	$2X10^7$	$5X10^5$
Sorted cell number	$1X10^6$	$2X10^5$
Purity	>98%	>91%

We kept the cells to recover for two weeks marking the wells that had green cells for further inspection. We allowed the cells to expand before we resumed G418 maintenance while growing (figure 3-6). Many of the colonies did not survive, 4 colonies were expanded that we labelled (A, B, C and D). Only clone D seemed to show persistent GFP expression. Therefore, we decided to use clone D for phenotype characterisation.


A) After FACS
GFP
Brightfield
Merged
AKN1 GFP
AKN1 GFP-PRH-Myc
500 µm
500 µm
500 µm

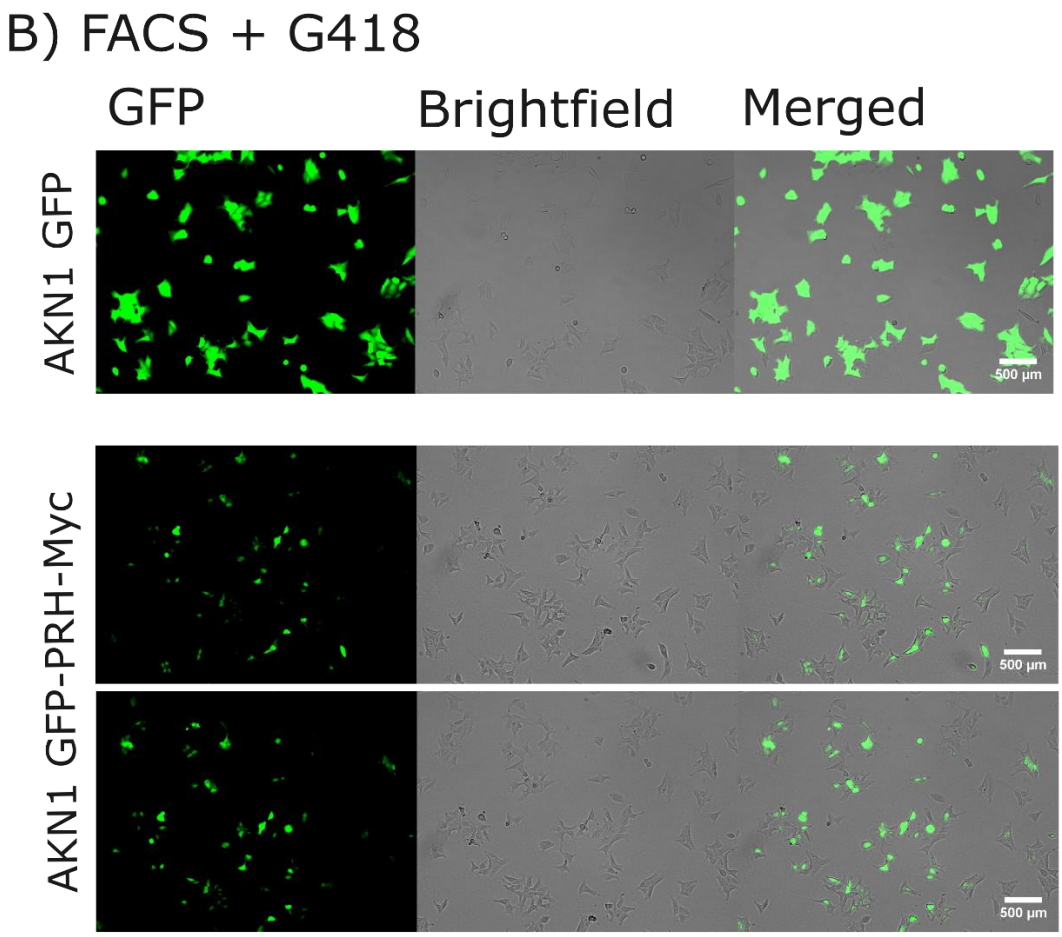


Figure 3-6 AKN1 GFP and AKN1 GFP-PRH-Myc cells after FACS single-cell sorting and growth with maintenance doses of G418 (D clone).

Cropped images of live AKN1 GFP and AKN1 GFP-PRH-Myc cells at a magnification of 2.5X acquired with a Leica DMI3000 B fluorescence microscope. They show the outcome of expanding AKN1 cells expressing either GFP or GFP-tagged PRH after FACS single-cell sorting and following the supplementation of conditioning growth media with G418. Scale bars 500 µm.

We harvested cells to make samples for RT-qPCR mRNA and Western blotting protein expression validation as detailed in the method chapter. Figure 3-7 shows that AKN1 GFP-PRH-Myc cells expressed significantly more PRH mRNA (p= 0.0025) and GFP-PRH-Myc protein (p= 0.0426) compared to AKN1 GFP cells.

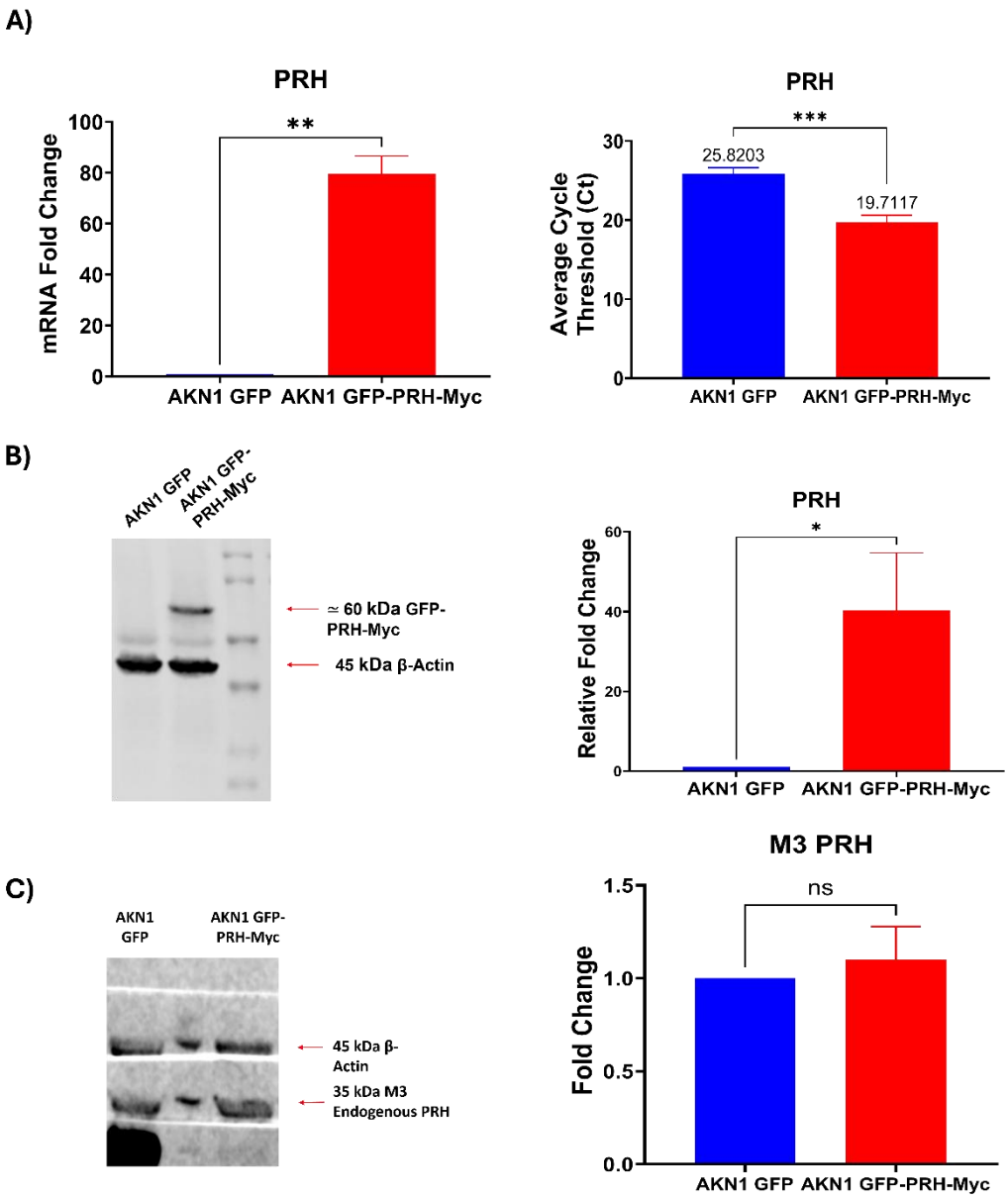


Figure 3-7 Validation of PRH mRNA and GFP-tagged PRH protein expression in AKN1 GFP and AKN1 GFP-PRH-Myc cells (D clone cell line).

(A) PRH mRNA relative fold expression showing PRH mRNA significantly more in AKN1 GFP-PRH-Myc cells compared to control cells (p= 0.0025). The average Ct value for PRH in AKN1 GFP control cells is 25.8203 reflecting the level of endogenous PRH mRNA in AKN1 cells. (B) A representative Western blotting image of AKN1 whole cell protein extract probed with a Myc Tag antibody with densitometry analysis showing a significant GFP-tagged PRH expression in AKN1 GFP-PRH-Myc cells (p= 0.0426). (C) Western blotting of endogenous PRH protein expression in AKN1 using M3 antibody against endogenous PRH. B-Actin was used as a loading reference. Both RT-qPCR and WB were performed in N=3 independent biological repeats and normalised to 1 for comparison. Two-tailed unpaired student t-test analysis was performed [P ≤ 0.05 (*), P ≤ 0.01 (**), P ≤ 0.001 (***)]. Error bars represent SD.

Altogether, the data confirmed the successful expression of either GFP or GFP-tagged PRH in AKN1 cell lines.

3.2.2 GFP-tagged PRH expression in AKN1 cells increases cell growth and proliferation

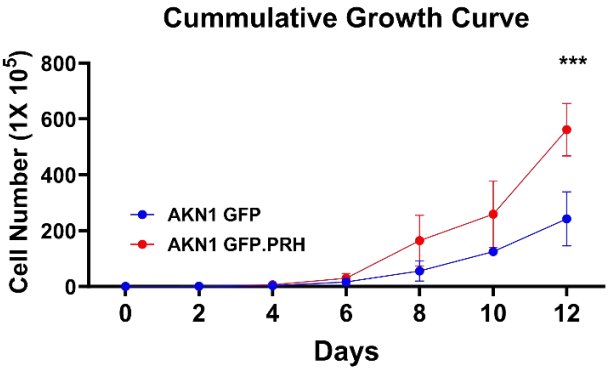
When the cells were growing, we noticed that AKN1 GFP-PRH-Myc cells needed more frequent passaging compared to AKN1 cells expressing GFP alone. To assess that observation *in vitro*, we used cumulative cell growth and EdU incorporation assays to monitor the population doubling time and cell proliferation, respectively. For cumulative cell growth, we seeded $2X10^4$ cells in 12-well plates as duplicates per condition. Over 2 weeks, we trypsinised, counted, and re-seeded the cells at known dilution every other day. We manually counted the cells using a haemocytometer and corrected the counted cells to the volume of cell suspension. We also added the counts of surplus cells that were not re-seeded to each time interval. This produced the cumulative cell number which we plotted against time to generate the final growth curve. We chose this continuous cell growth monitoring over the static snapshot time-point cell counting because we wanted to maintain the dynamics from cell proliferation over time. In addition, we wanted to have control over cell density to avoid potential growth inhibition effects from full confluency. Moreover, this method allows the flexibility of experiment progress and termination since there are no restriction on the number of time points. Cell culture maintenance was more cost-effective in our work than snapshot counting in regard to growth vessels and volumes of cell media Figure 3-8-A shows the average of N=2 independent growth experiments plotted as a

cumulative growth curve. Overall, we recorded more cell growth activity in the AKN1 GFP-PRH-Myc population from day 2 post seeding (not visible on the plot due to Y-axis scale but confirmed on the attached average cell count table). However, we started to see a difference on the plot from day 6 after which AKN1 GFP-PRH-Myc cells started exceeding the control cells in growth widening the gap between the two curves. By day 12, there were significantly more AKN1 GFP-PRH-Myc cells than AKN1 GFP cells (*p*= 0.0003). To calculate cell population doubling time from the growth curve, we used the formula below and plotted the results in a bar chart (449).

$$Doubling\ time\ (hours) = \frac{t2 - t1}{3.32\ \times (Log\ n2 - Log\ n1)}$$

Figure 3-8-B shows that AKN1 GFP-PRH-Myc cells needed on average 25.97 hours to double the original cell count which is significantly lower than the 48.37 hours required for AKN1 GFP (*p*= 0.0305).

A)



	Average Cell Count (1X10^5)	
Days	AKN1 GFP	AKN1 GFP-PRH-Myc
0	0.2	0.2
2	0.816	1.255
4	2.8	6.031
6	16.33	29.79
8	55.65	164.7
10	125.1	258.9
12	242.5	561.5

B)

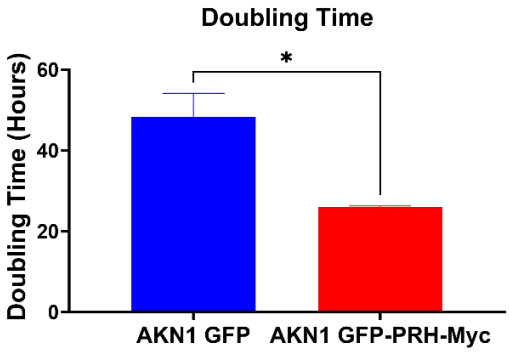


Figure 3-8 GFP-tagged PRH expression in AKN1 cells appears to significantly induce cell growth and shorten cell doubling time.

(A) A cumulative growth curve that shows total cell number at different time points over 12 days. Every 2 days after seeding the cells were trypsinised and counted in duplicates. The experiment was done as N=2 independent biological repeats. AKN1 GFP-PRH-Myc cells have significantly more cell number on day 12 compared to control cells (p=0.0003). Two-way ANOVA statistical analysis with multiple comparisons was performed. Error bars represent SEM. (B) Doubling time calculated from the linear part of the growth curve as N=2 independent biological repeats showing a significant shorter time required for cell doubling in AKN1 GFP-PRH-Myc, compared to control cells (p= 0.0305). Analysis was done using one-tailed unpaired student t-test. Error bars represent SEM.

To examine cell proliferation, we performed EdU incorporation assay. This assay relies on the cells incorporating synthetic EdU into their DNA during DNA synthesis while transitioning through the S phase of the cell cycle to reflect cell proliferation. We performed the assay as described in the method

chapter. The results indicated that over 2 hours AKN1 GFP-PRH-Myc cells showed a greater number of cells (≈50%) transitioning through the S phase than AKN1 GFP cells (*p*= 0.0044) (figure 3-9).

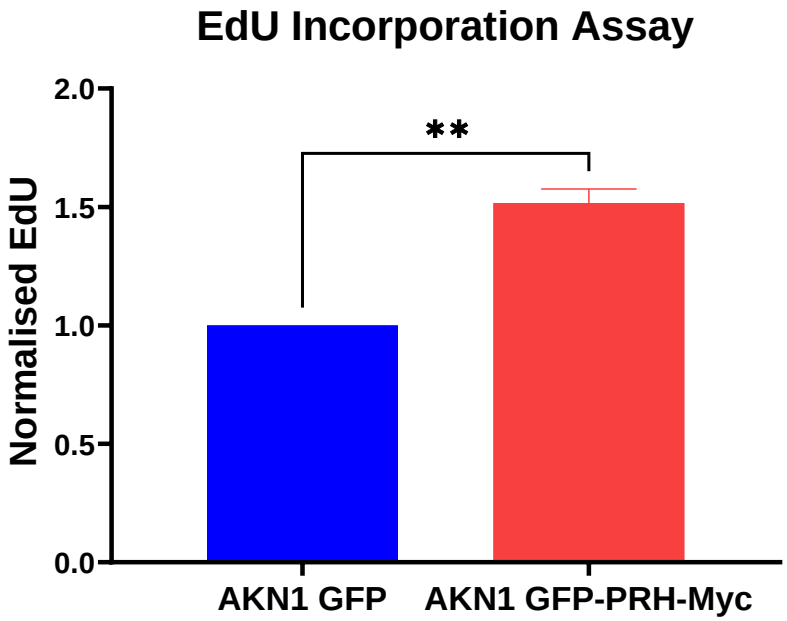


Figure 3-9 GFP-tagged PRH expression in AKN1 cells induces cell proliferation.

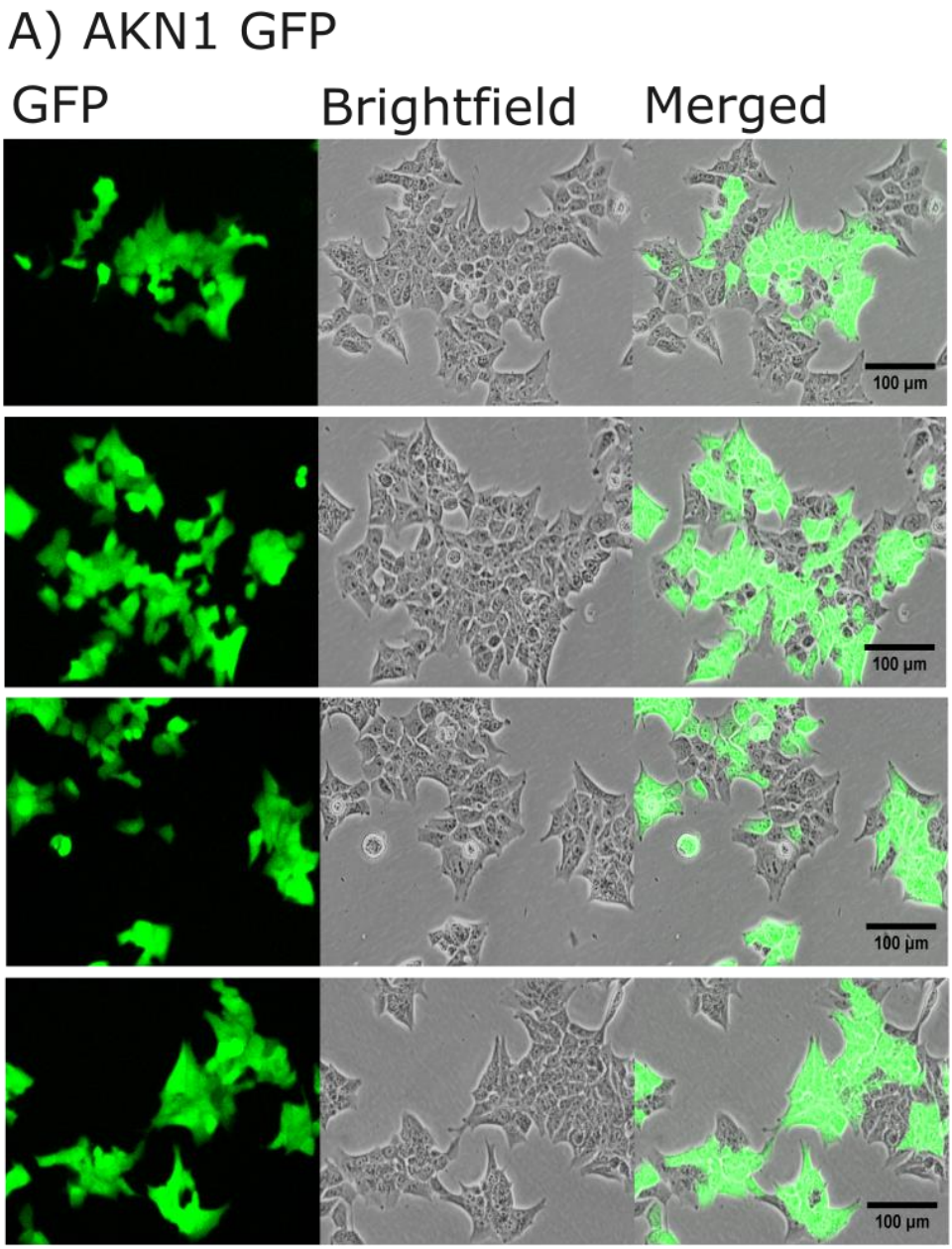
EdU incorporation assay shows significantly more proliferative activity in AKN1 GFP-PRH-Myc cells compared to control cells. Cells were incubated with EdU for 2 hours and the Click-IT reaction was detected by measuring the absorbance of the colour produced from the addition of an OPD substrate for HRP-conjugated antibody at 450 nm. Correction to cell number was achieved by crystal violet normalisation. Absorbance was normalised to cell number and then to 1 for comparison. The assay was performed as N=3 independent biological repeats and analysed using two-tailed unpaired student t-test. (*p*= 0.0044). Error bars represent SD.

We concluded that GFP-tagged PRH expression promotes cell proliferation in AKN1 cells.

3.2.3 Morphological characterisation of AKN1 GFP-PRH-Myc cells

Studying the morphology of cells gives an important indication of cell phenotype and biological activities. Brightfield imaging showed that AKN1

GFP cells were clustered and flattened, whereas AKN1 GFP-PRH-Myc cells were slightly elongated with less cell-cell cohesion (figure 3-10).



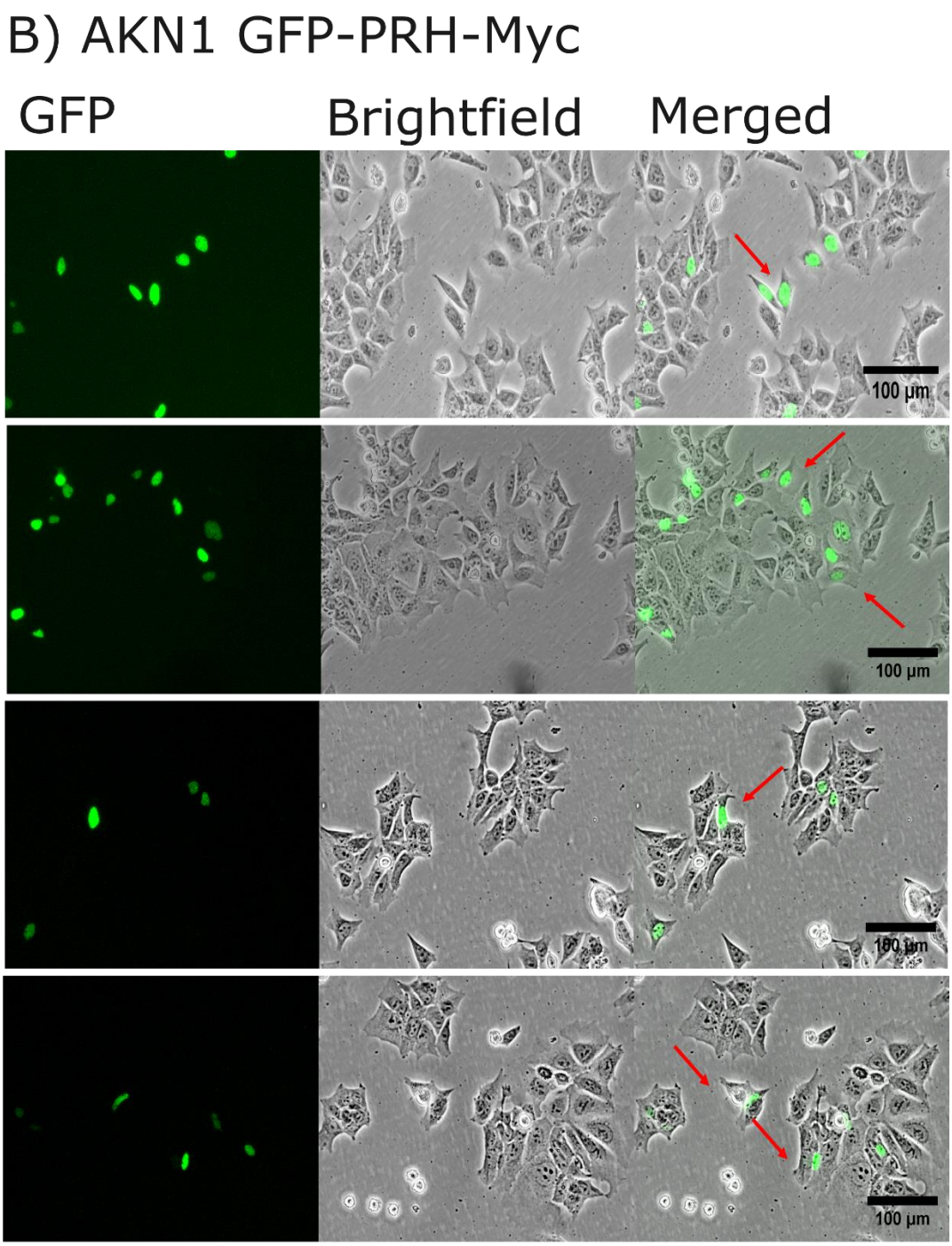


Figure 3-10 **GFP-tagged PRH expression induces mesenchymal-like morphology in AKN1 cells.**

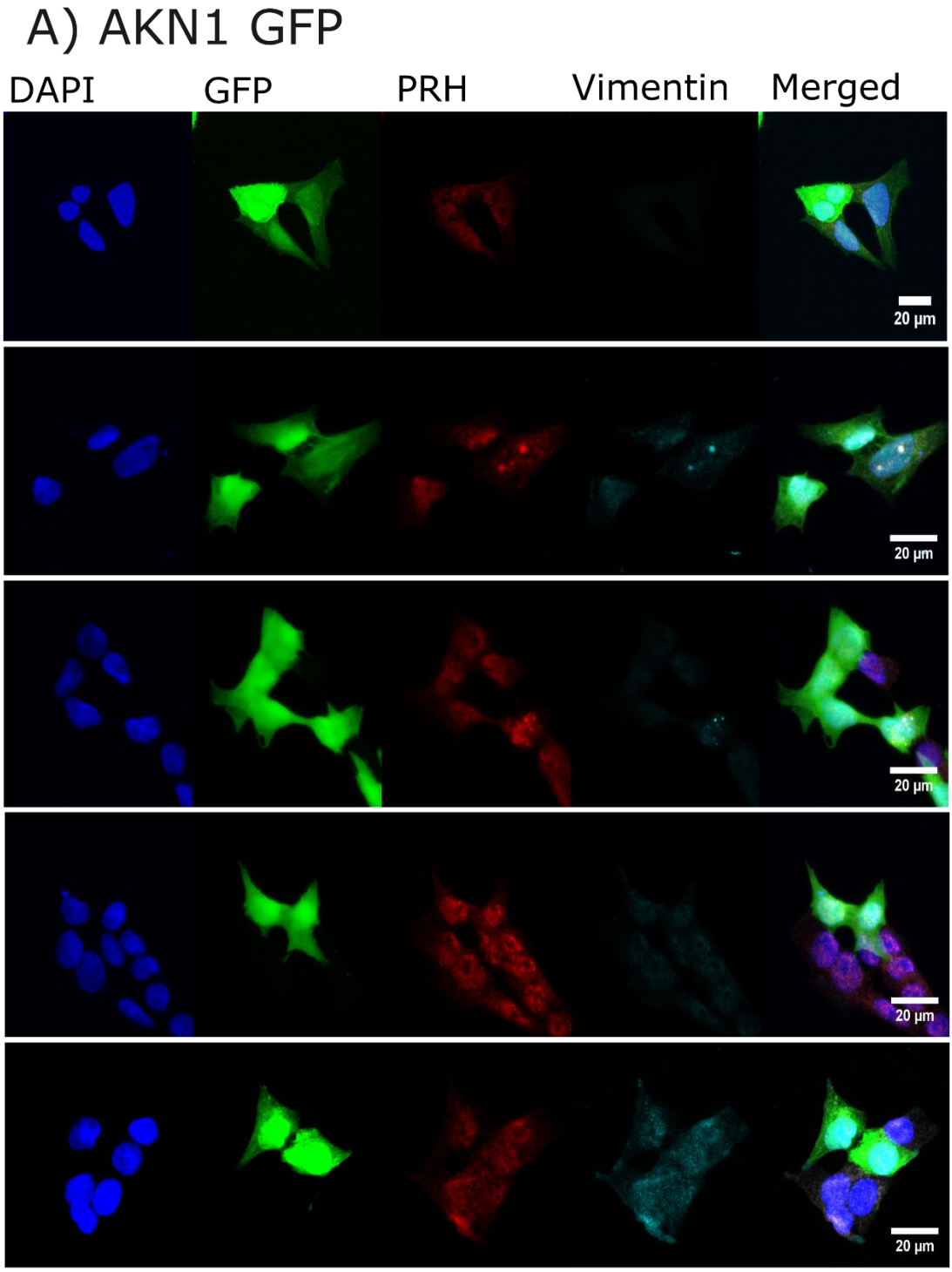
Morphology of clustered epithelial AKN1 GFP cells and mesenchymal-like AKN1 GFP-PRH-Myc cells (red arrows) on brightfield images acquired at 1.5X magnification using a Leica DMI3000 B fluorescence microscope. Scale bars 100 µm.

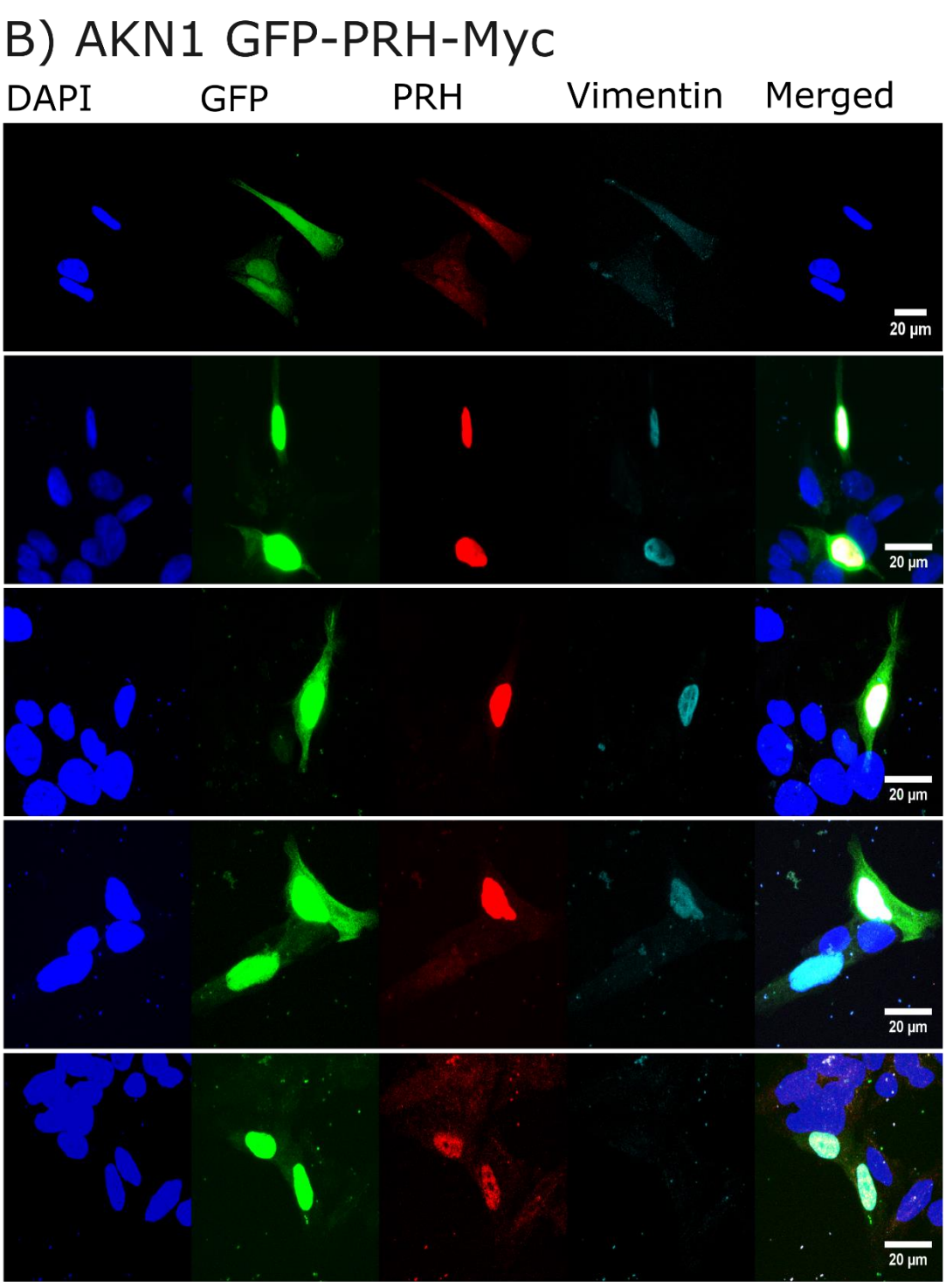
To have a better view of the impact of GFP-tagged PRH expression on AKN1 cells morphology, we probed fixed samples of AKN1 GFP and AKN1 GFP-PRH-Myc cells with PRH and Vimentin antibodies after which we acquired microscopic imaging using a confocal Leica DMI4000B microscope. To ensure obtaining a true expression intensity measurement from the images, we acquired the images using the same settings for laser excitation power, exposure, gain, pinhole size, and magnification. We also imaged cells of negative controls where a primary antibody was not used to exclude interfering signals from the secondary antibodies.

To statistically analyse the findings, we used the software FIJI to extract and analyse data from the images (450). We first processed all the images to remove the background and exclude false positive signals. We used the average intensity of at least five negative control images where no primary antibodies were added. This background subtraction created processed images ready for analysis. To produce representative images, we adjusted the brightness and contrast equally across the images. We generated a composite image of all the channels. Then, we used the plugin (Figure Combine pt2) created by Dr Kenton Arkil to make a compiled image (451).

As seen in (figure 3-11–A), our images included DAPI in blue, GFP in green, PRH in red, and Vimentin in cyan. Visual inspection of the cells revealed completely green cells for both cell lines, but less cytoplasmic and more nuclear for AKN1 GFP-PRH-Myc cells (figure 3-11-A-green). In addition, there was also a punctuated form of PRH expression in both cell lines (figure 3-11-A-red), more prominent in AKN1 GFP-PRH-Myc cells. Moreover, GFP-tagged

PRH expression induced a similar punctuated form of Vimentin (Figure 3-11-A-cyan) which seemed to appear when the cells adopt a mesenchymal shape.




B) AKN1 GFP-PRH-Myc
DAPI
GFP
PRH
Vimentin
Merged
20 µm
20 µm
20 µm
20 µm
20 µm

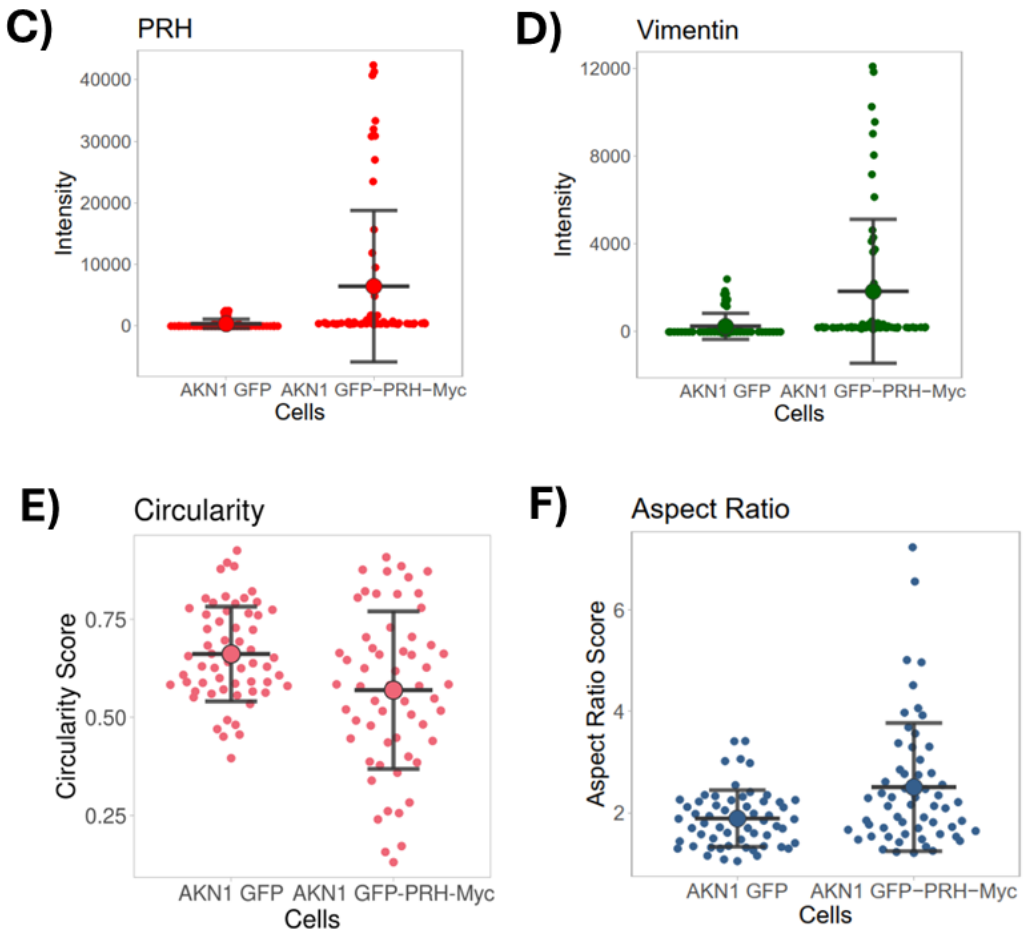


Figure 3-11 Morphological analysis of AKN1 GFP and AKN1 GFP-PRH-Myc cells using immunofluorescence imaging.

(A) AKN1 GFP images show polygonal cells. (B) AKN1 GFP-PRH-Myc images show elongated cells with more PRH and Vimentin visualised compared to control cells. Cells were fixed with 4% FA and labelled with PRH and Vimentin antibodies along with nuclear DAPI staining. Images were acquired sequentially for n= 59 cells using a confocal Leica DMI4000B microscope at 63X magnification. Scale bars 20 µm. Laser power, gain, exposure, and pinhole were the same throughout the imaging. Expression intensities represented in violin plots show significant increase in expressions of (C) PRH *p*= 4.00e-04 and (D) Vimentin *p*= 5.64e-04 in AKN1 GFP-PRH-Myc cells compared to control cells. Shape descriptor scores show AKN1 GFP-PRH-Myc cells significantly less (E) circular *p*= 0.003 compared to control cells. (F) GFP-tagged PRH expression induces cell stretching and elongation in AKN1 cells *p*= 0.001. Ordinary t-test analysis was performed, and error bars represent SD.

To statistically analyse the images, we relied on the combination of GFP and PRH channels to threshold (segment) the image and define our ROI as black regions over a white background. We measured the mean intensity for both PRH and Vimentin (n=59 cells). The data indicated a significant increase in

PRH staining (p= 4.00e-04) and Vimentin staining (p= 5.64e-04) expression in AKN1 GFP-PRH-Myc cells compared to control cells (figure 3-11 C and D).

To statistically analyse the change in cell morphology, we used shape descriptor scores circularity and aspect ratio available on FIJI (450). To analyse cell circularity, FIJI applies the formula [circularity=4π (area/perimeter^2)] on pre-selected ROI. A score value of 1.0 implies a perfect circle, while values closer to 0.0 indicates leaning towards an elongated shape. The other shape score we used on FIJI was aspect ratio which calculates the ratio of major axis (L) to minor axis (s) of a pre-selected ROI. Aspect ratio takes into consideration maximum and minimum Feret's diameters, i.e. longest and shortest distances between points on the selected boundary, respectively. Our results indicated that on average there were significantly fewer circular cells (p= 0.003) (figure 3-11-E) and more elongated cells (p= 0.001) (figure 3-11-F) in the AKN1 GFP-PRH-Myc population compared to AKN1 GFP population. This implied that GFP-tagged PRH expression induces a substantial increase in cell elongation.

Since Vimentin was seen induced in AKN1 GFP-PRH-Myc cells, we examined E-Cadherin expression using Western blotting on the same lysate from (figure 3-7-B). The results showed a small but statistically significant increase in E-Cadherin expression (p= 0.0135) in AKN1 GFP-PRH-Myc cells compared to control cells (figure 3-12).

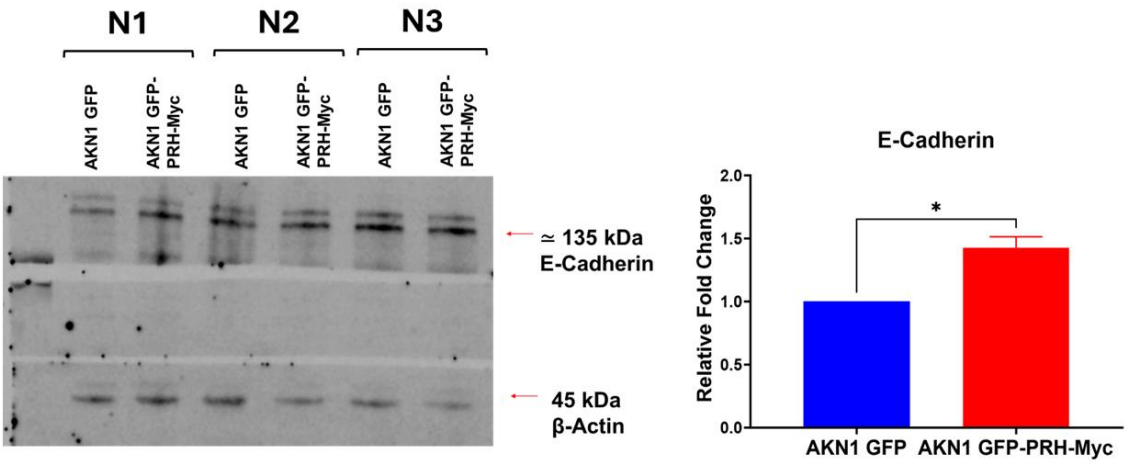


Figure 3-12 GFP-tagged PRH expression induces E-Cadherin expression in AKN1 cells.

E-Cadherin in AKN1 GFP and AKN1 GFP-PRH-Myc whole cell protein extracts probed with E-Cadherin antibodies. Densitometry analysis for both GFP-tagged PRH and E-Cadherin show a significant increase in AKN1 GFP-PRH-Myc compared to control cells (E-Cadherin *p*= 0.0135). B-Actin was used to show protein loading. WB was performed in N=3 independent biological repeats and normalised to 1 for comparison. Two-tailed unpaired student t-test analysis was performed [P ≤ 0.05 (*), P ≤ 0.01 (**), P ≤ 0.001 (***)]. Error bars represent SD.

Altogether, the data indicated that GFP-tagged PRH expression in AKN1 GFP-PRH-Myc cells induced some morphological changes.

3.2.4 AKN1 cells expressing GFP-tagged PRH increases cell migration and cell invasion

To assess whether we have a change in cell migration ability, we performed transwell migration (chemotaxis) experiments as described in the method chapter. In brief, we seeded $1X10^5$ cells on the apical side of Boydon chamber inserts that have porous membranes. We subjected the cells to a 2%-10% FBS gradient across the membrane in presence of 1 mM of Hydroxyurea which prevents cell proliferation. Hydroxyurea blocks ribonucleotide reductase which is important to produce nucleic acids. Therefore, Hydroxyurea arrests the cell cycle to exclude the effect of cell division on the number of migrated

cells. We estimated the number of migrated cells by interpolating the fluorescence of CalceinAM-stained cells on the basal side of the insert from a standard curve plotted from serial dilution of cells labelled with CalceinAM. To calculate the migration percentage, we divided the estimated cell number by the seeding quantity using the equation:

$$Migration\ \% = \frac{number\ of\ cells\ on\ basal\ side}{number\ of\ seeded\ cells} \times 100$$

Figure 3-13 shows that there was a significant increase in the migration rate of AKN1 GFP-PRH-Myc cells in comparison to the control cells (p= 0.0498).

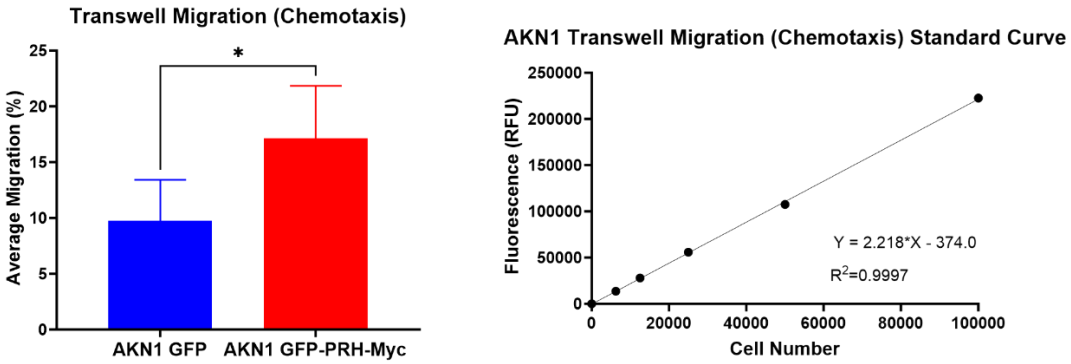


Figure 3-13 GFP-tagged PRH expression significantly increases AKN1 transwell migration (chemotaxis).

1X105[3] cells were seeded on a transwell insert and left to migrate under the influence of a (2%-10%) FBS gradient across a porous membrane for 72 hours in the presence of 1mM antiproliferative Hydroxyurea throughout the assay. Cells on basal side of the inserts (migrated cells) were labelled with 8 µM Calcein AM. Percentage of migrated cells were interpolated from a standard curve generated by using a serial dilution of AKN1 cells stained with 8 µM Calcein AM. The experiment was performed as N=4 independent biological repeats and analysed using a two-tailed unpaired student t-test. p= 0.0498 and error bars represent SD.

To investigate cell invasiveness, we performed N=2 independent biological transwell invasion experiments as detailed in the method section. In brief, we seeded 5X10[4] cells on the apical side of a Boyden chamber insert coated with a thin layer of 2 mg/ml Matrigel. As in the chemotaxis experiments, we created

a 2%-10% FBS gradient with 1 mM Hydroxyurea across the porous membrane. To count the number of invading cells, we fixed the cells on the basal side of the insert with 4% FA before staining them with DAPI. We acquired several images representing different views of the membranes using a Leica DMI3000 B fluorescence microscope. Then, we counted the total number of DAPI stained nuclei and divided it by the number of imaged fields using the equation:

$$Average\ Invasion = \frac{total\ number\ of\ cells\ on\ basal\ side}{number\ of\ fields}$$

Figure 3-14 shows that AKN1 GFP-PRH-Myc cells appeared significantly more invasive than the control AKN1 GFP cells (p= 0.0494).

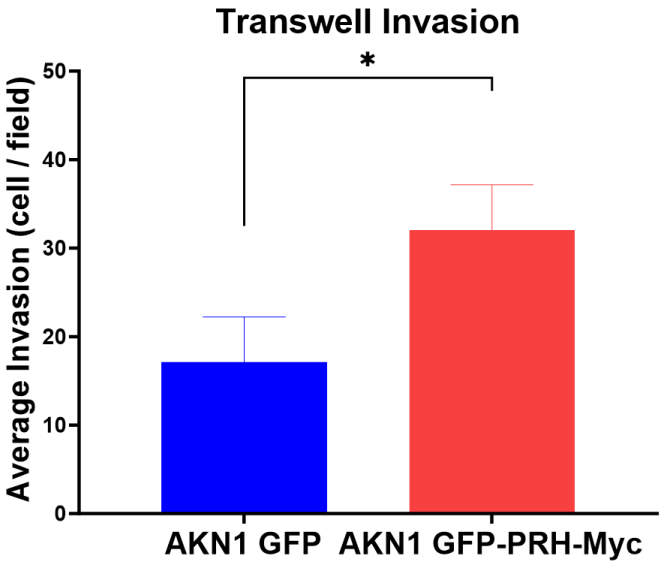


Figure 3-14 GFP-tagged PRH expression induces a significant increase in AKN1 transwell invasion.

5X10^4 cells were seeded on a transwell insert coated with a thin layer of 2 mg/ml Matrigel and left to migrate under the influence of a (2%-10%) FBS gradient across a porous membrane for 72 hours in presence of 1mM antiproliferative Hydroxyurea throughout the assay. Cells on basal side of the inserts (invaded cells) were fixed with 4% FA and labelled with DAPI. Average cell invasion per view field was calculated from N=2 independent biological repeats and analysed using a one-tailed paired student t-test. p= 0. 0.0494 and error bars represented SEM.

Taking the findings together, these data indicated that AKN1 GFP-PRH-Myc cells are more migratory and more invasive compared to the control cells that expressed GFP alone. Both biological processes are important for cancer cells to find their way through the surrounding ECM and metastasise.

3.2.5 Issues between early and late passages of AKN1 GFP and AKN1 GFP-PRH-Myc clone D

Earlier in this chapter we showed how we generated AKN1 cells expressing GFP-tagged PRH, and the measures we took trying to maintain GFP expression by G418 selection and FACS single cell sorting. We successfully characterised GFP-tagged PRH expression in AKN1 cells in comparison to the control cell line. However, in most of the experiments we needed early passages of the cell lines in which GFP-tagged PRH was expressed at high levels. We experienced a gradual decline in GFP-tagged PRH expression that started taking place after the fifth passage of the cells. We even had a marked reduction of GFP-tagged PRH expression after the seventh passage where the ratio of GFP-expressing to non-GFP expressing cells was very low. The control cell line AKN1 GFP was also going through a similar loss of GFP expression, but at a slower rate. To show examples of some challenges regarding GFP-tagged PRH expression in later passages, we included here the status of PRH expression at mRNA and protein levels. We also added data from EdU incorporation assays at early and late passages. For late passages experiments, we used cells from passages later than 7.

Previously, we presented data demonstrating PRH expression at mRNA and protein levels in samples prepared from early passages of sorted AKN1 GFP

and AKN1 GFP-PRH-Myc cells (section 3.2.1 - figure 3-7). However, when we tested cDNA and protein lysates prepared from late passages, we detected no significant difference at mRNA levels (p= 0.6148) and various levels of GFP-tagged PRH protein expressions in different passages (figure 3-15).

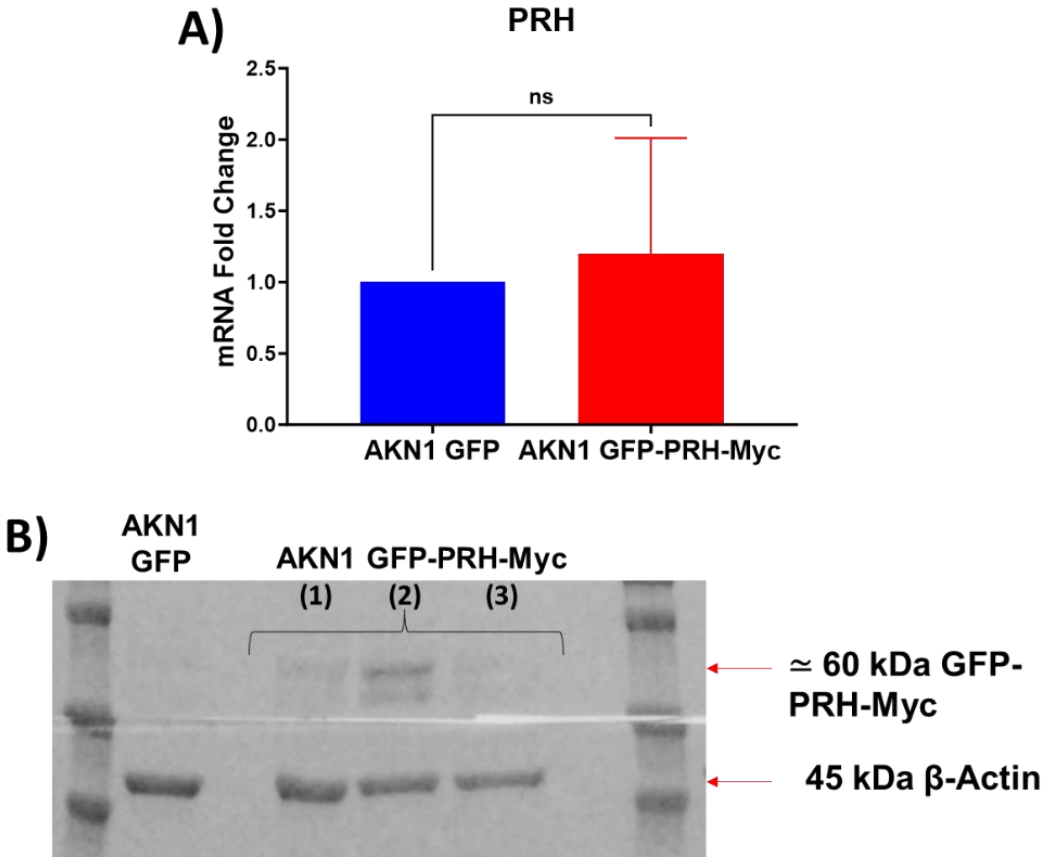


Figure 3-15 Late passages of clone D AKN1 GFP-PRH-Myc cells show no change in PRH mRNA levels and various GFP-tagged PRH protein levels.

(A) PRH mRNA relative fold expression showing little alteration in PRH mRNA in AKN1 GFP-PRH-Myc compared to control cells (n=5, p= 0.6148). RT-qPCR was performed in N=5 independent biological repeats and normalised to 1 for comparison. Two-tailed unpaired student t-test analysis was performed. Error bars represent SD. (B) A representative Western blotting image of AKN1 GFP and AKN1 GFP-PRH-Myc whole cell protein extract probed with GFP antibody. Protein extracts was prepared from 3 different AKN1 GFP-PRH-Myc cell populations. β-Actin was used to show protein loading.

To see an example of challenges we faced with investigating the GFP-tagged PRH expression phenotype in AKN1 cells, we used EdU incorporation assay

to examine cell proliferation. Figure 3-16 shows the difference in cell proliferation between early (A) and late (B) passages of the cells. There was a statistically significant increase in the proliferation of early passages AKN1 GFP-PRH-Myc cells compared to control cells (p= 0.0044), while cells from late passages showed a much smaller difference if any (p= 0.8904).

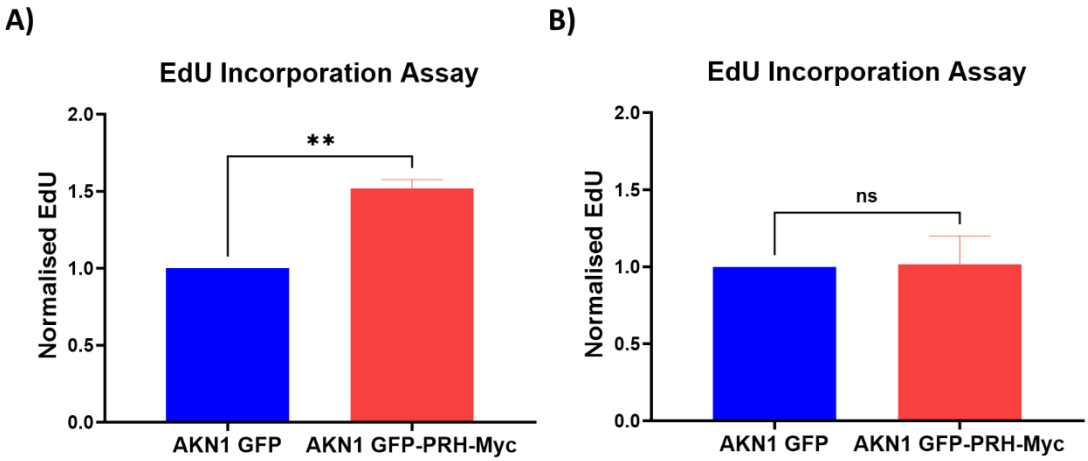


Figure 3-16 The reduction in cell proliferation of AKN1 GFP-PRH-Myc cells in later passages.

(A) shows the previous EdU incorporation assay of AKN1 stably expressing GFP-tagged PRH where they are significantly more proliferative compared to their control (p= 0.0044). (b) A late passage (beyond 7) EdU incorporation assay showing no difference in proliferation. Cells were incubated with and without EdU for 2 hours and the Click-IT reaction was detected by measuring the absorbance of the colour produced from the addition of an OPD substrate for HRP-conjugated antibody at 450 nm. Absorbance was normalised to cell number and then to 1 for comparison. EdU Assay was performed as N=3 independent biological repeats and analysed using two-tailed unpaired student t-test (p= 0.8904) error bars represent SD.

Altogether, these data suggested that GFP-tagged PRH expression is stable up to around passage 5 and then it starts to decline to undetectable levels. In general, we had mixed population AKN1 GFP and AKN1 GFP-PRH-Myc cells.

3.3 Discussion

The aim in this chapter was to generate a human immortalised AKN1 cells expressing GFP-tagged PRH for phenotype characterisation and further future experiments. This work was performed in parallel with collaborators (Kitchen et al.) who reported that PRH acts as an oncoprotein in CCA cells producing cellular phenotypes associated with tumour growth and progression (215). We used cumulative growth and EdU incorporation assays for cell growth and proliferation, respectively. We acquired immunofluorescence images to assess cell morphology. We also performed transwell migration and invasion assays for cell migration and cell invasion, respectively. Finally, we reported data reflecting the issue we encountered with GFP and GFP-tagged PRH expression in AKN1 cells.

AKN1 cells are immortalised cholangiocytes isolated from a normal liver initially meant for donation (406). Cell characterisation revealed that the cells underwent neoplastic transformation with abnormal karyotypic changes of chromosomes 2 and 8 that led to their immortalisation (406). Although AKN1 cells are reported to have an accumulation of p53 protein, there is no evidence of *p53* gene mutation (406). There are no reports on the cells originally harbouring *neoR* gene (confers resistance to neomycin or G418) or other gene mutations which may influence the genetic stability of AKN1 cells. Yet, the accumulation of p53 proteins may have an impact on cell cycle progress and apoptosis. In turn, this might reduce the efficacy of transient transfection in the cells, decrease the number of cells integrating the overexpressed transgene for stable transfection, and affect the response to G418 antibiotic selection.

These issues might have contributed to stability of GFP-tagged PRH expression in AKN1 cells.

The level of PRH overexpression in AKN1 GFP-PRH-Myc cell line was evident at both mRNA and protein levels. This was consistent with the level of expression reported by Kitchen et al. (215). We single-sorted the cells in addition to G418 antibiotic selection to maintain GFP-tagged PRH expression in a monoclonal cell line. Monoclonality in cell lines is associated with the benefits of genetic homogeneity which minimises the variability in the biological behaviours between the cells (452). In addition, monoclonality often ensures more expression stability which makes them preferred over polyclonal cell lines for long-term experiments (453). We only managed to characterise a single clone when it is ideal to characterise at least three separate clones to make an unequivocal conclusion in the context of GFP-tagged PRH expression. Arguably, polyclonal transfected cell lines have the advantage of reflecting the tumour heterogeneity through their diverse cellular makeup particularly after cell sorting. The cells are able to generate a wide range of responses to treatments making them fit for drug sensitivity studies (454). While such environment would have been a good option for our project, we needed the FACS single-cell sorting to maintain GFP expression in our cells. We still have the option to characterise the other clones for AKN1 expressing GFP-tagged PRH from FACS sorting to confirm our current findings and compensate for not using polyclonal transfected cell lines. FACS sorting itself has been reported to introduce oxidative stress and cause metabolic changes in the cells independent of the treatment (455). These changes may affect the

biological behaviour of cells including cells transformation. While this outcome is extremely unlikely, we still cannot exclude its occurrence in our cells post sorting.

GFP-tagged PRH significantly induced cell growth and cell proliferation in AKN1 cells. Sustained cell proliferation is an important hallmark of cancer cells (6). Our results came in agreement with the role of PRH as an oncoprotein in CCA cells where it was found to act as an oncoprotein promoting tumour growth and progression (215). PRH was also reported to act as an oncoprotein in acute T cell lymphoblastic lymphoma where PRH was upregulated by *LMO2* oncogene (266). Moreover, PRH interacted with the complex YAP/TEAD4 in HCT-116 cells and promoted their tumorigenesis (243). This interaction was associated with poor prognosis in colorectal cancer (243). In contrast, in other studies, PRH was seen to act as a TSG. We mentioned in the introduction that PRH expression also inhibits cell proliferation. PRH overexpression inhibited the action of EIF-4E and reduced cell proliferation in leukaemia cells (233, 255, 261). In addition, PRH overexpression in a hepatoma cell line resulted in increased expression of *TP53* and *RB*, and decreased expression of *Bcl2* proto-oncogenes. Subsequent *In vivo* work in nude mice revealed that these cells prevented tumour formation in the liver (230). Therefore, PRH can promote and inhibit cell growth and proliferation. Beside cell growth and cell proliferation, we examined the changes in cell shape and whether we had changes in cells migratory and invasive abilities. We showed that AKN1 GFP-PRH-Myc cells tended to depict mesenchymal-like cell shape with less cell-cell cohesion. Interestingly we only detected an upregulated Vimentin

expression through the immunofluorescence staining of AKN1 GFP-PRH-Myc cells. Vimentin appeared in cells that were elongated and strongly expressing PRH, but less prominent when cells were expressing PRH while in a transitioning shape between epithelial and mesenchymal-like cells. The expression of Vimentin in our cells, although mostly limited to the cell nucleus, was similar to the reports on early stages of Vimentin organisation when overexpressed in MCF7 cells (335). The authors reported that Vimentin was expressed as particles and squiggles (335). As we mentioned in the introduction, Vimentin goes through a multistep process to assemble its monomers into a cytoskeletal network (317). Vimentin is a well-known mesenchymal cell marker that when it is upregulated, there is usually an induction of EMT (273). Surprisingly, GFP-tagged PRH also significantly induced E-Cadherin expression at mRNA and protein levels. E-Cadherin is a well-known epithelial marker that is usually downregulated in EMT (83). Although GFP-tagged PRH slightly induced Vimentin while persistently expressing E-Cadherin, AKN1 GFP-PRH-Myc cells were showing increased migration and invasion rates. Our Vimentin and E-Cadherin expression levels were not consistent with the level of expression reported by Kitchen et al. in their study where they reported that PRH overexpression in AKN1 cells downregulated E-Cadherin and upregulated Vimentin at both mRNA and protein levels (215). However, GFP-tagged PRH still increased cell migration and cell invasion which agreed with the published data (215). Other published data reported that TGF-β downregulated PRH which was associated with decreased E-Cadherin expression in PC3 prostate cancer cells and normal

PNT2-C2 cells where that induced cell migration in both cell lines (456). In addition, PRH downregulates EMT in breast cancer by repressing *GSC* (264). PRH overexpression was also reported to inhibit cell migration and cell invasion of MDA-MB-231 breast cancer cells (260). PRH. AKN1 GFP-PRH-Myc cells were in transition between epithelial and mesenchymal states, with induced Vimentin and E-Cadherin expression. However, the literature reported that mixed findings of both EMT and MET markers is common in invasive cancer cells (93). Such a picture is referred to as partial EMT, hybrid EMT-MET, or incomplete EMT (93). We could take our cells for further aim to investigate whether our AKN1 GFP-PRH-Myc display this phenomenon. We could perform detailed molecular characterisation of more EMT markers using RT-qPCR and Western blotting. In addition, detailed transcriptomic analysis can identify gene set enrichment associated with partial EMT.

It is important to mention that cell morphology analysis through immunofluorescence staining requires staining the cell cytoskeleton to give an insight into the shape of the cells under investigation. We relied on GFP expression to define the ROI when we applied the shape descriptors on FIJI. GFP expression for cell morphology analysis was found valid before when it was exploited to analyse neuronal cells morphology (457).

We encountered issues regarding the longevity of GFP-tagged PRH expression beyond passage 5 to 7. We initially had high level of GFP-tagged PRH expression in our cells, but it declined as the cells grew older. However, Kitchen et al. managed to maintain a stable expression of PRH in their study (215). Although we used a common transfection method to express GFP-

tagged PRH, we cannot guarantee 100% level of success as cells tend to show variability in their responses to gene manipulation. AKN1 GFP-PRH-Myc cells at higher passages experienced a dramatic drop in PRH mRNA expression that closely matched the basal level of expression. That was associated with various degrees of reductions in GFP-tagged PRH at protein levels. This was reflected on the phenotype as well where there was a reduction in cell proliferation to a level that almost matched the control AKN1 GFP cells. Loss of GFP-expressing cells could be attributed to various reasons such as unstable genomic integration of the transgene, epigenetic modifications including promoter methylation, or loss of integrity of the transgene. Alternatively, this might be due to a post-translational modification leading to protein degradation.

One significant limitation in our approach is characterising a single clone of AKN1 expressing GFP-tagged PRH (clone D) due to time and fund restriction. Without confirming the findings by validating more clones of the cell lines we cannot guarantee observation reproducibility or generalise the findings to the whole population due to genetic and epigenetic variations between the clones. We previously mentioned that monoclonal cell lines ensure findings homogeneity within the single population, but they lack tumour diversity representation (454). Moreover, initial screening of all clones my induce bias selection toward the clone taken forward for further experiments. Yet, these limitations could be overcome by including the other clones for characterisation as well as the polyclonal sorted cells.

Also, we relied on a single method for plasmid transfection using PEI when there are alternative methods that might be superior to PEI transfection of AKN1 cells. Lipid-based reagents like Lipofectamine and Fugene can facilitate the uptake of plasmid into the cells for reliable results with minimised toxicity (458). Inducible viral transduction is another method to consider for inducing controllable levels of GFP-tagged PRH expression.

Another controversial limitation is the use of a plasmid containing a gene expressing GFP in its backbone. While it has the advantage of rapid monitoring the transfection in live cells using live microscopy, it caused some confusion with the immunofluorescence investigation of protein localisation. This could be avoided by either trying a control plasmid without GFP or label the cells with antibody against GFP to provide a reference on how the cells look after transfection prior to PRH and Vimentin immunofluorescence. In addition, GFP has been reported to induce cell toxicity and death (459).

When it comes to longevity of GFP-tagged PRH expression, our cells went through several days of recovery and growth expansion after FACS single-cell sorting. This was accompanied by loss of expression and phenotype toward later passages. This could be solved by using the cells after transient transfection or making a polyclonal cell line of the sorted cells for rapid growth expansion and heterogeneity preservation of the cell line. It would be beneficial to further investigate the reasons affecting the longevity of GFP-tagged PRH expression and to salvage the remaining stocks of monoclonal AKN1 GFP-PRH-Myc cells and use them for the detailed reassessment of EMT markers and cell morphology. Further experiments where AKN1 GFP-

PRH-Myc cells could be exploited include transwell extravasation experiments to study the effect of PRH on AKN1 cells ability to adhere and cross the endothelial cells. Moreover, the cells could be taken further to *in vivo* work to study the effect of GFP-tagged PRH in AKN1 cells on tumour growth.

To summarise, we generated an AKN1 GFP-PRH-Myc cell line that we used to validate published *in vitro* key results of PRH expression in CCA. In general, our findings have come in agreement with the published data. GFP-tagged PRH expression in AKN1 cells promotes growth and proliferation rates. It also induces cells migration and invasion. In addition, we added to the current knowledge microscopic data showing how PRH is expressed in immunofluorescence imaging and how that influenced cell morphology and the expression of the Vimentin EMT marker protein. Published data reported that PRH exerts oncogenic and tumour suppressing effects on cells. PRH is a transcription factor that acts as an oncoprotein with tumour suppressing functions.

Chapter 4: The investigation of PRH and Vimentin localisation in CCA cells

4.1 Introduction

In the previous chapter we generated AKN1 cell lines expressing either GFP alone or GFP-tagged PRH. We confirmed that PRH overexpression in AKN1 cells promotes cell growth and proliferation rates. We also showed that PRH overexpression drives phenotypic changes resembling partial EMT and increases cell migration and invasion. Published RNAseq data obtained following overexpression of PRH in CCA cells indicated that PRH enriches gene sets and pathways important for CCA progression including EMT and cell migration (215). One specific interesting target gene for PRH in CCA is *VIM,* which encodes the cytoskeletal protein Vimentin (273, 460). Vimentin is a well-known mesenchymal marker that is crucial for numerous cellular functions contributing to cancer metastasis, e.g. EMT regulation (359), cell migration (326, 461), and cell proliferation (383, 384, 462).

Dr Ka Ying Lee has demonstrated that PRH has a peak at the *VIM* gene in ChIPseq data (figure 4-1). Dr Lee also showed that Vimentin is a potential protein partner among several others for PRH using mass spectrometry (figure 4-2). Moreover, Dr Lee confirmed this PPI using a co-immunoprecipitation assay in CCLP Myc-PRH and CCLP AdEmpty cellular fractions (figure 4-3) (269). These data suggest that Vimentin may play a role in the regulation of gene expression by PRH. We mentioned in the introduction that Vimentin regulates the expression of *p21Waf1* in Neuroblastoma cells (463).

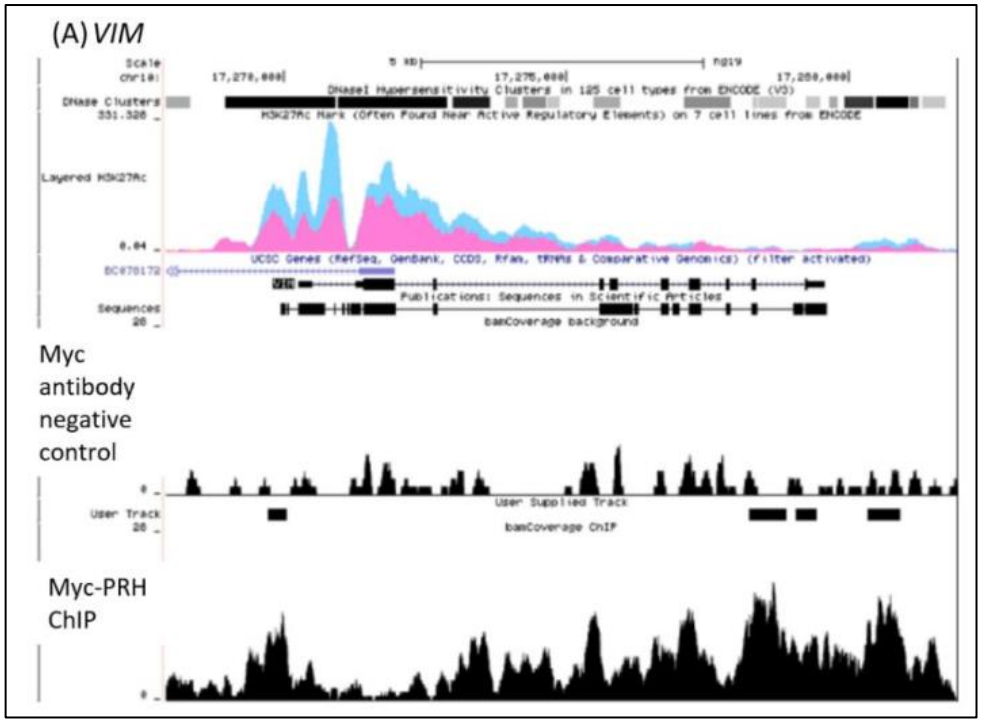
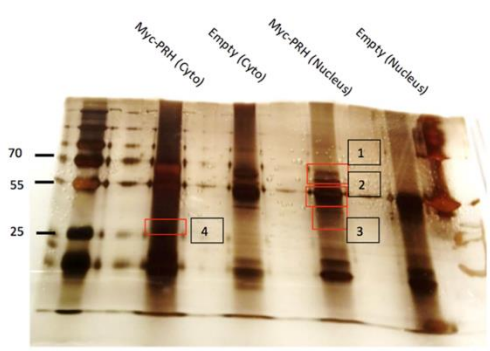


Figure 4-1 ChIPseq track of CCLP Myc-PRH visualised on UCSC genome browser.

There are multiple peaks for PRH at *VIM* gene. Data generated by Ka Ying Lee and Jayaraman.

A)



B)

Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# AAs	MW [kDa]	calc. pI
P04264	Keratin, type II cytoskeletal 1 OS=Homo sapiens GN=KRT1 PE=1 SV=6 - [K2C1_HUMAN]	29.32	19.10	2	7	7	8	644	66.0	8.12
P35527	Keratin, type I cytoskeletal 9 OS=Homo sapiens GN=KRT9 PE=1 SV=3 - [K1C9_HUMAN]	7.83	6.26	1	2	2	2	623	62.0	5.24
P35908	Keratin, type II cytoskeletal 2 epidermal OS=Homo sapiens GN=KRT2 PE=1 SV=2 - [K22E_HUMAN]	6.53	3.76	3	2	2	2	639	65.4	8.00
P08670	Vimentin OS=Homo sapiens GN=VIM PE=1 SV=4 - [VIME_HUMAN]	6.51	6.87	2	2	2	2	466	53.6	5.12
P13645	Keratin, type I cytoskeletal 10 OS=Homo sapiens GN=KRT10 PE=1 SV=6 - [K1C10_HUMAN]	6.41	3.25	1	2	2	2	584	58.8	5.21
P02768	Serum albumin OS=Homo sapiens GN=ALB PE=1 SV=2 - [ALBU_HUMAN]	4.67	2.46	1	1	1	1	609	69.3	6.28
P08779	Keratin, type I cytoskeletal 16 OS=Homo sapiens GN=KRT16 PE=1 SV=4 - [K1C16_HUMAN]	3.13	4.02	1	1	1	1	473	51.2	5.05
B2RXH8	Heterogeneous nuclear ribonucleoprotein C-like 2 OS=Homo sapiens GN=HNRNPCL2 PE=1 SV=1 - [HNRC2_HUMAN]	3.04	3.41	4	1	1	1	293	32.1	5.44
Q03014	Hematopoietically-expressed homeobox protein HHEX OS=Homo sapiens GN=HHEX PE=1 SV=1 - [HHEX_HUMAN]	3.01	4.07	1	1	1	1	270	30.0	7.24

Figure 4-2 Vimentin is identified as a protein partner for PRH in CCLP Myc-PRH mass spectrometry.

(A) Silver staining of CCLP Myc-PRH subcellular nuclear and cytoplasmic (cyto) fractions immunoprecipitation using anti Myc-tag antibody. (B) Mass spectrometry corresponding to band number 2 on the silver staining gel. Data generated by Ka Ying Lee and Jayaraman.

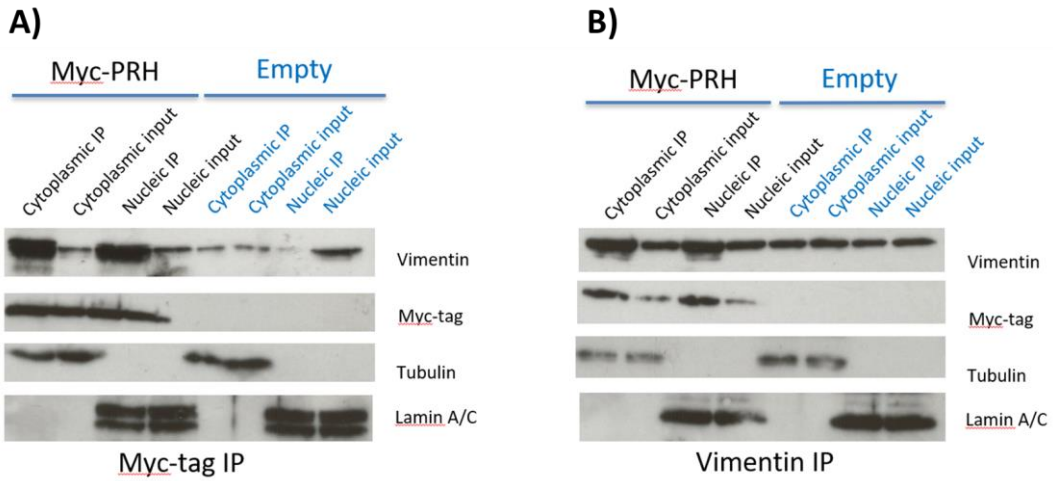


Figure 4-3 PRH interacts with Vimentin in the cytoplasmic and nuclear fractions of CCLP Myc-PRH cells.

(A) Western blotting showing an CoIP using Myc antibody in CCLP cells infected with an adenovirus expressing Myc-PRH or Ad-empty. An antibody against Vimentin was used to

probe for Vimentin. Antibodies against Tubulin and Lamin were used to confirm cytoplasmic and nuclear fractions, respectively. 20% of the input was loaded into the gel. The CoIP was done as N=2 with similar outcomes. (B) Western blotting showing an CoIP using Vimentin antibody in CCLP Myc-PRH or CCLP Ad-empty. An antibody against Myc-tag was used to probe for PRH. Antibodies against Tubulin and Lamin were used to confirm cytoplasmic and nuclear fractions, respectively. 20% of the input was loaded into the gel. The CoIP was done as N=2 with similar outcomes. Data generated by Ka Ying Lee and Jayaraman.

In this chapter, we show the stepwise approach we followed to visualise and explore the subcellular distribution of PRH and Vimentin, as well as investigate and characterise potential PRH-Vimentin adjacent localisation in CCLP1 cells. CCLP1 cells have high abundance of both PRH and Vimentin at mRNA and protein levels (215). This should make it easier to investigate and visualise PRH-Vimentin expression and distribution before examining PRH-Vimentin sub-cellular localisation upon overexpressing PRH e.g. AKN1 cells. We initially used high-resolution confocal microscopy for immunofluorescence imaging and escalated to using super-resolution PALM dSTORM microscopy.to acquire better quality images Providing visual evidence to pre-existing knowledge of PPI could provide further support for the physical interaction, and also elaborate on the molecular arrangements of the interacting proteins together and in relation to the surrounding components e.g. nuclear chromatin (413). Evidence of nuclear TF-protein localisation near an open chromatin is crucial to support hypotheses of the role of such interaction or the partner proteins in nuclear-specific biological process that are dysregulated in cancer e.g. transcription regulation, signal transduction, and DNA repair (464). In turn, this could serve future work aimed at uncovering biomarkers or therapeutic targets for cancer diagnosis and treatment.

Finally, we provide Fluorescence Correlation Spectroscopy (FCS) data on the dynamics of GFP and GFP-PRH in AKN1 cells.

4.2 Results

4.2.1 PRH is expressed in the nuclei and cytoplasm of CCLP1 cells

To pave the way for PRH-Vimentin spatial localisation studies in CCA cells, we wanted to first characterise the expression of PRH protein in these cells. We used immunofluorescence staining to label PRH in CCLP1 cells following the protocol detailed in the method chapter. We captured n=40 Z-stacks images of our stained CCLP1 cells as well as imaged cells that we did not incubate with PRH antibodies for background subtraction purpose. We analysed the images using FIJI (7).

For analysis purposes, whenever 2D images were needed, we converted the stacks into 2D images using the tool Z-projection (SUM) in Fiji to sum up the expression from all the stacks into a single flat image (7). To exclude false positive expression, we first processed all the images to remove the background. We used the average intensity of at least five negative control images where no primary antibodies were added. This background subtraction created processed images ready for any desired analysis. To measure nuclear and cytoplasmic PRH intensities, we segmented the PRH channel to define ROI using the tool threshold (7). First, we used the filters Gaussian Blur or Mean to smoothen the edges which makes it easier to threshold. Then, we used the threshold (Otsu mainly) to highlight the areas of analysis. We followed segmentation with “make binary” tool, binary filters, and “create mask” tool to create a black ROI over a white background for ROI selection. Whenever we had cellular overlaps, we used the tool watershed to separate

the clumping. We applied the process to nuclear DAPI staining and used the nuclei as masks to identify nuclear PRH in PRH channels. We also made masks for the entire cell minus the nuclear region for cytoplasmic PRH evaluation. We relied on the PRH channel to draw a tracing line around our cells for intensity measurement and statistical analysis purposes. Applying these steps should exclude the background areas that would be sources of noise signals. We measured mean grey intensity for nuclear and cytoplasmic PRH separately. To produce representative images, we adjusted the brightness and contrast of colours, after the background subtraction step, equally across the images. We generated a composite image of both DAPI (blue) and PRH (red). We used the plugin (Figure Combine pt2) created by Dr Kenton Arkil to make a compiled image (451). Finally, we added scale bars and exported the images for presentation.

Figure 4-4 shows a representative image of CCLP1 cells stained for DAPI and PRH illustrating the extent of PRH distributions in the cells where DAPI was presented in (blue) and PRH in (red). On visual inspection of the images, PRH staining was distributed in both nuclear and cytoplasmic regions of CCLP1 cells, but more strongly in the former. PRH appeared as a mix of foci and speckles of various sizes in both compartments, but more prominent in the nuclear regions (referenced by DAPI)

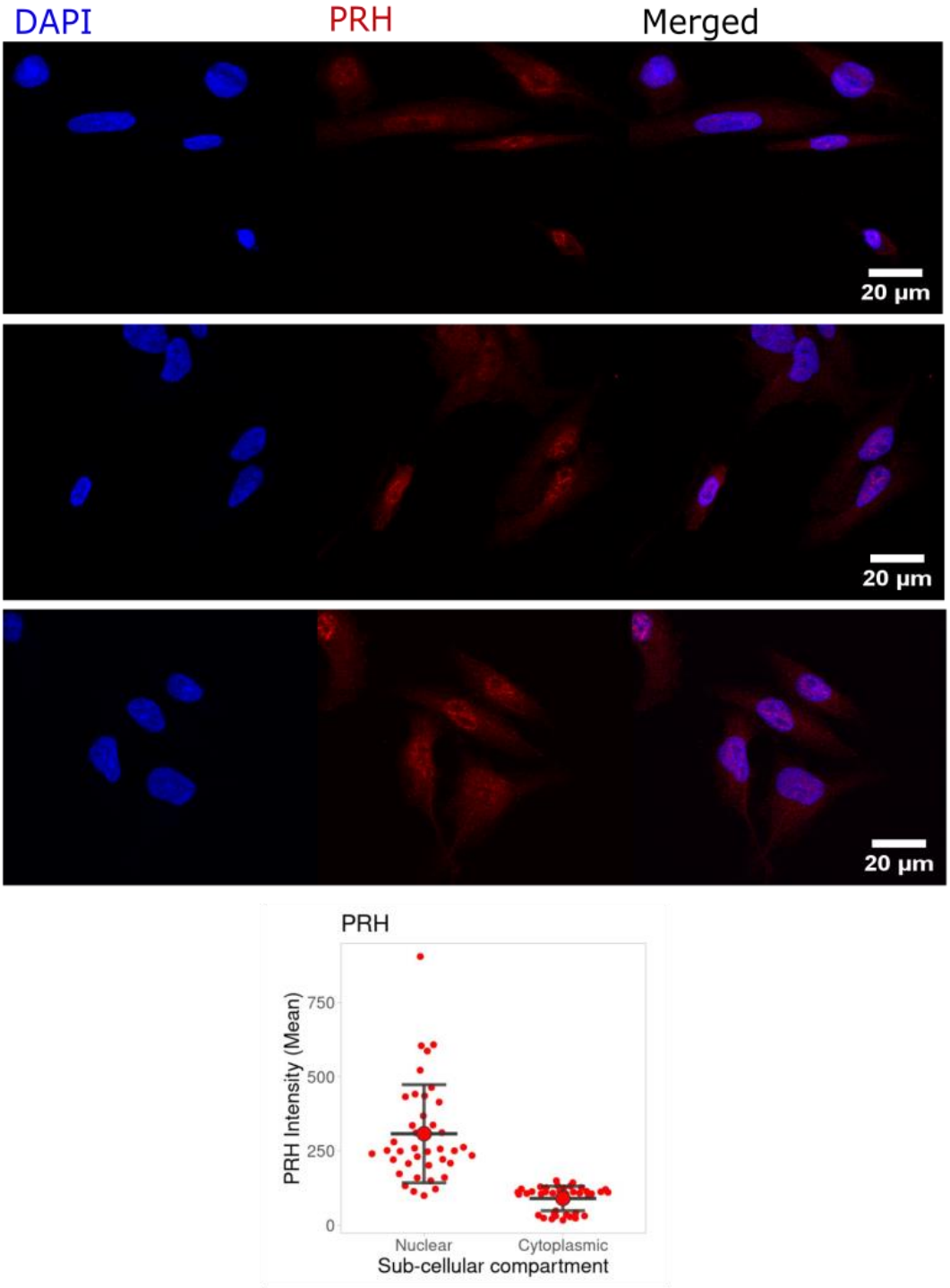


Figure 4-4 PRH expression in CCLP1 cells.

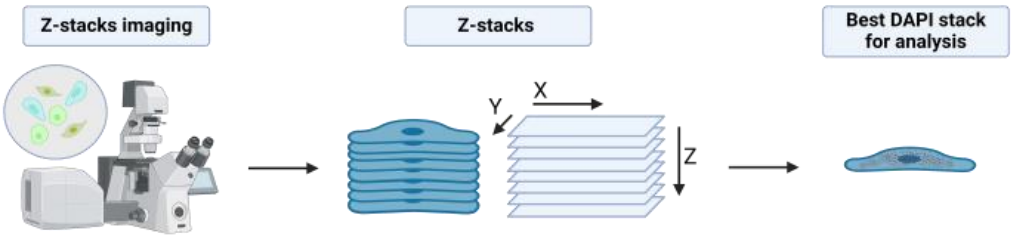
Images of CCLP1 cells stained for PRH and DAPI. Cells were fixed with 4% FA and labelled with PRH antibody along with nuclear DAPI staining. Images were acquired sequentially for n= 40 cells using a confocal Leica DMI4000B microscope at 63X magnification. Scale bars 20

µm. Laser power, gain, exposure, and pinhole were the same throughout the imaging. Expression intensities represented in violin plots show significantly higher in expression of nuclear PRH (p<1e-10) compared to cytoplasmic PRH. Ordinary t-test analysis was performed, and error bars represent SD.

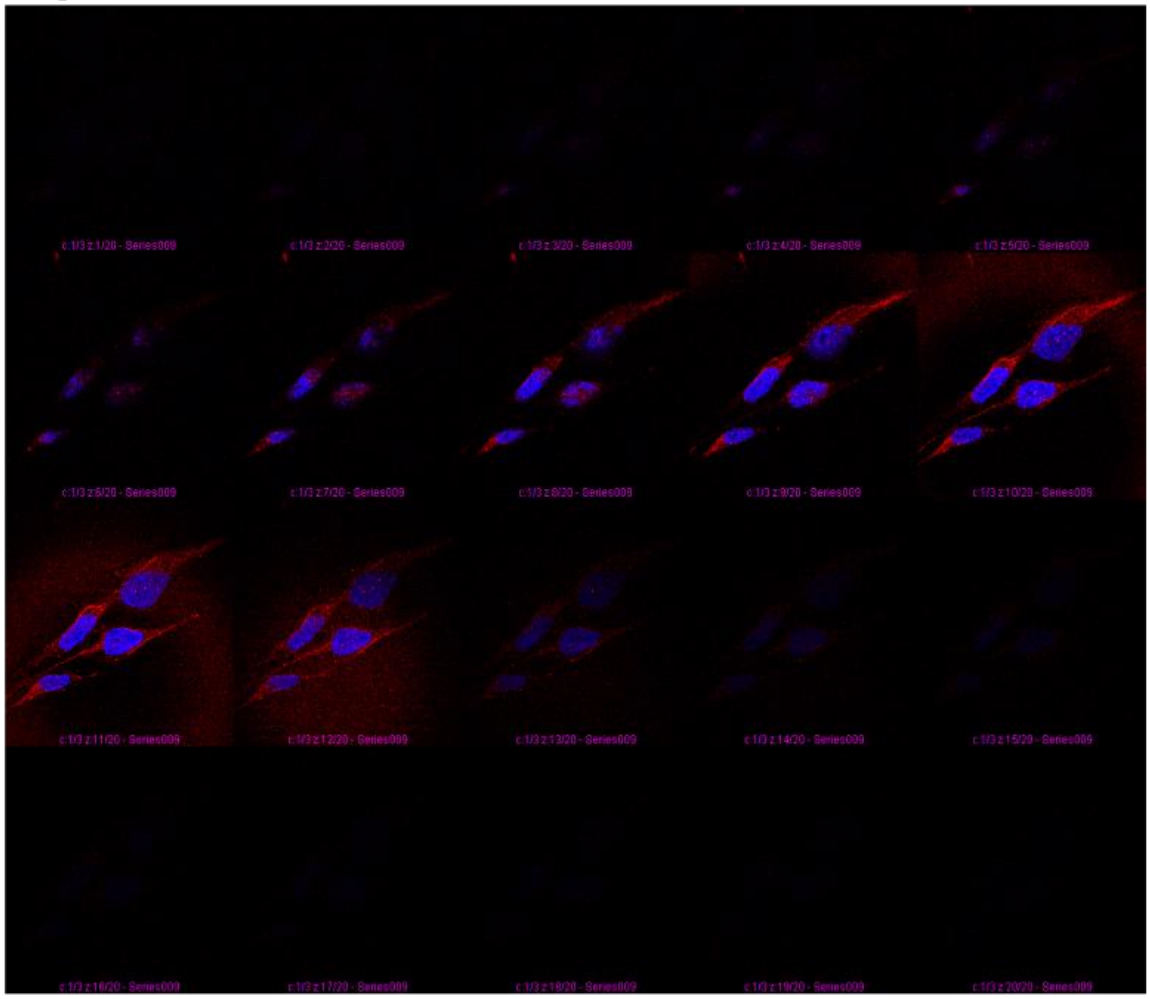
To statistically illustrate the difference of PRH expression between the nuclear and cytoplasmic regions in the cells, we used the nuclear and cytoplasmic masks we created to choose the areas where we measure the signal intensities in the desired ROI. The violin plot in (figure 4-4) shows a significant increase of nuclear PRH intensity compared to cytoplasmic PRH (p <1e-10). This confirms that PRH is a predominantly nuclear protein in these cells.

To better visualise the spatial distribution of nuclear PRH in relation to DNA (chromatin), we examined magnified images of the nuclear region. Figure 4-5-A shows an illustration of XYZ image acquisition of cells using a confocal Leica DMI4000B microscope at 63X magnification. This allowed us to examine the sub-planes of the imaged cells with the best DAPI staining to exclude noise from PRH expression localised outside the nuclear region. In (figure 4-5-B), we created a montage of the twenty Z-stacks using the tool (Make Montage (7)) where sections 9 to 11 represent the best nuclear planes to investigate nuclear PRH. Finally, in (figure 4-5-C-E) we provided representative images of DAPI (blue) and PRH (red) where PRH seemed to form spots of various sizes throughout the nuclear region away from condensed blue staining. This indicates that PRH aggregates away from heterochromatin in the nucleus.

A)



B)



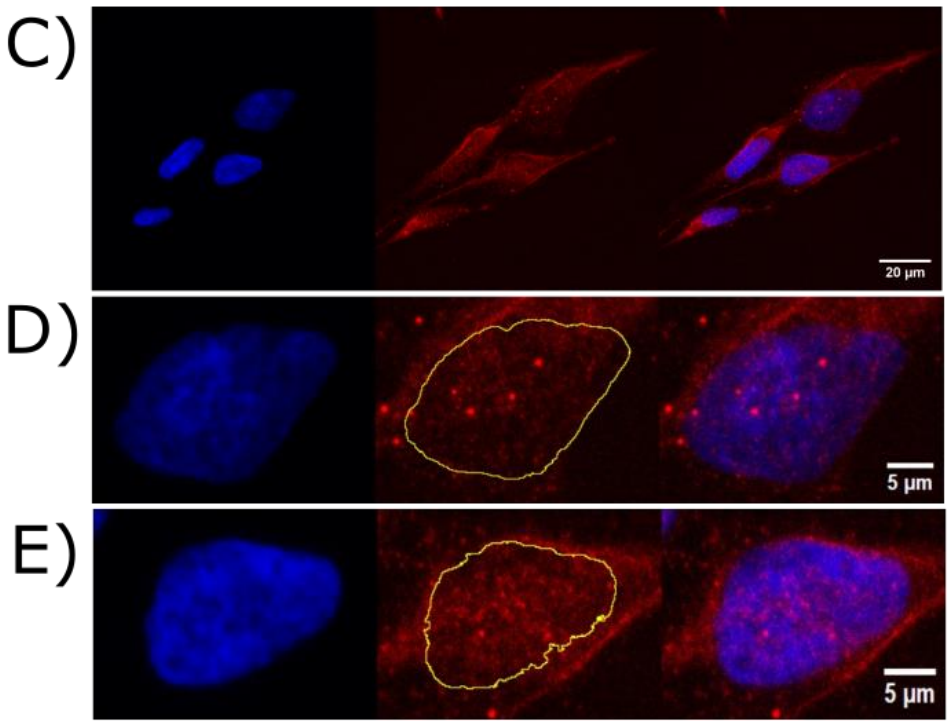


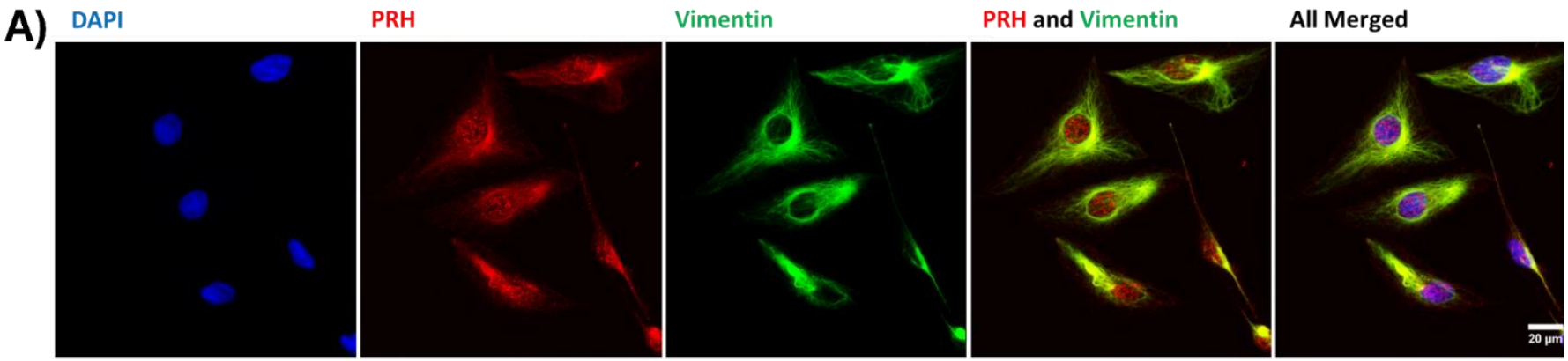
Figure 4-5 Nuclear PRH aggregates away from DAPI dark-stained regions (heterochromatin).

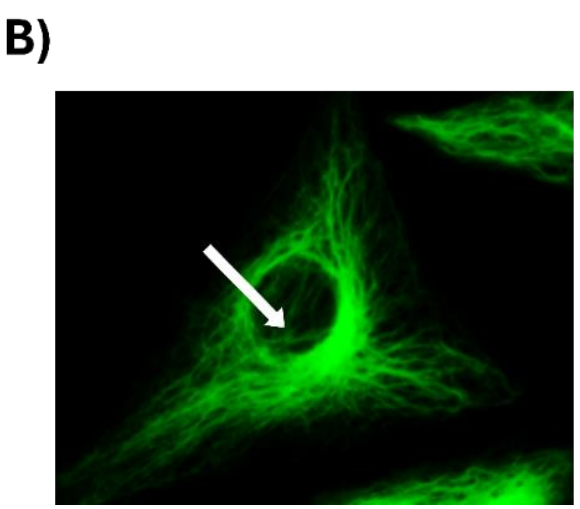
(A) An illustration of Z-stack images acquisition. (B) All the stacks of an example cell presented as a Montage. (C) Immunofluorescence images of CCLP1 cell stained for PRH and DAPI showing that PRH appears as speckles distributed throughout the cells. Scale bar 20 µm. (D) and (E) are enlarged views for two nuclei from (C) showing the strong expression of PRH within DAPI staining regions. PRH aggregates away from dark-stained DAPI (heterochromatin). Scale bar 5 µm. Cells were fixed with 4% FA and labelled with PRH antibody along with nuclear DAPI staining. Images were acquired sequentially for n= 40 cells using a confocal Leica DMI4000B microscope at 63X magnification. Laser power, gain, exposure, and pinhole were the same throughout the imaging.

4.2.2 PRH and Vimentin spatial localisation in the nuclear and cytoplasmic compartments of CCLP1 cells

To investigate the potential for PRH-Vimentin to spatially exist in close distance (overlap), we needed first to study the extent of Vimentin distribution in CCLP1 cells. We examined Z-stack images representing CCLP1 cells co-probed with PRH and Vimentin antibodies. We followed the same steps from the previous section to prepare the images for analysis; qualitative and

quantitative. Figure 4-6-A shows the sequential acquisition of DAPI (blue), PRH (red), Vimentin (green), merged PRH and Vimentin, and all three channels overlapped together (merged).


A)
DAPI
PRH
Vimentin
PRH and Vimentin
All Merged
20 µm


B)

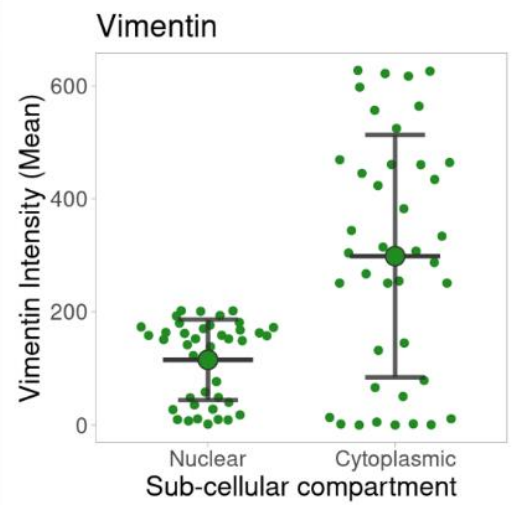

C)
Vimentin
Vimentin Intensity (Mean)
600
400
200
0
Nuclear
Cytoplasmic
Sub-cellular compartment

Figure 4-6 PRH and Vimentin show partial co-localisation in CCLP1 cells.

(A) Images of CCLP1 cell co-stained for PRH and Vimentin showing both PRH and Vimentin are distributed throughout the cells. Cells were fixed with 4% FA and labelled with PRH antibody along with nuclear DAPI staining. Images were acquired sequentially for n= 40 cells using a confocal Leica DMI4000B microscope at 63X magnification. Scale bars 20 µm. Laser power, gain, exposure, and pinhole were the same throughout the imaging. (B) Enlarged view of Vimentin green channel showing some filaments extending into the nuclear region (white arrow). C) Expression intensities represented in violin plots showing significantly increased expressions of Vimentin in the cytoplasm (p= 5.23e-06) compared to the nuclear Vimentin. Ordinary t-test analysis was performed, and error bars represent SD.

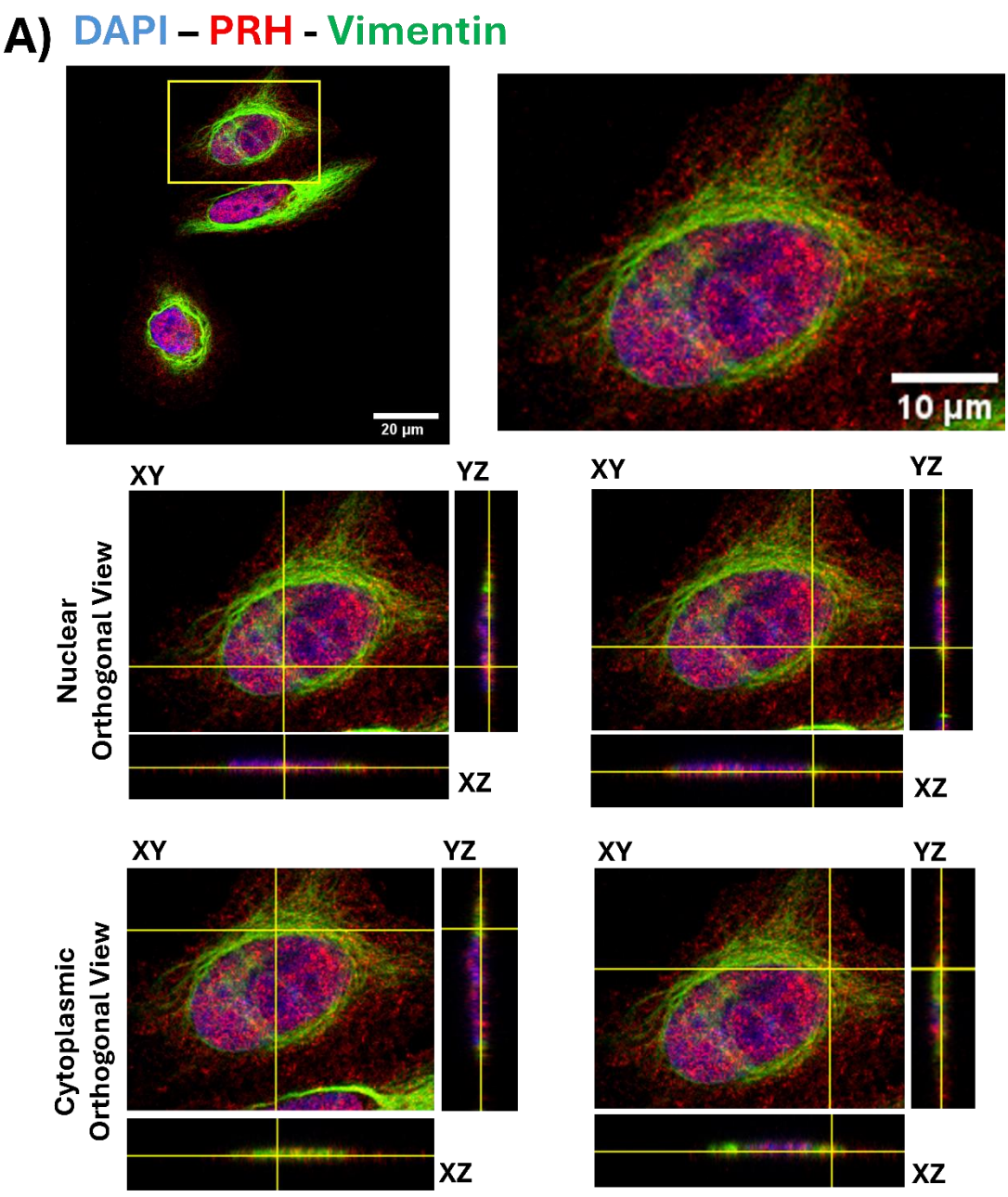
PRH (red) again appeared in both nuclear and cytoplasmic regions of the cells as punctuated foci and speckles of various sizes. Vimentin (green) assumed the classical filamentous form that extensively branched all over the cytoplasmic region representing the cytoskeleton and clearly demarcating the mesenchymal morphology of the cells. In addition, on a careful examination of the green channel Vimentin seemed to also extend into the nuclear region slightly (figure 4-6 – B white arrow).

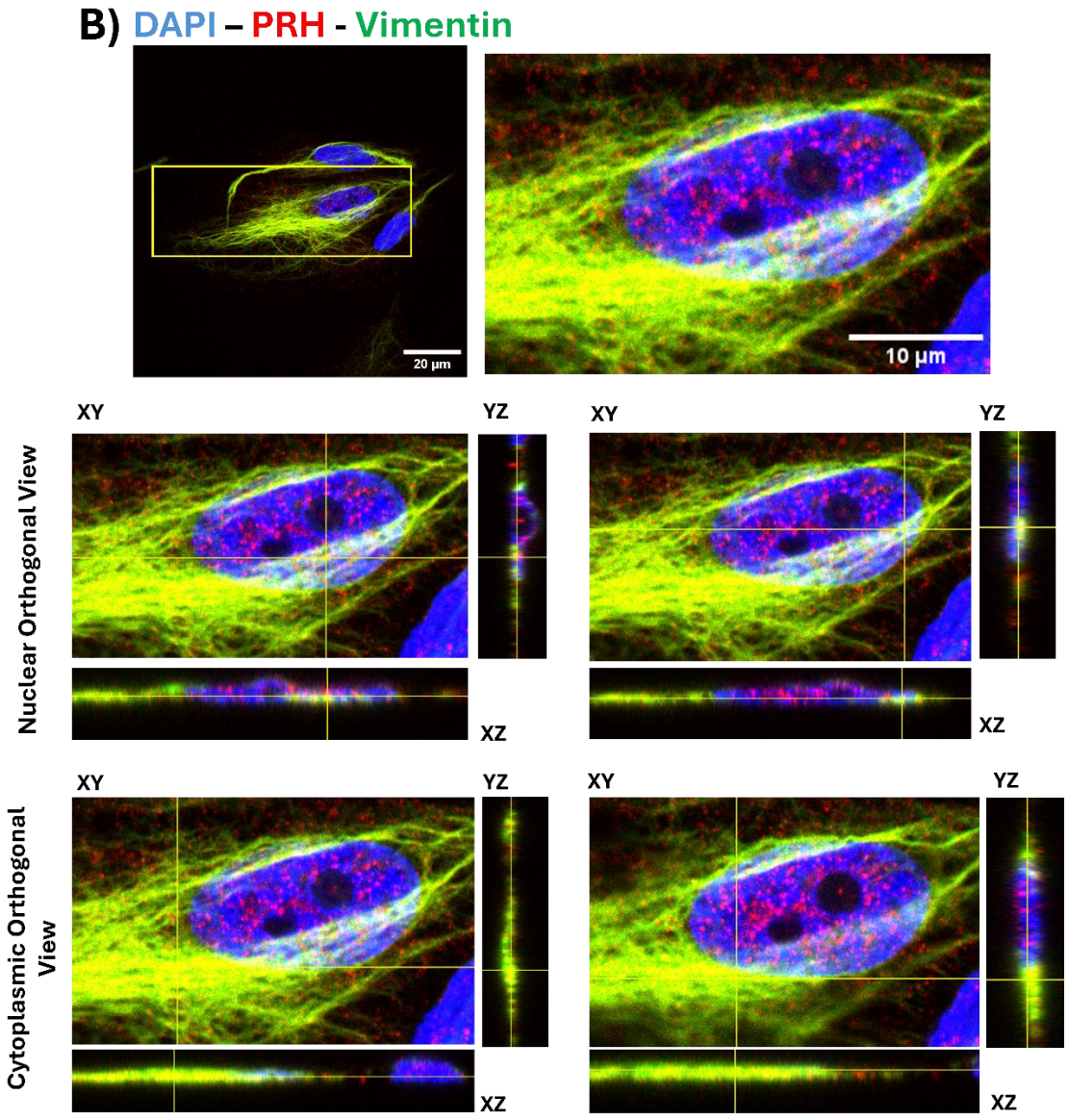
We previously reported in figure (4-4) that PRH is significantly more nuclear in CCLP1 cells (*p* <1e-10). To investigate the status of nuclear and cytoplasmic Vimentin, we used the nuclear and cytoplasmic masks we created to choose the areas where we measure the signal intensities in the desired ROI. The violin plot in (figure 4-6-C) shows significantly higher cytoplasmic Vimentin expression (p=5.23e-06). This confirms that Vimentin is a predominantly cytoplasmic protein in CCLP1.

To investigate whether PRH-Vimentin co-exist spatially together, we used FIJI XYZ-orthogonal view and RGB profiler analysis on FIJI (7). XYZ-orthogonal view tool allows us to examine Z-stacks in three different planes simultaneously: XY, XZ, and YZ. To apply this on our images, we utilised the pre-processed images saved as Z-stacks forms. Then, using the XYZ-orthogonal view tool in FIJI we traced PRH (red) and Vimentin (green) by locating yellow spots generated from the merge of red and green, hence PRH and Vimentin. This was done by controlling the sliding coordinate lines over the nuclear and cytoplasmic compartments in XY plane. If a yellow spot was

detected in the same location in all three planes: XY, YZ and XZ, it meant that PRH and Vimentin were located very close to each other (overlapped to the naked eye).

We performed XYZ-orthogonal view on n=40 cells. Figures 4-7-A to C show example traced yellow spots in all three spatial planes within the cytoplasmic and nuclear regions in three different cells. In each group, we show a widefield image of the cells under investigation accompanied by an enlarged square inset to focus on a certain cell. We, then, show all the three planes for two yellow spots in both cytoplasmic and nuclear regions. Cytoplasmic orthogonal tracing of potential adjacent PRH-Vimentin in all the analysed images was more prominent and easily achieved than nuclear tracing since both PRH and Vimentin are abundant here. However, XYZ orthogonal tracing of potential PRH-Vimentin closely positioned in the nucleus was challenging due to very low Vimentin nuclear expression in comparison to PRH. We were able to align PRH and Vimentin in all three planes (XY, YX, and YZ) in the analysed images by detecting yellow spots from PRH and Viemntin existing in proximity of each other. These yellow overlaps were observed in separate locations within DAPI region mostly existing towards the nuclear periphery close to the nuclear membrane. Unlike cytoplasmic yellow spots tracing, some of the nuclear tracing showed minimal (XY, YX, and YZ) overlaps. Overall, the data indicate that PRH and Vimentin are spatially more adjacent in the cytoplasmic subcellular compartments than within the nuclei of CCLP1 cells1


A) DAPI – PRH - Vimentin
20 µm
10 µm
Nuclear Orthogonal View
Cytoplasmic Orthogonal View
XY
YZ
XZ


B) DAPI – PRH - Vimentin
20 µm
10 µm
Nuclear Orthogonal View
Cytoplasmic Orthogonal View
XY
YZ
XZ

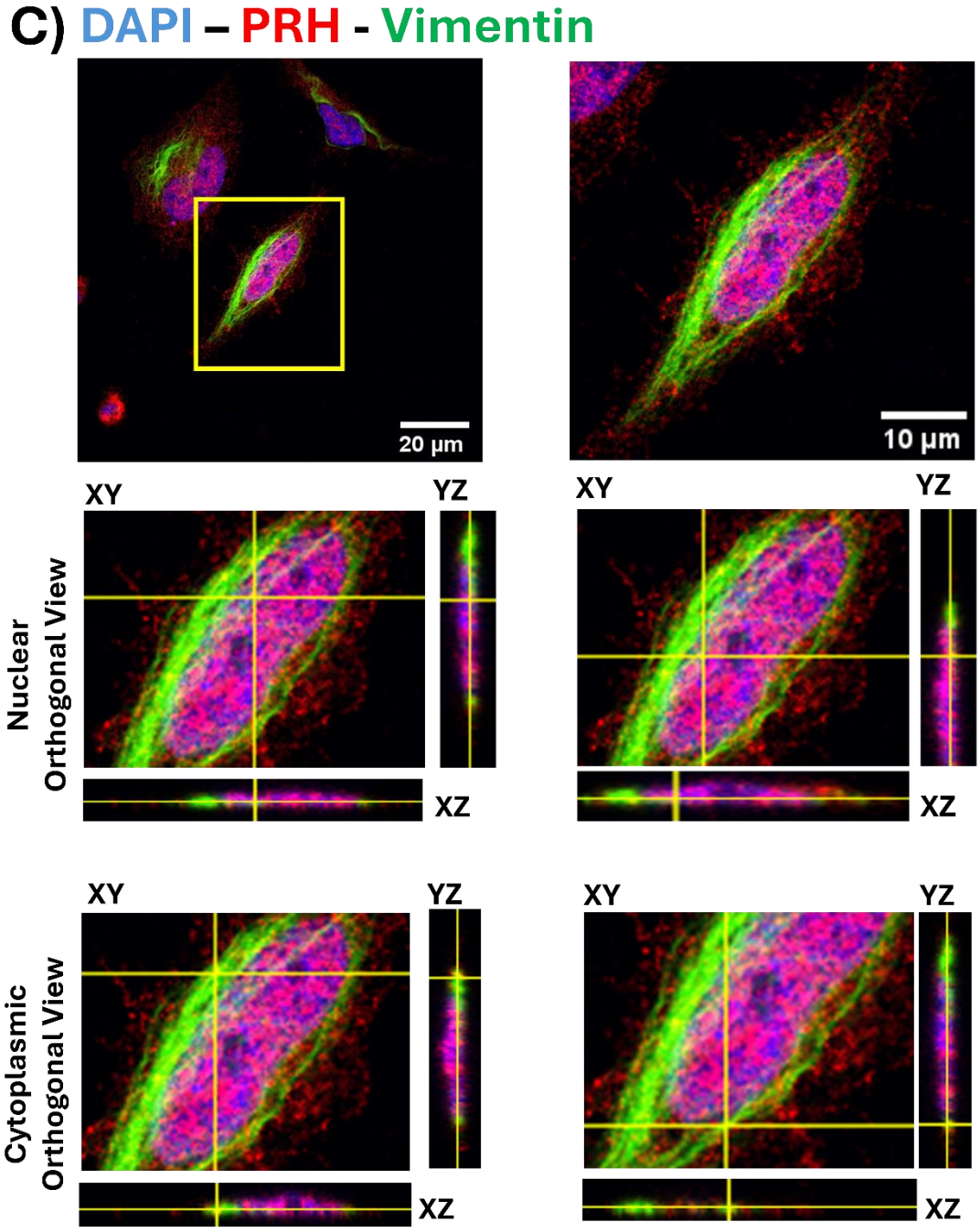


Figure 4-7 PRH and Vimentin overlap in the cytoplasmic and to a lesser extent in the nuclear subcellular compartments of CCLP1 cells.

Three examples of orthogonal view tracing lines in different planes (xy, xz and yz) of confocal microscope images used to analyse PRH-Vimentin overlapping. A ROI selection of a CCLP1 cell is added as an enlarged view beside each full image. Each cell has two examples of nuclear and cytoplasmic PRH-Vimentin overlapping (yellow) traced in all three planes (xy, xz and yz). Cells were fixed with 4% FA and labelled with PRH and Vimentin antibodies along with nuclear DAPI staining. Images were acquired sequentially for n= 40 cells using a confocal Leica DMI4000B microscope at 63X magnification. Scale bars 20 µm and 10 µm. Laser power, gain, exposure, and pinhole were the same throughout the imaging.

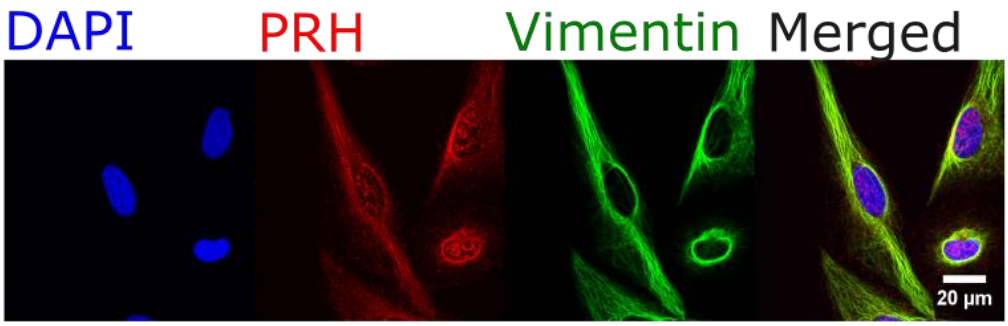
Because it was a challenge to confirm the nuclear overlapping from XYZ-orthogonal view examination, we used RGB profiler on FIJI as an additional method. This tool allows us to draw a line and generate a histogram profile across the cell where we visually see or suspect that there is PRH and Vimentin overlap (7). To apply the tool on our images, we again used pre-processed images saved as Z-stack. We then chose the layers that had the best DAPI intensity which represented the nuclear compartment of the cells. We drew a line across the area where we see or suspect yellow spots. Then, we examined the histogram for overlapped peaks. The peak height is correlated to the intensity of fluorophore signal. Overlapped peaks give insights on where the overlapping is taking place under the traced section of the cell.

Figures 4-8-A to D demonstrate the stepwise approach for RGB profiler analysis where (A) shows DAPI (blue), PRH (red), Vimentin (green), and all three channels overlayed together. Part (B) is a montage of the Z-stacks acquired showing planes 11 to 14 with best DAPI staining. We drew in section (C) of the figure the lines across three different nuclear regions for RGB profiling. Finally, part (D) shows each line accompanied by its intensity histograms. These histogram plots are coloured similar to the channels they represented; DAPI (blue), PRH (red), and Vimentin (green).

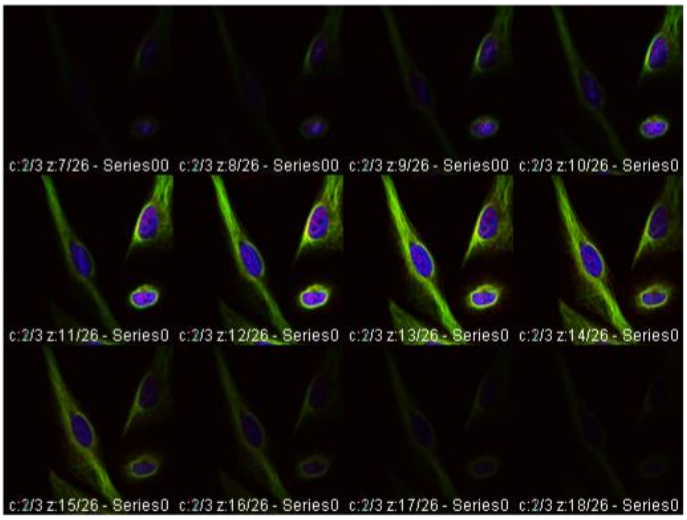
Although RGB profiler confirmed the level of nuclear PRH and Vimentin expression where the histograms portrayed more nuclear PRH intensities than vimentin, this analysis was not conclusive regarding the nuclear PRH-Vimentin overlapping. In each histogram, DAPI (blue) exhibited sharp peaks defining

the nuclear region of the plot, whereas the negligible flat DAPI signals on either side indicated the cytoplasmic areas of the plots. Red PRH lines in all the plots showed varied sharp deflections within DAPI region as well as beyond DAPI region. This indicated that PRH can present in both nuclear and cytoplasmic subcellular compartments. Whereas green Vimentin lines showed tall peaks outside DAPI areas and very small ones within the nuclear part of the plots which reflected more cytoplasmic to nuclear Vimentin. Visual inspection of PRH and Vimentin peaks overlap was prominent in the cytoplasmic portions of each plot. However, nuclear peaks overlaps were difficult to assess due to the nuclear Vimentin expression level. The middle and bottom parts of (figure 4-8-D) show cytoplasmic PRH-Vimentin overlaps extending slightly underneath DAPI lines at the periphery of the nuclear region that require more sensitive and higher resolution imaging techniques to confirm.

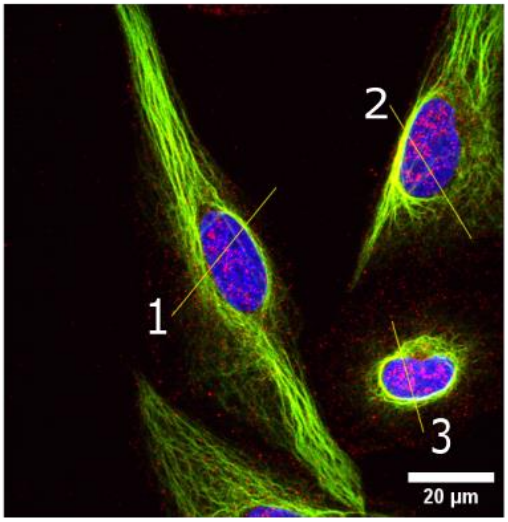
A)


DAPI
PRH
Vimentin
Merged
20 µm

B)


c:2/3 z:7/26 - Series00
c:2/3 z:8/26 - Series00
c:2/3 z:9/26 - Series00
c:2/3 z:10/26 - Series0
c:2/3 z:11/26 - Series0
c:2/3 z:12/26 - Series0
c:2/3 z:13/26 - Series0
c:2/3 z:14/26 - Series0
c:2/3 z:15/26 - Series0
c:2/3 z:16/26 - Series0
c:2/3 z:17/26 - Series0
c:2/3 z:18/26 - Series0

C)


1
2
3
20 µm

D)

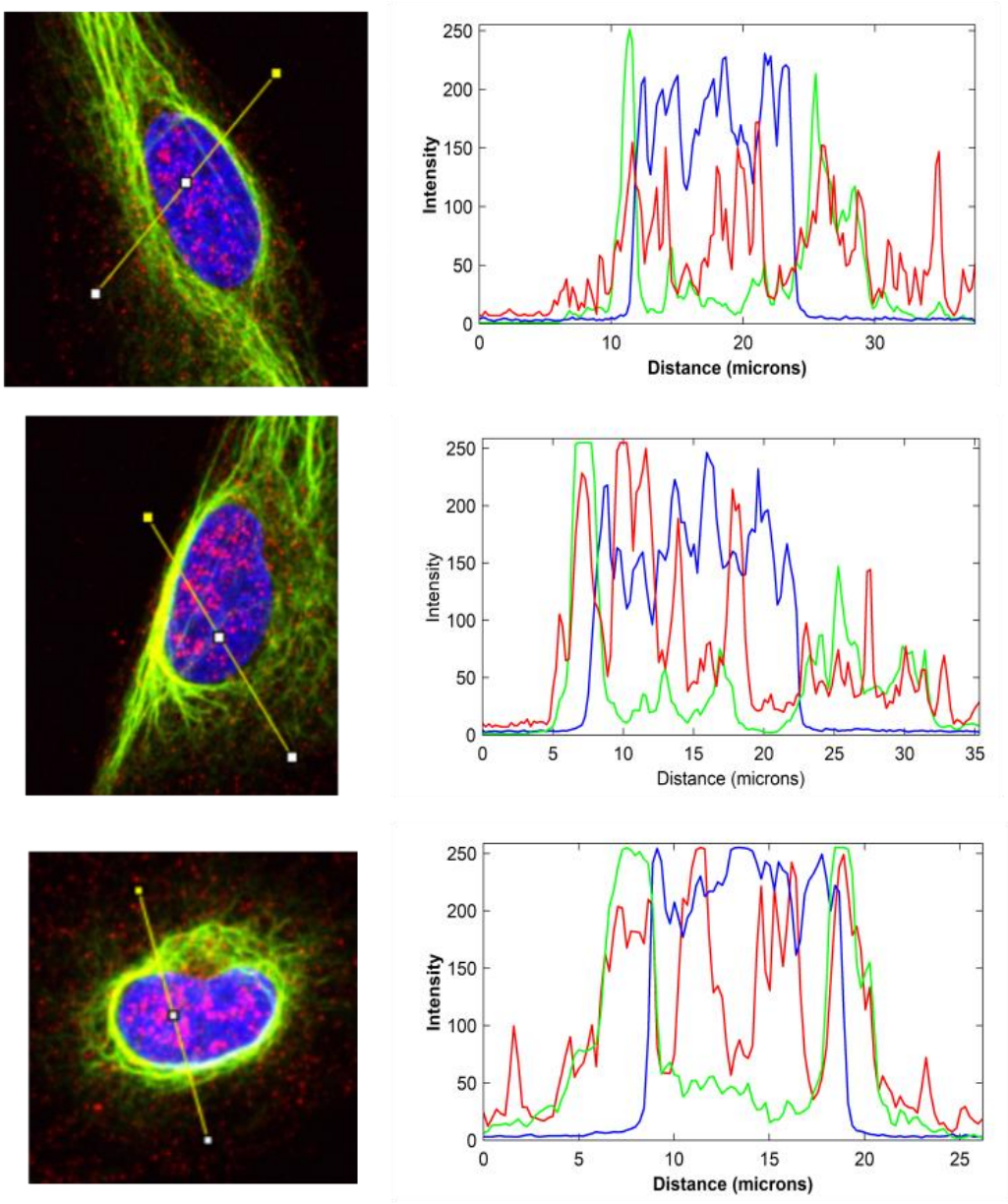


Figure 4-8 RGB profile analysis of PRH and Vimentin in the subcellular compartment of CCLP1 cells.

(A) CCLP1 cells labelled for DAPI (blue), PRH (red), and Vimentin (green) showing both PRH and Vimentin are distributed throughout the cells. Cells were fixed with 4% FA and labelled with PRH antibody along with nuclear DAPI staining. Images were acquired sequentially for n= 40 cells using a confocal Leica DMI4000B microscope at 63X magnification. Scale bars 20 µm. Laser power, gain, exposure, and pinhole were the same throughout the imaging. (B) Z-stack montage showing the different layers of Z plane. (C) Three examples of tracing lines for PRH-Vimentin co-localisation using RGB profiler plot in FIJI. (D) Histogram plots of RGB profiling lines showing DAPI (blue) defining the nuclear region of the plot, PRH (red), and Vimentin (green). There is a strong expression intensity of PRH (red) and a weak intensity for

Vimentin (green). The peaks overlap clearly in the cytoplasmic region, and some extends under DAPI region at the periphery of the nuclear portion of the plot.

Taking all the findings together, the data demonstrated that both PRH and Vimentin are distributed in CCLP1 nuclear and cytoplasmic compartments. PRH is predominantly nuclear, while Vimentin shows a predominantly cytoplasmic distribution with some nuclear protein. The data also revealed the strong chance of cytoplasmic PRH-Vimentin overlapping as well as suggested weaker overlapping in the periphery of the nuclear region. The latter, however, requires further investigation with better resolution imaging techniques.

4.2.3 Investigating PRH-Vimentin nuclear localisation in CCLP1 cells using photoactivated localisation microscopy – direct stochastic optical reconstruction microscopy (PALM dSTORM)

Previous findings from confocal imaging provided evidence regarding PRH and Vimentin expression, and their potential spatial adjacency. However, the theoretical diffraction limit in confocal microscopy is restricted around 250 µm which affects the resolutions of the acquired images. This might negatively influence image analysis and interpretation. To overcome these issues, we characterised CCLP1 cells for nuclear PRH and Vimentin expression using a super-resolution microscopy technique to image at single molecule level. Specifically, we opted for photoactivated localised microscopy, direct stochastic reconstruction microscopy (PALM dSTORM). This method relies on using a secondary antibody with a fluorophore capable of blinking, which means going through a long cycle of both emission and dark states. The more blinking and detection in multiple frames, the better the localisation (465). As

a result, we get to image individual molecules of PRH and Vimentin within the nucleus and assess their expression and behaviour.

4.2.3.1 Nuclear PRH characterisation in relation to chromatin in CCLP1 cells

We started with exploring the molecular behaviour of nuclear PRH in CCLP1 cells. We prepared and stained our cells as described in the method chapter. We stained the nucleus using SiR-DNA (Spirochrome SC501) and labelled the cells with PRH antibody. On the day of imaging, we incubated the cells in a solution of fucidal fluorescent beads (TetraSpeck™ Microspheres, 0.1 µm, fluorescent beads) in PBS prior to imaging. These beads are microspheres stained with fluorescent dyes and used here as a calibration reference for accurate localisation. To ensure our fluorophores maintain blinking for a long cycle and capture the molecules in as high number of frames as possible for precise localisation, we used a special imaging buffer for the coverslips as detailed in section (2.3.3.4 Imaging). We acquired the images with help from Dr Chris Toseland using Zeiss Elyra PS.1 system. We acquired 20,000 – 30,000 frames of 488 and 576 fluorophores using HiLo laser in continuous mode. We adjusted laser power and gain to 90 and 288, respectively, while observing the acquisition histogram. This ensured obtaining the images without damaging the camera.

We processed the images using ZENBLack and MATLAB for qualitative and quantitative analysis, respectively. In general, we used ZENBlack software to produce rendered images for display and generate data for quantitative analysis in MATLAB. First, we adjusted molecules settings to account for

molecules overlapping. In addition, we chose signal to noise ratio degree at 6 to accommodate for background noise. Also, under localisation settings, we selected both localisation algorithms and Gaussian. We scanned the frames rapidly under the previous settings before we apply the changes and produce a super-resolution image. We further rendered the image for best visual presentation. ZENBlack offers several filers that could be applied for rendering purposes. First, we corrected the drift according to the fluorescent fucidal beads. To remove poorly localised molecules and ensure high localisation accuracy, we adjusted the range of inclusion to 5 -75 nm precision. We changed PAL render settings whenever required to define the final image appearance. We made sure to observe the precision histogram as we modified the statistical settings. ZENBlack also generates a table of localisation that includes every recorded single molecule along with its spatial coordinates. We exported both the localisation table and histogram for quantitative analysis. We also snapped images at different magnifications to show molecules clustering whenever needed.

For quantitative analysis, we imported the precision tables into MATLAB. The package for co-clustering analysis for single molecule localisation microscopy (SMLM) data (ClusDoC) was downloaded from (https://github.com/PRNicovich/ClusDoC) (412).

To start quantitative analysis, we added ClusDoC package path and opened it in MATLAB. We initiated the tool by inserting the script ClusDoC in the script window. Then, we uploaded the localisation table and chose an output folder for the results. Using the tool selection, we traced nuclear ROI for analysis.

We used the option Ripley's *K* to obtain the linearised *Ripley's K* function for nuclear PRH and calculated *L(r)-r* (where *r* is radius) (figure 4-9). The quantitative analysis of PRH-SirDNA demonstrates that nuclear PRH in CCLP1 cells is organised into clusters and not random. A value of zero in *Ripley's* plot indicates random distribution, whilst positive values indicate clustering.

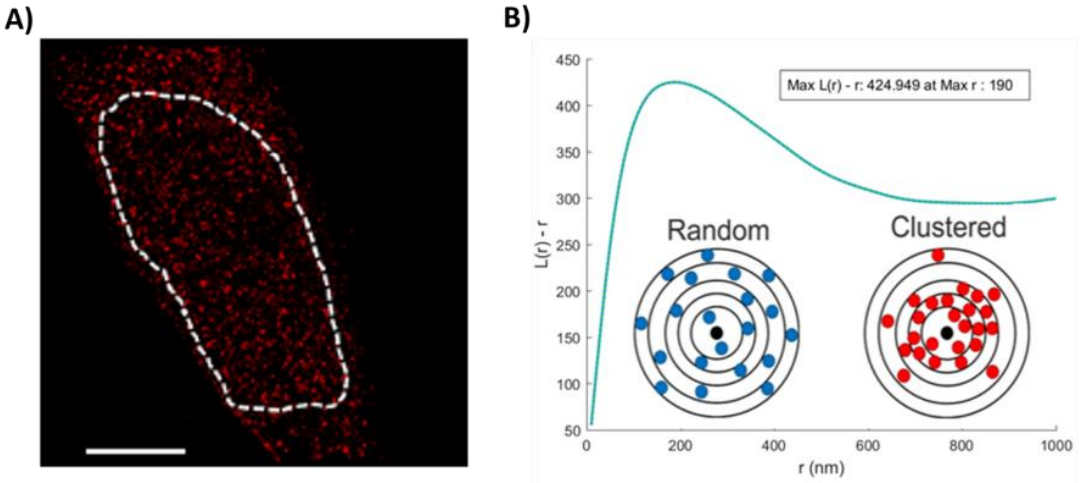


Figure 4-9 PRH is clustered in the nucleus.

(A) STORM image of endogenous PRH in CCLP1 cells. Dotted lines represent ROI used for cluster analysis. Scale bar 3 µm. (B) Linearised *Ripley's K* function for nuclear PRH. *L(r)-r* (where *r* is radius), calculated for selected ROI of STORM images. A value of zero in this plot indicates random distribution, whilst positive values indicate clustering.

Figure (4-10-A) shows the rendered STORM image for PRH (green) and SiR DNA (red) staining in CCLP1 cells. The close-up snapshot of an inset from (A) shows the specific arrangement of PRH around SiR DNA.

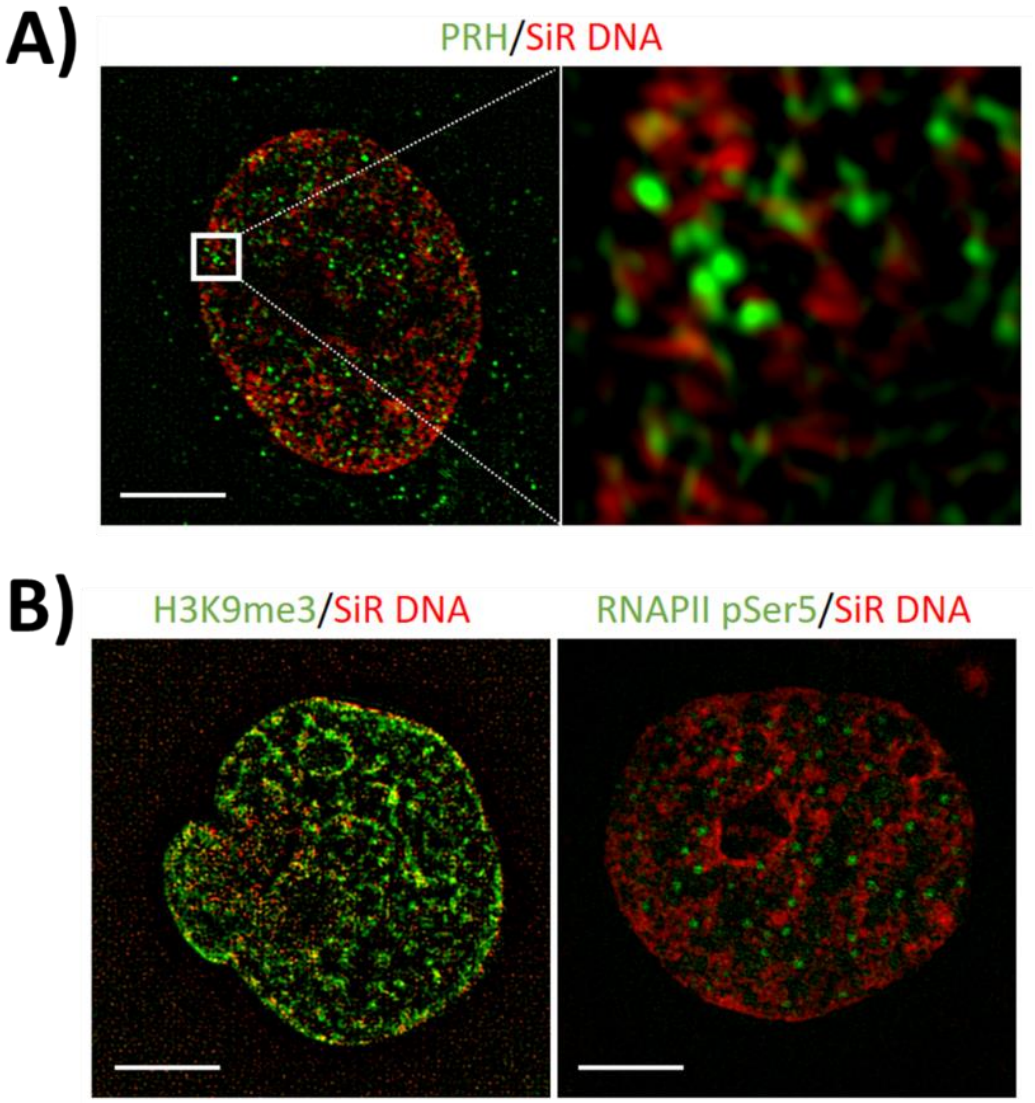


Figure 4-10 PRH is segregated from condensed heterochromatin and clustered in the nucleus.

(A) SIM image of endogenous PRH in CCLP1 cells. On the side: an enlarged ROI for PRH and SiR DNA staining from the highlighted regions in A. (B) This part was performed by Dr Chris Toseland. H3K9me3 and RNA Polymerase II Ser5 in HeLa cells, with DNA stained with SiR-DNA dye. N=5 images. Scale bar 3 µm. Heterochromatin mark H3K9me3 co-localised with dense DNA staining whilst RNAPII pSer5 forms large clusters away from dense chromatin regions. PRH is segregated from dense DNA staining regions which correlate with heterochromatin.

We compared the visual expression of PRH-SiR DNA to H3K9me3 and RNAPII pSer5 clustering in relation to heterochromatin as seen in (figure 4-10-B) (413). We concluded that PRH clusters are segregated from condensed

chromatin resembling the spatial organisation of RNAPII pSer5 away from dense heterochromatin.

Overall, the results confirmed the confocal imaging findings regarding the nuclear expression of PRH in CCLP1 nuclei. In addition, PALM dSTORM imaging of PRH and SiR DNA revealed more information about the spatial organisation of nuclear PRH. Quantitative image analysis revealed the ability of PRH single molecules to group into clusters of various sizes. Moreover, these clusters are arranged away from dense heterochromatin which resembles RNAPII pSer5 organisation in mammalian cells. This could imply the preference of PRH organisation around euchromatin to take part in DNA binding and regulation of target gene expression.

4.2.3.2 The spatial organisation of nuclear PRH and Vimentin in CCLP1 cells

To prepare the ground for PRH-Vimentin spatial localisation on a super-resolution scale, we repeated the STORM immunofluorescence assay replacing DNA staining with Vimentin labelling. We followed similar settings for images acquisition and ZENBlack processing to produce images for presentation.

The top part of (figure 4-11) shows STORM-rendered widefield images of a CCLP1 cell in Gaussian filter for presentation. In the displayed cell, we presented PRH in red, Vimentin in green, fucidal beads in white, merged PRH-Vimentin, and merged beads with PRH-Vimentin (scale bar 5µm). The bottom part of (figure 4-11) represents an enlarged view of the nuclear region since

we wanted to focus on nuclear PRH and Vimentin localisation. In general, the cell under investigation was mesenchymal. PRH was seen forming red foci of various sizes all over the cell, but more prominently within the nuclear region. Vimentin was arranged in long and extensive cytoplasmic filaments surrounding the nuclear region forming the typical nuclear cage. Some Vimentin filaments were also seen penetrating inside the nuclear region and appearing as sporadic foci. The beads we used to correct for drifting are visible in white. These findings are consistent with the morphology and expression profile we portrayed for CCLP1 cells using confocal microscopy.

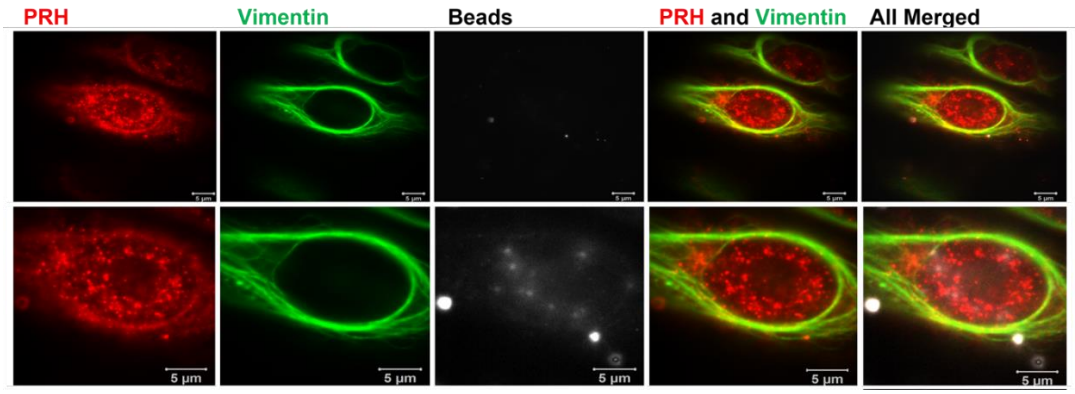


Figure 4-11 Spatial organisation of PRH and Vimentin in CCLP1 cells in PALM dSTORM images.

(Top) Widefield views of PRH and Vimentin expression. (Bottom) Nuclear views of a CCLP1 cell showing the expression of PRH (red), Vimentin (green) and drift-correcting fluorescent fucidal beads (white), merged PRH and Vimentin channels, and all three merged channels. Scale bar 5µm.

To show details of PRH-Vimentin molecular arrangements in CCLP1 cells, we used ZENBlack filters to view PRH-Vimentin as centroid. Figure 4-12 shows the nuclear snapshot of PRH (red) and Vimentin (green) as dots in various sizes (scale bar 5 µm). The histogram next to each image shows the peak at

15 nm for PRH and 25-30 nm for Vimentin indicating the precision of localisation for red and green fluorophores.

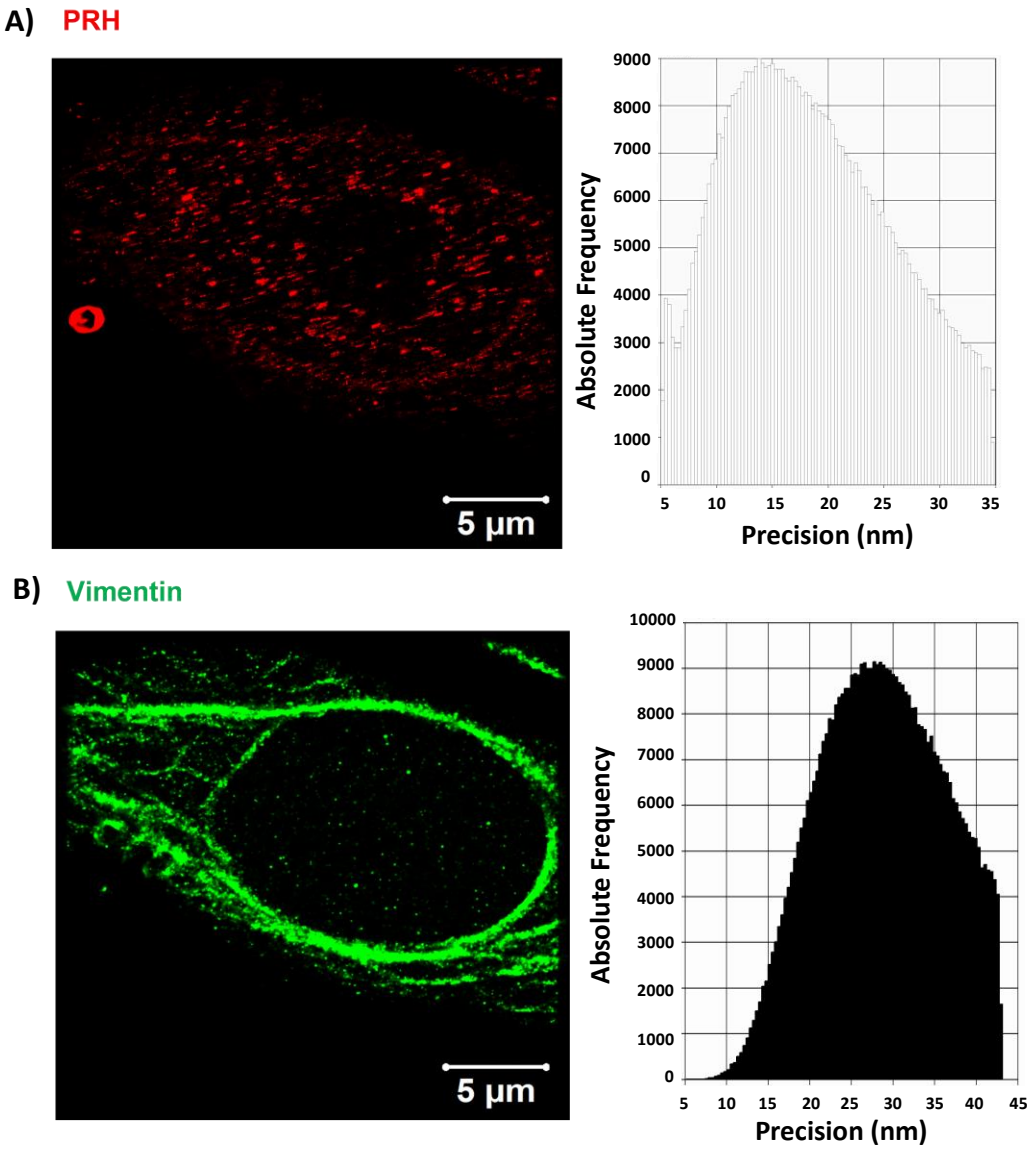
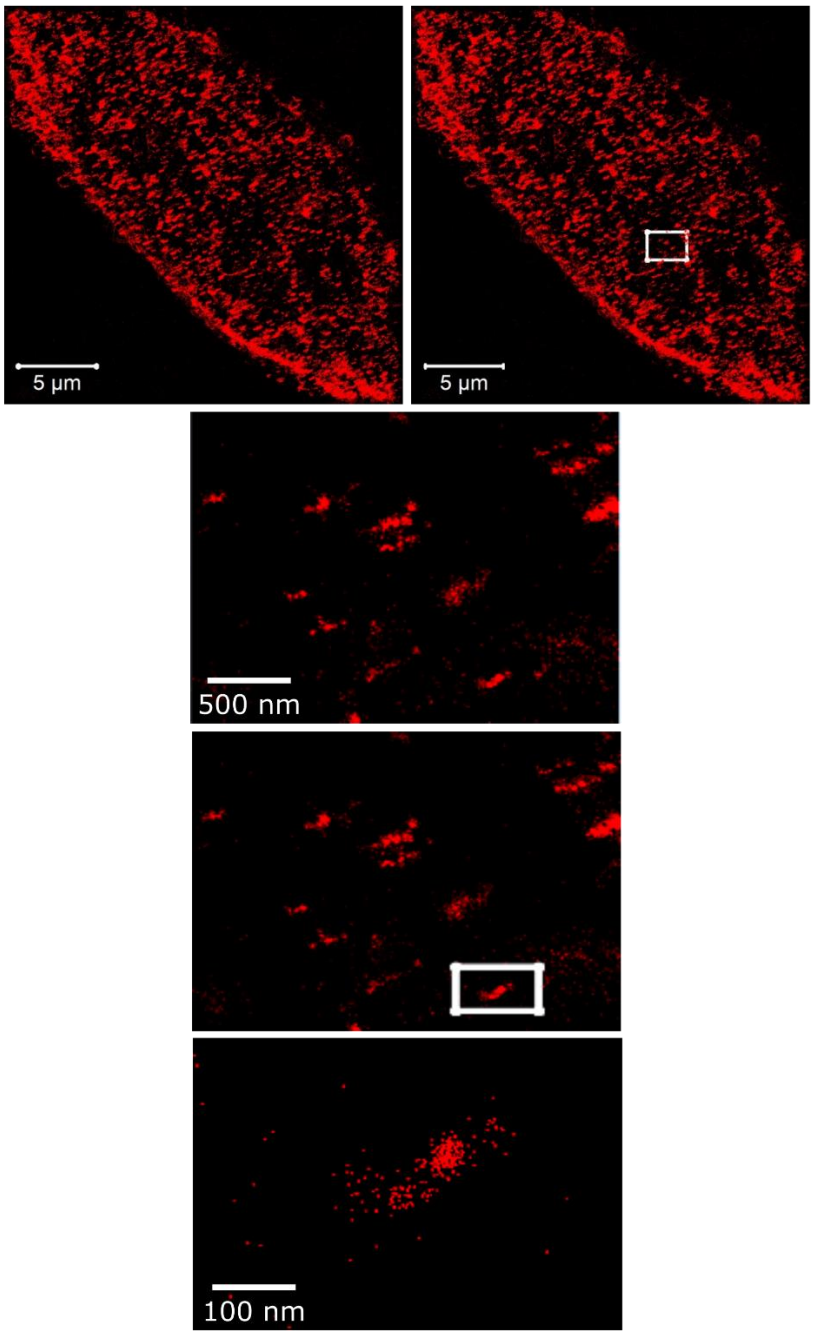


Figure 4-12 PRH and Vimentin single molecules are precisely localised using PALM dSTORM.

Precision histogram of molecules localisation precision distribution where the peak represents the central tendency of localisation. (A) STORM-rendered image of nuclear PRH in a CCLP1 cell. Scale bar 5 µm. The accompanying histogram represents the precision of molecule localisation peaking at 15 nm. (B) STORM-rendered image of nuclear Vimentin in a CCLP1 cell. Scale bar 5 µm. The accompanying histogram represents the precision of molecule localisation peaking between 25-30 nm.

In (figure 4-13) we show the actual makeup of (A) PRH and (B) Vimentin spots. We zoomed-in gradually into a certain inset (surrounded by a rectangle) to show that both PRH and Vimentin form clusters. A cluster is defined as a collection of grouped molecules. Each cluster can be seen composed of neighbouring clusters rather than being a single large molecule of protein (scale bar 500 nm). Further zooming-in revealed the single molecular formation of the clusters (scale bar 100 nm). These results unveiled more of the molecular base of PRH-Vimentin expression in CCLP1 cells than confocal imaging.

A) PRH



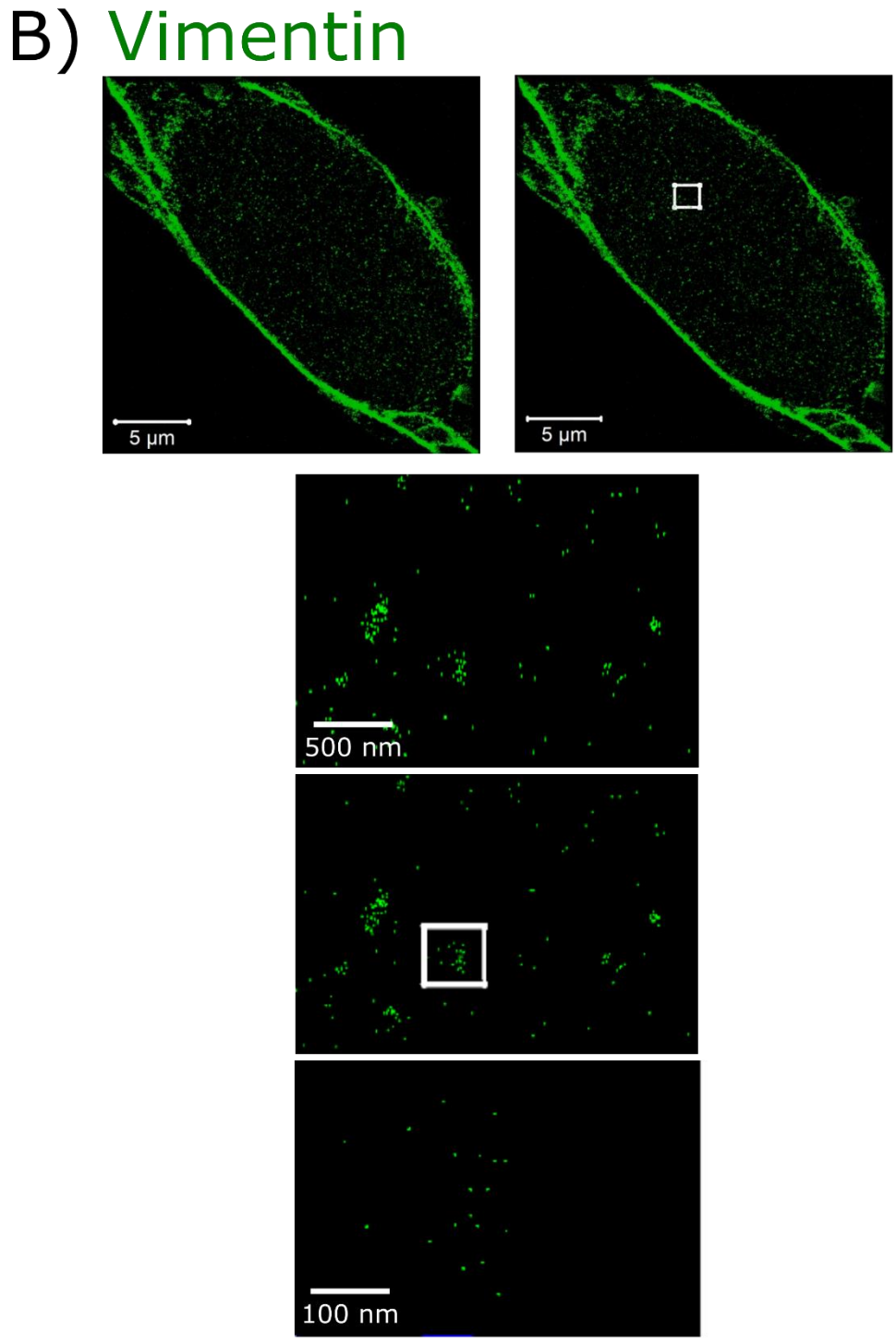


Figure 4-13 Example STORM-rendered images of nuclear PRH (red) and Vimentin (green) in CCLP1 cells showing various sizes of cluster formations at different magnifications.

A spot of either (A) PRH or (B) Vimentin is composed of several neighbouring single molecules (scale bars 5 µm, 500 nm, and 100 nm).

To analyse PRH-Vimentin nuclear clusters in CCLP1 cells, we followed the steps from the previous section. Because we wanted to perform PRH-Vimentin dual analysis in MATLAB, we needed to combine their localisation tables into a single file. In our files we put PRH coordinates first, and Vimentin second. This led to MATLAB assigning the colour green for the first protein and red for the second when it restored the image from its coordinates. Therefore, in this part of the results we show PRH in green and Vimentin in red. We show in (figure 4-14-B) the switch in PRH-Vimentin colours and the ROI divisions.

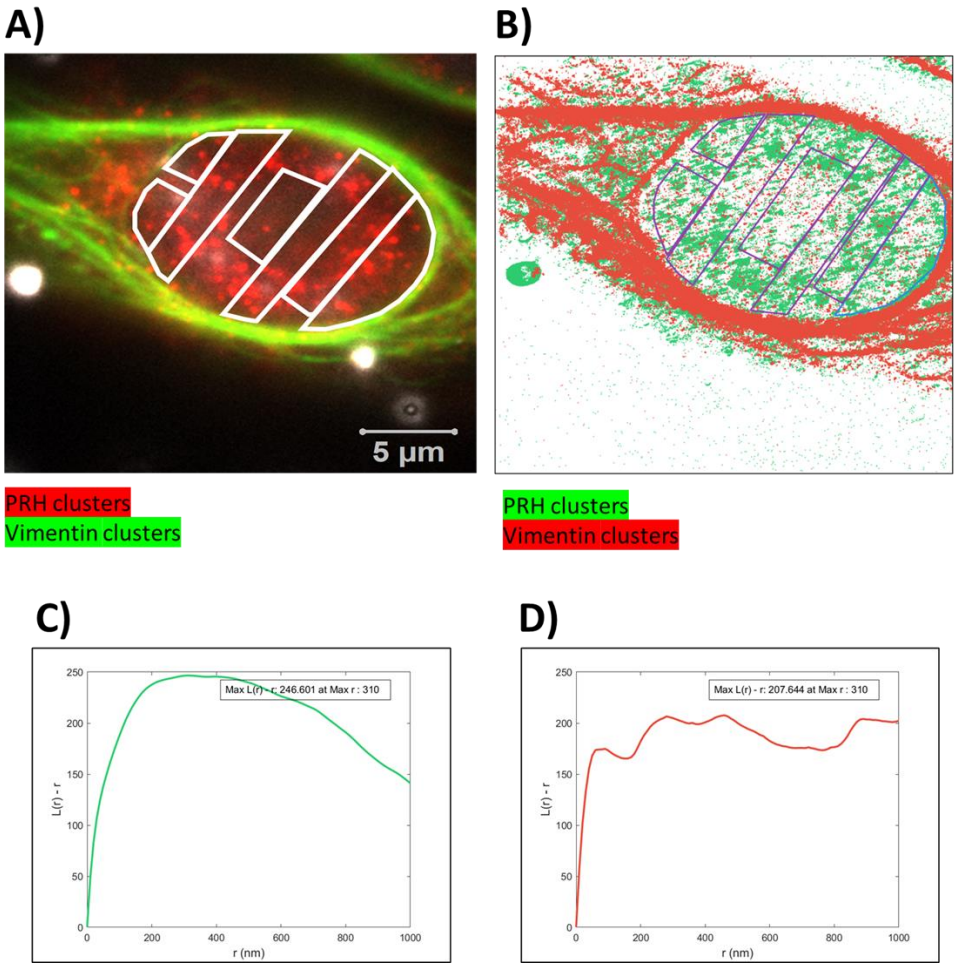


Figure 4-14 Cluster analysis of nuclear RPH and Vimentin.

(A) Example nuclear view of PRH (red) and Vimentin (green) showing the selected seven regions of interest avoiding the fucidal beads. Scale bar 5 µm. (B) The switch in colours after uploading the localisation table to MATLAB-ClusDoC tool showing PRH (green) and Vimentin

(red) along with the selected regions taken forward for cluster analysis. Selection was chosen inside the region bordered by the Vimentin cage. (C and D) Example depiction of molecular clustering and random distribution of one of the selected regions showing (c) PRH in green and (D) Vimentin in red. Cluster analysis was performed using the linearised form of *Ripley's K* function *(L(r)-r),* where *r* represents the radius. A plot of *L(r)-r* versus *r* gives a value of zero to show a random distribution, while positive values indicate clustering of molecules. PRH and Vimentin organisations peak at 310 nm.

We divided our nuclear region into multiple smaller ROIs to avoid the down crashes of ClusDoC on the computer we used for the analysis. Another benefit of breaking down the ROI is to avoid areas overlapping with the fucidal beads. Once the areas were selected, the file of selection could be saved in case it was needed for repeating the analysis. To find the linearised radius (*r*) at which most of the probable molecular aggregation occurs, we scanned ROIs for *Riply's K* (figure 4-14-C and D). Since ROI was segmented, we used the average *r* value in the following steps. The parameters for *Ripley's K* were adjusted to start at 0 nm to 1000 nm and calculate *Ripley's K* every 10 nm steps. The value for maximum radius (Max r), which represented the radius with the highest probability of aggregations, was recorded for the following ClusDoC step.

To scan PRH-Vimentin for clusters and characterise their clustering, we used the tool ClusDoC. This tool can perform two tasks simultaneously: cluster scanning with DBSCAN, and co-loclaisation with DoC.

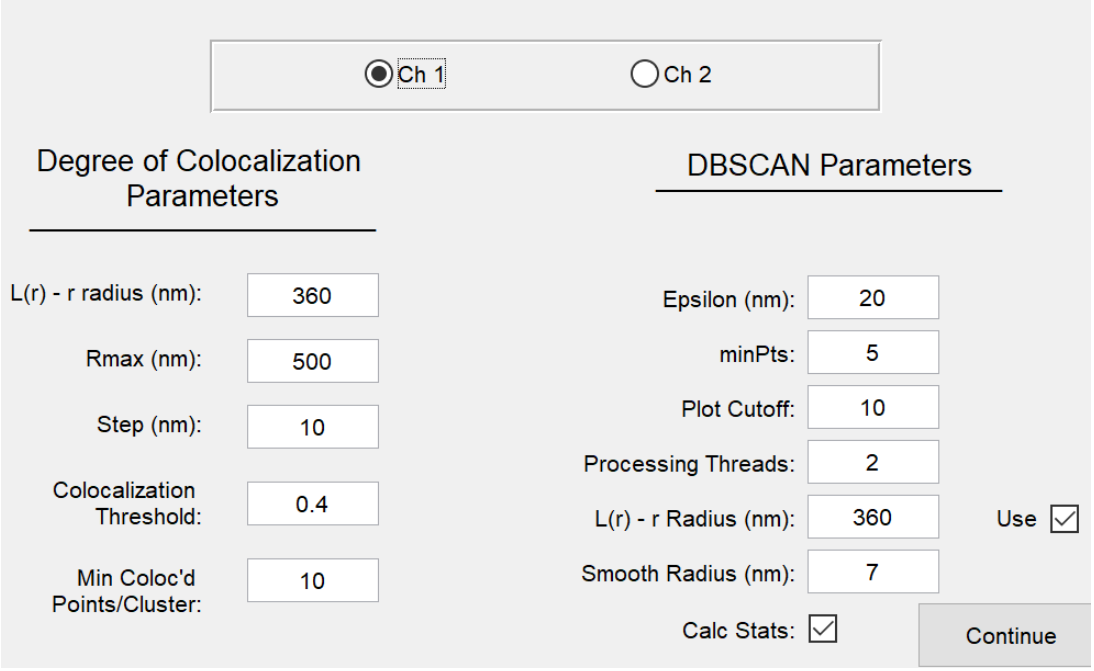


Figure 4-15 The window for ClusDoC tool showing the parameters of degree of co-localisation and DBSCAN.

In (figure 4-15), the parameters under the **degree of co-localisation** were submitted once for both channels. *L(r)-r* radius was chosen as the average of both channels taken from the *Ripley's* function step. Co-localisation threshold was adjusted to 0.4 according to the standard value used in the literature (413). The minimum number of co-localised clusters was chosen as 10. The rest of the settings were left as default (412, 413). We adjusted the **parameters for DBSCAN** for each PRH (Ch1) and Vimentin (Ch2) independently. Epsilon (nm) which is the radius at which the software searches for neighbours was set according to the peak value of the localisation precision histogram for each channel PRH (15 nm) and Vimentin (25-30 nm). min (Pts) which represented the minimum number of molecules that identified the formation of clusters was adjusted as 5. Plot cutoff and parenthesis threat

were kept as default (10 and 2, respectively), *L(r) -r* Radius (nm) value was taken from the *Ripley's* function step. Finally, a smooth radius of 7 nm was used to identify the radius of the boundary mask around the clusters and calculate the characteristics of the clusters (412, 413). Applying these settings generated two maps, one was for cluster identification, and the other was for degree of colocalisation. Additionally, we exported the outputs in txt format for EXCEL analysis.

The first outcome in (figure 4-16) shows PRH cluster maps. PRH clusters are illustrated as green areas of various sizes. The black edge represents those PRH molecules that failed to find further neighbours according to the parameters above, and therefore defined the periphery of the clusters. The grey areas are PRH outliers that did not exist closer to any other PRH molecules at all. Similarly, Vimentin clusters appear as red area rimmed with a black line representing the end of the cluster area. Gray areas are Vimentin outlier molecules. The cluster areas of both PRH and Vimentin are correlating to their abundance of expression in the nucleus.

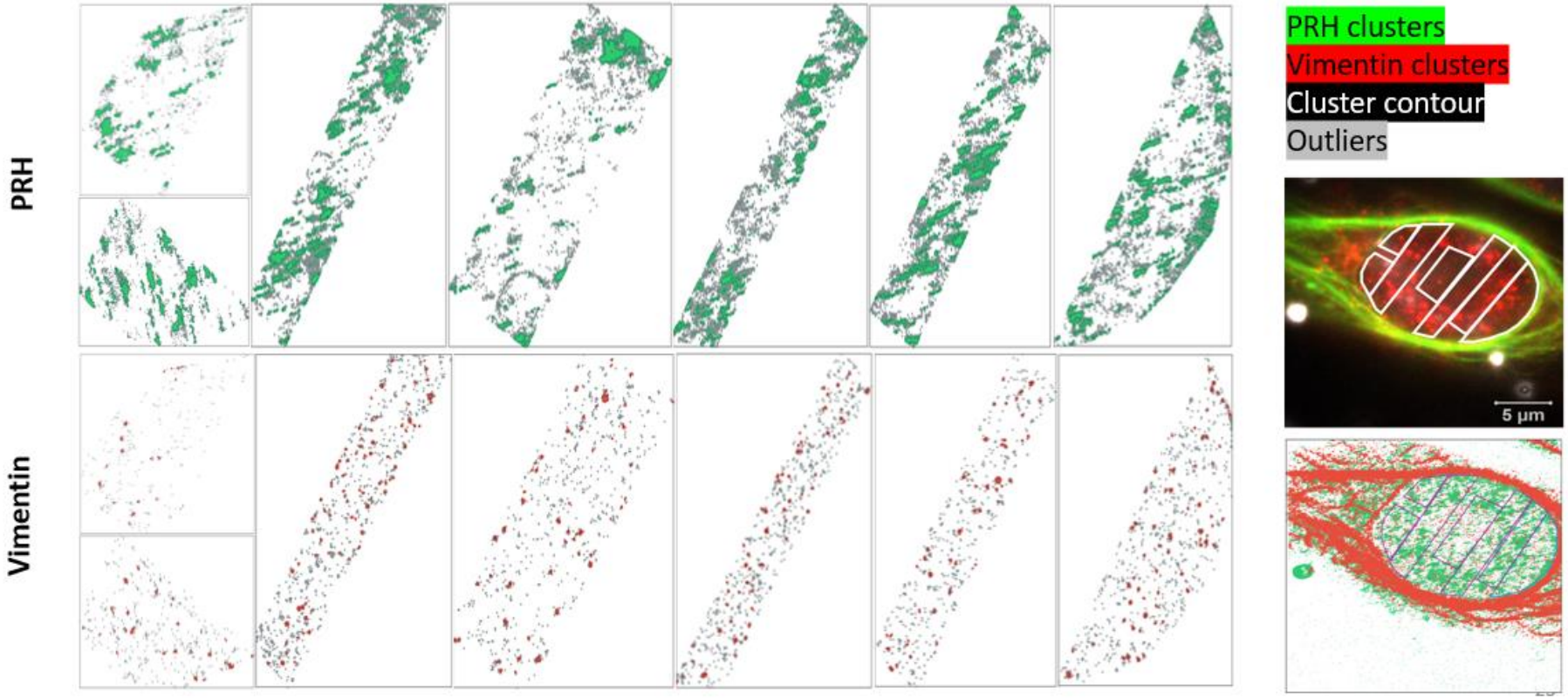


Figure 4-16 Cluster maps for PRH and Vimentin based on the selected regions for cluster analysis.

The maps show PRH clusters in green, Vimentin clusters in red, the boundaries of the clusters in black, and the outliers (non-clustered molecules) in grey. PRH clusters appear larger than Vimentin clusters.

To provide statistical overview about the differences in clustering between PRH and Vimentin, we used the numerical data output to find information about the number of the clusters, the number of molecules in them, their proportion, and their sizes. To achieve that, we analysed the output of n=14 cells. We prepared violin plots to represent the analysis using Graphpad PRISM (figure 4-17).

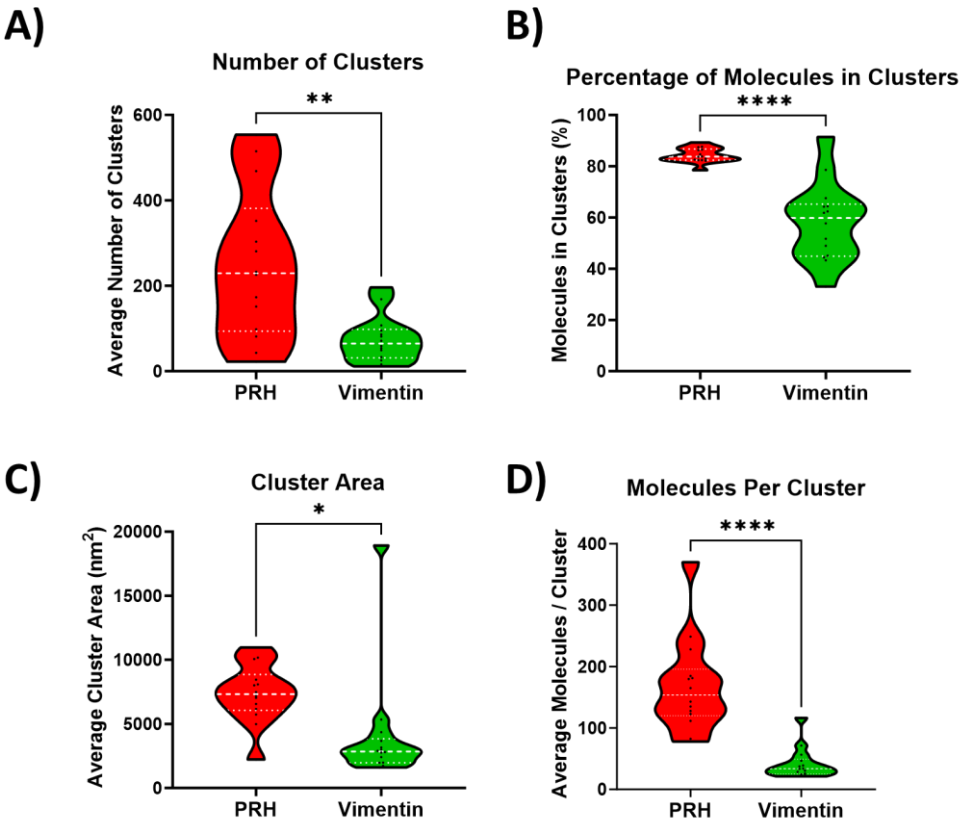


Figure 4-17 Statistical analysis of nuclear PRH and Vimentin cluster organisations in CCLP1 cells.

Individual data point represents the average value for a cell's selected ROIs (n = 14 cells). (A) PRH forms significantly more nuclear clusters than Vimentin (p= 0.0024). (B) 84.31% of the localised nuclear PRH is clustered, while only 58.19% of Vimentin molecules are in clusters (p<0.0001). (C) PRH forms significantly bigger nuclear clusters compared to Vimentin (p=0.0212). (D) PRH forms nuclear clusters containing significantly more molecules compared to Vimentin (p <0.0001).

On average, we had significantly more PRH (mean=250.2) clusters than Vimentin (mean=75.22) inside the nucleus (p=0.0024) (figure 4-17-A). To show the proportion of clustering PRH and Vimentin molecules, we calculated the average percentage of molecules in clusters. We found that PRH molecules were significantly closer and more grouped compared to Vimentin (PRH=84.31%, Vimentin=58.19%, p<0.0001, n=14) (figure 4-17-B). To check on the area occupied by the clusters, we calculated the average cluster area in nm^2 and found in (figure 4-17-C) that PRH clusters were significantly bigger than Vimentin (PRH=7367, Vimentin=4045, p=0.0212). To know how many protein molecules are found in the clusters, we calculated the average number of molecules per cluster. We found that PRH clusters had significantly more molecules compared to Vimentin (PRH=168.3, Vimentin=41.98, p<0.0001, n=14) (figure 4-17-D).

Taking these findings together, PRH molecules are visually and statistically more abundant than Vimentin in the nucleus of CCLP1 cells. This shaped the overall ability of cluster formation and characteristics of both PRH and Vimentin, i.e., bigger and more clusters for PRH compared to Vimentin.

4.2.3.3 PRH and Vimentin nuclear degree of co-localisation in CCLP1 cells using ClusDoC

After revealing how PRH and Vimentin are expressed in CCLP1 nuclei, we proceeded to explore their degree of co-localisation. The DoC task of ClusDoC generated the DoC map shown in (figure 4-18).

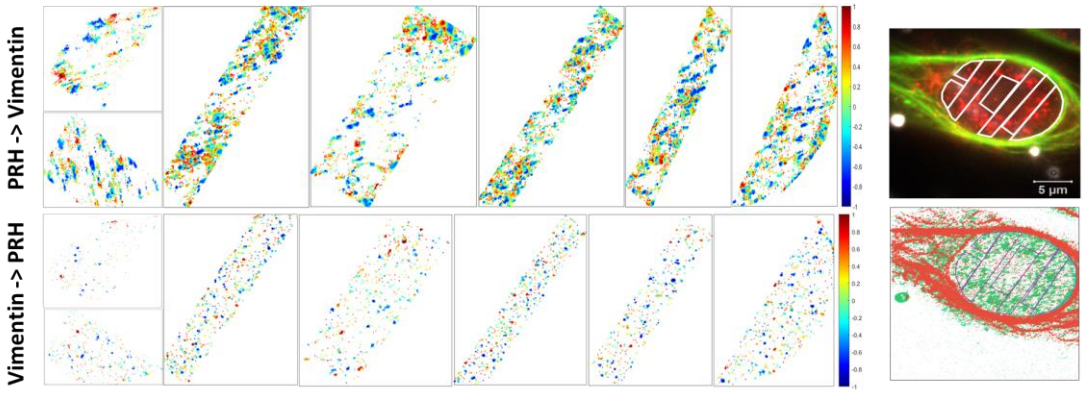


Figure 4-18 Degree of co-localisation analysis of nuclear PRH and Vimentin clusters in CCLP1 cells in seven selected ROIs.

Cluster degree of co-localisation heatmap showing the colour score ranging from -1 (blue) to 1 (red) where values of 1 (red) mean perfectly co-localised clusters and -1 (blue) means clusters separated from each other. Cut-off value is set to 0.4.

The DoC heatmap shows the outcome of bi-directional degree of co-localisation inspection, and the colour scale to interpret the result. The tool starts by picking up PRH molecules and scanning every neighbouring Vimentin molecule in an outward expanding mode. Then, the opposite is also done where the tool picks up every Vimentin molecule and scan for the surrounding PRH. The score ranges between 1 and -1, where 1 is a perfect colocalisation and -1 is no colocalisation. In the parameter, we chose 0.4 as cut-off value. Therefore, clustered molecules in deep orange to red represent a perfect colocalisation, whereas shades below 0.4 toward dark blue means less to no colocalisation. The degree of co-localised Vimentin is significantly more than PRH (p=0.0042) averaged from n=14 images (figure 4-19).

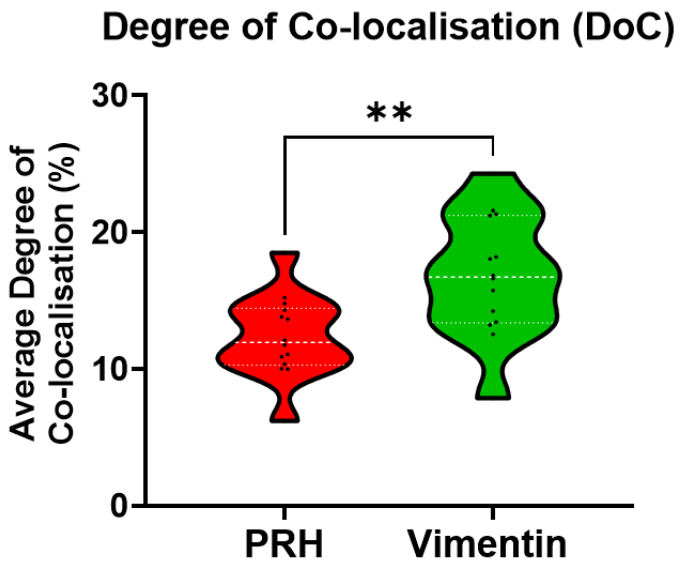


Figure 4-19 PRH nuclear clusters show a significant Degree of co-localisation with nuclear Vimentin (*p*= 0.0042).

Individual data points represent the average percent value of each protein that is localised in a cell's ROIs (n = 14 cells).

Overall, the data we showed from clustering and degree of co-localisation analyses provide a super-resolution insight into the spatial organisation of PRH and Vimentin in CCLP1 nuclei. They demonstrated the clustering capacity of PRH and Vimentin in the nuclear compartment. They also suggested Vimentin may exist spatially near PRH at single molecule level. There is more disposition of Vimentin clusters localisation compared to PRH. This is due to the low abundance of Vimentin in the nucleus compared to PRH. Vimentin is easily found close to PRH inside the nucleus, but more PRH will be found far away from Vimentin.

4.2.4 PRH and PRH-induced Vimentin spatial localisation in AKN1 cells

Kitchen et al. reported that PRH overexpression in AKN1 cells and BECs cells induced the expression of Vimentin at the protein level. RNAseq data from

those cell lines revealed the upregulation of Vimentin mRNA (215). Additionally, in the previous chapter we showed how we generated another AKN1 cells expressing GFP-PRH-Myc. We used these cells to investigate whether PRH-induced Vimentin can be characterised for spatial proximity with PRH by immunofluorescence confocal microscopy.

We examined Z-stack images representing AKN1 GFP and AKN1 GFP-PRH-Myc expressing cells co-probed with PRH and Vimentin antibodies. We followed the same steps of image processing from CCLP1 PRH and Vimentin localisation in section (4.2.2) to prepare the images for analysis; qualitative and quantitative.

Figure 4-20-A shows the sequential acquisition of DAPI (blue), PRH (red), Vimentin (green), and all three channels overlapped together (merged). PRH can be seen easily as small red foci in AKN1 cells despite the overall low expression. Vimentin shows similar pattern of focal expression in the GFP-PRH-Myc expressing cells. However, this differs from the cytoplasmic filamentous Vimentin seen in CCLP1 cells. This may be due to Vimentin being detected at its very early or very low stages of expression.

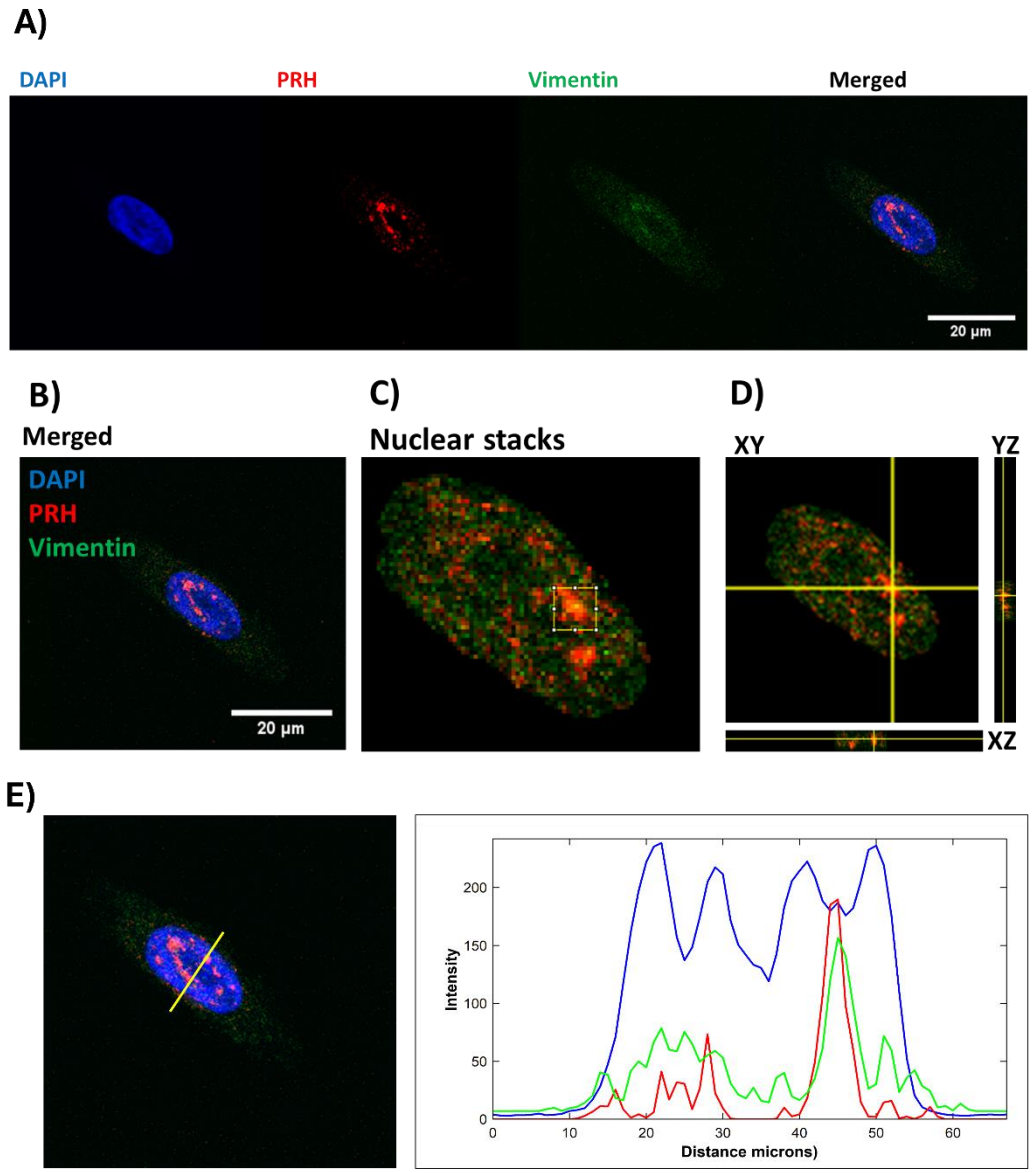


Figure 4-20 Weak PRH and PRH-induced Vimentin overlap in AKN1 cells expressing GFP-tagged PRH.

(A) Images of AKN1 GFP-PRH-Myc cell co-stained for PRH (red), Vimentin (green) and DAPI (blue) showing both PRH and Vimentin are mainly nuclear here. Cells were fixed with 4% FA and labelled with PRH antibody along with nuclear DAPI staining. Images were acquired sequentially for n= 59 cells using a confocal Leica DMI4000B microscope at 63X magnification. Scale bars 20 µm. Laser power, gain, exposure, and pinhole were the same throughout the imaging. (B) All channels merged. (C) Close-up view on the nuclear region of PRH (red) and Vimentin (green) showing a selected ROI taken for further exploration. (D) An example orthogonal view tracing lines in different planes (xy, xz and yz) of confocal microscope images used to analyse PRH-Vimentin overlapping. A nuclear PRH-Vimentin overlap (yellow) traced in all three planes (xy, xz and yz). (E) An example of PRH-Vimentin localisation in the nuclear regions of the cell using RGB profiler plot in FIJI. DAPI channel was used to trace the area of interest. There is a strong expression intensity of PRH (red) and a weak intensity for Vimentin (green).

To study PRH-Vimentin spatial localisation in AKN1 cells expressing GFP-PRH-Myc, we used FIJI XYZ-orthogonal view and RGB profiler analysis. Due to GFP brightness overshadowing PRH and Vimentin, we removed GFP channel to ease tracing the overlap. Since the expression for both PRH and Vimentin was seen better in the nuclear region, we focused on their nuclear spatial expression investigation. We used the DAPI channel to identify and extract the nuclear regions from the stacks. We followed the same steps stated in section (4.2.2). We performed this analysis on n=54 cells. Figure 4-20-C and D show a traced yellow spot in a box that is present in all three spatial planes. Like in CCLP1 cells, PRH and Vimentin overlap here was closer to the periphery of the nuclear region.

To support the XYZ-orthogonal view findings, we used the tool RGB profiler in FIJI on the best DAPI slices of the Z-stacks. Figure 4-20-E shows where we drew the line over DAPI region. The histogram plot on the side for the drawn line shows DAPI (blue), PRH (red) and Vimentin (green). There is one overlapped up-deflection around (45) µm from the left side of the plot. This suggests that PRH and Vimentin may present in close distance of each other in this location.

These data demonstrate that AKN1 GFP-PRH-Myc cells had focal pattern of nuclear PRH expressions. Unlike the staining of Vimentin in CCLP1 cells, Vimentin also appeared focal and nuclear in these cells. PRH-Vimentin overlap observations on XYZ-orthogonal views and RGB profiler, when detected, were consistent with the nuclear overlap findings seen in CCLP1 cells.

4.2.5 The use of fluorescence correlation spectroscopy (FCS) to study the dynamics of GFP-PRH fused protein in AKN1 cells for long term GFP-PRH expression

In chapter 3 we demonstrated how we encountered a reduction of GFP-PRH-Myc expression in later passages of the cells. This affected the phenotype we observed in these cells. In the previous section we used relatively late passaged cells to examine PRH-Vimentin spatial localisation. Although we observed few overlap events in some of the cells, it was relatively difficult to detect due to the level of GFP-tagged PRH expression and the level of Vimentin upregulation

We used the same lysate from section (3.2.5 Issues between early and late passages of clone D). Figure 4-21-A shows the expression of expressed (GFP-tagged PRH) and endogenous PRH (M3) with β-Actin for loading reference in 3 independent samples. The lane for N2 had two bands for GFP-tagged PRH where one was at 60 kDa, while the other was slightly lower. Figure 4-21-B shows a second Western blotting using the same lysate where N2 showed a band for GFP at 27 kDa. This suggests that we may have a stand-alone GFP expression along with the full-length fused protein or cleavage of the fusion protein.

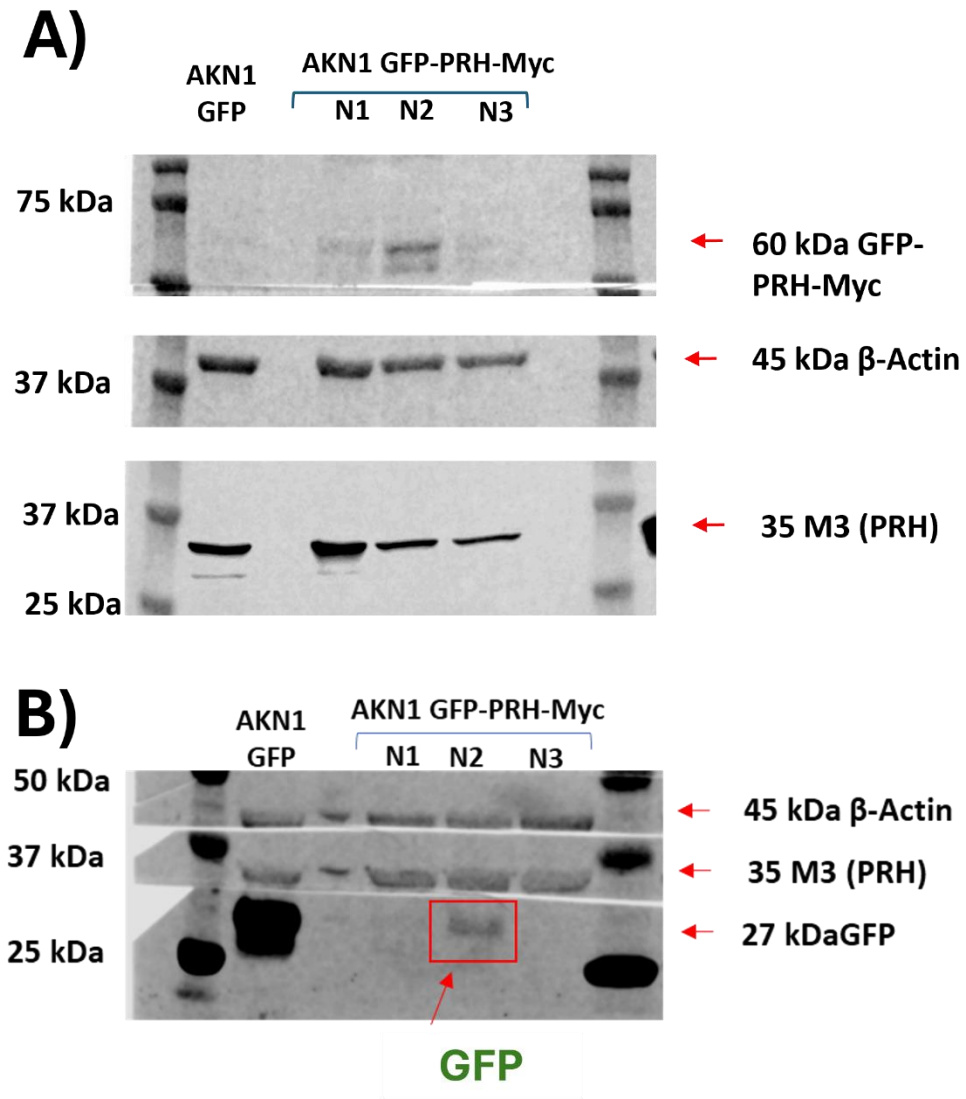


Figure 4-21 Western blotting of late passaged AKN1 GFP and AKN1 GFP-PRH-Myc cells.

(A) A representative Western blotting image of whole cell protein extract probed with GFP antibody. Protein extracts prepared from 3 different AKN1 GFP-PRH-Myc cell populations were compared. There are two bands in N2. (B) A second Western blotting of the same lysates showing N2 having a GFP band. β-Actin was used to show protein loading for both Western blotting.

To explore the stoichiometry and dynamics of GFP and GFP-tagged PRH fused protein in AKN1 cells and relate that to what we experienced at late passages of the cells, we used fluorescence correlation spectroscopy (FCS). FCS is a tool to study molecular dynamics via analysing the intensity

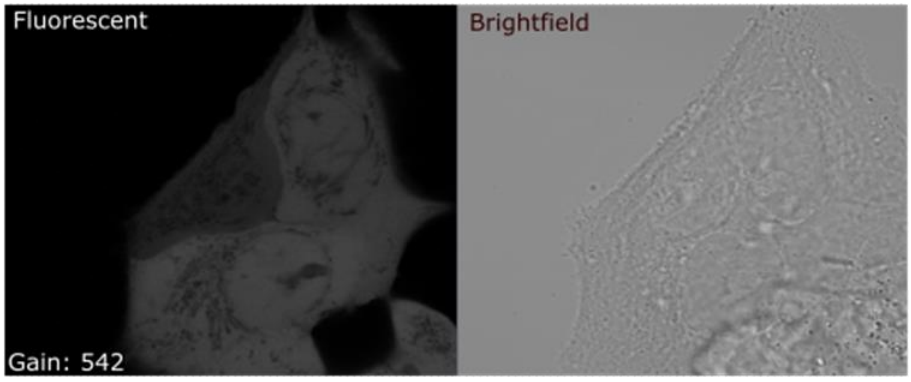
fluctuation of diffusing fluorescent molecules through a focused light (414). It can provide details on the concentration, diffusion, and brightness properties of cellular fluorescent components. The literature stated that FCS is better in fluorescent low-expression system (466). This experimental setting allows the detection of a large range of fluorescence fluctuation. This way FCS shows high sensitivity of monitoring small number of particles with low fluorescence level. To prepare cells for FCS, we seeded late passages of AKN1 GFP and AKN1 GFP-PRH-Myc cells at various densities in a Nunc Lab-Tek 8-well chamber slides gifted by Dr Joelle Goulding from the School of Life Science Imaging facility (SLIM). On the day of imaging, we replaced the media with an imaging buffer as detailed in section (2.3.4Fluorescent correlation spectrometry (FCS)).

Confocal and FCS work was conducted with help from Dr Goulding who also assisted with interpreting the results. We recorded GFP as the cells were passing through a confocal path positioned at defined regions of the cells (mainly cytosolic). This allowed us to obtain information on the average diffusion of GFP and GFP-labelled components.

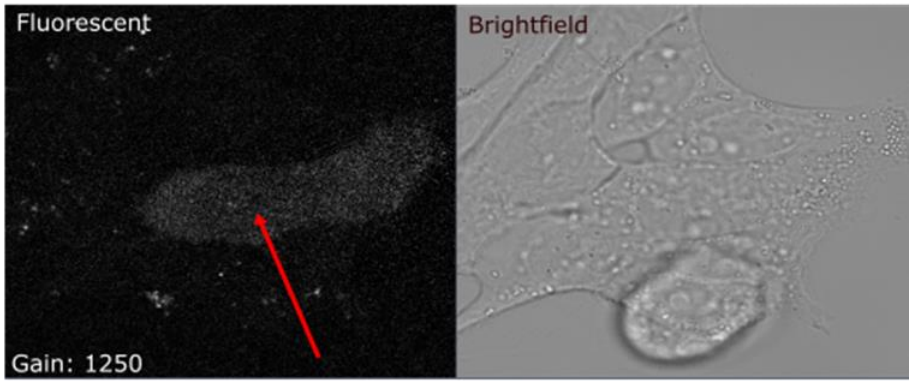
To have a general understanding of GFP and GFP-tagged PRH expression in our cells, we acquired confocal images prior to FCS. In (figure 4-22- A) AKN1 GFP displayed various level of GFP brightness, but overall, they are very bright. GFP is uniformly expressed throughout the cells and was easily detected at very low gain. We also had scattered non-GFP expressing cells indicating a mixed population. For AKN1 cells expressing GFP-tagged PRH, we observed three patterns of GFP expression. The first pattern (not shown

here) was quite similar to AKN1 GFP in which the cells were uniformly green. In the second form (figure 4-22- B), we had diffused cytosolic GFP. It was very dim that we needed to increase the gain to visualise them (red arrow). Lastly, we had cells expressing small foci and clusters of GFP also diffused in the cytoplasm (red arrow) (figure 4-22-C).

A) AKN1 GFP



B) AKN1 GFP-PRH-Myc



C) AKN1 GFP-PRH-Myc

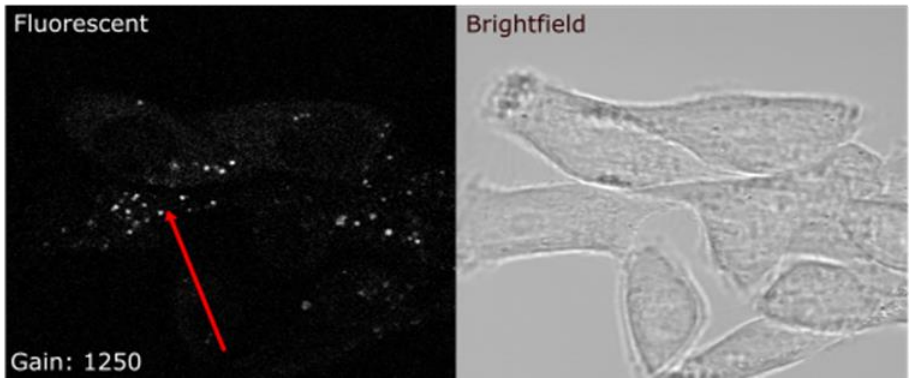


Figure 4-22 Confocal imaging of AKN1 GFP and AKN1 GFP-PRH-Myc showing the three observed patterns of GFP expression.

(A) Uniformed and very bright GFP detected at low gain (542) in AKN1 GFP cells (also in AKN1 GFP-PRH-Myc cells). (B) Diffused and very dim GFP (red arrow) detected at a high gain (1250). (C) Diffused and very dim GFP (red arrow) in foci and clusters detected at a high gain (1250). Images are representative of n=1 experiment.

To investigate GFP dynamics, we used FCS to trace GFP fluctuation starting with the control cells. Since FCS is preferred at very low fluorescence for faster tracing, we applied FRAP step. We bleached GFP using 488 nm at maximum power of 100% excitation for 10 seconds. Then, we recovered and recorded the fluorescence dynamics using 0.1% excitation for 30 seconds. Figure 4-23-A shows the Z-intensity of GFP as we scanned the cell from one edge to another. It reflected the uniform existence of GFP throughout the cell. Therefore, we positioned the observation volume just next to the peak representing a central cytosolic position. Figure 4-23-B shows the recorded fluctuation over 30 seconds using 0.1% 488 nm excitation at a count range between 160 – 1780 kHz. To see how our trace fit to a standard 2 X3D component model, we applied default settings for trace autocorrelation. In the trace to model fit (figure 4-23-C), we see GFP recorded fluctuation in blue line and the standard model in green line. There is a good fit between our recorded line and the theoretical model. The amplitudes of the two lines at the beginning of the recording were very close to each other (Circled in red). The two findings together indicate a low correlation. In FCS, there is an inverse relationship between the correlation of the fitted lines and the concentration of the fluorescent particles within the observed volume. In our case, the correlation between the two lines was at very low value indicating very high concentration (N=572) at a count value of 3.2 count per molecule (CPM). Closely inspecting

the recorded trace, we can pull out two components at extreme speeds: a faster one (421 microseconds µs) and a slower (775 milliseconds ms). The faster component represented almost 95% of the signal indicating free GFP. The slower component represented 5% of the signal which could be due to GFP aggregation or impedance from an organelle on GFP path.

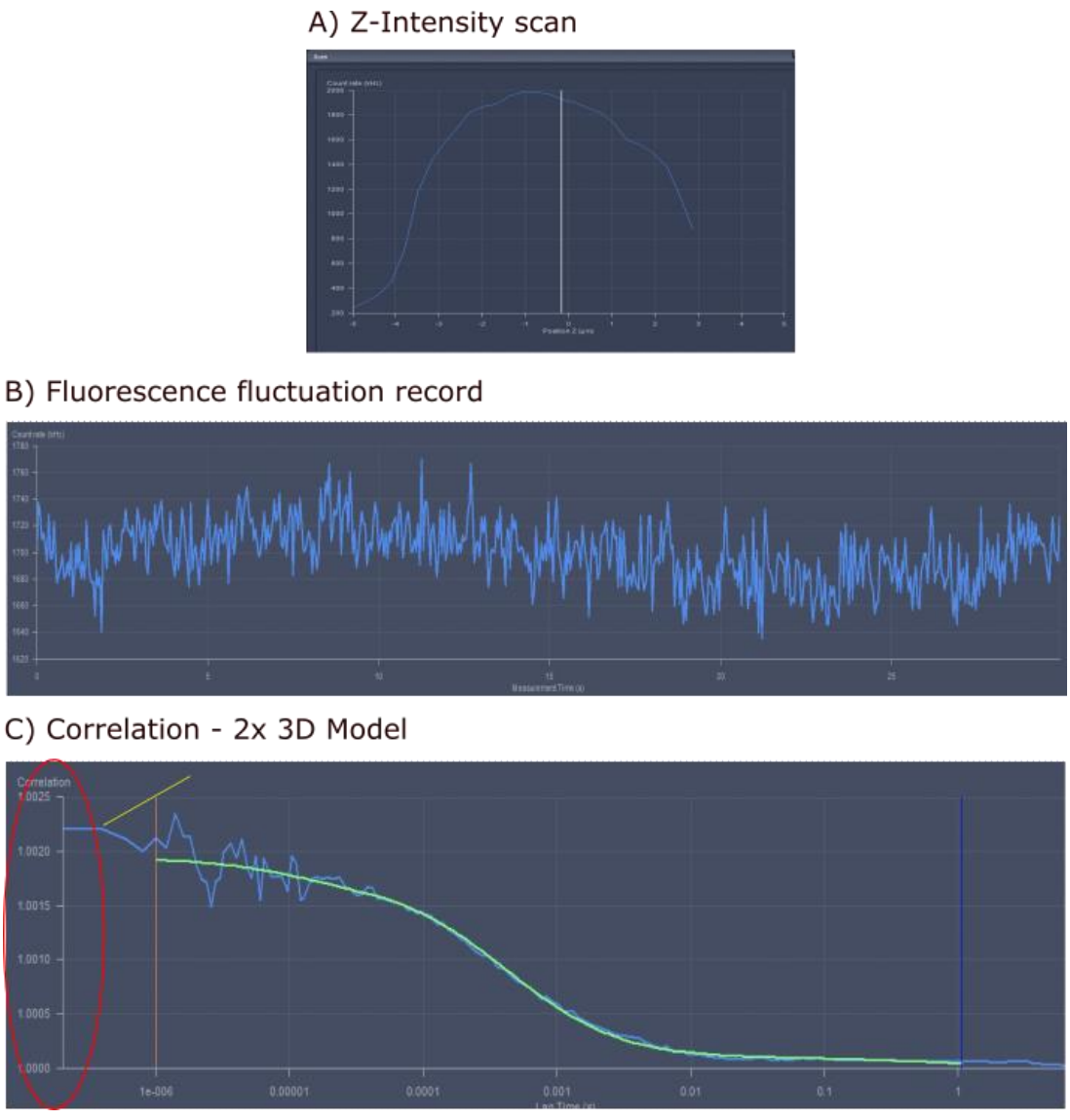


Figure 4-23 The recorded dynamic of the first pattern of GFP expression (uniformed and very bright) in AKN1 GFP cells.

(A) Z-intensity of GFP indicates the uniform existence of GFP throughout the cell. (B) The recorded fluctuation over 30 seconds using 0.1% 488 nm excitation at a count range between 160 – 1780 kHz. (C) GFP recorded fluctuation (blue) line and the standard model (green). It

shows a very low correlation between the two lines at the beginning of the recording (Circled in red) indicating very high concentration (N=572).

To investigate GFP-tagged PRH dynamics in AKN1 cells, we first traced the cytosolic-diffused GFP at 0.1% excitation of 488 nm for 30 seconds. We did not need to pre-bleach here since the fluorescence was not hindering the recording. However, there was some bleaching from excitation noticed by the drop in tracing that led us to omitting the highlighted portion of the graph before analysis in (figure 4-24-B). The Z-intensity scan on (figure 4-24-A) shows that we positioned our observation volume within the cytosolic region. The graph in (figure 4-24-B) shows the entire fluctuation record at very low count rate between 160 – 190 kHz (about 10X lower) compared to the control cells which confirms the findings from confocal imaging where we show how dim these cells were compared to control AKN1 cell expressing GFP alone. We also fitted our trace line (in blue) to a standard 2 X3D component model (in green) using the default settings again as seen in (figure 4-24-C). We can also see that we managed to extract information early in the track that we did not really need the entire duration of 30 seconds which indicates that this pattern of GFP was significantly fast.

The level of difference in amplitudes of the two lines at the beginning of the recording was big. In addition, the trace to fit model is weaker here. Together, they indicate that the correlation between the two lines is big correlation. This correlation here is reflecting low concentration of fluorescent particles within the observed volume (N=127) at an overall 1.36 CPM count rate. Like in control cells, we noticed here two components: a faster GFP at 398 µs speed

and a slower one at 424 ms. The first represented 93% of the signal showing very close resembling to the free GFP we detected in AKN1 GFP cells. The slower GFP composed around 6% of the signal and we assumed that was due to the slowness from GFP aggregation. We performed GFP dynamics analysis in AKN1 GFP-PRH-Myc cells at around 3375 CPSM brightness units.

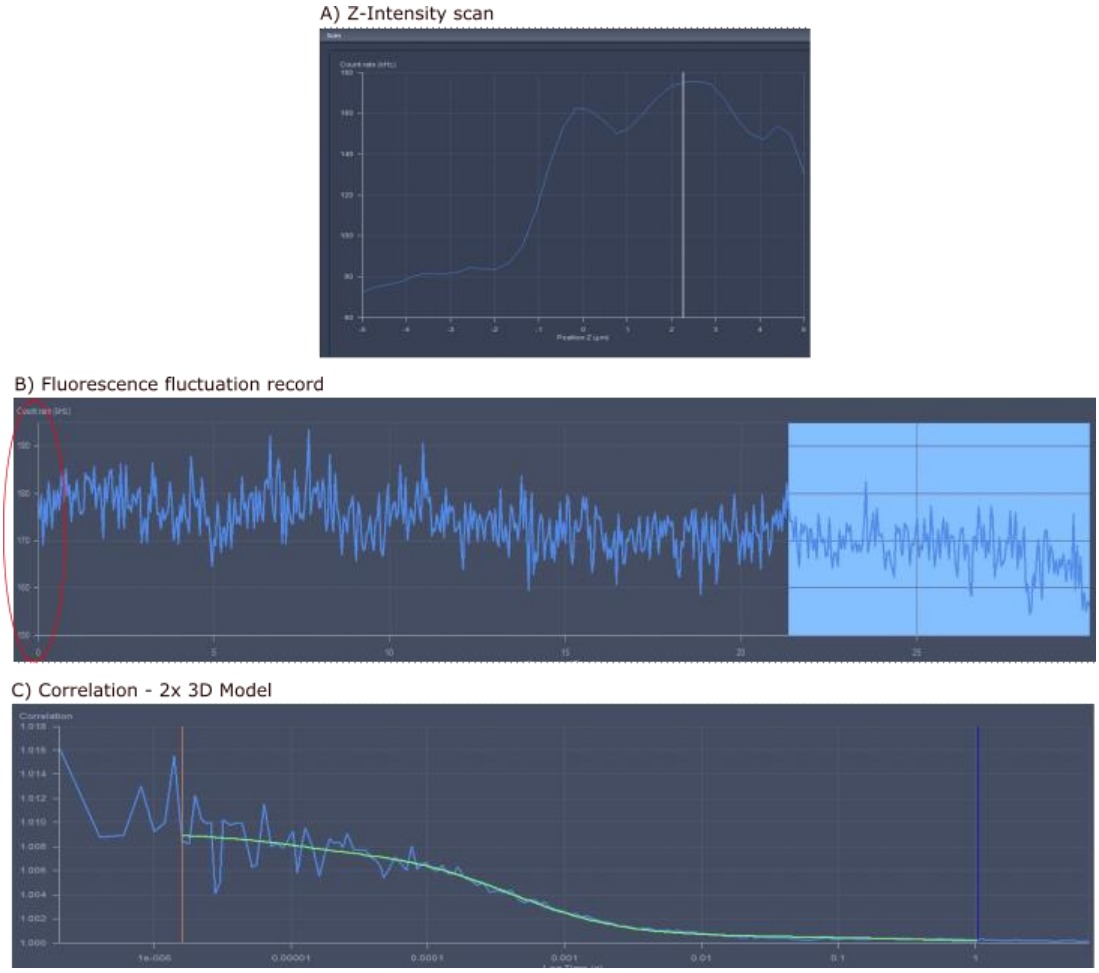


Figure 4-24 The recorded dynamic of the second pattern of GFP dynamics (diffused and very dim) in AKN1 GFP-PRH-Myc cells.

(A) Z-intensity of GFP indicates the cytosolic region of the cell. (B) The recorded fluctuation over 30 seconds using 0.1% 488 nm excitation at a count range between 160 – 190 kHz. (C) GFP recorded fluctuation (blue) line and the standard model (green). It shows a big correlation between the two lines at the beginning of the recording indicating very low concentration of GFP particles (N=127).

To explore clustered cytosolic-GFP in AKN1 GFP-PRH-Myc cells, we positioned the observation volume within green clusters in the cytosolic region as seen in the Z-intensity scan on (figure 4-25-A). Excitation settings were exactly the same that is 0.1% excitation of 488 nm for 30 seconds. Figure 4-25-B shows the full recording over 30 seconds at a range between 0 – 140 kHz that is lower than the range we saw in the previous two recordings. This is consistent with the need to increase the gain to see these dim clusters on confocal observation. Again, we excluded the two highlighted regions from the analysis due to the drop in average tracing resulted from excitation bleaching seen in (figure 4-25-B). Component analysis from fitting to model 1X3D in (figure 4-25-C) revealed that the GFP is significantly slower here than what we detected before. The correlation between the two lines indicates low concentration (N=43) at an overall rate of 2.5 CPM. The first component we noticed was at 415 µs and a query at 1 - 6 ms. GFP clusters were too slow that we could say they were relatively stationary which explained the increased drop in average tracing resulting from bleaching within our observation path. The entire trace was almost showing ~400us dwell time. We performed the analysis at around 2955 CPSM brightness units.

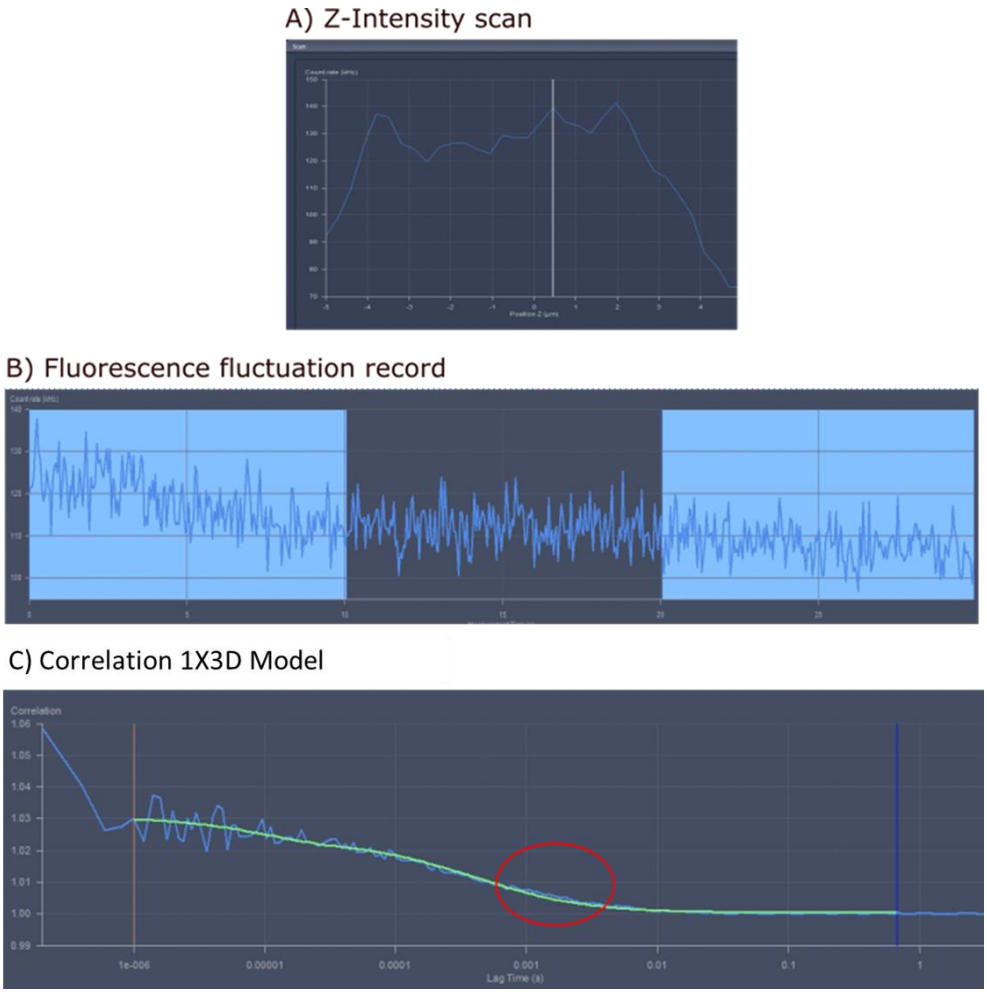


Figure 4-25 The recorded dynamic of the third pattern of GFP expression (diffused and very dim in foci and clusters) in AKN1 GFP-PRH-Myc cells.

(A) Z-intensity of GFP indicates the cytosolic region of the cell. (B) The recorded fluctuation over 30 seconds using 0.1% 488 nm excitation at a count range between 0 – 140 kHz. (C) GFP recorded fluctuation (blue) line and the standard model (green). It shows a big correlation between the two lines at the beginning of the recording indicating very low concentration (N=43).

Overall, we demonstrated the three forms of GFP dynamics in AKN1 GFP-PRH-Myc cells. The first form is, in fact, recorded in both AKN1 GFP-PRH-Myc and its control cell line AKN1 GFP which is the diffused bright pattern of GFP. This was seen uniformly existing throughout the cells. The other two forms were only recorded in AKN1 GFP-PRH-Myc; the diffusely dim GFP and the clustered very dim GFP.

To summarise, while maintaining cells growth we noticed that some cells in AKN1 expressing GFP-tagged PRH population were totally green resembling the control cell line. All the previous necessitated the need to explore potential factors accounting for these observations. We started by examining whether we had anomalous forms of GFP-tagged PRH in our cells using Western blotting focusing on late passages of the cell line. We previously provided data of plasmid diagnostic restriction digest to ensure we had the full-length transgene for the fused protein cloned in the correct site according to the plasmid map (figures 3-1 and 3-2). We further confirmed by sequencing the plasmid that we indeed had our transgene inserted correctly without a stop codon between GFP gene from the backbone plasmid and the PRH gene from the cloned transgene (sequencing available in the appendix). However, this still does not exclude the possibility of having alternative transcription start sites or differential promoter activities that eventually initiate the production of GFP independent of the fused protein.

Altogether, the data suggest we have different forms of protein products. One of the possibilities is expressing GFP independent of the complexed fused protein. Alternatively, the GFP-PRH fusion protein could be experiencing a degradation leading to expression of different forms of GFP including forms that suggest diffuse monomers of GFP-PRH-Myc rather than oligomers.

4.3 Discussion

Previously, Dr Ka Ying Lee conducted work for her PhD thesis where she identified Vimentin as a potential protein partner for PRH (269). Based on these findings, we aimed in this chapter to explore the potential of PRH-Vimentin close spatial localisation. PPI can be further confirmed if both proteins can be found together in subcellular compartments (467). Ideally, investigating protein localisation is better visualised in cells where both proteins are highly expressed. Since our AKN1 expressing GFP-tagged PRH cells were undergoing a reduction in PRH expression, we started by using CCLP1 cells where both PRH and Vimentin are abundant. While both environments contribute valuable insights into the biology of PPI, overlapping and PPI in cancer cells are more relevant to tumour biology, disease progress, and targeted therapies (468). In contrast, investigating protein overlapping upon protein overexpression in normal cells, although fundamental to understand cell biology, may not be directly related to cancer. We acquired images of cells co-probed for PRH and Vimentin using a high-resolution confocal microscopy or super-resolution single molecule microscopy. We, then, implemented several image analytical tools in order to characterise PRH and Vimentin distribution in CCA cells as well as study their spatial localisation. We also investigated whether we see a similar pattern in AKN1 cells expressing GFP-tagged PRH. Finally, we shed light on the dynamics of GFP and GFP-tagged PRH in the same cell lines.

We demonstrated that PRH in CCLP1 cells appears as punctuated foci of various sizes in the nuclear and cytoplasmic subregions of the cells. Although

PRH exhibits a recognisable cytoplasmic expression, it is mainly a nuclear protein in CCLP1 cells as indicated by the intensity measurements of PRH signals. We also showed that nuclear PRH can be seen arranged away from areas that are DAPI bright stained (heterochromatin). These findings are consistent with PRH nuclear roles. Findings related to PRH extent of distribution in subregions of cells are not confined to PRH in CCA cells, they were also seen in leukaemia, prostate, and breast cancer cells. Specifically, high nuclear to cytoplasmic PRH expression has been reported in U937 and K562 myeloid cells where PRH co-localises with eIF4E and inhibits its transport and consequently inhibit cell proliferation. In contrast, other cancers have been characterised with high cytoplasmic to nuclear PRH ratio where this shift in PRH localisation drives cell proliferation and migration in cancers of breast (214, 249), thyroid (245), and haematopoietic cancers (250). In regard to the appearance of PRH on immunofluorescence -stained images, PRH in CCLP1 cells has been also seen as foci in either subcellular compartment which is similar to how it was observed in AKN1 cells expressing GFP-tagged PRH in the previous chapter. It is also quite similar to how it was observed in K562 (lymphoblast cells from myelogenous leukaemia) (196), and U2OS (osteosarcoma cells) (252). To the best of our knowledge, this is the first time PRH was detected in CCLP1 context through immunofluorescence staining.

We also report for the first time the spatial distribution of Vimentin in CCLP1 cells via immunofluorescence staining. Vimentin as reported in the literature can be within different subcellular compartments in various forms ranging from

small squiggles to extensive filamentous network in the cytosol (317, 323). Our findings regarding Vimentin localisation in CCLP1 cells are consistent with the literature, Vimentin appeared as a classical filamentous cytoskeleton dominating the cytoplasmic region. In addition, we also observed the small extension of Vimentin as sporadic weakly stained foci in the nuclear regions where most of it was detected towards the periphery of the nuclear area. Furthermore, the latter resembled the forms of PRH-induced Vimentin in our AKN1 GFP-PRH-Myc cells as we reported in the previous chapter. It was reported that Vimentin goes through stage-wise formation when induced in Vimentin-null MCF7 cells where it was seen activating gene sets associated with cell migration (469). These stages started as small particles and squiggles that reached the filamentous form few days post induction (469).

The main reason behind exploring the extent and forms of expression for both PRH and Vimentin in this chapter was to set the ground for co-localisation analysis. The term co-localisation in the literature was found to refer to either spatial overlapping of multiple fluorophores (qualitative analysis), or correlation in expression of at least two spatially overlapped proteins (quantitative analysis) (470). For our confocal immunofluorescence image analysis, we focused on PRH-Vimentin spatial overlapping investigations. Simple approaches to investigate spatial overlapping are multi-channel overlay, XYZ-orthogonal views and RGB plot profile on FIJI (7). In our data we demonstrated that endogenous Vimentin could exist near endogenous PRH spatially in the cytoplasmic subregions of CCLP1 cells. We were also able to detect some peripheral nuclear PRH-Vimentin overlaps although to a lesser

extent compared to the cytoplasm. In addition, we also attempted to investigate the localisation of exogenous PRH in relation to PRH-induced Vimentin in AKN1 cells expressing GFP-tagged PRH. We applied the same analytical approach only in the nuclear subregion where we demonstrated very few events of potential PRH-Vimentin spatial proximity. The visual observations of PRH-Vimentin fluorophores overlap may appear co-localised to the naked eyes. Nevertheless, they are not definitive evidence for protein spatial co-localisation.

PRH was previously reported to overlap with various proteins using multi fluorophore confocal imaging and channel overlaps to serve certain biological functions. For instance, PRH was found to overlap with Abl Interactor 2 protein in hepatocellular carcinoma cell lines. This overlap was later confirmed via CoIP experiments as PPI that regulate the expression of SLC17A9 and promote tumourigenesis (471). Vimentin was also found to overlap and interact with sodium/hydrogen exchangers protein 7 (NHE7) via double fluorophore immunofluorescence imaging, channel overlaps, and CoIP in focal complexes of migrating MDA-MB-231 breast cancer cells (472). However, no prior data on PRH-Vimentin immunofluorescence overlaps or spatial co-localisation was reported in the literature.

Unlike the previous studies that relied on fluorophore channel overlaps and CoIP to prove PPI and co-localisation, we relied only on qualitative analysis of confocal imaging to show that PRH and Vimentin might be found in close distance of each other. Although it was quite easy to perform the analysis when it comes to cytoplasmic localisation, this was a challenge when we

wanted to examine nuclear localisation which is attributed to the low nuclear Vimentin expression. Such struggle to visualise nuclear Vimentin was reported previously in immunofluorescence staining of SK-N-SH and GI-M-EN cells in a study conducted to investigate *p21^{Waf1}* in neuroblastoma (348). Like our nuclear PRH-Vimentin peripheral observation, the authors also reported Vimentin-Lamin A/C existing towards the inner side of the nuclear membrane (348). We also faced challenges to analyse nuclear PRH-Vimentin localisation in AKN1 cell expressing GFP-tagged PRH. This was due to the overall level of Vimentin in the cells. We attributed this to factors including Vimentin multistep organisation and reduced PRH expression in higher passages of these cells. To detect Vimentin in AKN1 cells, we relied on PRH upregulating the expression of Vimentin at protein levels as reported by Kitchen et al. (215). This probably affected the Vimentin we detected in AKN1 cells since we only had high passages of these cells by the time we conducted immunofluorescence staining and imaging. In addition, Vimentin expression is a complex process where non-filament Vimentins (ULF (318)) come together to make intermediate filaments (316), before the final filamentous cytoskeleton (317). In our cells, we managed to visualise dot-like Vimentin which is probably the simplest form of Vimentin. The alternative to using plasmid-generated AKN1 GFP-PRH-Myc cell lines for induced-Vimentin visualisation would be inducible PRH viral transduction. This method, when successful, ensures controllable high-level of PRH expression. It would probably avoid the impact of low PRH levels on Vimentin expression, which in turn would have aided the investigation of PRH-Vimentin localisation in the AKN1 cell line.

In our analytical approach, we coupled channels overlay with XYZ-orthogonal view and RGB profiler analyses. Utilising the combination of these tools may positively contribute to the spatial characterisation and adjacency of PRH and Vimentin proteins in CCLP1 cells and in AKN1 expressing GFP-tagged PRH cells. XYZ-orthogonal view allows a better visualisation of fluorophores arrangements and identify regions of potential PPI in subcellular context compared to single-plane images, but the tool on its own lack the necessary resolution for definitive spatial co-localisation and indication of actual PPI (450). The RGB profiler (line profiler) may support the previous tool by overlapping the intensity peaks of protein expression which again suggest spatial proximity of fluorophores (450). However, this tool lacks the ability to account for complex protein interaction and co-localisation on its own, or in situation where the cells highly express the proteins of interest (450). Overall, while we collected data that suggest the close spatial localisation of PRH and Vimentin, additional quantitative methods are required to further elaborate on the nature of PRH-Vimentin overlaps.

Unfortunately, we did not support our findings with quantitative measurement of PRH-Vimentin overlap using correlation analysis such as calculating Pearson's coefficient partly because we already planned for super-resolution imaging experiments. Acquiring Pearson's coefficient might have predicted whether PRH-Vimentin expressions and overlap abundance are related functionally (470). In addition, we would have benefited from live imaging studies to prove that PRH-Vimentin overlap findings could be more of protein

co-localisation and interaction in dynamic environment rather than fixed cells and static proteins.

In the literature, microscopic qualitative and quantitative studies of protein overlaps and spatial co-localisation are often accompanied by biochemical experiments such as CoIP and Western blotting to identify co-immunoprecipitated proteins and confirm PPI (471). Unfortunately, we did not manage to have that work done for this thesis. But we previously mentioned that our collaborator (Dr Lee) generated data from mass spectrometry and CoIP in her work on PRH and Vimentin as potential partner proteins (269). Dr Lee's work included using Adeno virus to overexpress Myc-PRH and identify Vimentin as an interacting protein and a partner for exogenous PRH. We both had the common aim of investigating PRH and Vimentin colocalisation and interaction in CCA cells. We could say that our work is complementary to Dr Lee's findings. For this thesis, we generated data that suggested endogenous PRH and Vimentin could be both found in very close locations within the subcellular compartments of CCLP1 cells. Whereas, our collaborators managed to confirm that Vimentin is a an interacting partner protein for overexpressed PRH in CCLP1 cells (269). Even the ChIPseq analysis where PRH was found to have a peak at Vimentin genes was performed after expressing Myc-PRH (269). In other words, even if we assumed that both parts could complement each other, there is a distinct difference in the contexts within which we performed our experiments. This might have an impact on the generated data and how they were interpreted. Analysing endogenous proteins in their natural expression and localisation within the cells reflects

their native physiological behaviour and takes into consideration any post-translational modifications that could occur for the proteins. In contrast, cellular context where we overexpress PRH for example may induce artifacts such as exogenous protein aggregations or inducing non-specific protein interactions that influence the overall interpretation of findings (473). The latter is more related to studying the effect of altering protein expression on proteins interactions and biological processes. We mentioned that our AKN1 cells expressing GFP-tagged PRH were suffering declined levels of PRH towards later passages. If PRH-Vimentin immunofluorescence investigation worked as we expected, we could have correlated the findings to Dr Lee's work since both relied on exogenous PRH. To address the issues between our work and the collaborators' work, we could execute further experiments to link all the findings together and effectively cross-analyse the overall results. The minimum work to consider at this stage are biochemical experimental validation (CoIP) of the immunofluorescence findings along with integrated bioinformatics analysis (ChIPseq).

We also addressed the organisation of nuclear PRH and Vimentin in CCLP1 cells at single molecule levels utilising the super-resolution imaging technique PALM dSTORM. We adopted our approach here from the developer of ClusDoC tool (412), as well as from the work of RNAPII and nuclear myosin VI co-localisation (413). Dual channel imaging with confocal microscopy for the purpose of spatial localisation, is amenable to misinterpretation. This is due to the limited resolution in addition to bleed-through issues during image acquisition. There are several articles that reported nuclear Vimentin

expression, e.g. with Lamin A/C (243), several DNA regions in mice (e.g. mtDNA – centromeric and telomeric regions) (474). They mainly used fluorescent imaging or biochemical methods to reach their conclusions.

Here we demonstrated for the first time that PRH and Vimentin are arranged spatially as nuclear clusters in CCLP1 cells. The details from nuclear cluster analysis support our findings from confocal imaging. PRH is widely abundant in CCLP1 nuclei compared to Vimentin. This resulted in numerous and big PRH clusters with high molecules to cluster ratio. Cluster characterisation of both PRH and Vimentin molecules and the visualisation of various-sized aggregates in CCLP1 cells is consistent with the high-order assembly and organisation of both PRH and Vimentin we detailed in the introduction. PRH protein has a dimerising motif within its N-terminal domain which drives its oligomerisation with other PRH proteins to produce several forms ranging from dimers to hexadecameric spheres (198). We also explained how Vimentin monomers form ULFs which in turn interlink laterally to form longer filaments (317).

We further revealed the similarities in distribution between PRH and RNAPII clusters around the euchromatin. Comparing Vimentin clusters to PRH, we found that Vimentin forms few and small aggregations with small molecules to cluster ratio. Quantitative cluster characterisation, using the tool ClusDoC, showed a small degree of bidirectional PRH-Vimentin cluster co-localisation. Like our previous findings from confocal immunofluorescence imaging, this co-localisation result requires further experimental confirmation before making a definitive conclusion. However, we can state that PALM dSTORM ClusDoC

analysis provided significant data suggesting the adjacency of PRH and Vimentin nuclear spatial localisation at nm scale. Taking the spatial proximity of PRH and Vimentin into account, we assumed that Vimentin may also exist near euchromatin. However, this must be properly addressed by specifically staining Vimentin and DNA to be visualised using the same approach we followed for PRH and DNA.

PRH-Vimentin nuclear close localisation, specifically when existing closer to the inner nuclear membrane, suggests a significant implication on PRH-Vimentin role in the context of chromatin organisation, gene regulation, and interplay with other inner nuclear membrane proteins in CCA. Several protein complexes, e.g. Lim2-Nur1, are localised within or closer to inner nuclear membrane and are implicated in heterochromatin gene silencing (464). Yet, functional gene reporter assay and ChIP experiments on common target genes are needed to investigate the functions of the PRH-Vimentin complex.

Taking all PALS dSTORM findings together, coupling qualitative and quantitative image analysis, i.e. cluster characterisation and ClusDoC analysis, refined our understanding of the spatial molecular localisation and arrangements of nuclear PRH and Vimentin. While our data suggested PRH and Vimentin occupy adjacent subnuclear spaces in CCLP1 cells with more Vimentin localised near PRH, these findings are might be limited or highly dynamic due to the stochastic nature of single molecule observation using PALM dSTORM (465). Despite the novel insight we revealed, we acknowledge the limitation of lacking PPI and co-localisation with biophysical techniques such as CoIP and live imaging.

To address the stand alone GFP expression in later passages of AKN1 GFP-PRH-Myc cells, we re-checked on plasmid integrity since it is most likely the reason behind the observation. While our previous plasmid restriction digest and plasmid sequencing data ensured the full-length transgene was inserted in the right site without a stop codon between GFP genes from the plasmid backbone or PRH gene from the transgene, we did not exclude the existence of alternative transcription starting sites or differentiated promoter activity. Both may produce GFP independently beside the complete fused protein. We also hypothesised that GFP cleavage might be another reason behind losing the expression in AKN1 GFP-PRH-Myc cells which were important for Vimentin expression and localisation studies. We principally followed the FCS approach, described by Goulding et al. (415), to study the dynamics of very low GFP fluorescence in our cells. We managed to trace 3 patterns of GFP forms: bright and diffuse, dimly diffuse, and very dim clusters. The first type is similar between AKN1 GFP-PRH-Myc and control cells which indicates that the cells are producing GFP in addition to the fused GFP-PRH in AKN1 GFP-PRH-Myc cells. The other two forms were seen only in AKN1 GFP-PRH-Myc cells. We complemented our fluctuation analysis with representative confocal images. We predicted that these GFP are representing monomer PRH. We would expect to see several fold-increase in fluorescence intensity to say that we detected oligomers since we assumed every induced PRH to be tagged by GFP, i.e. direct proportional relationship between the intensity of fluorescence and the size of oligomer. The overall detected GFP intensity in AKN1 GFP-PRH-Myc required increasing the gain during the imaging. In addition, we did

not need to include a pre-bleach step for GFP-PRH dynamic recording. In addition, the amplitude of fluctuation in fluorescence was quite low for GFP-PRH between 0-140 kHz. All three points suggest the detection of monomer GFP-tagged PRH proteins. Several studies have used FCS to correlate the level of fluorescence to the size of the detected protein. For instance, Herrick-Davis et al. reported that the induced receptor 5-HT2C-GFP in HEK293 cells exist predominantly as dimers as they compared its level of GFP brightness to HEK293 cells co-expressed with 5-HT2C-GFP and un-tagged 5-HT2C (475). Collectively, the data are consistent with our observations that GFP on its own in AKN1 cells is abundant, stable, and persists longer compared to GFP-tagged PRH despite the mixed population. They are also consistent with the difficulties we faced to detect GFP-PRH-Myc expression in later passages of the cells, or even see their effect on phenotype as we reported in the previous chapter. This could also explain the low levels of PRH-Vimentin proximity overlap using XYZ-orthogonal views and RGB profiler in AKN1 GFP-PRH-Myc cells. In addition to the loss of visible GFP and GFP-PRH-Myc expression and loss of apparent mesenchymal phenotype, FCS revealed findings suggested standalone GFP production monomers of GFP-PRH-Myc. These findings supported the spiculations of AKN1 GFP-PRH-Myc experiencing cleavage of the fused protein GFP-PRH-Myc producing PRH fragments. Such observation has been reported before to produce defective by-products of ribosomal proteins tagged with GFP (DRiPs) due to auto fragmentation which interfered with the overall experimental plan (476). Other explanations for fragmented GFP-protein side products include enzymatic recognition of specific peptide

sequences between GFP and the protein, GFP-protein misfolding, post-translational modification of the tagged protein, or cell culture environmental stress which may activate degradation pathways.

Taking FCS data together, we followed the standard settings for best model fitting and trace analysis in our FCS experiment. However, we could benefit from repeating the analysis after optimising the settings. Specifically, recording free GFP pattern in more AKN1 GFP cells is needed to set a proper baseline for GFP to confidently report on GFP typical speed and brightness.

The most obvious limitation in our approach in this chapter is the number of cell lines and cells included during the analysis although with a super-resolution technique an n=14 is considered acceptable for cell number (Chris Toseland personal communication). Additional cell lines plus a bigger number of cells for each cell line would have provided a greater insight onto the localisation. The use of images and immunoblotting experiments, in which the cells are co-probed for PRH and Vimentin and subcellularly fractionated, would have been a good addition to simple image analyses like the ones we performed for our confocal images. That would have provided additional information to support our findings.

Considering the previous altogether, we provided primary evidence on PRH-Vimentin spatial adjacent localisation using two levels of resolution in microscopic imaging: confocal and PALM dSTORM imaging. PALM dSTORM of PRH with SiR DNA raises the consideration for the future need to replicate the approach with mutant forms of PRH (e.g. PRH-N187A). This construct encodes a mutated homeodomain that was previously shown to have less

Vimentin interaction. Alternatively, PALM dSTORM could also be performed in cells with knocked down PRH or Vimentin. Both options would reveal the influence of having less PRH-Vimentin interaction on the level of spatial localisation. In addition, we would also benefit from expanding the previous plan to include imaging Vimentin and SiR DNA to support our suggestions that Vimentin in the nucleus is distributed adjacent to the chromatin. This would agree with previous report on Vimentin being capable of binding to DNA to regulate gene expression (348). Including future work to study the perturbation of PRH and Vimentin in relation to the chromatin could reveal a potential behaviour mimicking pioneer transcription factors. Furthermore, knowing exactly how Vimentin is being trafficked into the nucleus in CCA cells would complement and further support our current findings. FRAP imaging of fluorescent-tagged Vimentin and monitoring the recovery of emission overtime can provide insights into the cytoplasmic-nuclear movements of Vimentin (446). Live cell imaging of fluorescent-tagged Vimentin is also another experiment to trace the translocation of Vimentin in the cells. Additionally, nuclear import assay using permeabilising agents like Digitonin can help track the movement of fluorescent-tagged Vimentin (477).

To advance the investigation of PRH-Vimentin localisation into PPI, we mentioned earlier the benefits from including CoIP experiments. In addition, FRET imaging to calculate energy transfer between two close molecules within 1-10 nm further support molecules interaction (447). Moreover, proximity ligation assays can also prove protein interaction by detecting a single emitted fluorescence signal generated from two neighbouring

molecules that are more likely interacting structures (478). Lastly, *in silico* approaches to study PPI coupled with immunoprecipitation experiments of mutated forms of PRH and Vimentin could reveal the exact a.a. sequences required for the binding. Protein-protein molecular docking is a computational method where the binding interface is predicted, simulated artificially, and scored relying on protein sequences to show the bound a.a. residues (479). This could be performed between the homeodomain of PRH and head domain of Vimentin.

Chapter 5: The investigation of gene expression co-regulation by Vimentin and PRH in CCA cells

5.1 Introduction

In this project, we have provided data attempting to expand our current knowledge of the role of PRH in CCA. In the previous chapter, we investigated the spatial localisation of PRH with one of its protein partners, Vimentin. We established that Vimentin in CCA cells can exist spatially within the nuclear region in clusters that might be found near PRH clusters. Such a combination of characteristics for Vimentin, along with the extensive roles Vimentin undertakes in the cytoplasm, offers an interesting axis for investigation.

Previously, Dr Ka Ying Lee reported in her thesis that both Vimentin and PRH promoted cell proliferation in CCA cells (269). Dr Lee also revealed using RT-qPCR experiments that changes in Vimentin expression induced changes in the expression of important cell cycle genes that were also known to be regulated by PRH via DNA binding e.g. *CCND2* and *CDKN1B* (269). Here, we extend upon the preliminary work conducted by Dr Lee through employing RNAseq analysis incorporated with published ChIPseq data to investigate hallmarks and pathways co-regulated by Vimentin and PRH in CCA. Our approach also offers an insight into a set of genes harbouring a high propensity of being regulated by Vimentin through DNA binding. Thus, our work is a significant contribution to the frontier of nuclear Vimentin and set the ground for future Vimentin ChIP and ChIPseq investigations.

In brief, we depleted Vimentin in CCA cell lines (CCLP1 and CCSW1) using si*VIM5* reverse transfection. Both cell lines were used to conduct the work that introduced PRH as an oncoprotein in CCA (215). After ensuring a satisfactory depletion of Vimentin at mRNA and protein levels, we investigated the expression of PRH at mRNA and protein levels. We characterised the phenotype of Vimentin knockdown cells using EdU incorporation assays for cell proliferation, brightfield and confocal imaging for morphological changes, and transwell cell migration (chemotaxis) assays for cell migration. Then, we proceeded to explore Vimentin's capabilities in gene regulation by RNAseq, GO, and GSEA. In addition, we conducted several overlap investigations with published RNAseq data from PRH knockdown and overexpression in CCA cell lines to explore the possibility of Vimentin-PRH gene expression co-regulation in CCA. Moreover, we revealed the functional involvement of the co-regulated genes in cancer-related hallmarks and gene sets. Finally, we used RT-qPCR to validate interesting genes we highlighted from RNAseq analysis. Employing cellular phenotypic investigations integrated with bioinformatics analyses can delineate the roles Vimentin exhibits in CCA.

5.2 Results

5.2.1 *VIM* and *HHEX* genes are highly expressed in CCA

To investigate how Vimentin and PRH are expressed in CCA tumours at the mRNA level, we used an online platform for Gene Expression Profiling Interactive Analysis (GEPIA2) developed by the Zhang lab (1, 2). The server uses public RNA sequencing data of 9,736 tumours and 8,587 normal samples to give an overview of gene expression profiling as well to apply different analytical tests including gene correlation analysis. For gene expression purpose, GEPIA2 uses a standard pipeline to match data from The Cancer Genome Atlas (TCGA) and the genotype-tissue expression project (GTEx). We input *VIM* and *HHEX* independently to plot their expressions across multiple cancers datasets. ANOVA test was used to find the differential expression of both genes in tumour vs. paired normal samples with Log_2(FC) cut-off value equal to 1 and *q*-value cut-off equal to 0.01. PRH and Vimentin genes were considered overexpressed if the Log_2(FC) was more than 1 and *q*-value was less than 0.01.

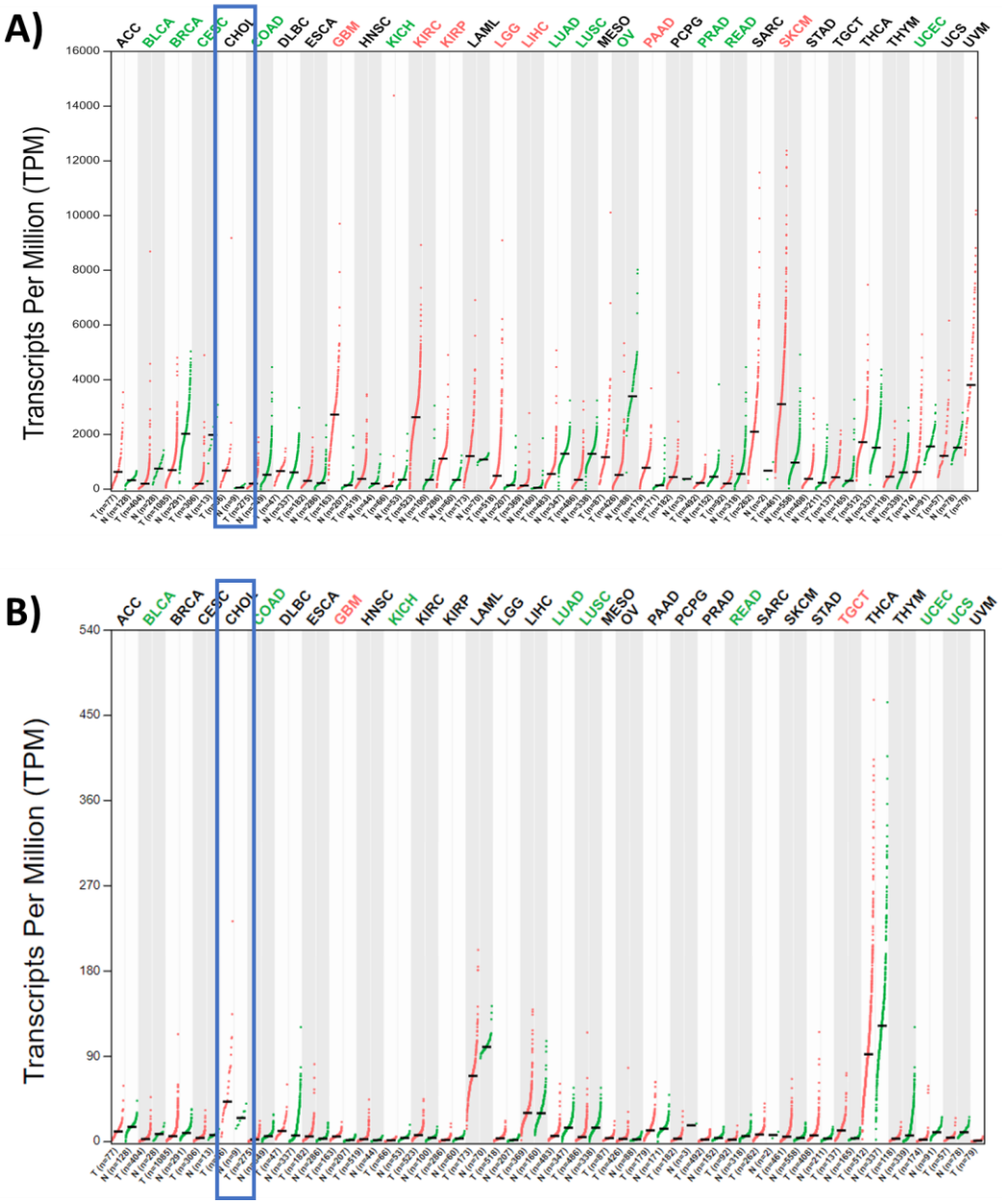


Figure 5-1 Dot plot gene expression profiling of *VIM* and *HHEX* displayed as transcripts per million in different cancer (red) and normal tissues (green).

(A) Blue box: *VIM* expression across 36 CCA tissues compared to 9 normal tissues. (B) Blue box: *HHEX* expression across 36 CCA tissues compared to 9 normal tissues. (1, 2).

The dot plot (figure 5-1 a & b = top *VIM* and bottom *HHEX*) above represented the tissue-wise genes expression of *HHEX* and *VIM* in n=36 tumour and n=9 normal samples reported as transcripts per million (TPM). Inside the blue boxes, both genes were found highly expressed (red dots) in CCA compared

to their control normal samples (green dots) with each dot representing an input cancer or normal samples.

To statistically compare the averages of both *VIM* and *HHEX* individually between CCA and normal samples, we used OncoDB developed by Tang et al. (480). This tool allows the analysis of RNAseq, TCGA, and clinical data to explore gene expression in cancer. The box plot (figure 5-2-A) shows the student's t-test analysis where the average *VIM* (710.7) in CCA tumour samples is significantly higher than average *VIM* in normal tissues (72.7) with Log_2(FC) of 3.74 (*p*= 3.3e-04). Similarly, the average *HHEX* expression (111.1) in CCA patients is significantly more than the average of normal samples (51.9) with Log_2(FC) of 0.8 (*p*= 9.2e-05) (figure 5-2–B).

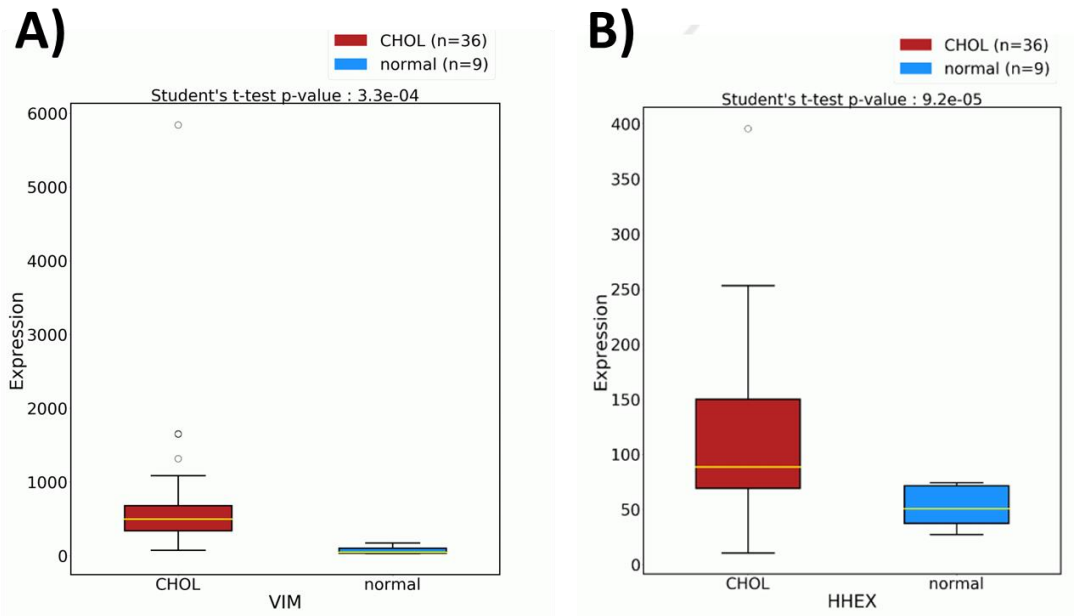


Figure 5-2 Box plot indicating the high expression of *VIM* and *HHEX* in 36 cancer tissues compared to 9 normal samples.

(A) *VIM* expression is significantly higher in CCA tissues compared to normal samples (*p*= 3.3e-04). (B) *HHEX* expression is significantly higher in CCA tissues compared to normal samples (*p*= 9.2e-05). Statistical analysis was done using two-tailed student t-test. Error bars represent SD.

To see if *VIM* and *PRH*/*HHEX* genes were clinically correlated, we used GEPIA2 to perform a pair-wise gene expression analysis of *VIM* and *PRH*/*HHEX* in CCA and using Spearman correlation (figure 5-3).

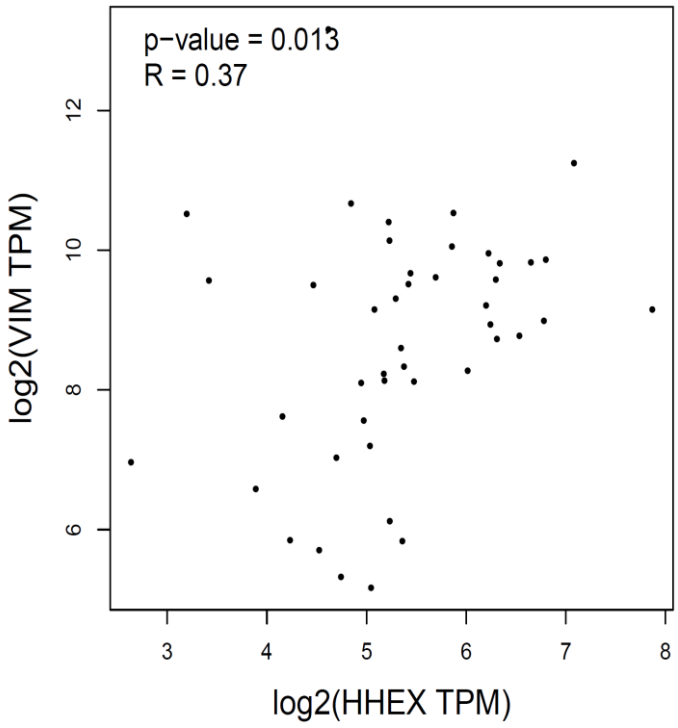


Figure 5-3 *VIM* expression correlates with *HHEX* in CCA

36 cancer tissues compared to 9 normal samples. p= 0.013.

The data indicated that *VIM* and *HHEX* expressions are statistically correlated in CCA (*p*=0.013). Altogether, GEPIA2 and OncoDB provided evidence that both *VIM* and *HHEX* were significantly highly expressed in CCA, and that they were significantly correlated.

5.2.2 Characterising Vimentin depletion in CCA cell lines

5.2.2.1 Efficient Vimentin knockdown in CCA cells downregulates the expression of PRH

To deplete Vimentin in CCLP1 and CCSW1 cell lines, we used siRNA AllStar negative control (1027418, Qiagen) and si*VIM5* (Hs_VIM_5 FlexiTube siRNA, NM_003380, Qiagen) to generate CCLP1 and CCSW1 control and siVimentin cells. We followed the protocol detailed in the method chapter. After 48 hours of transfection, we collected cell pellets for RT-qPCR and Western blotting Vimentin expression validation. Both experiments were performed in three independent biological repeats.

RT-qPCR results show that we had a significant reduction in Vimentin mRNA (over 95%) in both cell lines 48 hours post transfection; CCLP siVimentin (p <0.0001) and CCSW siVimentin (p= 0.0001) (figure 5-4).

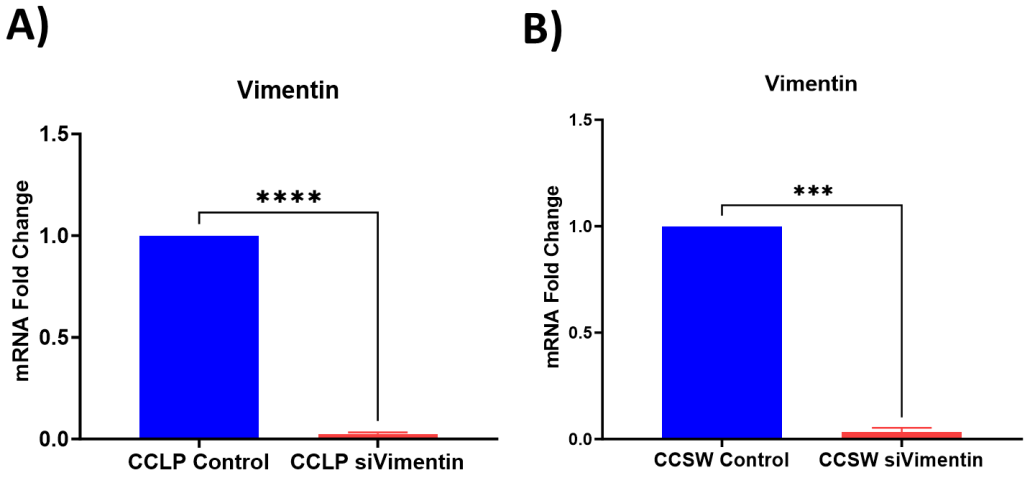


Figure 5-4 Vimentin mRNA fold change using RT-qPCR after si*VIM5* transfection.

(A) Vimentin mRNA relative fold expression showing Vimentin mRNA significantly reduced in CCLP siVimentin compared to control cells (p<0.0001). (B) Vimentin mRNA relative fold expression showing Vimentin mRNA significantly reduced in CCSW siVimentin compared to control cells (p= 0.0001). Both RT-qPCR were performed in N=3 biological repeats and

normalised to 1 for comparison. Two-tailed unpaired student t-test analysis was performed [P ≤ 0.05 (*), P ≤ 0.01 (**), P ≤ 0.001 (***)]. Error bars represented SD.

Similarly, there was also a significant loss of Vimentin expression when examined by Western blotting at protein level. We needed to load low protein lysate (10 µg) of control and siVimentin samples for SDS-PAGE to visualise Vimentin knockdown. The percentage of Vimentin loss was around 93% for CCLP siVimentin (*p*= 0.0011) and 87% for CCSW siVimentin (*p*= 0.0051) (figure 5-5).

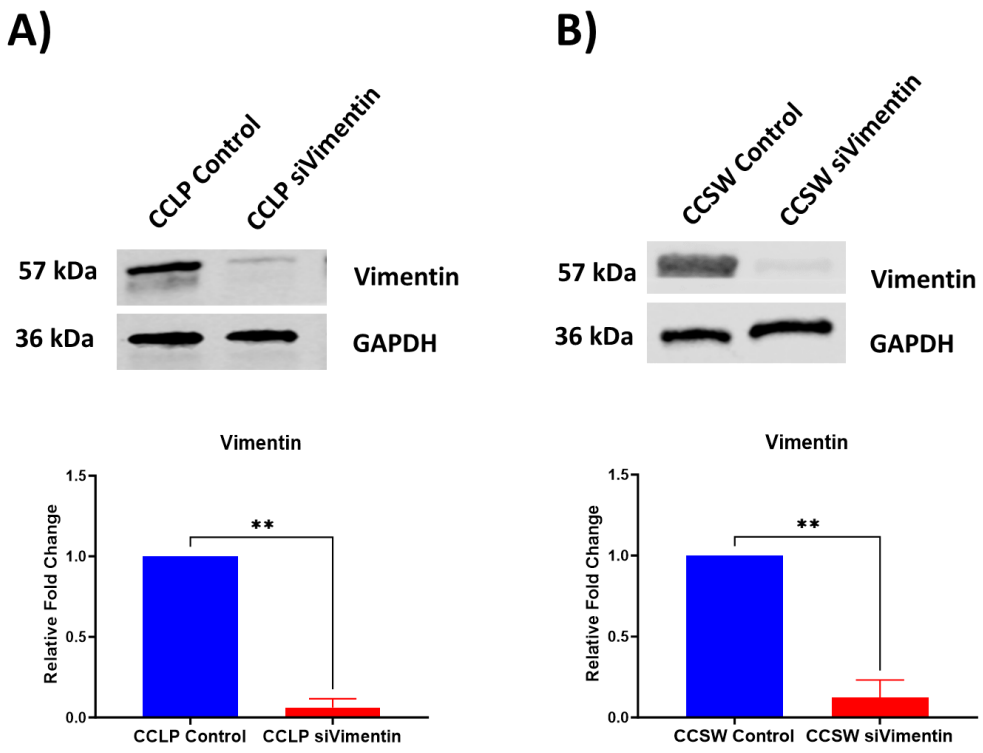


Figure 5-5 Determination of Vimentin relative fold change using Western blotting(48 hours after *siVIM* transfection).

(A) A representative Western blotting image of CCLP control and siVimentin whole cell protein extracts probed with a Vimentin antibody with densitometry analysis showing a significant Vimentin knockdown in CCLP siVimentin cells (*p*= 0.0011). (B) A representative Western blotting image of CCSW control and siVimentin whole cell protein extracts probed with a Vimentin antibody with densitometry analysis showing a significant Vimentin knockdown in CCSW siVimentin cells (*p*= 0.0051). GAPDH was used to show equal loading. Both Western

blottings were performed in N=3 independent biological repeats and normalised to 1 for comparison. Two-tailed unpaired student t-test analysis was performed [P ≤ 0.05 (*), P ≤ 0.01 (**), P ≤ 0.001 (***)]. Error bars represented SD.

Both CCLP1 and CCSW1 cell lines have elevated levels of PRH at the mRNA and protein levels (215). Since PRH loss in both cell lines was associated with a downregulation of Vimentin expression (215), we investigated the effect of Vimentin loss on PRH expression utilising the same cDNA from the siVimentin RT-qPCR verification above. After 48 hours of siVIM5 transfection, there was around half a fold reduction in PRH mRNA compared to control cDNA (CCLP siVimentin p= 0.0246, CCSW siVimentin p=0.0271) (figure 5-6).

To see whether we detect a change in PRH expression at protein levels, we stripped the same membranes from (figure 5-5) and re-probed them with antibodies targeting PRH. There was no significant change in PRH expression (CCLP siVimentin p= 0. 9248) and (CCSW siVimentin p= 0. 8498) (figure 5-6).

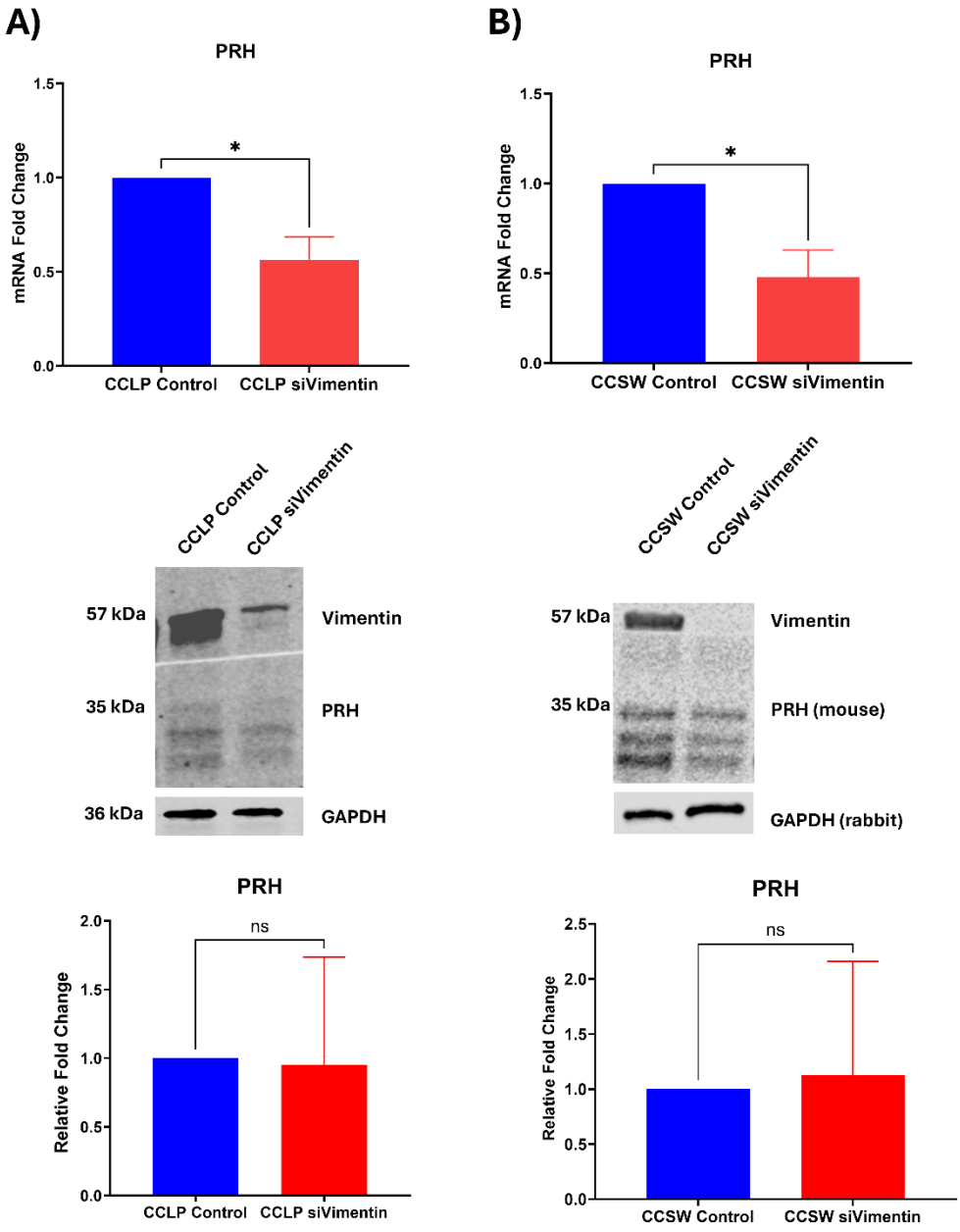


Figure 5-6 PRH mRNA and protein expression after knocking down Vimentin.

(A) PRH mRNA relative fold expression showing PRH mRNA significantly reduced in CCLP siVimentin compared to control cells (p=0.0246). A representative Western blotting image of CCLP control and siVimentin whole cell protein extracts re-probed with a PRH antibody with densitometry analysis showing no PRH change in CCLP siVimentin cells (p= 0.9248). (B) PRH mRNA relative fold expression showing PRH mRNA significantly reduced in CCSW siVimentin compared to control cells (p= 0.0271). A representative Western blotting image of CCSW control and siVimentin whole cell protein extracts re-probed with a PRH antibody with densitometry analysis showing no change in PRH expression in CCSW siVimentin cells (p= 0.8498). GAPDH was used to show equal loading. Both RT-qPCR and Western blotting were performed in N=3 biological repeats and normalised to 1 for comparison. Two-tailed unpaired student t-test analysis was performed [P ≤0.05 (*), P ≤ 0.01 (**), P ≤ 0.001 (***)]. Error bars represented SD.

Since verifying PRH protein levels using Western blotting gave non-significant results, we decided to re-check the change in PRH protein expression relying on immunofluorescence imaging of the cells (data available in section 5.2.2.3 Vimentin knockdown in CCA cells induced cellular morphological changes – figure 5-9).

Taking all the data together, we show that an efficient Vimentin knockdown can be achieved at the mRNA and protein levels after 48 hours of si*VIM5* transfection. Vimentin knockdown was associated with a significant reduction of PRH mRNA, but not PRH protein.

5.2.2.2 Vimentin knockdown in CCA cells reduces cell proliferation

To examine the effect of Vimentin depletion on cell proliferation in CCLP1 and CCSW1 cell lines, we used EdU incorporation assay as described in the method section. We used the same si*VIM5* transfections from the previous section to setup the experiments for cell proliferation. We show here the average of two independent biological repeats for CCLP siVimentin and three for CCSW siVimentin (figure 5-7). The results indicated that Vimentin loss significantly decreased cell proliferation in both cell lines. After 2 hours of EdU incubation, there was a significant reduction in cell proliferation by more than 70% in Vimentin knockdown cells compared to control cells (CCLP siVimentin p = 0.0232, and CCSW siVimentin p=0.0175).

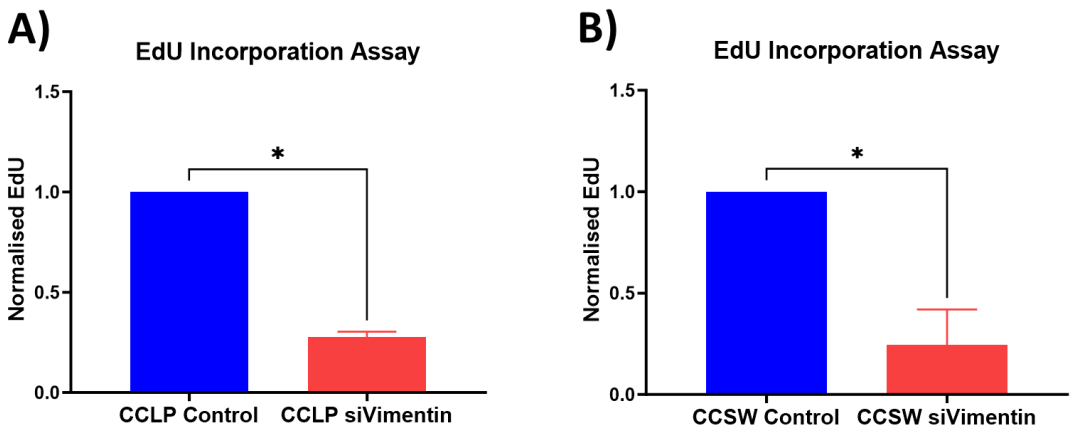


Figure 5-7 Vimentin loss in CCA cells significantly decreases cell proliferation.

(A) EdU incorporation assay shows significantly less proliferative activity in CCLP siVimentin cells (N=2 biological repeats, *p*= 0.0232, and error bars represent SEM) in comparison to control cells. (B) EdU incorporation assay also shows significantly less proliferative activity in CCSW siVimentin cells compared to control cells (N=3 biological repeats, *p*= 0.0175, and error bars represent SD). Cells were incubated with EdU for 2 hours and the Click-IT reaction was detected by measuring the absorbance of the colour produced from the addition of an OPD substrate for HRP-conjugated antibody at 450 nm. Correction to cell number was achieved by crystal violet normalisation. Absorbance was normalised to cell number and then to 1 for comparison. Statistical analysis was done using unpaired two-tailed student t-test.

We concluded that loss of Vimentin expression lowers cell proliferation in CCA cell lines.

5.2.2.3 Vimentin knockdown in CCA cells induces cellular morphological changes

Because Vimentin is a cytoskeletal protein that has a role in maintaining the shape of the cells (334, 353-355), we expected to see a change in cell morphology upon knocking down Vimentin in CCA cells. Cell morphology has an impact on its phenotype and functional biology.

We initially inspected the cells using brightfield microscopy. In either cell line, the control cells appeared elongated and mesenchymal. Despite the high confluency, we can still see the cells keeping a small distance from each other,

i.e. less cell-cell cohesion. In the transfected population, we observed cells that were small and polygonal gathering in groups (red arrows in figure 5-8) in addition to the mesenchymal cells.

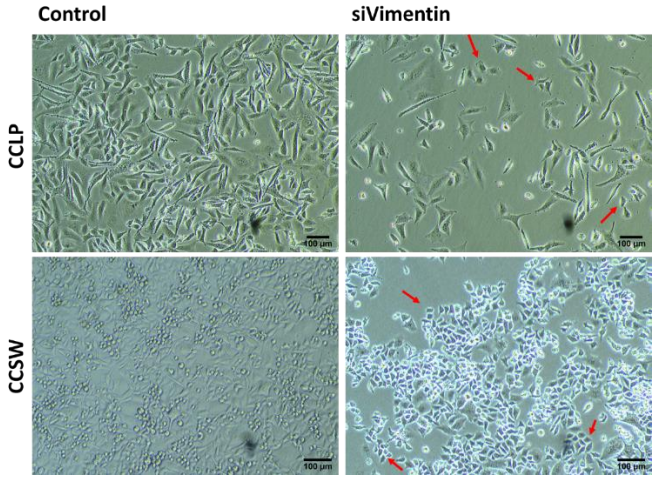


Figure 5-8 Brightfield images of the control and Vimentin knockdown cells.
Images were acquired using a Leica DMI3000 B fluorescence microscope. Scale bars 100 µm.

To have a better look into the impact of Vimentin knockdown on cell morphology, we used one of the transfections in section (5.2.2.1) to seed for immunofluorescence staining. We used a confocal Leica DMI4000B microscope to image fixed control and siVimentin cells (CCLP1 and CCSW1) probed with Vimentin and PRH antibodies as detailed in section (2.3.2). We maintained the same settings for laser excitation power, exposure, gain, pinhole size, and magnification. We also included images of negative controls without primary antibodies. For image analysis, we pre-processed the images to eliminate background noise. Then, we threshold the images choosing our regions of interests over the background before we proceeded to intensity measurement and shape analysis. The analysis was performed using FIJI (7).

As seen in (figure 5-9-A), our images included DAPI in blue channel, PRH in red channel, Vimentin in green channel, and a final merged image. We used Vimentin expression to assess cell morphology. In general, the change in shape is consistent with the results from the previous brightfield imaging. With Vimentin knockdown the cells transformed from elongated mesenchymal cells to small and round cells. In the control cells, Vimentin (green) assumed a filamentous arrangement throughout the cells portraying the classical cytoskeleton, while PRH (red) was distributed in speckles all over the cells.

In CCLP siVimentin cells, Vimentin (green) appeared less bright and less filamentous compared to control cells. Vimentin knockdown did not seem to affect PRH (red) brightness on visual inspection.

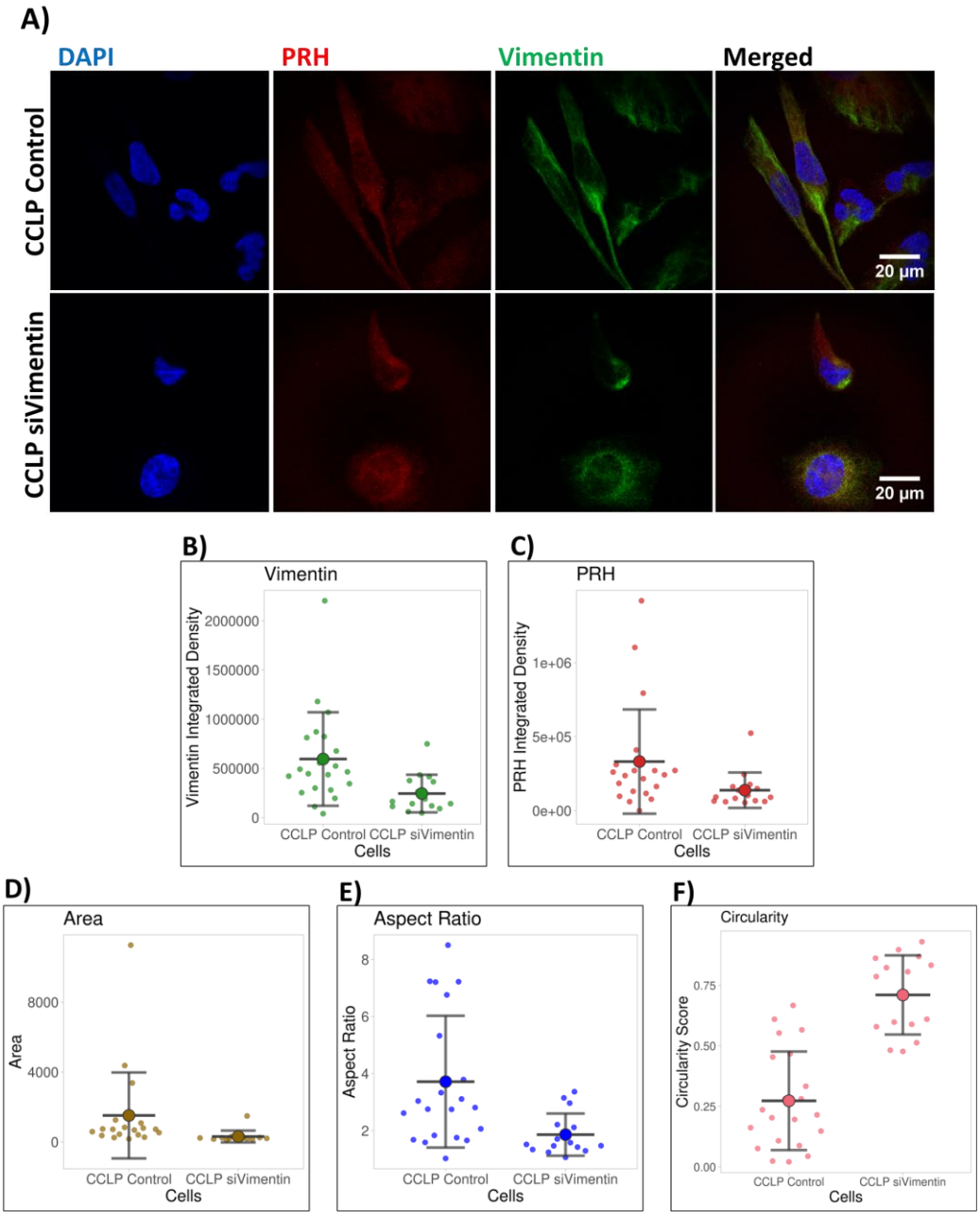


Figure 5-9 Morphological analysis of CCLP siVimentin cells using immunofluorescence imaging.

(A) CCLP siVimentin cells showed an epithelial-like morphology with less Vimentin visualised in comparison to control cells. Cells were fixed with 4% FA and labelled with PRH and Vimentin antibodies along with nuclear DAPI staining. Images were acquired sequentially for n= 21 CCLP control cells and n=15 CCLP siVimentin cells using a confocal Leica DMI4000B microscope at 63X magnification. Scale bars 20 µm. Laser power, gain, exposure, and pinhole were the same throughout the imaging. Expression intensities represented in violin plots showed significant reduction in expressions of (B) Vimentin (p= 0.004) and (C) PRH (p= 0.027) in CCLP siVimentin cells compared to control. Shape descriptor scores showed CCLP siVimentin (D) cells significantly smaller (p= 0.03) (E) less elongated (p= 0.002), and (F) more circular (p= 3.00e-08) compared to control CCLP cells. Ordinary t-test analysis was performed, and error bars represent SD.

To statistically confirm the observations above, we measured the intensity of Vimentin and PRH expression and analysed the cells morphologically using different shape description scores in FIJI **(450)**. We used integrated density instead of the mean grey value to detect Vimentin and PRH expressions due to the striking difference in cell size (area) after Vimentin depletion. Mean grey intensity values are influenced by the area of the examined regions. We performed the analysis on n=21 CCLP control cells and n=15 CCLP siVimentin cells. The data indicated a significant loss in Vimentin staining (p= 0.004) (figure 5-9-B) accompanied by a significant decrease in PRH staining (p= 0.027) (figure 5-9-C) in CCLP siVimentin cells compared to control cells. To statistically analyse the change in morphology after Vimentin depletion, we measured cell area, circularity and aspect ratio in FIJI (450). FIJI has all the three parameters under “set measurement”. Cells area was measured in square micrometre. Cell circularity was calculated based on the formula [circularity=4π (area/perimeter^2)] where a value of 1.0 means a perfect circle and values closer to 0.0 means there is an elongation in shape. Aspect ratio is based on the formula [Aspect ratio = Long axis / Short axis] where both axes are defined according to Feret’s diameters, i.e. longest and shortest values. Our results indicated that, Vimentin knockdown produced significantly smaller cells (p= 0.03) (figure 5-9-D), less elongated cells (p= 0.002) (figure 5-9-E), and more circular cells (p= 3.00e-08) (figure 5-9-F) compared to the control population.

The findings from CCSW siVimentin immunofluorescence images were also consistent with the findings from CCLP siVimentin immunofluorescence

imaging (figure 5-10-A). On visual inspection, Vimentin loss induced the production of small and round cells compared to the mesenchymal morphology of their control cells. In addition, the extent of Vimentin (green) filaments was less than the control cells. PRH again did not seem to change in appearance but showed a reduction in staining density.

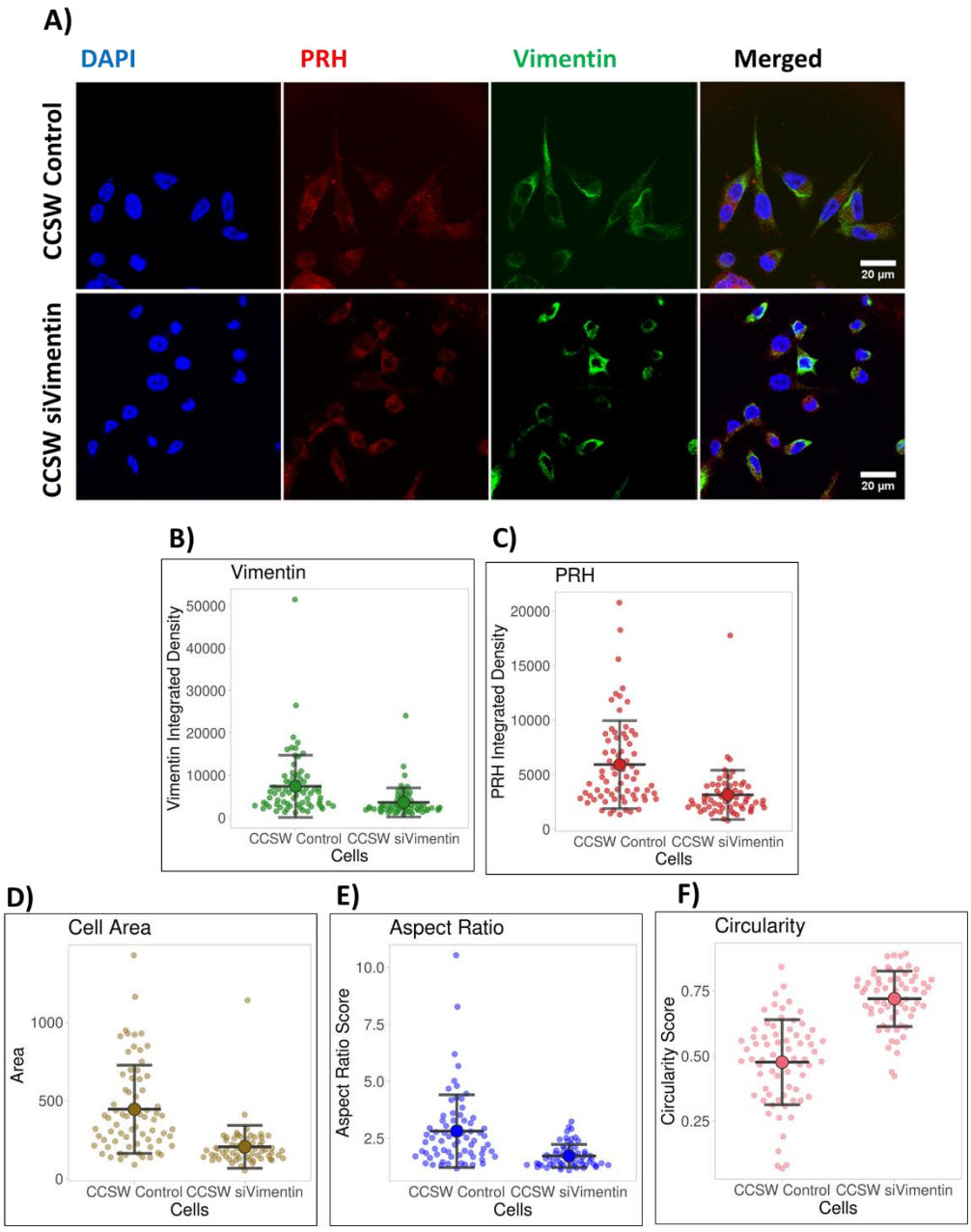


Figure 5-10 Morphological analysis of CCSW siVimentin cells using immunofluorescent imaging.

(A) CCSW siVimentin cells images showed an epithelial-like morphology with less PRH and Vimentin visualised in comparison to control cells. Cells were fixed with 4% FA and labelled with PRH and Vimentin antibodies along with nuclear DAPI staining. Images were acquired sequentially for n= 70 CCSW control cells and n=63 CCSW siVimentin cells using a confocal Leica DMI4000B microscope at 63X magnification. Scale bars 20 µm. Laser power, gain, exposure, and pinhole were the same throughout the imaging. Expression intensities represented in violin plots showed significant reduction in expressions of (B) Vimentin (*p*= 1.85e-04) and (C) PRH (*p*= 2.54e-06) in CCSW siVimentin cells compared to control. Shape descriptor scores showed CCSW siVimentin (D) cells significantly smaller (*p*= 1.00e-08) (E) less mesenchymal (*p*= 5.70e-07), and (F) more circular (p<1e-10) compared to control CCSW cells. Ordinary t-test analysis was performed, and error bars represent SD.

We performed the statistical analysis on n=70 CCSW control cells and n=63 CCSW siVimentin cells. Intensity of expression analysis for Vimentin and PRH integrated density indicated a significant decrease in both Vimentin expression (*p*=1.85e-04) and PRH expression (*p*=2.54e-06) in CCSW siVimentin cells (figure 5-10-B and C). Morphology analysis revealed that on average we had significantly smaller cells (*p*= 1.00e-08) (figure 5-10-D), less elongated cells (*p*= 5.70e-07) (figure 5-10-E), and more circular cells (*p*<1e-10) (figure 5-10-F) in CCSW siVimentin cells compared to the control population.

The level of Vimentin depletion across the analyses of Western blotting, RT-qPCR, and immunofluorescence imaging revealed a consistent trend in Vimentin expression after si*VIM5* transfection. In addition, this trend was also consistent in both CCLP1 and CCSW1 cell lines. The magnitude of Vimentin depletion in transfected cells was robust in comparison to control samples demonstrating a striking Vimentin reduction in Western blotting and RT-qPCR results. In contrast, visualising the cells using immunofluorescence revealed various levels of Vimentin knockdown. Vimentin cytoskeleton was still detected in most of the cells, yet the level of reduction was statistically significant as indicated from intensity measurements. Morphology analyses in both cell lines revealed consistent changes in cell shapes and areas. Such

observations are expected from manipulating the expression of a major cytoskeletal protein like Vimentin. Altogether, the data indicated that Vimentin depletion induced a change of morphology in CCA cells

5.2.2.4 Vimentin knockdown in CCA cells decreases transwell cell migration (chemotaxis)

To investigate whether the change in morphology of CCLP siVimentin and CCSW siVimentin cells altered cell migration ability, we performed transwell migration (chemotaxis) assays. Vimentin was extensively studied in cell migration before (378, 379). We expected Vimentin loss to impact the ability of CCA cells to migrate. We performed this assay in three independent biological repeats using cells from the transfections mentioned in section (5.2.2.1) We followed the protocol detailed in the method chapter. We estimated the number of migrated cells from a standard curve created from serial dilution of cells labelled with CalceinAM. We used the following equation to find the percentage of migration:

$$Migration\ \% = \frac{number\ of\ cells\ on\ basal\ side}{number\ of\ seeded\ cells} \times 100$$

Our results indicates that Vimentin loss in CCLP1 and CCSW1 led to a significant drop in the migratory activity of the cells, (CCLP p= 0.0066, CCSW siVimentin p= 0.0270) compared to control cells (figure 5-11).

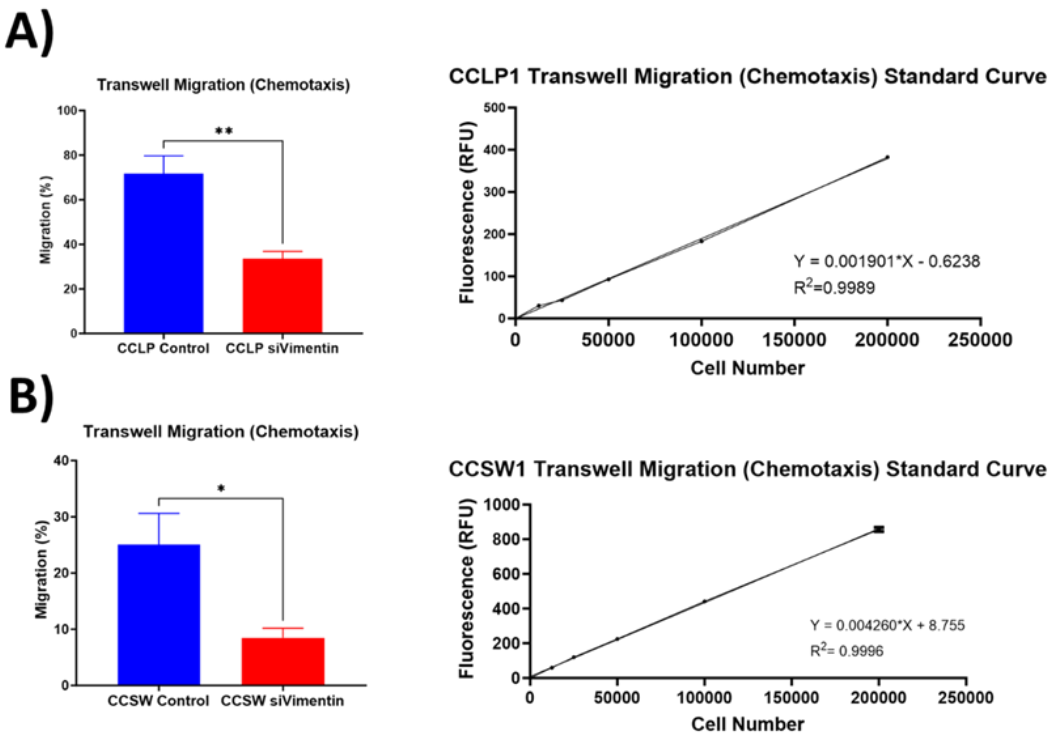


Figure 5-11 Vimentin knockdown significantly decreases cell migration in CCA cells.

(A) A significant reduction in CCLP siVimentin cells migratory activity (*p*= 0.0066) in comparison to control cells. (B) A significant reduction in CCSW siVimentin cells migratory activity (*p*= 0.0270) in comparison to control cells. $1X10^5$ cells were seeded on a transwell insert and left to migrate under the influence of a chemo-gradient (2%-10%) FBS across a porous membrane for 48 hours in the presence of 1mM antiproliferative Hydroxyurea throughout the assay. Cells on basal side of the inserts (migrated cells) were labelled with 8 µM Calcein AM. Percentage of migrated cells were interpolated from a standard curve generated by using a serial dilution of either CCLP1 or CCSW1 cells stained with 8 µM Calcein AM. The experiments were performed as N=3 independent biological repeats and analysed using a two-tailed unpaired student t-test. Error bars represented SD.

Taking the findings together we concluded that when Vimentin is lost, the cells become less migratory.

5.2.3 Vimentin Knockdown RNAseq and enrichment analysis in CCA cell lines

5.2.3.1 High-quality total RNA sequencing

After we successfully reached satisfactory levels of Vimentin depletion in CCA cell lines and observed phenotypes consistent with what is commonly reported

in the literature e.g. a reduction in cell proliferation, change in cell morphology, and decreased cell migration, we proceeded to investigate whether there is a correlated change in gene expression. We isolated RNA from two independent Vimentin si*VIM5* transfections in CCLP1 and CCSW1 cell lines following the protocol from the method chapter. To ensure we had high quality RNA, we sent samples of RNA to Deep Seq Next Generation Sequencing facility at the university of Nottingham for quality control check-up using Agilent TapeStation 4200.

Table 5-1 RNAseq samples prepared for pre-sequencing quality check-up.

QC report labels	Sample details
C1	G316 CCLP Control
D1	G316 CCLP siVimentin
E1	G317 CCLP Control
F1	G317 CCLP siVimentin
G1	G313 CCSW Control
H1	G313 CCSW siVimentin
A2	G314 CCSW Control
B2	G314 CCSW siVimentin

The main part of quality check-up was to obtain an RNA integrity number (RIN) which is calculated by finding the ratio of 28S rRNA to 18S rRNA. To say that we had a high-quality RNA, we needed the RIN score to be above 9. Samples with yellow marks (figure 5-12) indicated high RNA concentration outside the requested range by the facility. These samples still gave high RIN scores.

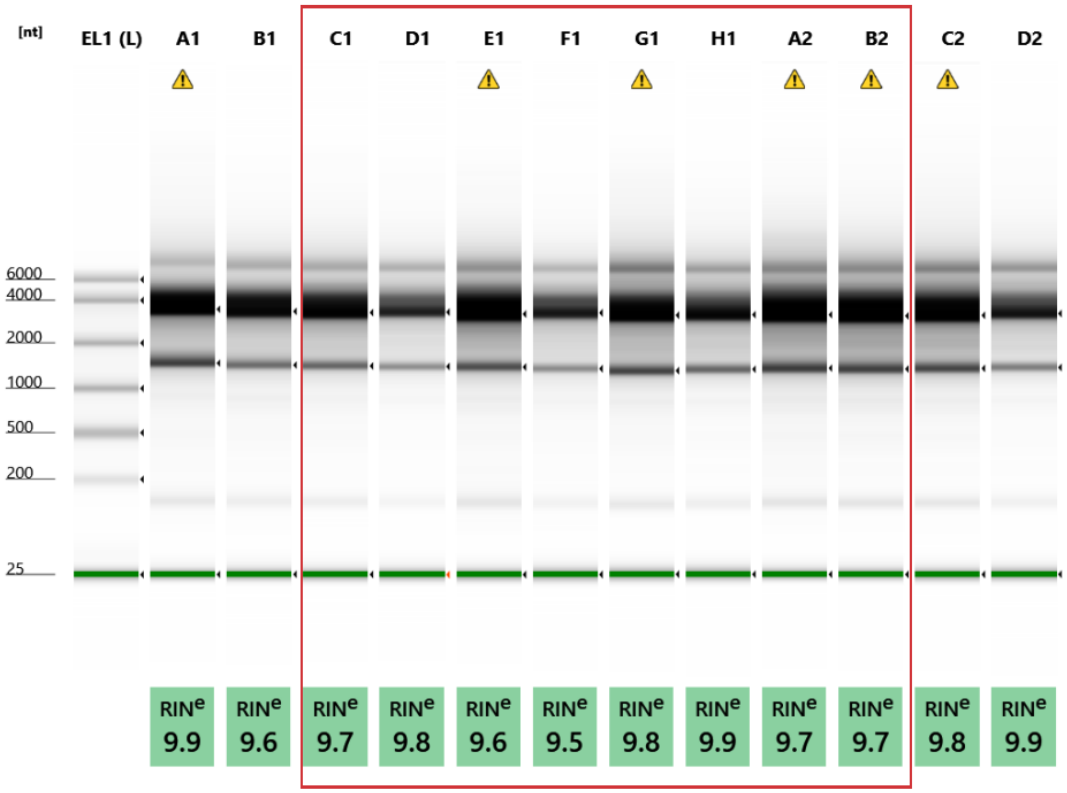


Figure 5-12 RNA samples integrity was assessed using gel electrophoresis.

5.2.3.2 Illumina RNA sequencing using NovaSeq PE150

We prepared N=2 independent biological repeats of CCLP siVimentin and CCSW siVimentin RNAs with their respective controls. Sequencing was performed by Illumina for mRNA using NovaSeq PE150 platform to build a poly A enriched RNA library and generate 5 G of raw sequences per sample. Sequences were of 150 bp length and pair ended. Raw sequencing data were shared as FASTQ files.

5.2.3.3 Gene Ontology and Gene Set Enrichment Analysis of CCLP siVimentin RNAseq identified hallmarks and gene sets associated with cell proliferation, EMT, and cell migration

We initially analysed the raw data from Illumina using the Galaxy web platform as detailed in the method chapter. In brief, we used the tool FASTQC to check on the quality of the sequences (424). We only proceeded when the QC report showed no critical red flags. We trimmed low-quality reads below 30 Phred and the synthetic adaptors using TrimGalore (425). We aligned the sequences to Human Genome 38 (HG38) using the TopHat aligning tool (427). We, then, imported an annotated HG38 reference from Galaxy's built-in library to count the features (428). Later, we analysed the two independent biological repeats using the tool DESeq2 to generate the list of differentially expressed genes (DEG), normalised gene counts, and DESeq2 plots (429).

To prepare DESeq2 output for downstream analysis, we converted the genes Ensemble identities to gene Symbols using MART_export obtained from GitHub Bioinformatics forum (430). We applied cut-off measures for the adjusted *p*-value and the gene expression fold change (FC). Adjusted *p*-value takes into consideration the FDR. To say that a gene has an adjusted *p*-value below 0.05 means that the chance of having a false positive regulated gene is below 5%. We kept all the genes with adjusted *p*-value below 0.05 and referred to them as **significant DEG**. To sub-classify this list of genes into **upregulated** and **downregulated** genes, we extracted them using FC cut-off values of **FC >1.5** and **FC <0.66**, respectively.

Vimentin knockdown in CCLP1 cells resulted in a total of 2860 significant DEG where 305 were significantly upregulated genes, and 280 genes were downregulated (figure 5-13-A). The volcano plot illustrated the combined 585 significant and regulated DEG created by plotting Log_2(FC) vs. -Log_{10}(adjusted *p*-value) (figure 5-13-B).

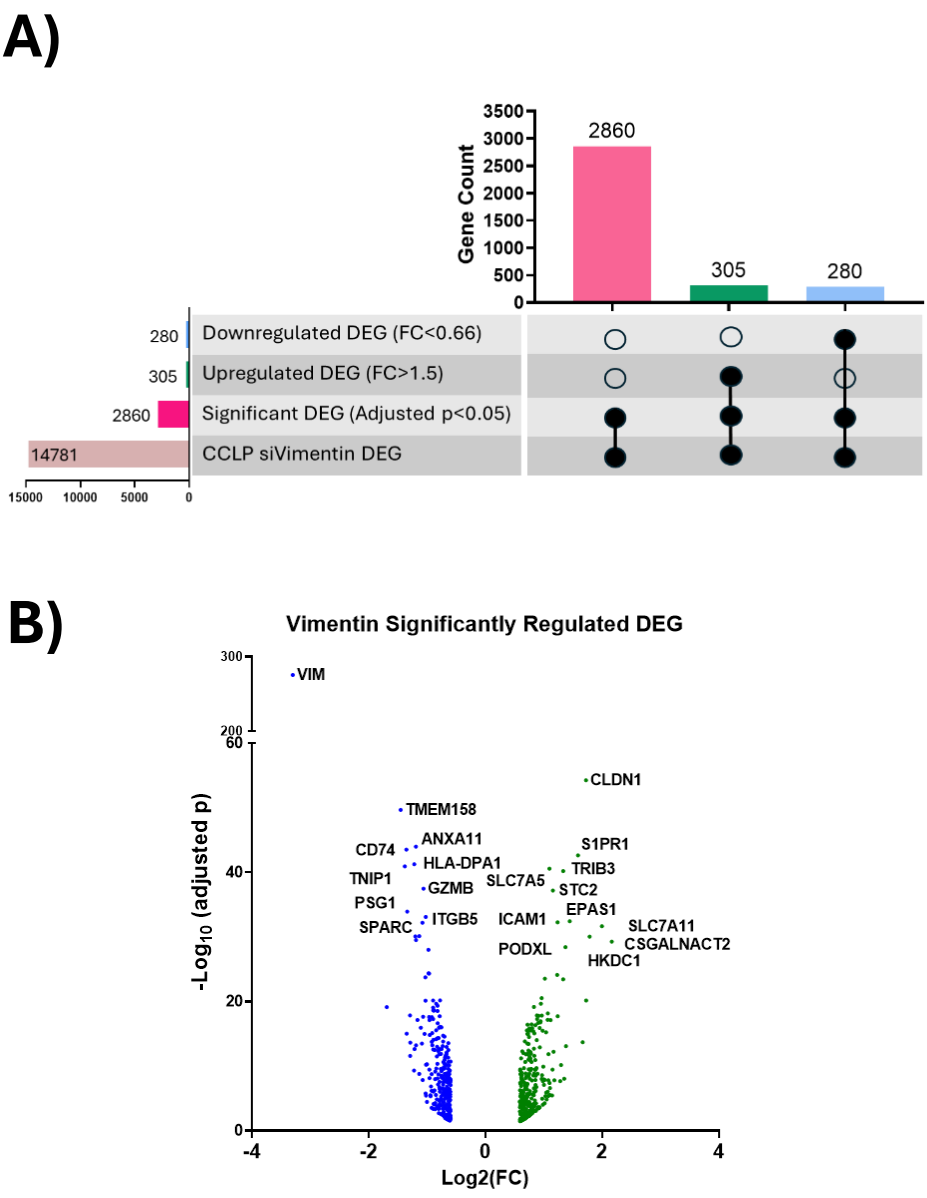


Figure 5-13 Differentially expressed genes in CCLP siVimentin RNAseq.

(A) UpSet plot of significant DEG (adjusted p<0.05) further classified into upregulated genes (FC>1.5) and downregulated genes (FC<0.66). (B) Volcano plot of the 585 significant and regulated DEG (Log_2(FC) vs. -Log_{10}(adjusted p)).

To have a biological insight into the functional importance of significant and regulated DEG, we carried out two types of analyses: GO analysis and GSEA. GO is a standardised tool to annotate input genes and investigate their overrepresentation in predefined terms to describe their association in biological processes, cellular components, or molecular functions (481). We performed GO analysis using the Enrichr web platform. Enrichr is an integrative tool that include over a hundred of gene set libraries (436). Enrichr can perform enrichment and ORA as well as provide various visualisation approaches to display the results inferring functional knowledge of the input genes. Enrichr uses four scores when it generates the final report: *p*-value*, q*-value*,* odd ratio, and combined score. Enrichr acquire *q*-values from adjusted *p*-values corrected to account for multiple hypotheses using Benjamini-Hochberg method (435, 436). For CCLP siVimentin RNAseq GO analysis, we used the significant and regulated 585 DEG. The outcome contained different terms classified under biological processes (GO-BP), cellular components (GO-CC), and molecular functions (GO-MF). We only showed terms that had an adjusted *p*-< 0.05. We had 17 biological processes terms, 3 cellular component terms, and a single molecular function term (figure 5-14).

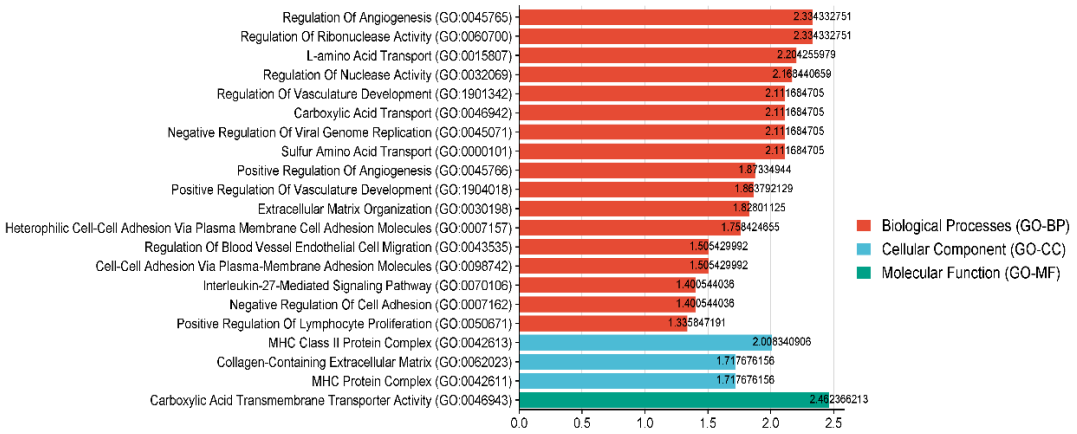


Figure 5-14 Gene ontology analysis on CCLP siVimentin RNA seq data.

GO analysis using Enrichr of CCLP siVimentin significant and regulated DEG. Only the significantly enriched pathways were listed in order of -Log_{10}(adjusted *p*).

Majority of the terms under GO-BP indicated that Vimentin knockdown regulated gene sets associated with the progress and regulation of cell proliferation and cell migration.

We further investigated CCLP siVimentin significant and regulated DEG using GSEA MSigDB Hallmarks. MSigDB Hallmarks offered several collections of curated gene sets that infer biological information on key functions and pathways often implicated in cancer (433, 434). Input genes were analysed by hypergeometric overlapping analysis, or over-representation analysis (ORA), to highlight the gene sets that were significantly enriched in hallmark pathways (433, 434). Since the limit for gene input on MSigDB online computation was limited to 500 genes, we again used Enrichr. Enrichr used a borrowed MSigDB library to investigate input genes against the Hallmarks libraries (435, 436). The outcome was presented in a horizontal bar chart illustrating the hallmark gene sets and their -Log_{10}(*p*-value). Hallmarks were ranked based on their *p*-

values where statistically significant terms were identified when p <0.05 (figure 5-15).

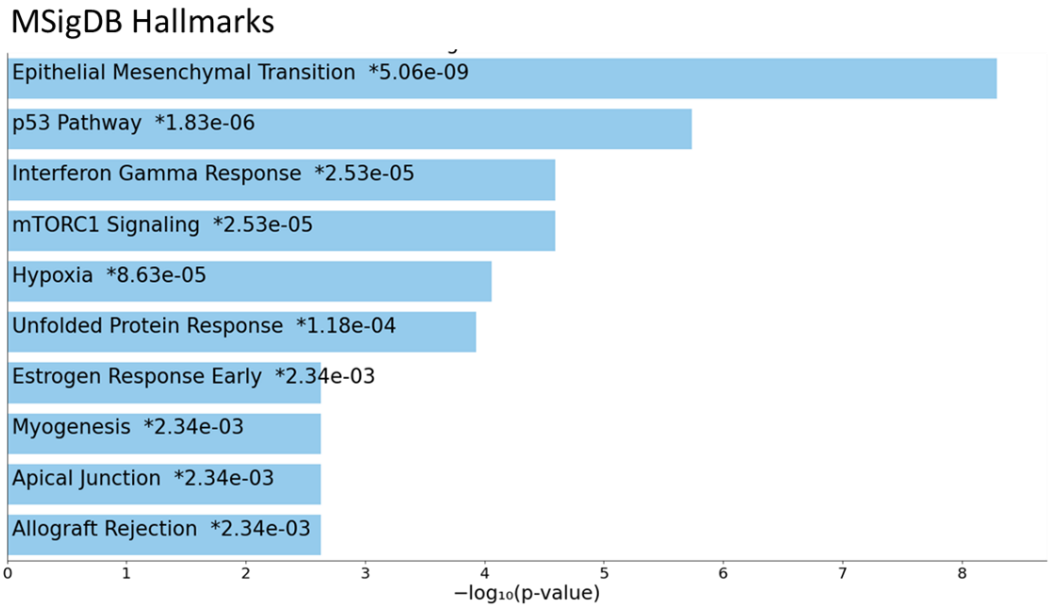


Figure 5-15 GSEA using MSigDB Hallmarks on CCLP siVimentin significant and regulated DEG.

The list represented the top 10 significantly enriched hallmarks in order of their -Log_{10}(*p*-value).

The top 10 hallmarks here indicated that Vimentin knockdown significantly enriched genes associated with cell growth, cell proliferation and cell migration e.g. EMT, p53 pathway, mTORC1, hypoxia, Oestrogen response, and apical junctions. In fact, the EMT hallmark was the hallmark with the lowest *p*-value.

Since EMT was the most significantly enriched hallmark, we wanted to have a deeper look into the enriched EMT genes. We proceeded with pre-ranked GSEA using hallmarks gene sets. Pre-ranked GSEA is a variant of GSEA that allowed the users to choose the parameter of interest to rank the gene input instead of the default differential expression used in traditional GSEA (433). Pre-ranked GSEA have the advantage of taking into consideration whether

the pre-determined metric parameter was associated with the investigated gene sets (433). We used a Log_2(FC)-ranked DEG list as an input and performed pre-ranked GSEA choosing the most recent hallmark dataset. Although the literature recommended using the entire DEG list for pre-ranked GSEA, we applied *p*-value cut-off for the DEG because we wanted to eliminate the noise from genes that were non-significant.

The enrichment plot is one of the primary outputs from pre-ranked GSEA which provides a visual view of gene set enrichment. The top part of enrichment plot had a green curve generated from running the enrichment score calculation down the ranked gene input. The curve went upward in response to an increase in the running-sum statistics whenever a gene from the ranked list was found in the EMT gene set, while the drop in the curve occurred when a gene from the ranked list was not in the EMT gene set. The magnitude of increment or decrement relied on the degree of gene correlation with the examined phenotype. The maximum deviation on the curve (peak) indicated the enrichment score (ES) which reflected the gene overrepresentation at either the top (positive ES) or bottom (negative ES) of a ranked gene list. The result was considered most interesting when the peak was at either the top (positive enrichment) or bottom (negative enrichment) of the ranked list. The middle part of the plot contained the overlapped genes from the ranked list with the EMT gene set. The gene members that contributed the most in making the ES are called the leading-edge genes. They appeared prior to a positive ES peak and after a negative ES. The bottom

part of the plot showed the pre-ranking metric for input genes (Y-axis) and the ranking in ordered EMT dataset (X-axis).

EMT hallmark was identified as one of the positively enriched gene sets. EMT enrichment plot (figure 5-16-A) indicated an overall enrichment score (ES=0.18 – Normalised=ES 1.14). A total of 74 genes from CCLP siVimentin ranked genes were found in the EMT gene set. However, 21 genes contributed to the enrichment signal (or core enrichment) at the top of the ranked list correlating with the upregulated genes.

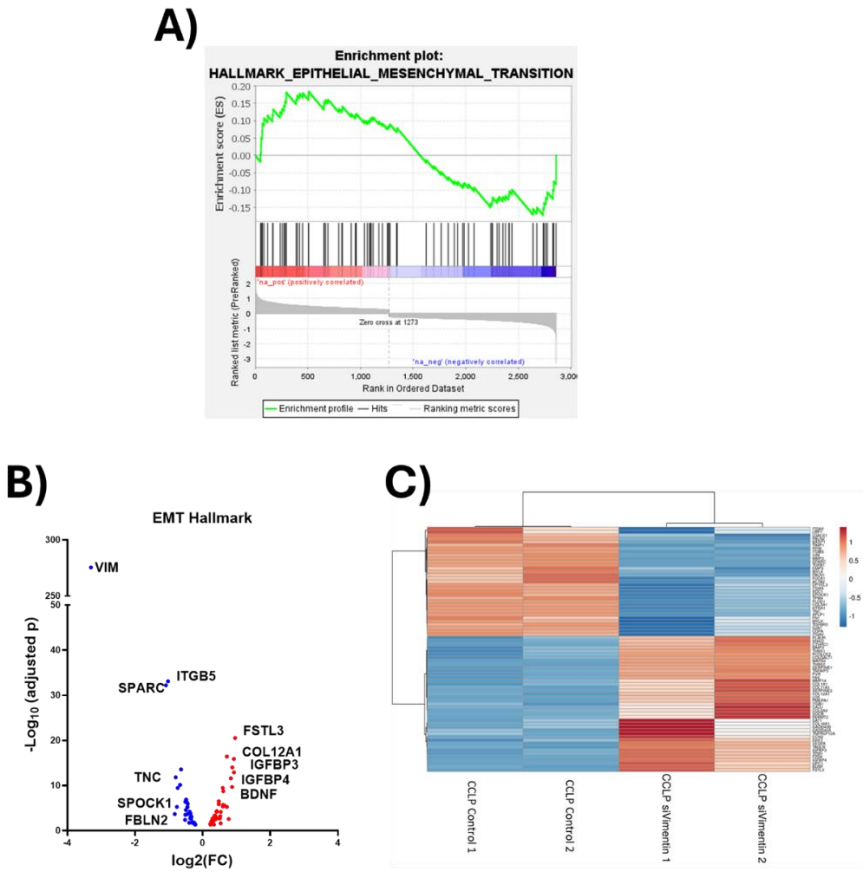


Figure 5-16 Pre-ranked GSEA analysis using Hallmarks on CCLP siVimentin significant DEG.

(A) Enrichment plot of EMT hallmark. (B) Volcano plot of CCLP siVimentin DEG enriched in EMT hallmark showing genes at the top and bottom of the ranked list. (C) Heatmap of all CCLP siVimentin DEG enriched in EMT hallmark showing the normalised counts from the different conditions of RNAseq experiment.

The volcano plot for EMT genes showed the relationship between genes Log_2(FC) and -Log_{10}(adjusted *p*-value) (figure 5-16-B). The plot also showed that some of the key EMT genes enriched by Vimentin knockdown were at the top of the ranked list e.g. *FSTL3, IGFBP4, COL12A1, IGFBP3*, and *BDNF*. While those at the bottom of the ranked list included: *SPOCK1, TNC, FBLN2*, *ITGB5,* and *SPARC*. We also presented the genes in a clustered heatmap using the equivalent DESeq2 normalised counts from the different conditions of the RNAseq experiment (figure 5-16-C).

Taking the findings together, our results indicated that Vimentin perturbation in CCLP1 cells induced gene expression changes that were associated with the phenotype we observed *in vitro.* Vimentin knockdown regulated gene sets that were related to cell proliferation, EMT, and cell migration.

5.2.3.4 Gene Ontology and Gene Set Enrichment Analysis of CCSW siVimentin RNAseq identified hallmarks and gene sets associated with cell proliferation, EMT, and cell migration

We followed the same approach from the previous section to interrogate CCSW siVimentin RNAseq data starting with processing raw FASTQ files using Galaxy web platform to pre-ranked GSEA.

When we applied the cut-off for adjusted *p*-value, we had a total of 1274 DEG. 243 of the genes were significantly upregulated and 387 were downregulated bringing down the total number to 630 genes (figure 5-17-A). The volcano plot shows the combined 630 regulated genes created by plotting Log_2(FC) vs. -Log_{10}(adjusted *p*) (figure 5-17-B).

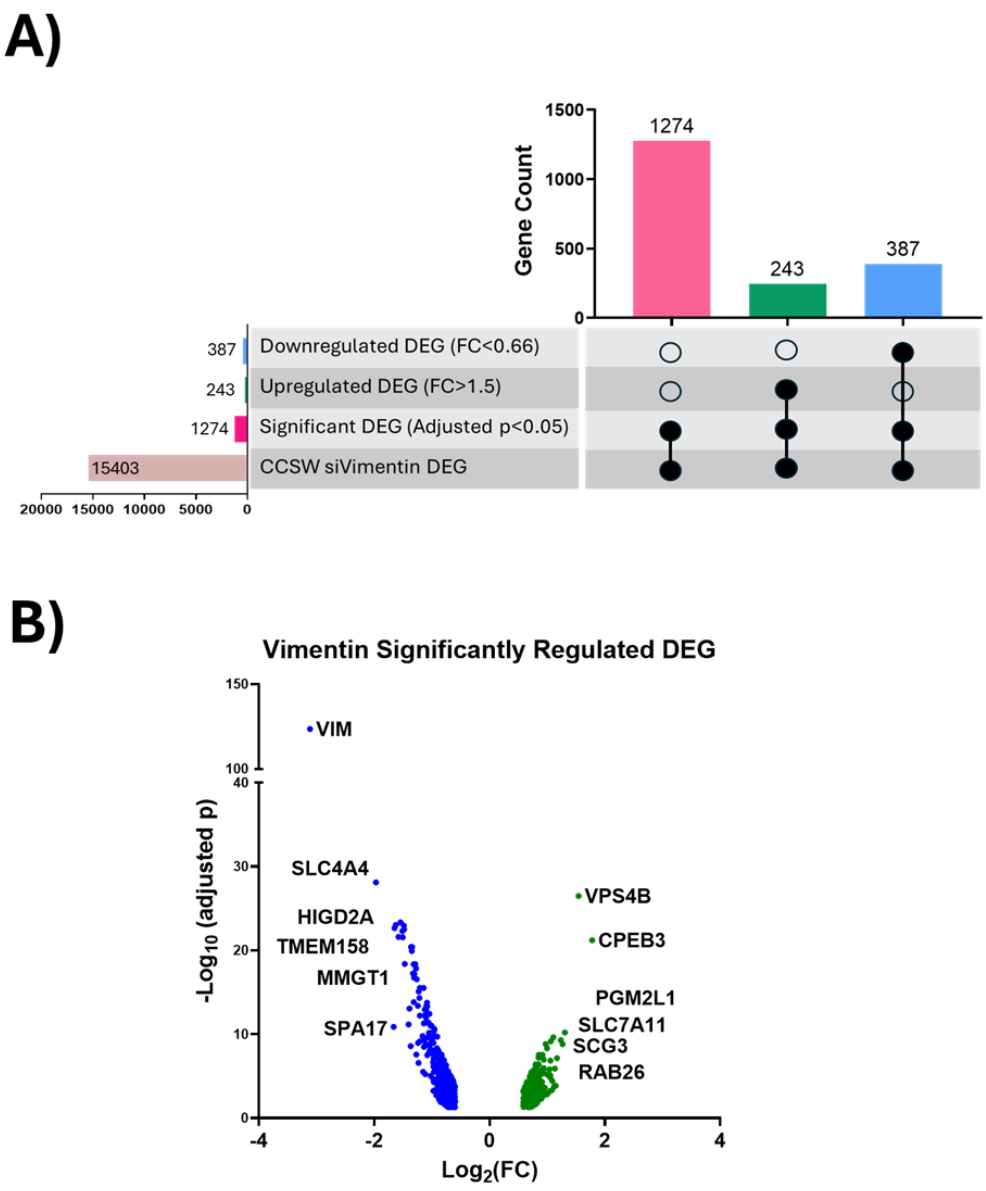


Figure 5-17 Differentially expressed genes in CCSW siVimentin RNAseq.

(A) UpSet plot of significant DEG (adjusted p<0.05) further classified into upregulated genes (FC>1.5) and downregulated genes (FC<0.66). (B) Volcano plot of the 630 significant and regulated DEG (Log_2(FC) vs. -Log_{10}(adjusted p)).

For CCSW siVimentin RNAseq GO analysis, we used the significant and regulated 630 DEG. The outcome contained different terms classified under biological processes (GO-BP), cellular components (GO-CC), and molecular

functions (GO-MF). Only one term under GO-MF had an adjusted *p*-value below 0.05 which is Ephrin Receptor Binding. However, we still presented the top 5 terms in each GO category: 5 biological processes terms, 5 cellular component terms, and 5 molecular function terms (figure 5-18).

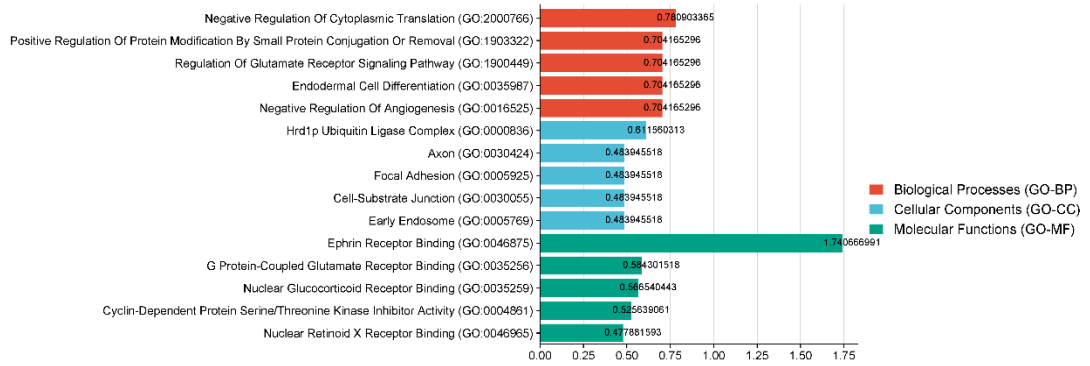


Figure 5-18 Gene ontology analysis on CCSW siVimentin RNA seq data.

GO analysis using Enrichr of CCSW siVimentin significant and regulated DEG. Only the significantly enriched pathways were listed in order of -Log_{10}(adjusted *p*).

We further investigated CCSW siVimentin DEG using GSEA with MSigDB Hallmarks. We plotted the outcome in a horizontal bar chart illustrating the hallmark gene sets and their -Log_{10}(*p*-value). Statistically significant terms were identified when $p < 0.05$ (figure 5-19).

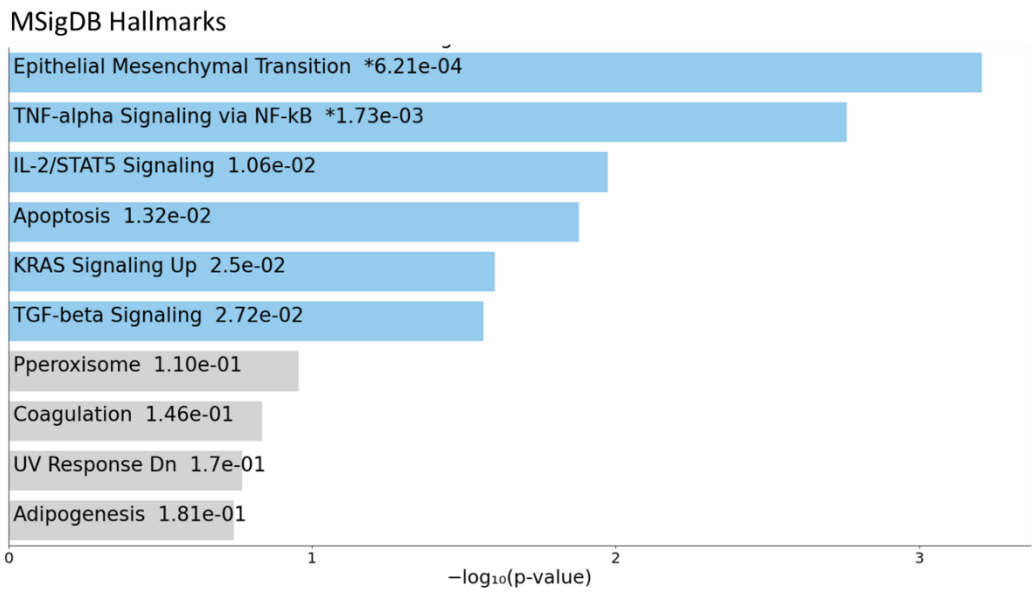


Figure 5-19 GSEA using MSigDB Hallmarks on CCSW siVimentin significant and regulated DEG.

The list represented the top 10 significantly enriched hallmarks in order of their -Log_{10}(*p*-value).

Only 6 out of the top 10 hallmarks had an adjusted *p*-value below 0.05. Vimentin knockdown significantly enriched genes associated with cell growth, cell proliferation, cell apoptosis or cell migration e.g. EMT, apoptosis pathway, KRAS signalling, and TGF-β signalling pathway. As in the CCLP siVimentin RNAseq GSEA data, EMT hallmark was the hallmark with the lowest *p*-value.

We also performed pre-ranked GSEA using hallmarks gene set to get more details on the involved genes. The input ranked gene list contained significant DEG. Unlike CCLP siVimentin RNAseq, EMT hallmark here was identified as one of the negatively enriched gene sets. In fact, it was the one with the most significant adjusted *p*-value (*p*= 0.003). EMT enrichment plot (figure 5-20-A) indicated an overall enrichment score of -0.53 (ES=-0.53 – Normalised=ES 2.41). A total of 24 genes from the ranked gene list were found in EMT gene

set. However, 12 genes contributed to the enrichment signal (or core enrichment) at the bottom of the ranked list correlating with the downregulated genes.

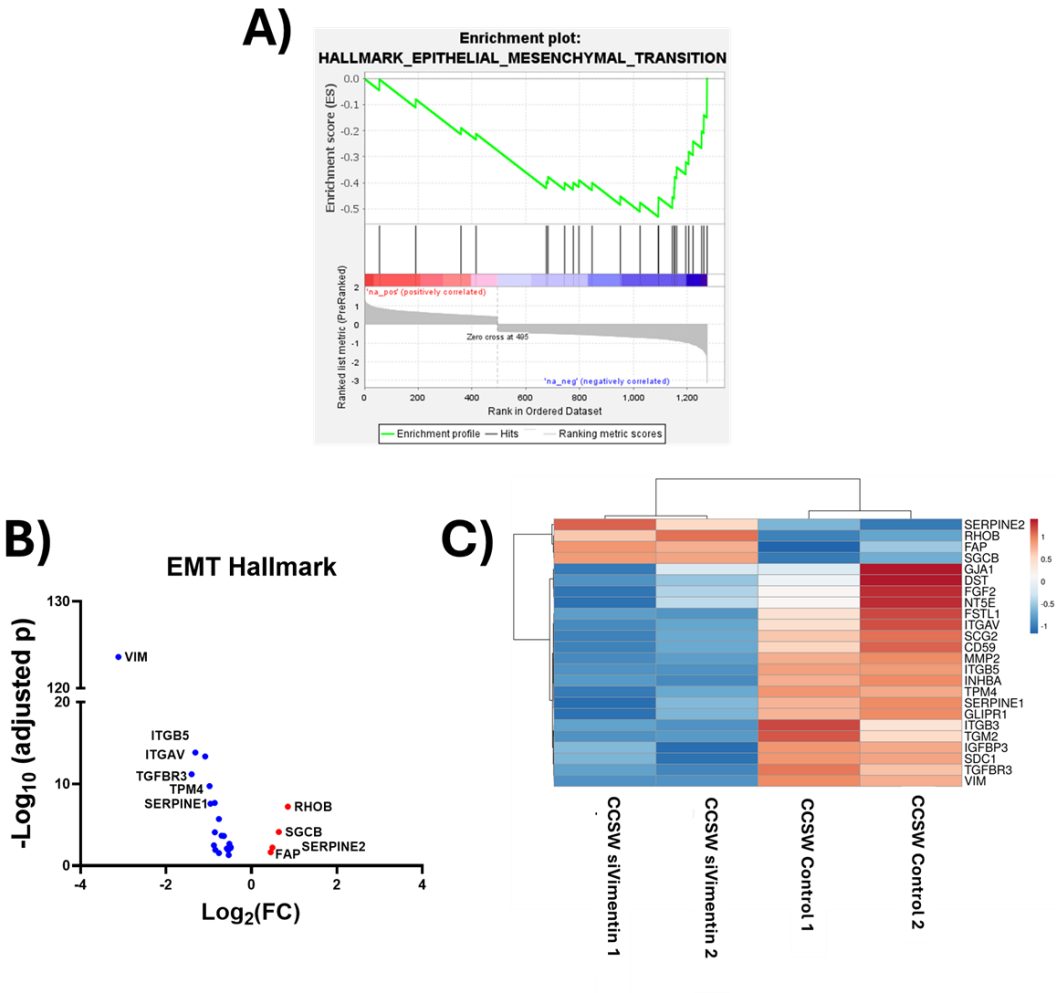


Figure 5-20 Pre-ranked GSEA analysis using Hallmarks on CCSW siVimentin significant DEG.

(A) Enrichment plot of EMT hallmark. (B) Volcano plot of CCSW siVimentin DEG enriched in EMT hallmark showing genes at the top and bottom of the ranked list. (C) Heatmap of all CCSW siVimentin DEG enriched in EMT hallmark showing the normalised counts from the different conditions of RNAseq experiment.

The volcano plot for EMT genes showed the relationship between genes Log_2(FC) and -Log_{10}(adjusted *p*-value) (fig 5-20-B). The plot also highlighted some of the key EMT genes enriched by Vimentin knockdown at the top of the

ranked list e.g. *RHOB*, *SGCB*, *SERPINE2*, and *FAP*. While those at the bottom of the ranked list included: *ITGB3*, *SERPINE1*, *TPM4*, *ITGB5*, *ITGAV*, and *TGFBR3*. We also presented the genes in a clustered heatmap (figure 5-20-C) using the equivalent DESeq2 normalised counts from the different conditions of the RNAseq experiment.

Taking the findings together, our results indicated that Vimentin perturbation in CCSW1 cells induced gene expression changes that were associated with the phenotype we observed in *vitro*. Vimentin knockdown regulated gene sets that are related to cell proliferation, EMT, and cell migration.

5.2.3.5 Common enriched EMT genes in CCLP siVimentin and CCSW siVimentin RNAseq data sets

Both RNAseq data sets in CCLP siVimentin and CCSW siVimentin showed the EMT hallmark as the most significant pathway on GSEA. However, pre-ranked GSEA revealed that EMT hallmark genes contributed to a positive enrichment score (ES=0.18 – Normalised=ES 1.14) in CCLP siVimentin, and a negative enrichment score (ES=-0.53 – Normalised=ES 2.41) in CCSW siVimentin. To see whether we had common enriched EMT genes, we intersected the two lists using Venn diagrams (figure 5-21-A).

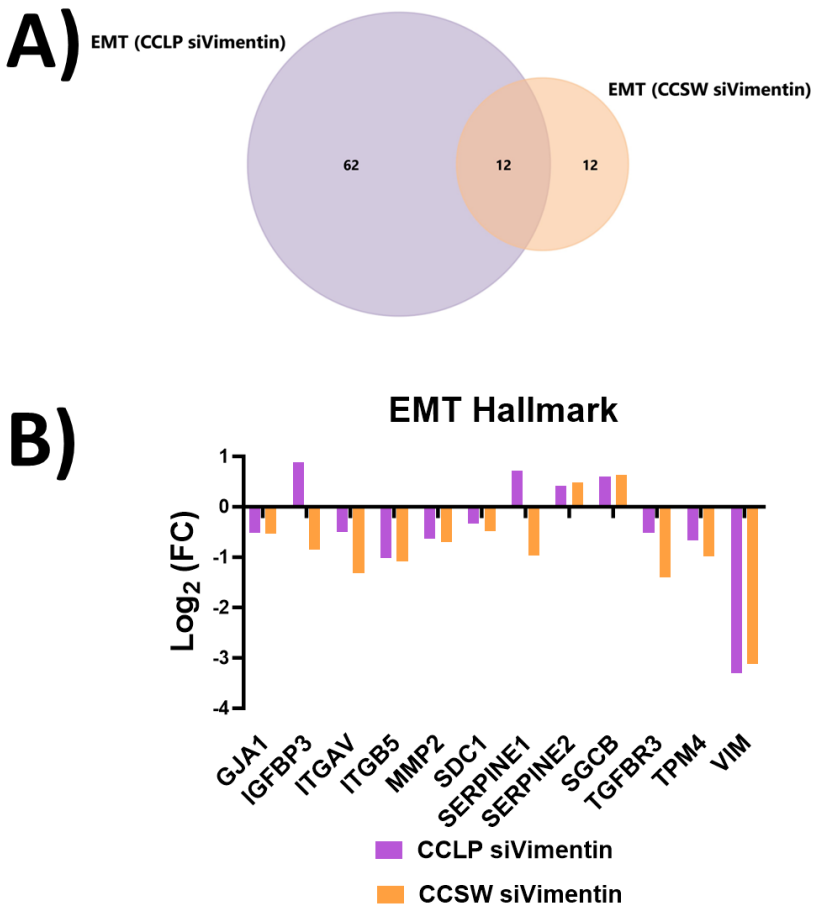


Figure 5-21 Common DEG enriched in EMT hallmark in CCLP siVimentin RNAseq and CCSW siVimentin RNAseq.

(A) Venn diagram showing 12 overlapped DEG enriched in EMT hallmark between the two RNAseq data sets. (B) Bar chart showing the expression fold change as (Log_2(FC)) of all the 12 common DEG.

The results indicated that a total of 9 genes (excluding *VIM*) were regulated in the same direction by Vimentin knockdown in both cell lines: *GJA1*, *ITGAV, ITGB5, MMP2, SDC1, SERPINE2, SGCB, TGFBR3*, and TPM4. The two enriched EMT genes regulated in opposite direction by Vimentin were *SERPINE1* and *IGFBP3*. They were upregulated in CCLP siVimentin and downregulated in CCSW siVimentin RNAseq (figure 5-21-B).

5.2.3.6 Summary on differentially expressed gene counts by Vimentin knockdown in CCA cells

Vimentin knockdown resulted in slightly fewer differentially expressed genes in CCLP1 cells (14781 DEGs) compared to CCSW1 cells (15403 DEGs) regardless of the cut-offs in adjusted *p*-value or FC. When we took adjusted *p*-value cut-off in consideration, we had more genes for CCLP siVimentin (2860) compared to CCSW siVimentin (1274). By applying FC cut-off to CCLP siVimentin, we had 585 DEGs where 305 genes were upregulated, and 280 genes were downregulated. Applying the same for CCSW siVimentin RNAseq gave a greater number of genes (630 DEGs) compared to CCLP1. 243 genes of the 630 were upregulated, and 387 genes were downregulated. (figure 5-22)

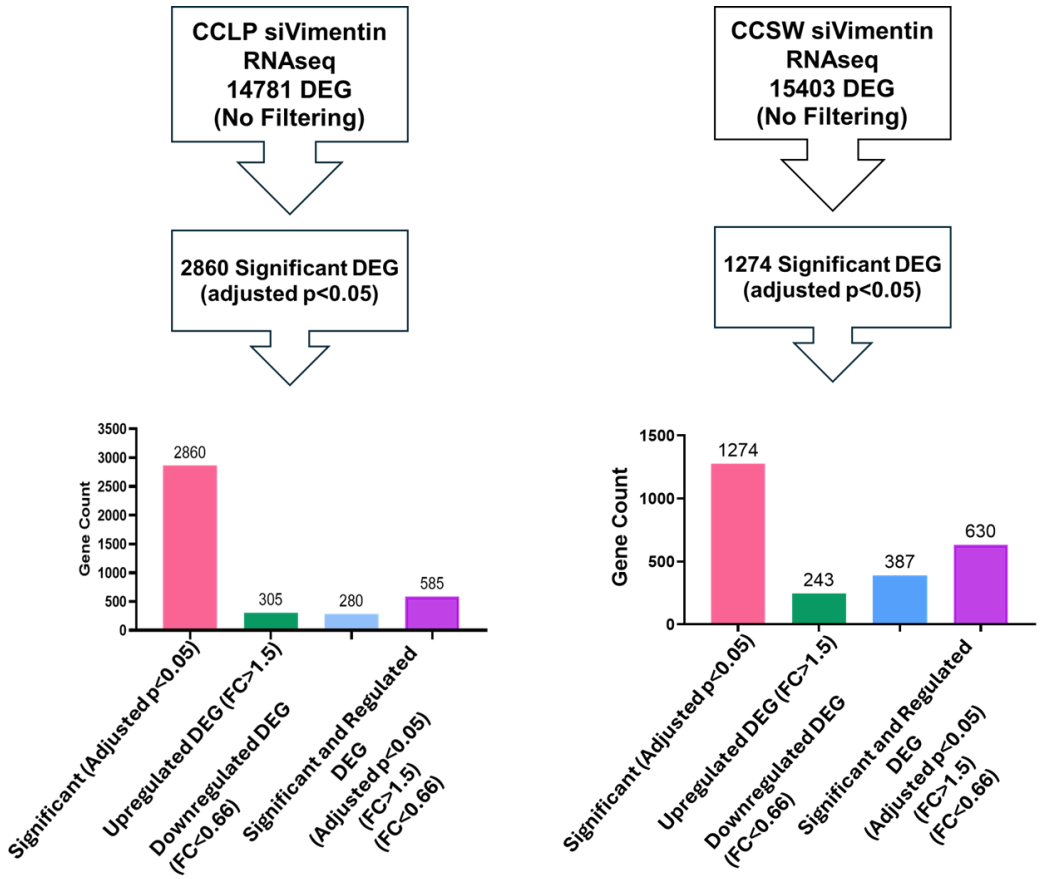


Figure 5-22 Summary on DEG counts in CCLP siVimentin and CCSW siVImentin RNAseq.

DEG were filtered based on their adjusted *p* and FC values into significant up and downregulated DEG.

To find the number of common DEGs with *p*-value and FC cut-offs, we first overlapped the DEGs with adjusted *p*<0.05 from CCLP siVimentin (2860) with CCSW siVimentin (1274) (figure 5-23). The overlap revealed 558 common DEGs for the two RNAseq data sets. This represented up to 19.5% of CCLP siVimentin DEGS (adjusted *p*<0.05). For CCSW siVimentin, 558 genes accounted for 43.8% of CCSW siVimentin DEGS (adjusted *p*<0.05).

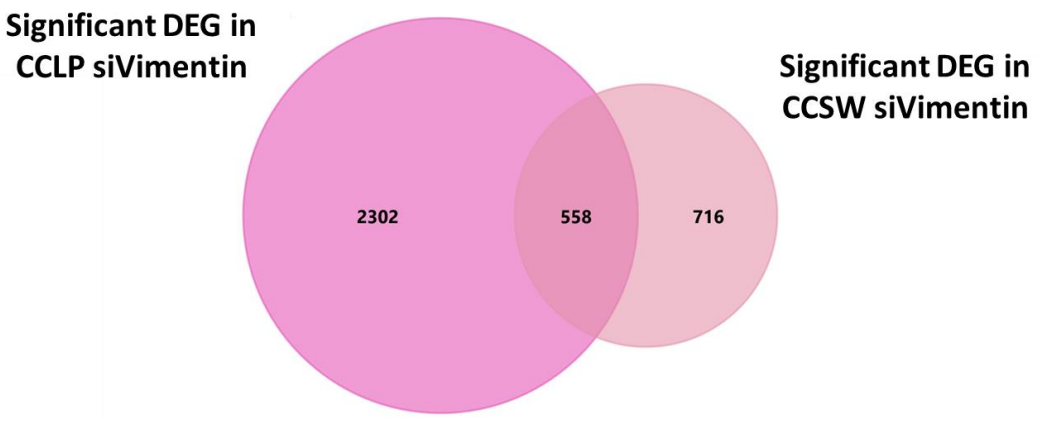


Figure 5-23 Venn diagram showing 558 significant DEG were regulated by Vimentin in both RNAseq data sets.

We further extracted the upregulated and downregulated genes from the 558 common genes by applying FC cut-off. We had a total of 118 genes where 33 genes were upregulated, and 85 genes were downregulated. This represented 21.15% of the common genes. We proceeded to perform GSEA using the common genes regulated in the same direction by vimentin in both cell lines to see the hallmarks in which the genes were enriched (figure 5-24). Only 3 hallmarks were significant. Again, EMT hallmark was the most significantly enriched pathway.

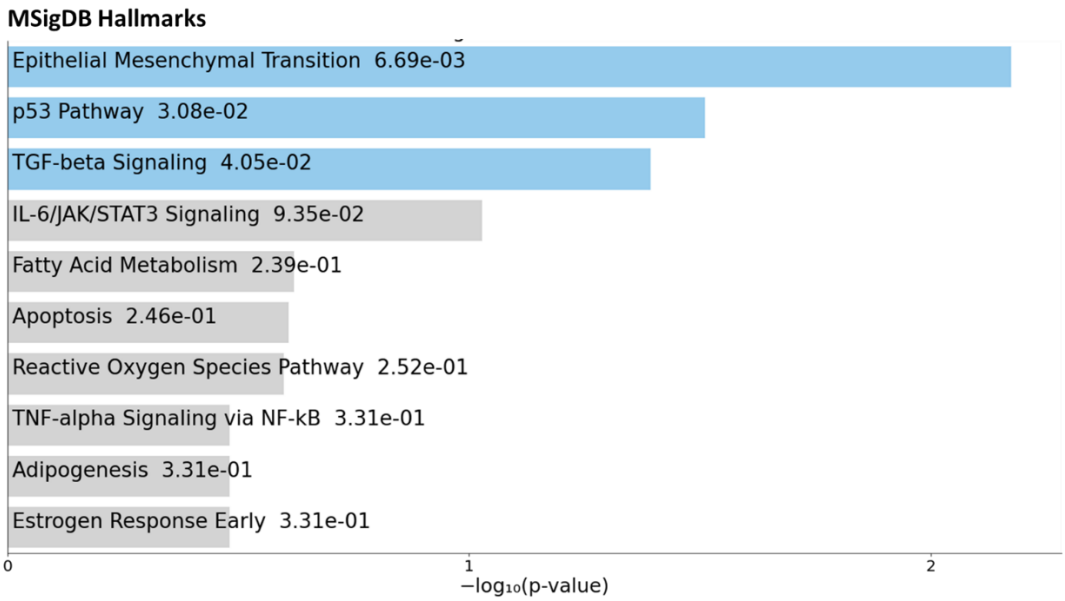


Figure 5-24 GSEA using MSigDB Hallmarks on 118 significant and regulated DEG by Vimentin in the same direction.

The list represented the top 10 significantly enriched hallmarks in order of their -Log_{10}(*p*-value). Overall, the results indicate that Vimentin knockdown resulted in changes in the expression of common genes in CCLP1 and CCSW1 CCA cell lines that were also enriched in pathways associated with the phenotype we observed *in vitro*. Vimentin knockdown in CCA cell lines significantly regulated genes that were enriched in gene sets associated with cell proliferation, EMT, and cell migration.

5.2.3.6.1 Overview of Vimentin knockdown impact on the expression of important cell cycle genes

Previously, Dr Lee reported in her thesis that knocking down Vimentin in CCLP1 cells upregulated *CDKN1B* and downregulated *CCND2* which is also observed after knocking down PRH in CCLP1 cells (269). Both genes were also reported as direct targets for PRH on CCLP Myc-PRH ChIPseq data. To have a closer look at the changes brought about by Vimentin depletion in our

CCLP1 and CCSW1 cells on *CCND2, CDKN1A, CDKN1B and CDKN2B* genes, we investigated their Log_2(FC) and normalised counts in both RNAseq data sets.

Figure 5-25-A shows that Vimentin knockdown in CCLP1 cells significantly upregulated the expression of *CCND2* (adjusted p = 0.008) and *CDKN1A* (adjusted p = 0.0227). Whereas CDKN1B and CDKN2B were differentially expressed, but non-significantly (adjusted p = 0.69) and (adjusted p = 0.64), respectively. In CCSW1 cells, (figure 5-25-B) shows that all the genes were differentially expressed although non-significantly. Only *CDKN1B* (adjusted p = 0.519) expression seemed to increase while *CCND2* (adjusted p = 0.971), *CDKN1A* (adjusted p = 0.980) and *CDKN2B* (adjusted p = 0.814) decreased.

Overall, our data indicated that Vimentin regulates the expression of some genes involved in the progression of the cell cycle in CCA cells. However, most of the findings were neither significant nor similar to the previous results.

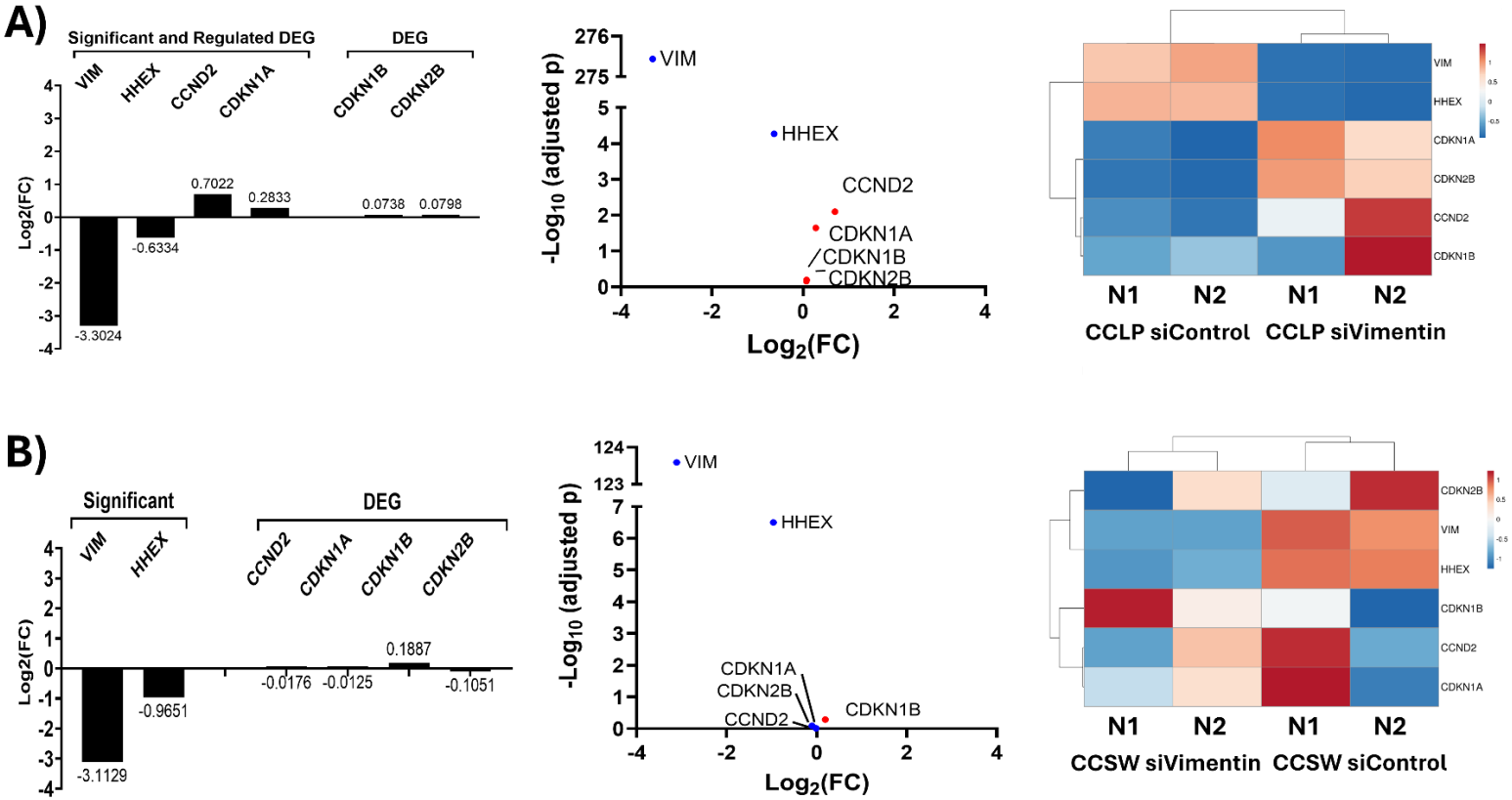


Figure 5-25 The effect of Vimentin knockdown on the expression of genes with roles in the progression of the cell cycle.

(A) Vimentin knockdown in CCLP1 cells significantly upregulated the expression of *CCND2* and *CDKN1A*.
(B) Vimentin knockdown in CCSW1 cells did not significantly alter the expression of *CCND2, CDKN1A, CDKN1B*, or *CDKN2B*.

5.2.4 RNAseq overlap analysis to investigate co-regulation of gene expression by Vimentin and PRH in CCA

5.2.4.1 Genes changing in the same direction when co-regulated by Vimentin and PRH

We established that Vimentin could regulate gene expression when it is altered in CCA cell lines. In order to investigate the genes co-regulated in the same direction by Vimentin knockdown and PRH knockdown in CCLP1 cells, we utilised published CCLP PRH knockdown RNAseq data generated by Kitchen et al. in conjunction with CCLP siVimentin RNAseq data (215). We considered the DEG to be significantly co-upregulated by PRH and Vimentin if they had an adjusted p<0.05 and FC>1.5 in both RNAseq data sets, and vice versa for co-downregulated genes. We analysed PRH knockdown RNAseq following the same approach we used in the previous section. For intersection analysis, we overlapped DEG with adjusted p<0.05 from both RNAseq data sets which produced 131 genes (figure 5-26).

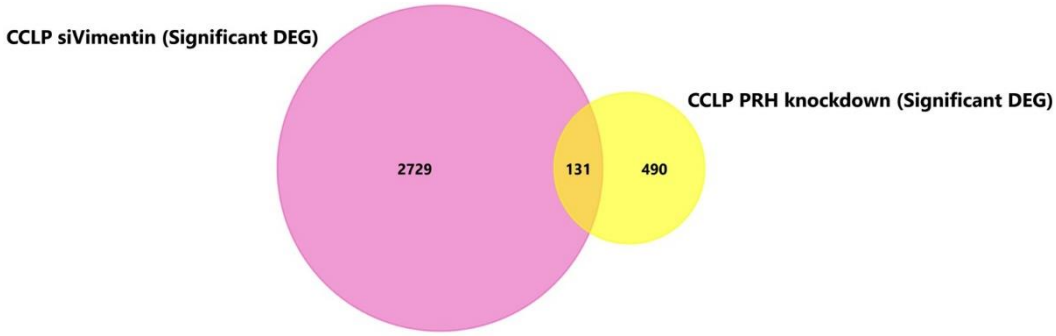


Figure 5-26 Venn diagram showing that 131 significant DEG were co-regulated by Vimentin and PRH.

To see whether these 131 genes were likely to be regulated via PRH-DNA binding, we utilised published CCLP PRH-Myc ChIPseq data (215). We analysed ChIPseq data as detailed in the method section. We overlapped the 131 significant DEG with genes that have peaks for PRH in CCLP PRH-Myc ChIPseq regardless of the peak number or the distance from TSS. The intersection produced 33 genes where 7 genes were co-downregulated by Vimentin and PRH, and 11 genes were co-upregulated by Vimentin and PRH. (figure 5-7).

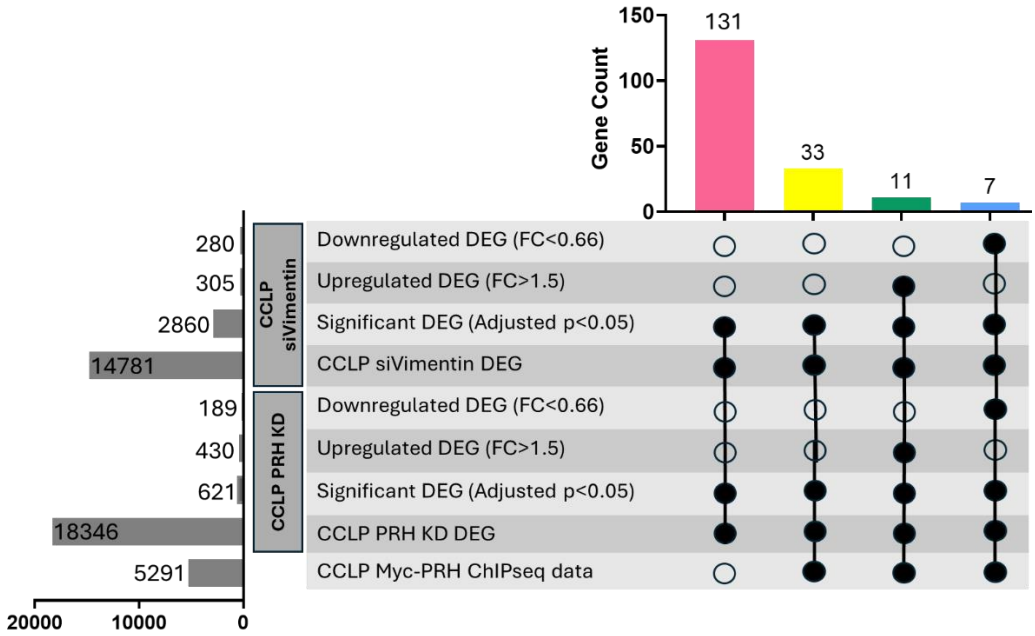


Figure 5-27 UpSet plot showing the number of extracted significant DEG co-regulated by Vimentin and PRH in the same direction. These DEG also have peaks for PRH on ChIPseq.

DEG co-upregulated by PRH and Vimentin = common DEG with adjusted p<0.05 and FC>1.5. DEG co-downregulated by PRH and Vimentin = common DEG with adjusted p<0.05 and FC<0.66.

To have a biological insight into the functions of these genes, we investigated the hallmarks pathways in which these 18 genes were enriched. The top enriched hallmarks were all associated with cellular growth, proliferation, EMT,

migration and invasion (figure 5-28). This indicates that both Vimentin and PRH significantly co-regulate genes that enriched cancer associated pathways.

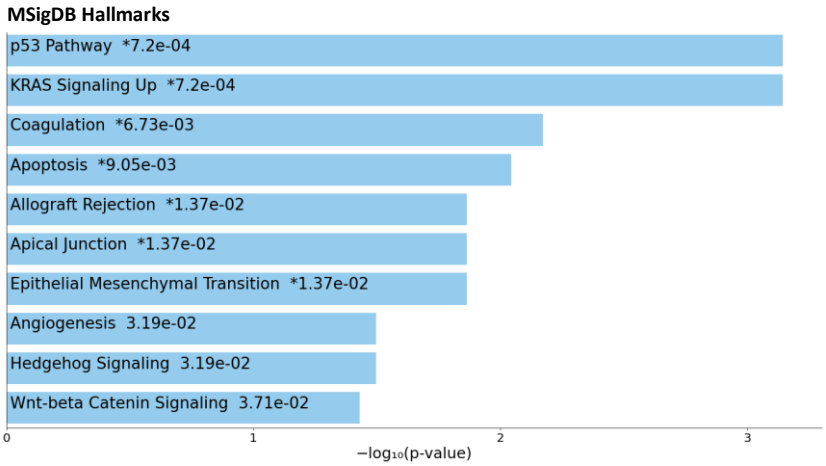


Figure 5-28 GSEA using MSigDB Hallmarks on 18 significant DEG co-regulated by Vimentin and PRH in the same direction.

The list represented the top 10 significantly enriched hallmarks in order of their -Log_{10}(*p*-value).

5.2.4.2 Genes changing in opposite directions when co-regulated by Vimentin and PRH

To investigate the genes co-regulated in opposite direction by Vimentin and PRH in CCLP1 cells, we used CCLP siVimentin RNAseq with CCLP PRH overexpression RNAseq generated by Kitchen et al. (215). We overlapped significant DEG from CCLP siVimentin with significant DEG from CCLP PRH overexpression RNAseq. This produced a set of 599 common significant DEGs (figure 5-29).

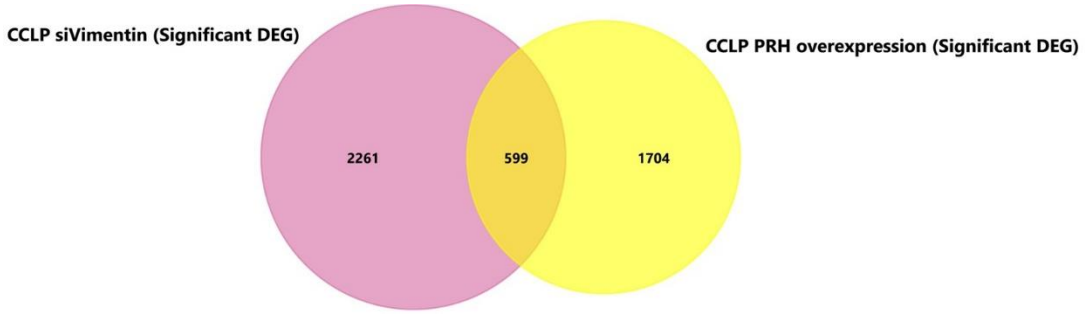


Figure 5-29 Venn diagram showing that 599 significant DEG were co-regulated by Vimentin and PRH.

Then, to examine whether the above 599 DEGs were likely to be directly regulated via PRH-DNA binding, we overlapped them with gene list from CCLP PRH-Myc ChIPseq data regardless of the peaks number or distance from TSS. This narrowed down the list into 187 significant DEG with ChIP peaks for PRH. Then we used these 187 DEG to extract the genes that were changing in opposite directions (i.e. upregulated by Vimentin and downregulated by PRH and vice versa). Out of these 187 genes, 7 genes were co-regulated by Vimentin and PRH (upregulated by Vimentin – downregulated by PRH), and 19 genes were co-regulated by Vimentin and PRH (downregulated by Vimentin – upregulated by PRH) bringing down the final gene count to 26 (figure 5-30).

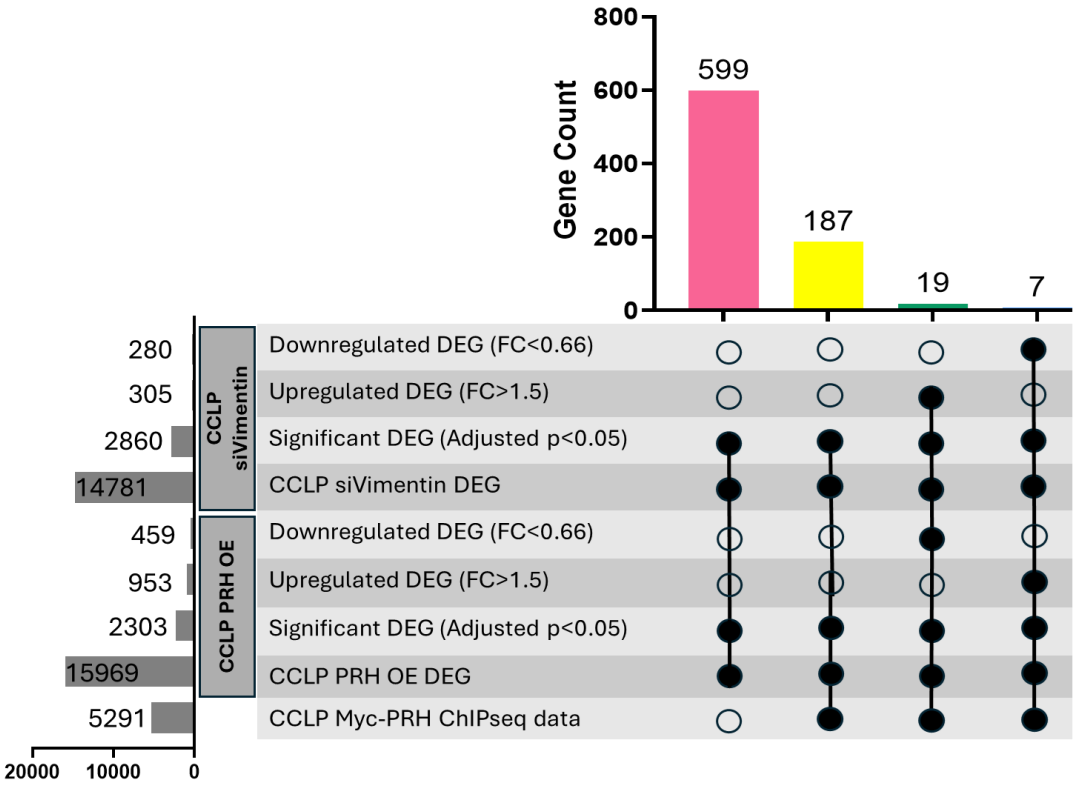


Figure 5-30 UpSet plot showing the number of extracted significant DEG co-regulated by Vimentin and PRH in opposite direction after applying FC cut-off on the common DEG.

To explore the functional importance of these 26 significant DEG co-regulated by Vimentin and PRH, we performed GSEA using hallmarks terms (figure 5-31-A). Out of the top 10 enriched hallmarks, 5 showed a *p*-value below 0.05 which were as follow: coagulation, mitotic spindle, apoptosis, p53 pathway, and EMT. All of which indicate that the Vimentin regulates gene sets that were enriched in pathways associated with cancer progress. These genes were also regulated by PRH via PRH-DNA binding.

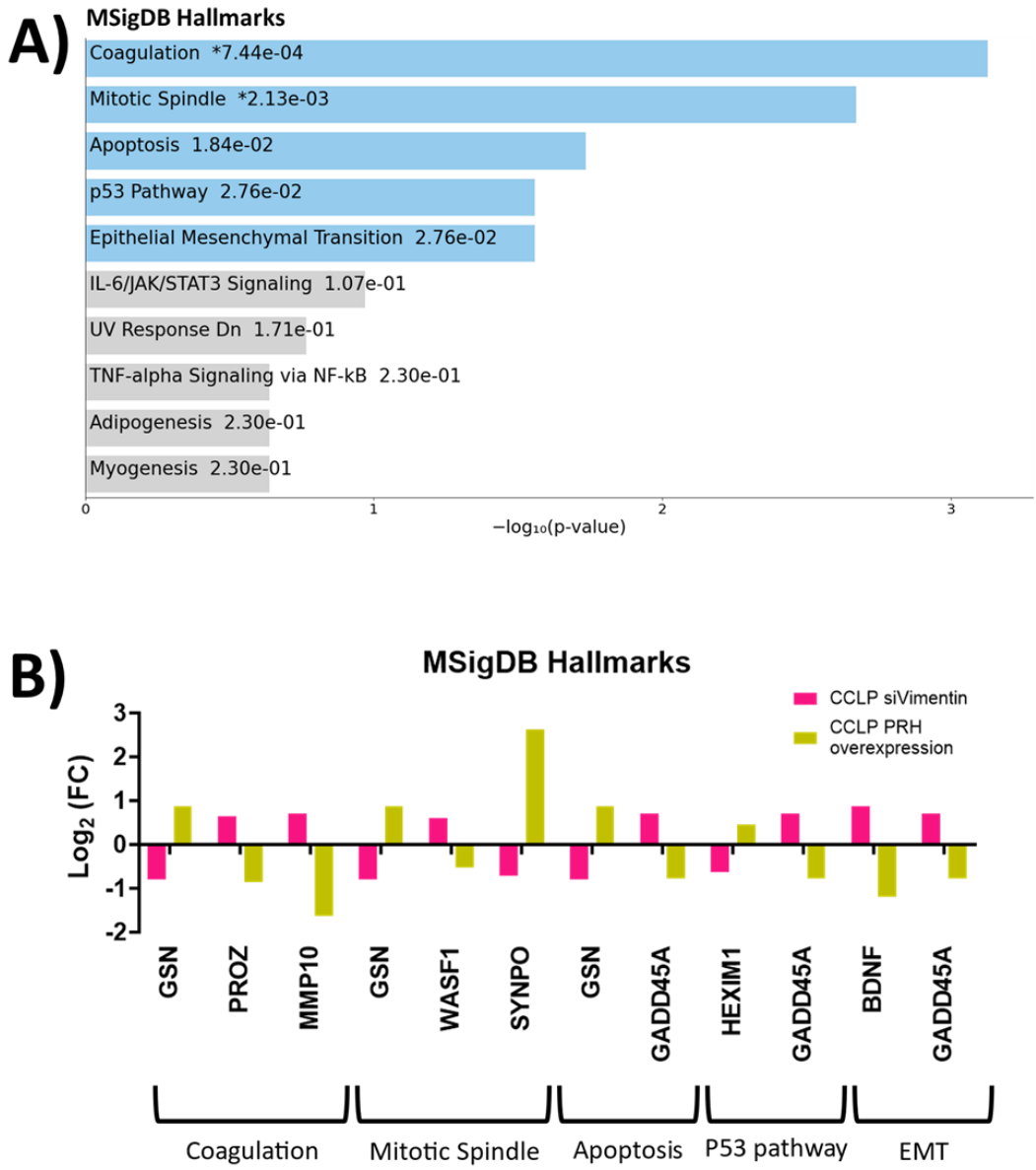


Figure 5-31 GSEA using MSigDB Hallmarks on 26 significant DEG co-regulated by Vimentin and PRH in opposite directions.

(A) The list represented the top 10 significantly enriched hallmarks in order of their -Log*10*(*p*-value). (B) Bar chart showing the expression fold change as (Log_2(FC)) of the enriched significant DEG in MSigDB Hallmarks.

To validate RNAseq findings, we further expanded on exploring the genes co-regulated by Vimentin and PRH that were enriched in the above hallmarks taking into consideration PRH ChIP peak distance and peak numbers. We highlighted the genes with PRH peaks within 100 Kb distance of the TSS as

well as quantify the number of PRH peaks at each gene (figure 5-32). We identified the genes *GADD45A*, *GSN*, *MMP10*, *STAT3* and *WASF1* to have between 1-10 peaks at different locations within 100 Kb distance of the TSS.

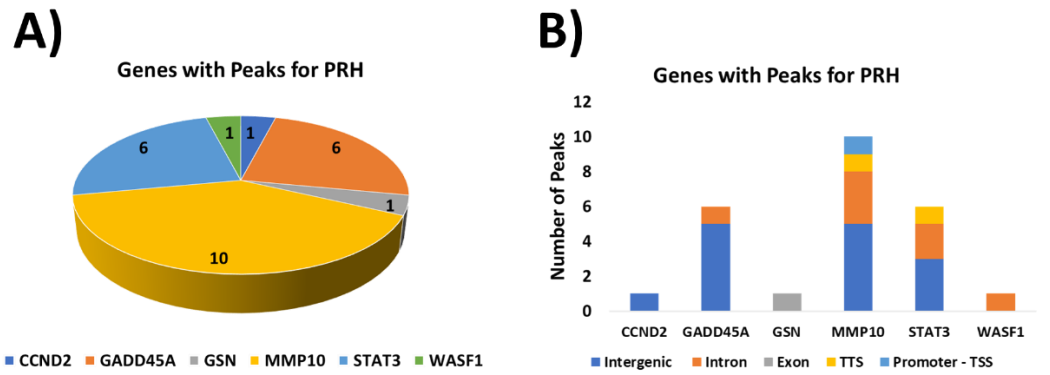


Figure 5-32 Significant DEG co-regulated by Vimentin and PRH and the number of PRH peaks.

(A) A pie chart representing the number of PRH peaks at each identified DEG. (B) The epigenomic distribution of PRH peaks at each identified DEG.

When we further narrowed the screening distance to within 2 Kb of TSS we found that *MMP10*, *STAT3* and *WASF1* genes each had a single peak for PRH at the promoter, intergenic, and intronic regions, respectively. Therefore, we decided to proceed with examining their expression using RT-qPCR. Since *GADD45A* and *GSN* genes were enriched in 3 hallmarks each, we added them as candidates for RT-qPCR validation.

We designed and optimised RT-qPCR primers for *GADD45A*, *GSN*, *MMP10*, *STAT3* as detailed in the method chapter. Out of all the primers, 3 primer pairs were qualified for RT-qPCR validation: *GSN*, *STAT3*, and *WASF1* (Results from RT-qPCR primer optimisation are available as supplementary data in the appendix). *GADD45A* and *MMP10* primers were left for future further

optimisation due to time restriction. We performed RT-qPCR in N=3 independent biological repeats using cDNA from CCLP siVimentin (figure 5-33).

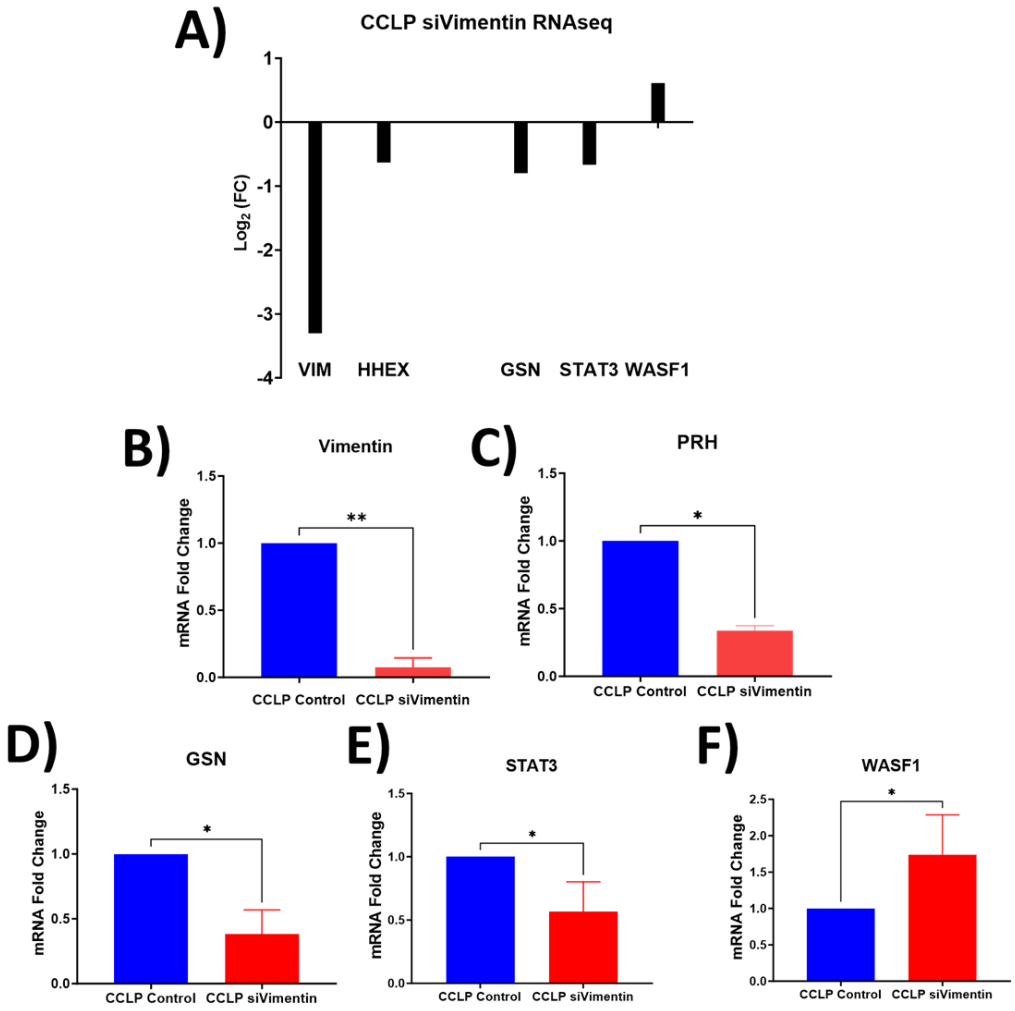


Figure 5-33 Validation of RNAseq findings using RT-qPCR.

(A) Genes Log_2(FC) from RNAseq data. (B) Vimentin mRNA relative fold expression (p= 0.0002). (C) PRH mRNA relative fold expression (p= 0.0353). (D) GSN mRNA relative fold expression (p=0. 0.0406). (E) STAT3 mRNA relative fold expression (p= 0.0428). (F) WASF1 mRNA relative fold expression (p= 0.0500). All RT-qPCR were performed in N=3 biological repeats and normalised to 1 for comparison. Two-tailed unpaired student t-test analysis was performed (for WASF1 One-tailed student t-test) [P ≤ 0.05 (*), P ≤ 0.01 (**), P ≤ 0.001 (***)]. Error bars represented SD.

Overall, the fold changes in gene expressions of *GSN*, *STAT3*, and *WASF1* on RT-qPCR agreed with their differential expression on CCLP siVimentin

RNAseq. *GSN* and *STAT3* were significantly downregulated when Vimentin was knocked down *GSN* (*p*=0.0406), *STAT3* (*p*=0.0428). Whereas *WASF1* was upregulated (*p*=0.0500).

5.3 Discussion

Vimentin is a major cytoskeletal protein that is widely recognised for its extensive structural roles in the cells (10). We discussed in the introduction the subcellular localisation of Vimentin in the cells and how that influences its functions in each compartment. For example, the dynamic cytoplasmic filaments of Vimentin contribute to the re-organisation of the cytoskeleton serving cell shape preservation (334, 353-355), EMT regulation (359), cell motility and migration (378), and cell proliferation (384). We also discussed the existence of Vimentin in the nuclear subregion of cells where it was suggested to regulate gene transcription, particularly *p21*^{Waf1} gene expression (348).

The aim of this chapter is to explore the potential role of Vimentin in conjunction with PRH to regulated gene expression in CCA. This was achieved by examining the roles of Vimentin in CCA; as a structural protein that influences CCA cells phenotype, and as a regulator of gene expression. Figure 5-34 shows the experimental model we followed in this chapter.

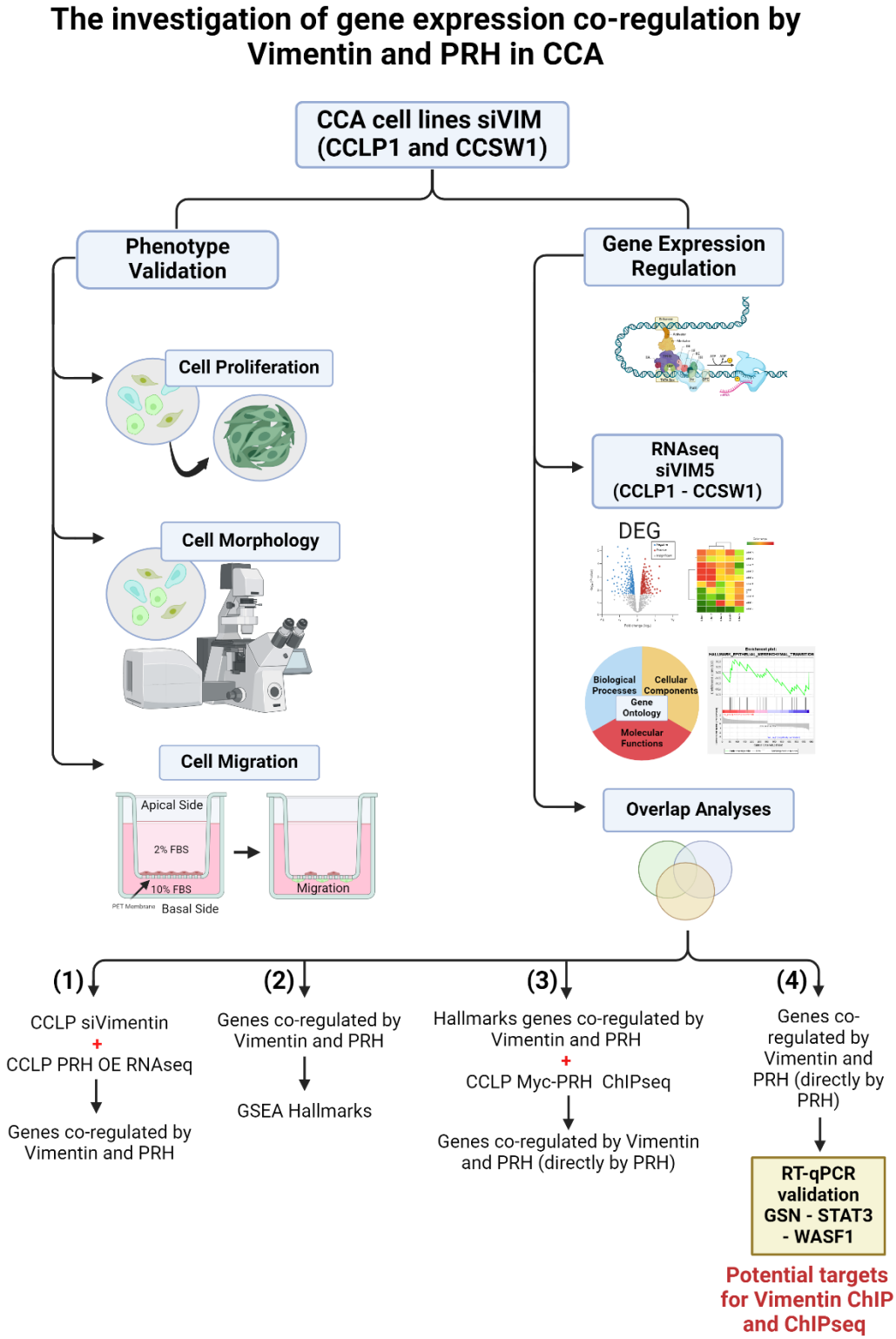


Figure 5-34 Flowchart analysis model for the investigation of gene co-regulation by Vimentin and PRH in CCA

Previously, Dr Ka Ying Lee (in her PhD work) drew attention to Vimentin as an important player in CCA when she first reported it as a protein partner for PRH in CCA cells (269). She subsequently revealed that both Vimentin and PRH promoted cell proliferation in CCA as well as regulated the expression of *CCND2* and *CDKN1B* cell cycle genes (269). In response, we followed up with co-localisation investigations to further confirm the spatial co-existence of Vimentin and PRH in CCA cells, particularly in the nucleus.

To fulfil our aim of exploring Vimentin as a structural protein in CCA, we needed to knockdown Vimentin in CCA cells (CCLP1 and CCSW1) and confirm that the resulting phenotype is in line to what is already known about Vimentin in the literature. We focused on the changes in cell proliferation, cell morphology, and cell migration. We used EdU incorporation assays, transwell migration (chemotaxis) assays, and microscopy imaging. To investigate the potential roles of Vimentin as a regulator of gene transcription, we used RNAseq analysis to identify the differentially expressed genes (DEG), analyse the biological impact of Vimentin knockdown on the functional enrichment of target genes in cancer using GO and GSEA, and validate the expression of target genes using RT-qPCR.

We demonstrated using GEPIA2 that *VIM* and *HHEX* genes are upregulated and exhibited a correlated expression in CCA samples. This indicates that Vimentin is also important for CCA tumorigenesis. We also showed RT-qPCR and protein expression results demonstrating that an efficient depletion of Vimentin in CCA cells resulted in the downregulation of PRH expression. This is in agreement with the data Dr Ka Ying Lee generated for her thesis (269).

PRH was found to regulate the expression of Vimentin at both mRNA and protein levels in CCA cells (215). This further strengthened the potential role for Vimentin in CCA.

As part of Vimentin knockdown phenotype verification, we showed that Vimentin depletion significantly reduced CCA cell proliferation. Vimentin expression has been seen to be induced in response to cell injury to promote cell proliferation via TGF-β signalling pathway (380). In addition, Vimentin importance for cell proliferation has been further illustrated in the impaired wound healing of (*VIM* $^{-/-}$) mice (300). Moreover, Vimentin was reported to regulate the expression of *p21* cyclin dependent kinase inhibitor in the nucleus (348). Our findings confirmed the initial work conducted by Dr Ka Ying Lee in which she reported both PRH and Vimentin promoted cell proliferation in CCA cells (269). It is important to say that we technically used a similar approach to suppress Vimentin expression in our cells and display its effect on cell proliferation. Our data were also consistent with the reduction in cell proliferation seen when Vimentin was depleted in HN12 cells (nodal metastatic squamous cell carcinoma) *in vitro* (482). The same study also confirmed the reduction in proliferation by reporting smaller tumour growth in xenograft models *in vivo* (482). Similarly, interfering with Vimentin expression in colorectal cancer cells, where AP-1 was overexpressed, was also reported to diminish the cell proliferation upregulated by AP1 in xenograft mice models (280). In contrast, Vimentin loss was linked to increased HepG2 proliferative activity in liver cancer via Rictor/AKT/β-catenin signalling pathway (384). Our finding do not align with a recently reported result of Vimentin knockdown in

KKU-214L5 CCA cell line (lung metastatic CCA cell line) where no influence on cell proliferation was seen (483). It worth mentioning that in the latter two studies, cell proliferation was assessed using a colorimetric cell viability assay relying on the metabolic reduction of WST-8 (water soluble tetrazolium salts) which is different from our approach in which we detected the number of cells incorporating synthetic EdU as they go through the S phase to reflect the change in cell proliferation (484). While both methods are end point assays, they have different principles. Although we combined cell counting with EdU assay in chapter 3, we did not conduct cumulative growth curve here because it required counting over two weeks. This would necessitate consecutive Vimentin knockdowns or stable Vimentin knockdown cells. It would be beneficial to combine several methods to measure the proliferative activity of the cells in the future to gain more comprehensive results. In general, the conflicting findings of Vimentin influence on cell proliferation has been reported several times in the literature perhaps indicating the complexity of the regulation of cell proliferation.

Beside cell proliferation, we chose to examine the phenotype of our Vimentin knockdown cells by addressing the changes in cell morphology. We followed similar approach from chapter 3 where we inspected the changes in cell morphology using brightfield imaging followed by an immunofluorescence imaging confirmation combined with shape descriptor scores. In chapter 4, we co-stained CCLP1 cells for PRH-Vimentin spatial localisation. Our results indicated that CCLP1 cells were spindle-like and mesenchymal in appearance displaying a filamentous expression of Vimentin in the cytoplasm representing

the cytoskeleton with some fibres extending to appear in the nuclear region. Our results in this chapter for CCLP1 control cells were consistent with our previous findings. Our data also showed that control CCSW1 cells displayed similar elongated morphology and a Vimentin cytoplasmic staining pattern to CCLP1 control cells. There was a striking reduction in cell area upon repressing Vimentin expression in CCA cells (CCLP siVimentin and CCSW siVimentin) on both brightfield and immunofluorescence imaging. We supported those visual findings with analysis of cells area and shape using the following parameters on FIJI: area, aspect ratio, and circularity (450). All three measurements indicated that we had significantly smaller, polygonal, circular cells (i.e. epithelial-like cells) with Vimentin knockdown compared to control cells (CCLP1 and CCSW1). Vimentin is a well-known canonical EMT marker (359), and it is important to preserve the shape of cells (334, 353-355). Our findings were consistent with the transitioning into epithelial-like morphology seen in KKU-214L5 cells when they were stained for Vimentin 48 hours post Vimentin depletion (483). As in spatial localisation imaging, the number of imaged cells in our experiment was small. However, we partially compensated that by including two cell lines. We were not able to show whether there are changes in other EMT markers using RT-qPCR or WB due to time restrictions and this would be an important area for future studies. However, overall, our results further strengthened the importance of Vimentin in preserving the elongated morphology of mesenchymal cells.

Quantitative analysis of Vimentin expression from confocal immunofluorescence imaging drew the attention to the discrepancy in

Vimentin expression across multiple techniques; Western blotting, RT-qPCR and immunofluorescence imaging. Although Vimentin reduction was overall consistent and statistically significant, results from RT-qPCR and Western blotting were more aligned compared to immunofluorescence. Visual inspection and intensity measurements of CCA cells after transfection revealed various degrees of Vimentin depletion. While the inherent difference in experimental principles of each technique is a major reason behind our observations, we cannot neglect the fact that Vimentin is major cytoskeletal protein that is challenging to completely deplete. All the methods we used required optimisation of the various steps in the protocols. In both Western blotting and RT-qPCR, Vimentin is detected in isolated total protein and mRNA from the cells, respectively. Densitometric quantification of protein bands are affected by the variability in antibody specificity and affinity (485). While mRNA quantification in RT-qPCR is dependent on sample integrity and primer efficiency (419, 420). In contrast, sample preparation in fluorescent imaging involves manipulating cell permeability to allow accessing target protein at various subcellular regions, antibodies specificity, and background of images (486). All the previous could serve as sources of variation in Vimentin biological data indicating the importance of integrating multiple techniques to assess the level of expression at protein or mRNA levels.

We observed a drastic change in cell shape and size following the loss of Vimentin expression in CCLP1 and CCSW1 cells and Vimentin is a major intermediate filament contributing to the structural maintenance of cells. The cells lost their elongated mesenchymal form to become rounder and more

epithelial like. This effect is not unique to Vimentin since reducing the levels of cytoskeletal proteins in cancer cells is expected to bring about profound changes in cancer cell shape and morphology. This comes in support of the distortion of cell morphology and changes in functional phenotypes reported after disorganising Tubulin microtubules using an antimitotic drug (Nocodazole), or when Tubulin- associated proteins (stathmin) underwent downregulation (487). Similarly, disrupting actin filaments via downregulating its binding proteins (e.g. cofilin) caused loss of cells polarity (488). Since manipulating the expression of Vimentin induced a significant change in cell shape, there is a possibility that it also affected the organisation of several other cytoskeletal proteins. We generated control cells using non-targeting siRNA to exclude the non-specific effect of siRNA in our approach. In the future, control cell lines where Vimentin depletion is restored are important to include as control cells to ensure that the morphological changes are related directly and solely to Vimentin expression. Additionally, adding wildtype cells in all the experiments sets the basal morphology without any interference from siRNA manipulation. Furthermore, it is important to avoid using cytoskeletal housekeeping genes or proteins to monitor cDNA input or protein loading. We detailed in the introduction how Vimentin-induced EMT is essential for many cellular biological processes regulated by different signalling pathways in which Vimentin is an upstream or downstream player (359). Examples of these biological processes include: cells adhesion (369), motility (365), migration (376, 377), and invasion (335). In this chapter, we demonstrated that Vimentin depletion in CCA cells significantly reduced cell migration in Boyden transwell

(chemotaxis) assays. Reduced migratory ability is a characteristic of epithelial cells (489). This is consistent with the roles of Vimentin we reported in the introduction highlighting the need for Vimentin to induce and regulate cell migration. Vimentin regulates directed migration via inducing actin filaments flow rates (375), and by stabilising microtubules through direct interaction (376, 377). Vimentin upregulation in gastric cancer was reported to increase transwell migration of GES-1 cells (379). Vimentin overexpression in CCA was associated with increased metastasis in CCA patients (490, 491). Recently, Vimentin suppression in KKU-214L5 (CCA cells) was showed to significantly decrease cell migration and invasion (483). Taking the data from the changes in cell morphology and cell migration together we confirmed that Vimentin is required to maintain the mesenchymal morphology which is essential to potentiate cell migration in CCA cells. Adding a transwell invasion assay for future follow up work would show us whether Vimentin depletion affects cells invasiveness. After we confirmed that Vimentin knockdown in CCA cells produced a phenotype that is in line with the published literature, we proceeded with transcriptomic RNAseq analysis. To the best of our knowledge, RNAseq of altered Vimentin has not been conducted in CCA cells so far. Here, we provided RNAseq results from depleting Vimentin in two CCA cell lines; CCLP1 and CCSW1. The results we obtained from CCLP siVimentin and CCSW siVimentin RNAseq analysis and their subsequent GO and GSEA analysis demonstrated that Vimentin has substantially greater roles in CCA than what is already reported in the literature. Vimentin differentially regulated the expression of genes that were also enriched in hallmarks associated with

cell proliferation, EMT, and cell migration. The EMT hallmark was the most significantly enriched pathway for both RNAseq data sets individually as well as for the common DEG between the two data sets. Altogether, the enriched pathways were associated with the change in phenotypes we observed *in vitro*.

While examining and validating every enriched gene identified in the EMT hallmark as well as the other hallmarks is important, this was beyond the time and capability of this PhD project. Instead, we identified 12 common genes (including VIM) enriched in the EMT hallmark in both RNAseq data sets. 9 genes were downregulated in both data sets (*GJA1, ITGAV, ITGB5, MMP2, SDC1, SERPINE2, SGCB, TGFBR3*, and *TPM4*), while (*SERPINE1* and *IGFBP3*) were upregulated in CCLP siVimentin RNAseq. Gap Junction Protein Alpha 1 (*GJA1*) encodes a gap junction protein which regulates the intercellular diffusion of low molecular weight materials (492). Its expression was linked to tumour progression and metastasis of gastric cancer (493). Integrin Subunit Beta 5 (*ITGB5)* encodes Integrin cell-surface receptors which are important for cell adhesion and cell signalling (492). Matrix Metallopeptidase 2 (*MMP2*) encodes gelatinase A which can breakdown components of the ECM and other molecules important for cell signalling *(492). Syndecan* 1 *(SDC1)* encodes a proteoglycan protein which is important for cell proliferation and migration. It also mediates cell binding and cell signalling *(492).* Serpin Family E Member 2 *(SERPINE2)* encodes protease inhibitor proteins (492). *SERPINE2* was reported to promote cancer cell proliferation and invasion in pancreatic and breast cancer (494, 495).

Sarcoglycan Beta (*SGCB)* encodes *Sarcoglycans transmembrane proteins* which facilitates the link between ECM and muscle cells cytoskeletal *(492).* Tropomyosin 4 (*TPM4*) encodes a member of tropomyosin proteins which is involved in the contractility of the cytoskeleton of cells (492). Integrin Subunit Alpha V (*ITGAV*) also encodes a protein member of the integrin family which may regulates angiogenesis and drives cancer progression (492). Transforming Growth Factor Beta Receptor 3 (*TGFBR3*) encodes the transforming growth factor (TGF)-beta type III receptor that is linked to cancer progression when its expression is suppressed (492). Serpin Family E Member 1 (*SERPINE1*) is an inhibitor of fibrinolysis, while Insulin Like Growth Factor Binding Protein 3 (*IGFBP3*) regulates growth of cells in response to Insulin Like Growth Factor. *ITGB5*, *MMP2* and *ITGAV* were linked to cancer progression and metastasis in lung cancer (496), and glioblastoma (497). The changes in expression of these genes comes in support of Vimentin regulating EMT genes related to several cellular processes that are important in cancer e.g. cells adhesion (369), migration (376, 377), and invasion (335).

Interestingly, RNAseq analysis revealed that Vimentin only exerted a significant differential expression on 19.51% of CCLP siVimentin DEGs and 43.80% of CCSW siVimentin DEGs. Nonetheless, this percentage of genes still showed the EMT hallmark as the most significantly enriched pathway, followed by p53 and TGF-β pathways. Interestingly, the effect of Vimentin on the expression of *CCND2*, *CDKN1A*, *CDKN1B*, and *CDKN2B* cell cycle genes in both RNAseq data was not strongly significant despite the crucial roles Vimentin has in regulating pathways related to cell division (312)., and to cell

proliferation (384). It is important to note that we sequenced RNA samples from cells harvested 48 hours after transfection. This was probably a factor affecting the count of genes with adjusted $p<0.05$ and the fold changes in expression. If possible, it would be useful to sequence RNA from samples harvested after prolonged Vimentin depletion to see if more genes would be significant and displaying greater fold changes in expression. This may influence the functional examination of gene enrichment via GO and GSEA to reveal more about the regulation of gene expression by Vimentin in CCA.

Overlap analysis between RNAseq data sets generated from altering Vimentin and PRH expression in a CCA cell line (CCLP siVimentin, CCLP PRH knockdown, and CCLP PRH overexpression) showed that Vimentin and PRH co-regulated gene expression in the same direction and in opposite directions. The subsequent GSEA using Hallmarks gene sets revealed the significant enrichment of genes in EMT, p53, mitotic spindle, and KRAS signalling pathways. This indicates that Vimentin and PRH co-regulate a common set of genes (in either direction) important for CCA tumour initiation and progression. This is consistent with previous data on PRH promoting the mesenchymal phenotype in CCA and regulating pathways like EMT, Wnt/β-catenin, TGF-β, (215).

We expanded on the investigation of gene co-regulation by Vimentin and PRH in opposite directions. We extracted the genes with adjacent PRH peaks according to CCLP PRH-Myc ChIPseq data. That further narrowed down the gene count to 26 genes. These genes were significantly enriched in the following hallmarks: coagulation, mitotic spindle, apoptosis, p53 pathway, and

EMT. We chose (*GSN, STAT3,* and *WASF1*) genes for further RT-qPCR validation. Results from RT-qPCR mRNA relative expression analysis agreed with the change in expression observed in RNAseq. *GSN* and *STAT3* were down regulated in CCLP siVimentin, while *WASF1* was upregulated. Gelsolin (*GSN*) regulates calcium binding to actin monomers in order to modulate their assembly into filaments *(492)*. Downregulation of GSN in gastric cancer was linked to poor survival (498). *STAT3* encodes a protein that is a member of the STAT family displaying a significant role in cell growth and apoptosis *(492).* WASP Family Member 1 (*WASF1*) regulates actin filaments organisation. WASF1 was reported to induce the proliferation and invasiveness of epithelial ovarian cancer *(492)*. The downregulation we observed in *STAT3* and *WASF1* mRNAs along with the phenotype produced in CCLP siVimentin cells are in support with Vimentin regulating genes important for tumour growth and invasion. These genes were upregulated in CCLP1 cells by PRH which correlates with the increase in cell proliferation and invasion seen when PRH was overexpressed in CCA cells (215). Taken these together, this highlights the intricate nature of gene expression and the interplay in their co-regulation by PRH and Vimentin. Collectively, the results from this chapter substantially expanded our understanding of gene regulation by Vimentin and PRH in CCA which was the main aim of this part of the project.

One significant limitation in this part of the project was the limited number of biological repeats for RNAseq in either CCLP siVimentin or CCSW siVimentin. However, we had the advantage of including two independent cell lines for all the experiments. Similarly, the number of cells imaged for the effect of

Vimentin loss on cell morphology is also limited. Nonetheless, the results were statistically significant. Yet, this still could be confirmed with more cell number. Another limitation was using siRNA transient depletion of Vimentin in our cells. This required multiple transfections followed by RT-qPCR and Western blotting validations, in addition to careful timing of the seeding for the different assays we performed. Using cell lines with stable Vimentin knockdown or Vimentin knockout would have avoided these problems and could be an aim of future work. The benefits would also extend to performing more assays to investigate other common Vimentin-related phenotypes such as apoptosis, adhesion, and invasion assays. In addition, we could investigate the influence of long-term Vimentin suppression on differential gene expression. Although we came across several interesting genes that are potentially regulated by Vimentin and PRH, time restriction was an obstacle against going further with more gene validation using RT-qPCR and western blotting. It is important to carry out these validations in the future. Lastly, integrating future work of Vimentin ChIPseq with the current work we performed in this chapter is needed to study the direct regulatory relationship between Vimentin and PRH in gene co-regulation. ChIP sequencing should be performed on ChIP of endogenous Vimentin in CCA cells, ChIP after altering Vimentin levels in the cells as well as ChIP in cells depleted of Vimentin followed by restoration of PRH expression which theoretically would restore Vimentin expression. Comparing ChIPseq results of before and after Vimentin levels manipulation can provide knowledge on the specificity of Vimentin bindings to target genes. In addition, this would explore the effect of Vimentin levels on chromatin

organisation and accessibility. Integrating results from Vimentin ChIP with PRH ChIP could identify potential target genes co-regulated by PRH and Vimentin via direct DNA binding in CCA. Moreover, overlapping analyses could unveil the effect of Vimentin levels on PRH occupancy at Vimentin target genes. The next step would be to design ChIP qPCR primers targeting the same identified genes co-regulated by PRH and Vimentin where we had ChIP peaks for PRH. There is a high chance that these genes would have peaks for Vimentin when Vimentin ChIP and ChIPseq experiments are carried out.

Chapter 6: Conclusion

6.1 Overview

The general focus of this thesis is to explore the roles of the PRH oncoprotein in cholangiocarcinoma (CCA). Due to the vital regulatory functions PRH displayed in normal development (217, 218) and in cancer (214, 215, 243, 245, 259, 260), an important pre-requisite was to generate immortalised cholangiocytes (AKN1 cells) expressing GFP-tagged PRH. Chapter 3 carried out and discussed the establishment of an in vitro 2D model for investigating PRH in CCA. To this end, such a system has been produced in parallel with our work and this project sought to advance the work and take the cells forward for further investigations (215). The initial consistency and efficiency of AKN1 cells expressing GFP- tagged PRH helped to unveil and track the changes brought about by PRH in the cells. For this reason, the cells were to be selected to acquire a monoclonal stable AKN1 cell line expressing GFP-tagged PRH and referred to as AKN1 GFP-PRH-Myc cells with a control AKN1 GFP cell line. Challenges to maintaining a constitutive expression of GFP-tagged PRH were not fully circumvented. Nonetheless, the current cell line allowed exploration in different areas of the project. In line with collaborators, it was found that AKN1 GFP-PRH-Myc cells, in early passages, are more proliferative. Additionally, partial EMT-like morphological changes in AKN1 GFP-PRH-Myc cells can induce cell migration and likely cell invasion as well. However, further work is warranted to examine the changes in EMT markers and cell morphology.

Following the acquisition of evidence on PRH inducing the transformation of AKN1 cells to tumour-like cells expressing Vimentin, investigating PRH and Vimentin spatial subcellular localisation in CCA cells is the main aim of chapter 4. This work was also a follow up on the discovery of Vimentin as a protein partner for PRH in CCA cells using CoIP and mass spectrometry (269). PRH is known to exist in the nuclear and cytoplasmic subcellular compartments of the cells (196, 241). Whenever PRH was altered or mis-localised, a collection of phenotypic cellular changes was detected (233). Such changes were related to either tumour cell progression or suppression (214, 245, 249, 250). For localisation investigation purpose, an important pre-requisite was to characterise PRH expression via immunofluorescence staining in CCA cells. This is essential to serve as a reference for Vimentin distribution in CCA cells. The exploitation of coupled confocal and PALM dSTORM imaging techniques is a common approach to explore spatial localisation of proteins at high and super-resolution levels. In fact, challenges of detecting and quantifying PRH-Vimentin adjacent localisation were partially surpassed by utilising PALM dSTORM. In agreement with PRH staining in published works, PRH in CCA cells depicted nuclear and cytoplasmic distributions assuming a mix of foci, speckles, and homogenous form of expression. Also, in support of its role as a transcription factor regulating gene expression, there was an adjacency between PRH and euchromatin staining. In addition, there was a similarity in expression pattern between PRH-DNA staining and RNAPII-DNA staining. The clustering of PRH molecules in the nucleus resembled the clustering of RNAPII which is away from densely stained DNA. On that basis, Vimentin-

nuclear localisation in relation to PRH was taken forward. There was an insightful discovery of spatial proximity between PRH and Vimentin in the nucleus. In this project, the localisation investigations at the single molecule level were limited to the nuclear compartments of CCA cells.

PRH-Vimentin localisation was attempted in AKN1 GFP-PRH-Myc at high resolution level (confocal imaging). In chapter 3, PRH was found to induce the expression of Vimentin on immunofluorescence imaging. The very low level of PRH-Vimentin spatial localisation is attributed to several reasons including the level of Vimentin induction by PRH in the cells. This arises from the level of GFP-tagged PRH in AKN1 GFP-PRH-Myc cells.

In line with previous studies, GFP-tagged PRH in AKN1 GFP-PRH-Myc is amenable to GFP stand-alone expression or cleavage interfering with long-term expression and phenotype observations (475). Applying FCS fluctuation analysis on GFP and GFP-tagged PRH in AKN1 cells drew the attention to the complexity of GFP-tagged PRH behaviour in AKN1 cells. GFP and GFP-tagged PRH dynamics revealed several forms of expression in AKN1 cells proving the occurrence of GFP-tagged PRH cleavage in our cells. Hence, GFP-tagged PRH in AKN1 GFP-PRH-Myc cells warrants the re-assessment of plasmid integrity as well as the need to investigate GFP-tagged PRH dynamics. Studying AKN1 GFP-PRH-Myc cells remains a worthy goal as the workup for phenotype characterisation in AKN1 GFP-PRH-Myc cells was conducted in a single clone due to time restrictions.

Recent studies suggested that Vimentin could bind to DNA and regulate gene expression (348). Extensive research has taken place to investigate the role

of Vimentin in the cytoplasm owing to its predominant cytoplasmic expression (317, 323). The aim of chapter 5 is to explore the gene regulation potential of Vimentin in CCA in conjunction with PRH. The prerequisite was to characterise the phenotype from altering Vimentin in CCA cells. Vimentin siRNA depleted its expression 48 hours after transfection accounting for over 95% reduction in Vimentin at mRNA level and approximately 90% reduction at protein level. Vimentin brought about phenotypic changes aligned with the literature. There was a significant reduction in the proliferative activity of the cells. This also came in support of pilot experiments conducted on the regulation of proliferation and cell cycle genes by PRH and Vimentin using RT-qPCR (269). Concurrent acquisition of epithelial-like morphology in CCA cells upon Vimentin repression resulted in decreasing the migratory activity of the cells. As discussed in chapter 5, a reduction in cell migration is a phenomenon of epithelial cells. A similar finding was reported in a study conducted to investigate the importance of Vimentin to cell migration in KKU-214L5 CCA cell line (483).

PRH and Vimentin are both highly expressed in CCA tissues, and their expression is correlated according to GEPIA2 and OncoDB. PRH is a regulator of Vimentin expression in CCA (215), and together, they regulated important cell cycle genes (*CCND2* and *CDKN1B*) using RT-qPCR (269). Vimentin downregulation brought about PRH expression changes at mRNA and protein levels shown using RT-qPCR and immunofluorescence imaging, respectively.

Given the phenotypes identified, follow up with RNA sequencing and analysis was warranted. Pathways and hallmarks enrichment in GO and GSEA analysis of Vimentin knockdown RNAseq came in support of the observed phenotypes. EMT hallmark genes are the most significantly enriched differentially expressed genes. Several enriched EMT genes were common between RNAseq data of (CCLP1 and CCSW1) CCA cells. Among the common enriched genes, the expression of *GJA1*, *ITGAV*, *ITGB5*, *MMP2*, and *SERPINE2* genes was linked to the progression of breast cancer (494, 495), gastric cancer (493), lung cancer (496), and glioblastoma (497). The downregulation in their expression subsequent to Vimentin downregulation may explain the suppression of proliferation and migration in CCA cells. Despite the relatively low overlap between DEG counts in RNAseq of the two CCA cell lines, there are abundant similarities in the regulated pathways and hallmarks worth of detailed examination. However, time restriction hindered further analysis for this thesis. This could be taken forward in the future to further study the role of Vimentin in CCA.

In parallel, analysis of published CCA PRH overexpression and knockdown RNAseq as well as ChIPseq data allowed the examination of the potential for gene expression co-regulation by PRH and Vimentin in CCA. Directionality of regulation is essential to highlight when gene co-regulation is surveyed. Although screening for gene expression co-regulation in both similar and opposite directions was commenced, only gene co-regulation in opposite directions was taken further. The relatively large number of common genes in the latter situation implied higher chances of identifying interesting target

genes. *GSN*, *STAT3*, and *WASF1* were some of the co-regulated genes by PRH and Vimentin in opposite direction and these were validated using RT-qPCR. These genes have been recommended as good potential targets to investigate in Vimentin ChIP and ChIPseq experiments.

To explore the importance of the PRH-Vimentin interaction in regulating the common target genes and their associated functional pathways, further experimental work is warranted to exclude unrelated shared functions. Sequential ChIP at common target genes using antibodies against PRH followed by an antibody for Vimentin can suggest PRH-Vimentin co-occupancy is required to regulate a specific target gene. Additionally, transcriptomic analysis of differentially expressed genes upon depleting both PRH and Vimentin simultaneously, or depletion followed by restoration of expression, can further support that the shared biological functions are more related to the protein interaction rather than individually regulated.

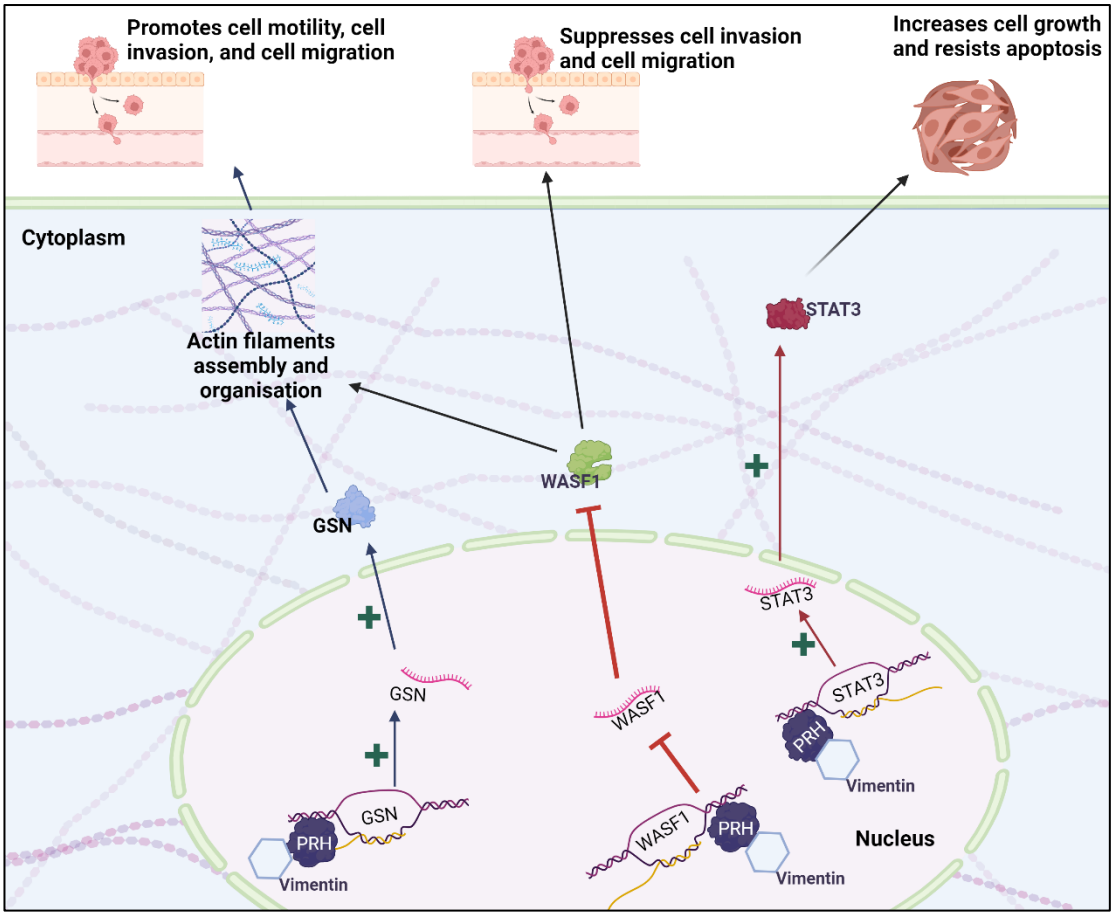


Figure 6-1 Schematic illustration of gene regulation by PRH and Vimentin.

Conclusion Vimentin knockdown down regulates the expression of GSN and STAT3 in CCLP1 cells and upregulates the expression of WASF1. All three genes are directly regulated by PRH.

6.2 Conclusion

Overall, the results obtained in this project expand our current understanding of the role of PRH in CCA. We characterised the tumorigenic effects of PRH expression in immortalised cholangiocytes. We extended the benefits from the generated cell line to exploit the cells for visual characterisation of PRH expression as well as PRH possibly inducing morphological changes that resembles partial EMT. We provided data of PRH and Vimentin nuclear clustering and subcellular localisation in CCA cells. We also revealed the

substantial roles for Vimentin in CCA cells and revealed new insights into the regulation of gene expression in CCA by PRH and Vimentin. The combination of subcellular PRH and Vimentin clustering and organisation along with the identified potential target genes co-regulated by PRH and Vimentin could serve as new targets for future interventional studies.

Appendix

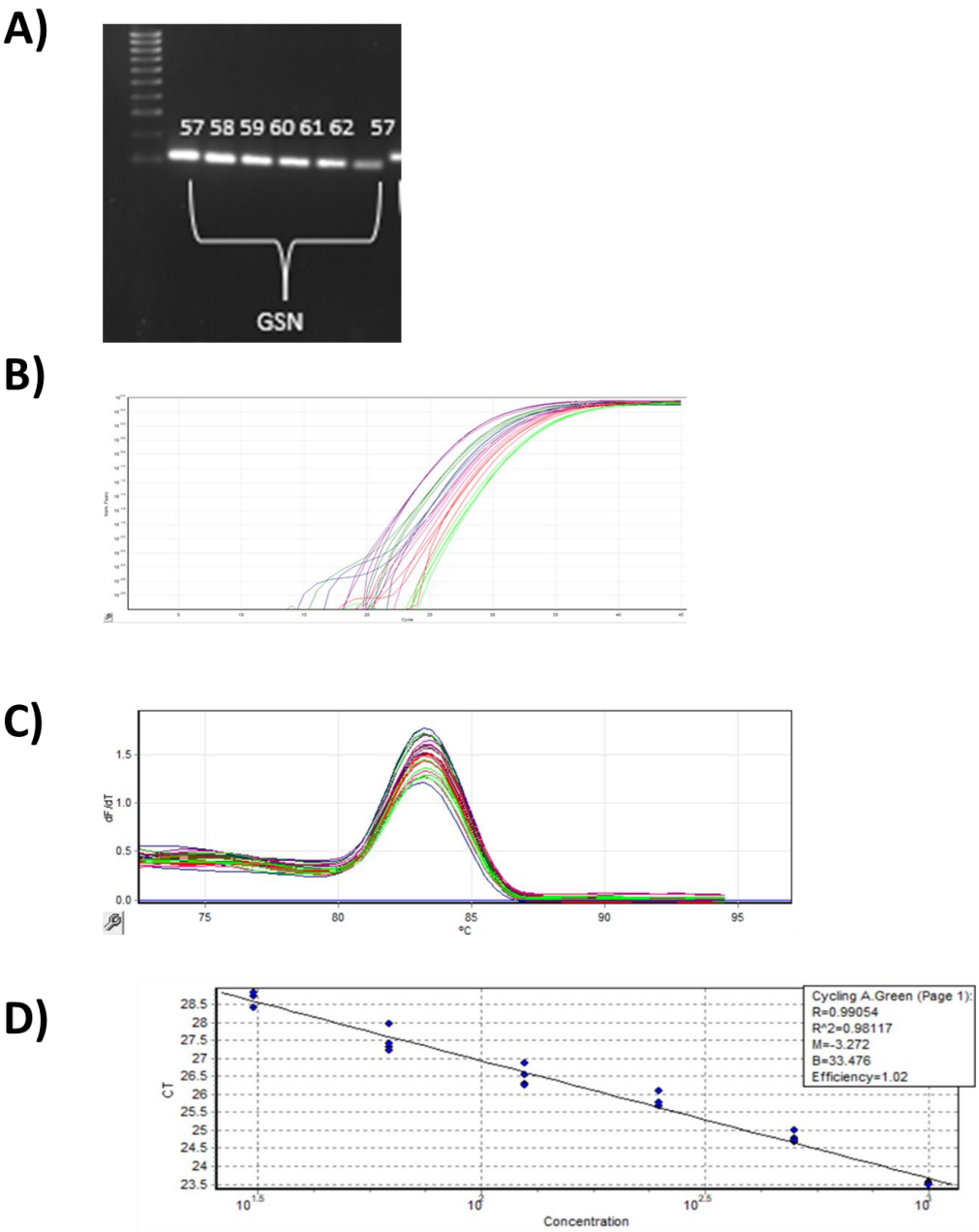


Figure 7-1 GSN RT-qPCR primer optimisation.

A. DNA gel electrophoresis for GSN Ta optimisation on gradient PCR.
B. Amplification plot of serial dilution cDNA tested for GSN primers efficiency.
C. RT-qPCR melting curve showing a single peak for GSN amplification.
D. GSN primers standard curve.

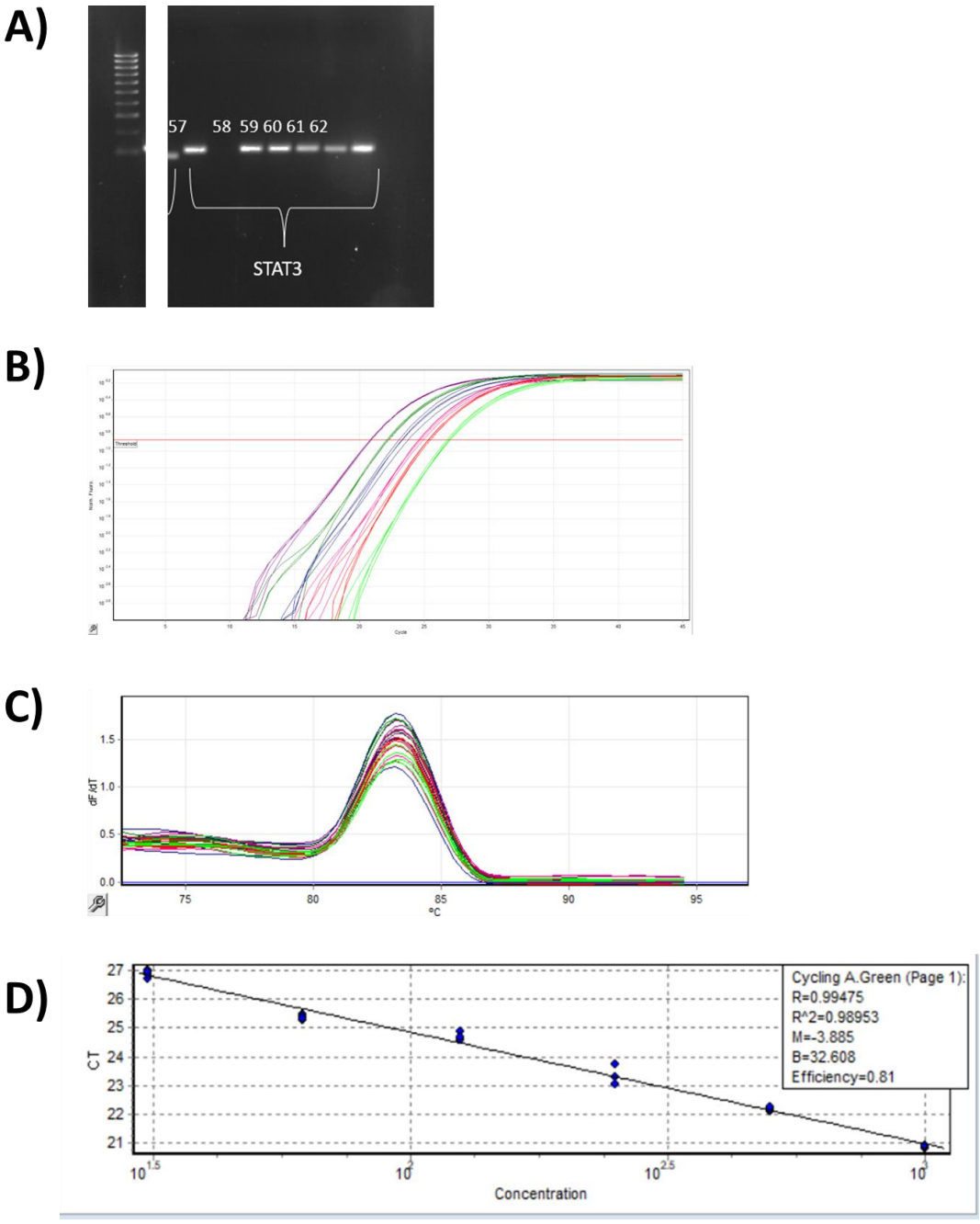


Figure 7-2 STAT3 RT-qPCR primer optimisation.

A. DNA gel electrophoresis for STAT3 Ta optimisation on gradient PCR.

B. Amplification plot of serial dilution cDNA tested for STAT3 primers efficiency.

C. RT-qPCR melting curve showing a single peak for STAT3 amplification.

D. STAT3 primers standard curve.

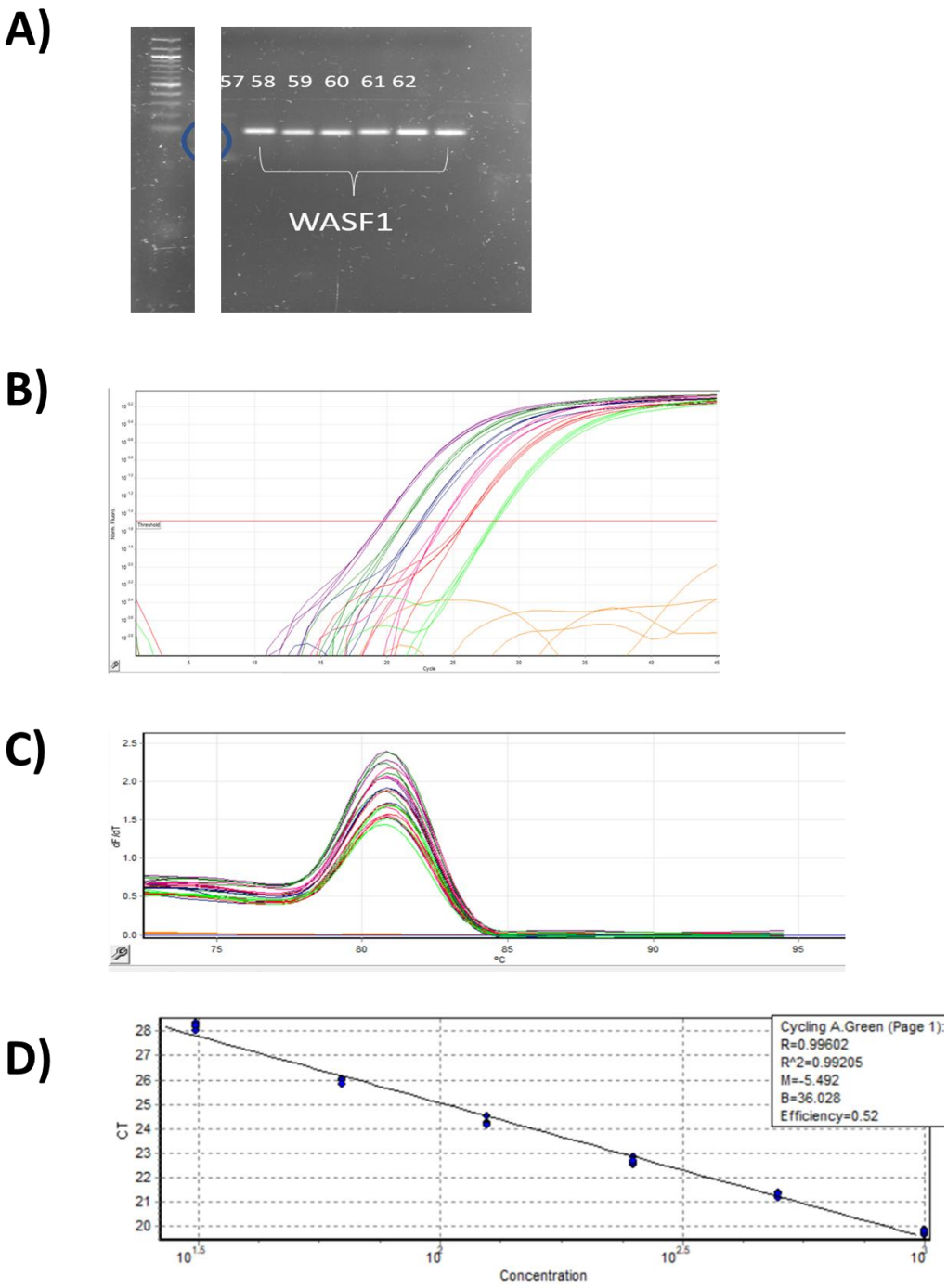


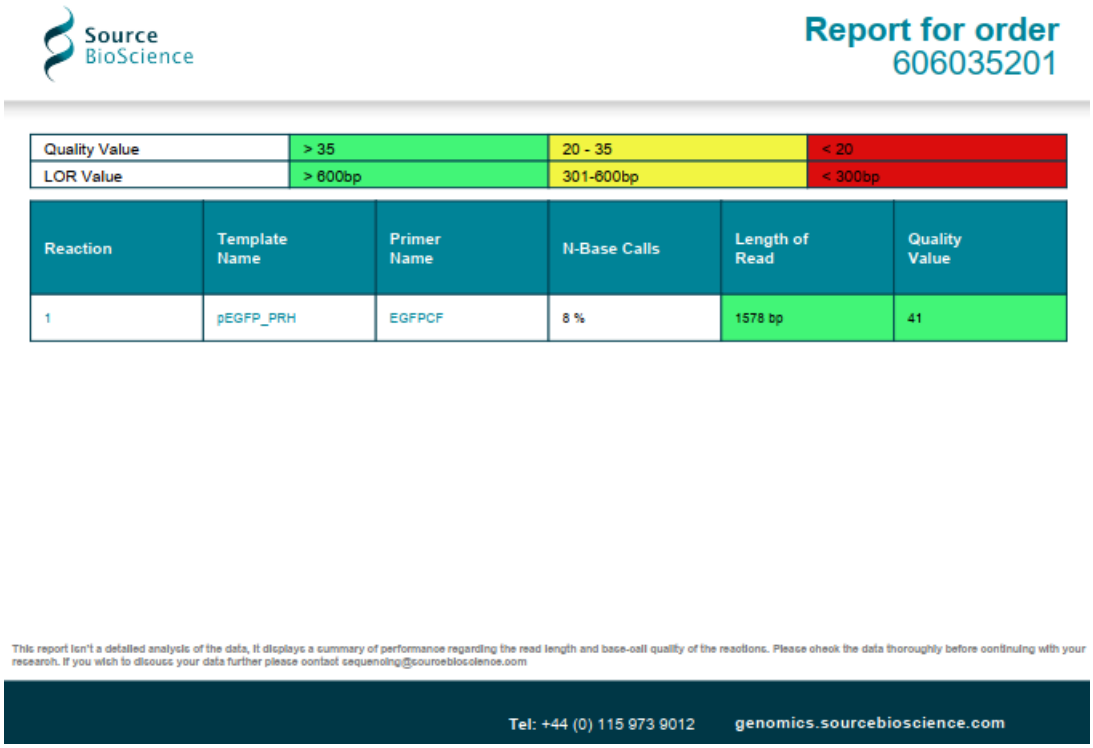
Figure 7-3 WASF1 RT-qPCR primer optimisation.

A. DNA gel electrophoresis for WASF1 Ta optimisation on gradient PCR.

B. Amplification plot of serial dilution cDNA tested for WASF1 primers efficiency.

C. RT-qPCR melting curve showing a single peak for WASF1 amplification.

D. WASF1 primers standard curve.



Source BioScience

Report for order 606035201

Quality Value	> 35	20 - 35	< 20
LOR Value	> 600bp	301-600bp	< 300bp

Reaction	Template Name	Primer Name	N-Base Calls	Length of Read	Quality Value
1	pEGFP_PRH	EGFPCF	8 %	1578 bp	41

This report isn't a detailed analysis of the data, it displays a summary of performance regarding the read length and base-call quality of the reactions. Please check the data thoroughly before continuing with your research. If you wish to discuss your data further please contact sequencing@sourcebioscience.com

Tel: +44 (0) 115 973 9012 genomics.sourcebioscience.com

Figure 7-4 Plasmid sequencing repor.t

Sequencing:

NNNNNNNNNNNCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGA
TCTCGAGCTCAAGCTTCGAATTCGGGCGGAGC
CATGCAGTACCCGCACCCCGGGCCGGCGGCGGGCGCCGTGGGGGTG
CCGCTGTACGCGCCCACGCCGCTGCTGCAACCCG
CACACCCGACGCCCTTTTACATCGAGGACATCCTGGGCCGCGGGCCC
GCCGCGCCCACGCCCGCCCCCACGCTGCCGTCC
CCCAACTCCTCCTTCACCAGCCTCGTGTCCCCCTACCGGACCCCGGTG
TACGAGCCCACGCCGATCCATCCAGCCTTCTC
GCACCACTCCGCCGCCGCGCTGGCCGCTGCCTACGGACCCGGCGGCT
TCGGGGGCCCTCTGTACCCCTTCCCGCGGACGG
TGAACGACTACACGCACGCCCTGCTCCGCCACGACCCCCTGGGCAAAC
CTCTACTCTGGAGCCCCTTCTTGCAGAGGCCT
CTGCATAAAAGGAAAGGCGGCCAGGTGAGATTCTCCAACGACCAGACC
ATCGAGCTGGAGAAGAAATTCGAGACGCAGAA
ATATCTCTCTCCGCCCGAGAGGAAGCGTCTGGCCAAGATGCTGCAGCT
CAGCGAGAGACAGGTCAAAACCTGGTTTCAGA
ATCGACGCGCTAAATGGAGGAGACTAAAACAGGAGAACCCTCAAAGCA
ATAAAAAAGAAGAACTGGAAAGTTTGGACAGT
TCCTGTGATCAGAGGCAAGATTTGCCCAGTGAACAGAATAAAGGTGCTT
CTTTGGATAGCTCTCAATGTTCGCCCTCCCC
TGCCTCCCAGGAAGACCTTGAATCAGAGATTTCAGAGGATTCTGATCAG
GAAGTGGACATTGAGGGCGATAAAAGCTATT

TTAATGCTGGATGATGACCACTGGCATTGGCATGTTCAGAAAACTGGAT
TTAGGAATAATGTTTTGCTACAGAAAATCTT
CATAGAAGAACTGGAAGGCTATATAAGAAAGGGAATCAATTCTCTGGTA
TTCTGGAAACCTAAAAATATTTGGTGCACTG
CTCAATTAACAAACCTACATGGAGACCTTAATTTTGACTTAACAAATAGT
TTATGTACTGCTCTTAGGTTGTTTTGATAA
AGTGACATTATAGTGATTAAATTCTTCCCCCTTTAAAAAAACAGTTAGTG
GTTTTCACTATTTATAAAAAATTAATTTTG
AACTTTTTGTTAAATTTTTAAGTTANAGCTTTAAAGGTTTTAATAGGACCT
TCTTGAACGACTTTTCTGTAATCTGTTTA
TCTCCCACTTAATGGAAAGGCAAAGGGGTACCNNNGGGCCCCGGGATC
CACCGGATCNAGATAACTGANNTAANTCNNCC
ATACCANATTTGTNNNGGTTTTACTTGCTTNAAAAAACNNNCCCACCNN
NCCCCCNNANNCNNNAANNNAAAAAAANNNN
NNNANNNNTTTTTTNNNANTTTTNTTTTNNNNNCCTTNNANNGGNNNNA
NANAAAAANNNANNNNNCCNNNAANTNTTNC
ANAANANAAAANTNNNNTTTNNNNNNNNNNNNNNNNNNGNNNNNNNN
CNCCCNNNAA

References

1. National Cancer Institute (NIC), Understanding Cancer 2015 [Available from: https://www.cancer.gov/about-cancer/understanding/what-is-cancer.
2. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2024;74(3):229-63.
3. American Cancer Association (ACA), World Cancer Day 2012 [Available from: http://www.worldcancerday.org/sites/default/files/private/UICC_Final_web_print%281%29.zip.
4. Chen S, Cao Z, Prettner K, Kuhn M, Yang J, Jiao L, et al. Estimates and Projections of the Global Economic Cost of 29 Cancers in 204 Countries and Territories From 2020 to 2050. JAMA Oncology. 2023;9(4):465-72.
5. Vogelstein B, Kinzler KW. The multistep nature of cancer. Trends in genetics : TIG. 1993;9(4):138-41.
6. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.
7. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022;12(1):31-46.
8. Kovalchuk I. Chapter 1 - Genome Stability: An Evolutionary Perspective. In: Kovalchuk I, Kovalchuk O, editors. Genome Stability. Boston: Academic Press; 2016. p. 1-18.
9. Cooper GMA, K. W. . The cell: A molecular approach. New York, NY: Oxford University Press; 2019.
10. O'Connor CMAJU. Essentials of Cell Biology. Cambridge: MA: NPG Education; 2010.
11. Elmore S. Apoptosis: a review of programmed cell death. Toxicologic pathology. 2007;35(4):495-516.
12. Berdasco M, Esteller M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. Developmental cell. 2010;19(5):698-711.
13. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61(5):759-67.
14. Lord CJ, Ashworth A. The DNA damage response and cancer therapy. Nature. 2012;481(7381):287-94.
15. Clarke TL, Mostoslavsky R. DNA repair as a shared hallmark in cancer and ageing. Mol Oncol. 2022;16(18):3352-79.
16. Shoshani O, Brunner SF, Yaeger R, Ly P, Nechemia-Arbely Y, Kim DH, et al. Chromothripsis drives the evolution of gene amplification in cancer. Nature. 2021;591(7848):137-41.
17. Korbel Jan O, Campbell Peter J. Criteria for Inference of Chromothripsis in Cancer Genomes. Cell. 2013;152(6):1226-36.
18. Junttila MR, Evan GI. p53 — a Jack of all trades but master of none. Nature Reviews Cancer. 2009;9:821.

19. Adams JM, Cory S. The Bcl-2 apoptotic switch in cancer development and therapy. Oncogene. 2007;26(9):1324.
20. Lowe SW, Cepero E, Evan G. Intrinsic tumour suppression. Nature. 2004;432(7015):307-15.
21. Qian S, Wei Z, Yang W, Huang J, Yang Y, Wang J. The role of BCL-2 family proteins in regulating apoptosis and cancer therapy. Front Oncol. 2022;12:985363.
22. Galluzzi L, Kroemer G. Necroptosis: a specialized pathway of programmed necrosis. Cell. 2008;135(7):1161-3.
23. Festjens N, Vanden Berghe T, Vandenabeele P. Necrosis, a well-orchestrated form of cell demise: signalling cascades, important mediators and concomitant immune response. Biochimica et biophysica acta. 2006;1757(9-10):1371-87.
24. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. Cell. 2010;140(6):883-99.
25. Zapletal E, Vasiljevic T, Busson P, Matijevic Glavan T. Dialog beyond the Grave: Necrosis in the Tumor Microenvironment and Its Contribution to Tumor Growth. International Journal of Molecular Sciences. 2023;24(6):5278.
26. Levine B, Kroemer G. Autophagy in the pathogenesis of disease. Cell. 2008;132(1):27-42.
27. White E, DiPaola RS. The double-edged sword of autophagy modulation in cancer. Clinical cancer research. 2009;15(17):5308-16.
28. Gewirtz DA. The four faces of autophagy: implications for cancer therapy. Cancer research. 2014;74(3):647-51.
29. Witsch E, Sela M, Yarden Y. Roles for growth factors in cancer progression. Physiology. 2010;25(2):85-101.
30. Darzynkiewicz Z, Gong J, Juan G, Ardelt B, Traganos F. Cytometry of cyclin proteins. Cytometry. 1996;25(1):1-13.
31. Fisher RP. The CDK Network: Linking Cycles of Cell Division and Gene Expression. Genes Cancer. 2012;3(11-12):731-8.
32. Simmons Kovacs LA, Orlando DA, Haase SB. Transcription networks and cyclin/CDKs: the yin and yang of cell cycle oscillators. Cell Cycle. 2008;7(17):2626-9.
33. Gao X, Leone GW, Wang H. Chapter Four - Cyclin D-CDK4/6 functions in cancer. In: Tew KD, Fisher PB, editors. Advances in Cancer Research. 148: Academic Press; 2020. p. 147-69.
34. Kim S, Leong A, Kim M, Yang HW. CDK4/6 initiates Rb inactivation and CDK2 activity coordinates cell-cycle commitment and G1/S transition. Scientific Reports. 2022;12(1):16810.
35. Gavet O, Pines J. Activation of cyclin B1–Cdk1 synchronizes events in the nucleus and the cytoplasm at mitosis. Journal of Cell Biology. 2010;189(2):247-59.
36. Barbiero M, Cirillo L, Veerapathiran S, Coates C, Ruffilli C, Pines J. Cell cycle-dependent binding between Cyclin B1 and Cdk1 revealed by time-resolved fluorescence correlation spectroscopy. Open Biology. 2022;12(6):220057.
37. Glotzer M, Murray AW, Kirschner MW. Cyclin is degraded by the ubiquitin pathway. Nature. 1991;349(6305):132-8.
38. Jin C, Samuelson L, Cui CB, Sun Y, Gerber DA. MAPK/ERK and Wnt/β-Catenin pathways are synergistically involved in proliferation of Sca-1 positive hepatic

progenitor cells. Biochemical and biophysical research communications. 2011;409(4):803-7.
39. Wee P, Wang Z. Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways. Cancers. 2017;9(5):52.
40. Sherr CJ, Roberts JM. Inhibitors of mammalian G1 cyclin-dependent kinases. Genes & development. 1995;9(10):1149-63.
41. Pavletich NP. Mechanisms of cyclin-dependent kinase regulation: structures of cdks, their cyclin activators, and cip and INK4 inhibitors1, 211998 Awardee, Walter J. Johnson Prize for the Encouragement of Research in the Life Sciences2Edited by P. E. Wright. Journal of Molecular Biology. 1999;287(5):821-8.
42. Russo AA, Tong L, Lee J-O, Jeffrey PD, Pavletich NP. Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. Nature. 1998;395(6699):237-43.
43. Ortega S, Malumbres M, Barbacid M. Cyclin D-dependent kinases, INK4 inhibitors and cancer. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer. 2002;1602(1):73-87.
44. Kamb A, Gruis NA, Weaver-Feldhaus J, Liu Q, Harshman K, Tavtigian SV, et al. A Cell Cycle Regulator Potentially Involved in Genesis of Many Tumor Types. Science. 1994;264(5157):436-40.
45. Curto M, Cole BK, Lallemand D, Liu C-H, McClatchey AI. Contact-dependent inhibition of EGFR signaling by Nf2/Merlin. J Cell Biol. 2007;177(5):893-903.
46. Dillekås H, Rogers MS, Straume O. Are 90% of deaths from cancer caused by metastases? Cancer Medicine. 2019;8(12):5574-6.
47. Bielenberg DR, Zetter BR. The Contribution of Angiogenesis to the Process of Metastasis. The Cancer Journal. 2015;21(4).
48. Lambert AW, Pattabiraman DR, Weinberg RA. Emerging biological principles of metastasis. Cell. 2017;168(4):670-91.
49. Zhuyan J, Chen M, Zhu T, Bao X, Zhen T, Xing K, et al. Critical steps to tumor metastasis: alterations of tumor microenvironment and extracellular matrix in the formation of pre-metastatic and metastatic niche. Cell & Bioscience. 2020;10(1):89.
50. Fidler IJ. Metastasis: Quantitative Analysis of Distribution and Fate of Tumor Emboli Labeled With 125I-5-Iodo-2′ -deoxyuridine23. JNCI: Journal of the National Cancer Institute. 1970;45(4):773-82.
51. Butler TP, Gullino PM. Quantitation of cell shedding into efferent blood of mammary adenocarcinoma. Cancer Res. 1975;35(3):512-6.
52. Gao D, Joshi N, Choi H, Ryu S, Hahn M, Catena R, et al. Myeloid progenitor cells in the premetastatic lung promote metastases by inducing mesenchymal to epithelial transition. Cancer Res. 2012;72(6):1384-94.
53. Guan X. Cancer metastases: challenges and opportunities. Acta Pharm Sin B. 2015;5(5):402-18.
54. Hanahan D, Folkman J. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. Cell. 1996;86(3):353-64.
55. Ferrara N, Henzel WJ. Pituitary follicular cells secrete a novel heparin-binding growth factor specific for vascular endothelial cells. Biochemical and biophysical research communications. 1989;161(2):851-8.

56. Liu Y, Li Y, Wang Y, Lin C, Zhang D, Chen J, et al. Recent progress on vascular endothelial growth factor receptor inhibitors with dual targeting capabilities for tumor therapy. Journal of Hematology & Oncology. 2022;15(1):89.
57. Rastinejad F, Polverini PJ, Bouck NP. Regulation of the activity of a new inhibitor of angiogenesis by a cancer suppressor gene. Cell. 1989;56(3):345-55.
58. Good DJ, Polverini PJ, Rastinejad F, Le Beau MM, Lemons RS, Frazier WA, et al. A tumor suppressor-dependent inhibitor of angiogenesis is immunologically and functionally indistinguishable from a fragment of thrombospondin. Proceedings of the National Academy of Sciences. 1990;87(17):6624-8.
59. Dameron KM, Volpert OV, Tainsky MA, Bouck N. Control of angiogenesis in fibroblasts by p53 regulation of thrombospondin-1. Science. 1994;265(5178):1582-4.
60. Mescher AL. Junqueira's Basic Histology Text and Atlas. United States: McGraw-Hill Education; 2013.
61. Walko G, Castañón MJ, Wiche G. Molecular architecture and function of the hemidesmosome. Cell Tissue Res. 2015;360(2):363-78.
62. Legerstee K, Houtsmuller AB. A Layered View on Focal Adhesions. Biology (Basel). 2021;10(11).
63. Giepmans BN, van IJzendoorn SC. Epithelial cell–cell junctions and plasma membrane domains. Biochimica et Biophysica Acta (BBA)-Biomembranes. 2009;1788(4):820-31.
64. Plotnikov SV, Pasapera AM, Sabass B, Waterman CM. Force fluctuations within focal adhesions mediate ECM-rigidity sensing to guide directed cell migration. Cell. 2012;151(7):1513-27.
65. Bouchard V, Demers MJ, Thibodeau S, Laquerre V, Fujita N, Tsuruo T, et al. Fak/Src signaling in human intestinal epithelial cell survival and anoikis: differentiation state-specific uncoupling with the PI3-K/Akt-1 and MEK/Erk pathways. J Cell Physiol. 2007;212(3):717-28.
66. Cicchini C, Laudadio I, Citarella F, Corazzari M, Steindler C, Conigliaro A, et al. TGFbeta-induced EMT requires focal adhesion kinase (FAK) signaling. Exp Cell Res. 2008;314(1):143-52.
67. Bakir B, Chiarella AM, Pitarresi JR, Rustgi AK. EMT, MET, Plasticity, and Tumor Metastasis. Trends in Cell Biology. 2020;30(10):764-76.
68. Pastushenko I, Blanpain C. EMT Transition States during Tumor Progression and Metastasis. Trends Cell Biol. 2019;29(3):212-26.
69. Geiger B, Ayalon O. Cadherins. Annual review of cell biology. 1992;8(1):307-32.
70. Corso G, Intra M, Trentin C, Veronesi P, Galimberti V. CDH1 germline mutations and hereditary lobular breast cancer. Familial Cancer. 2016;15(2):215-9.
71. Hay ED. An overview of epithelio-mesenchymal transformation. Cells Tissues Organs. 1995;154(1):8-20.
72. Kalluri R, Neilson EG. Epithelial-mesenchymal transition and its implications for fibrosis. The Journal of clinical investigation. 2003;112(12):1776-84.

73. Rout-Pitt N, Farrow N, Parsons D, Donnelley M. Epithelial mesenchymal transition (EMT): a universal process in lung diseases with implications for cystic fibrosis pathophysiology. Respir Res. 2018;19(1):136.
74. Zhang Y, Weinberg RA. Epithelial-to-mesenchymal transition in cancer: complexity and opportunities. Front Med. 2018;12(4):361-73.
75. Goncharov AP, Vashakidze N, Kharaishvili G. Epithelial-Mesenchymal Transition: A Fundamental Cellular and Microenvironmental Process in Benign and Malignant Prostate Pathologies. Biomedicines. 2024;12(2).
76. Howard AGA, Uribe RA. Hox proteins as regulators of extracellular matrix interactions during neural crest migration. Differentiation. 2022;128:26-32.
77. Marconi GD, Fonticoli L, Rajan TS, Pierdomenico SD, Trubiani O, Pizzicannella J, et al. Epithelial-mesenchymal transition (EMT): the type-2 EMT in wound healing, tissue regeneration and organ fibrosis. Cells. 2021;10(7):1587.
78. Gavert N, Ben-Ze'ev A. Epithelial–mesenchymal transition and the invasive potential of tumors. Trends in molecular medicine. 2008;14(5):199-209.
79. Yeo CD, Kang N, Choi SY, Kim BN, Park CK, Kim JW, et al. The role of hypoxia on the acquisition of epithelial-mesenchymal transition and cancer stemness: a possible link to epigenetic regulation. Korean J Intern Med. 2017;32(4):589-99.
80. Gonzalez DM, Medici D. Signaling mechanisms of the epithelial-mesenchymal transition. Sci Signal. 2014;7(344):re8.
81. Lindsey S, Langhans SA. Crosstalk of Oncogenic Signaling Pathways during Epithelial-Mesenchymal Transition. Front Oncol. 2014;4:358.
82. Lin Y-T, Wu K-J. Epigenetic regulation of epithelial-mesenchymal transition: focusing on hypoxia and TGF-β signaling. Journal of Biomedical Science. 2020;27(1):39.
83. Zeisberg M, Neilson EG. Biomarkers for epithelial-mesenchymal transitions. The Journal of clinical investigation. 2009;119(6):1429-37.
84. Scanlon CS, Van Tubergen EA, Inglehart RC, D'Silva NJ. Biomarkers of epithelial-mesenchymal transition in squamous cell carcinoma. J Dent Res. 2013;92(2):114-21.
85. Hajra KM, Chen DY, Fearon ER. The SLUG zinc-finger protein represses E-cadherin in breast cancer. Cancer research. 2002;62(6):1613-8.
86. Thiery JP. Epithelial–mesenchymal transitions in tumour progression. Nature Reviews Cancer. 2002;2(6):442.
87. Georgakopoulos-Soares I, Chartoumpekis DV, Kyriazopoulou V, Zaravinos A. EMT Factors and Metabolic Pathways in Cancer. Front Oncol. 2020;10:499.
88. Klymkowsky MW. Beta-catenin and its regulatory network. Human pathology. 2005;36(3):225-7.
89. Vignjevic D, Montagnac G, editors. Reorganisation of the dendritic actin network during cancer cell migration and invasion. Seminars in cancer biology; 2008: Elsevier.
90. Buck E, Eyzaguirre A, Barr S, Thompson S, Sennello R, Young D, et al. Loss of homotypic cell adhesion by epithelial-mesenchymal transition or mutation limits sensitivity to epidermal growth factor receptor inhibition. Molecular cancer therapeutics. 2007;6(2):532-41.

91. Kurrey NK, Jalgaonkar SP, Joglekar AV, Ghanate AD, Chaskar PD, Doiphode RY, et al. Snail and slug mediate radioresistance and chemoresistance by antagonizing p53-mediated apoptosis and acquiring a stem-like phenotype in ovarian cancer cells. Stem cells (Dayton, Ohio). 2009;27(9):2059-68.
92. Jordan NV, Johnson GL, Abell AN. Tracking the intermediate stages of epithelial-mesenchymal transition in epithelial stem cells and cancer. Cell Cycle. 2011;10(17):2865-73.
93. Jolly MK, Somarelli JA, Sheth M, Biddle A, Tripathi SC, Armstrong AJ, et al. Hybrid epithelial/mesenchymal phenotypes promote metastasis and therapy resistance across carcinomas. Pharmacology & Therapeutics. 2019;194:161-84.
94. Sahoo S, Nayak SP, Hari K, Purkait P, Mandal S, Kishore A, et al. Immunosuppressive Traits of the Hybrid Epithelial/Mesenchymal Phenotype. Frontiers in Immunology. 2021;12.
95. Mehdi S, Macdonald E, Galpin K, Landry DA, Rodriguez G, Vanderhyden B, et al. LY75 Suppression in Mesenchymal Epithelial Ovarian Cancer Cells Generates a Stable Hybrid EOC Cellular Phenotype, Associated with Enhanced Tumor Initiation, Spreading and Resistance to Treatment in Orthotopic Xenograft Mouse Model. International Journal of Molecular Sciences. 2020;21(14):4992.
96. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, et al. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. Modern pathology. 2006;19(2):264-71.
97. George JT, Jolly MK, Xu S, Somarelli JA, Levine H. Survival outcomes in cancer patients predicted by a partial EMT gene expression scoring metric. Cancer research. 2017;77(22):6415-28.
98. Grigore AD, Jolly MK, Jia D, Farach-Carson MC, Levine H. Tumor budding: the name is EMT. Partial EMT. Journal of clinical medicine. 2016;5(5):51.
99. Hendrix M, Seftor EA, Seftor R, Trevor KT. Experimental co-expression of vimentin and keratin intermediate filaments in human breast cancer cells results in phenotypic interconversion and increased invasive behavior. The American journal of pathology. 1997;150(2):483.
100. Strauss R, Li Z-Y, Liu Y, Beyer I, Persson J, Sova P, et al. Analysis of epithelial and mesenchymal markers in ovarian cancer reveals phenotypic heterogeneity and plasticity. PloS one. 2011;6(1):e16186.
101. Huang RY, Wong M, Tan T, Kuay K, Ng A, Chung V, et al. An EMT spectrum defines an anoikis-resistant and spheroidogenic intermediate mesenchymal state that is sensitive to e-cadherin restoration by a src-kinase inhibitor, saracatinib (AZD0530). Cell death & disease. 2013;4(11):e915-e.
102. Wolf K, Wu YI, Liu Y, Geiger J, Tam E, Overall C, et al. Multi-step pericellular proteolysis controls the transition from individual to collective cancer cell invasion. Nature cell biology. 2007;9(8):893.
103. Liotta LA, Stracke ML, Aznavoorian SA, Beckner ME, Schiffmann E. Tumor cell motility. Semin Cancer Biol. 1991;2(2):111-4.
104. Talkenberger K, Cavalcanti-Adam EA, Voss-Böhme A, Deutsch A. Amoeboid-mesenchymal migration plasticity promotes invasion only in complex heterogeneous microenvironments. Scientific Reports. 2017;7(1):9237.

105. Emad A, Ray T, Jensen TW, Parat M, Natrajan R, Sinha S, et al. An epithelial-mesenchymal-amoeboid transition gene signature reveals molecular subtypes of breast cancer progression and metastasis. bioRxiv. 2017:219410.
106. Wu JS, Jiang J, Chen BJ, Wang K, Tang YL, Liang XH. Plasticity of cancer cell invasion: Patterns and mechanisms. Transl Oncol. 2021;14(1):100899.
107. Yamaguchi H, Condeelis J. Regulation of the actin cytoskeleton in cancer cell migration and invasion. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2007;1773(5):642-52.
108. Carman CV, Springer TA. Trans-cellular migration: cell–cell contacts get intimate. Current opinion in cell biology. 2008;20(5):533-40.
109. Reymond N, d'Água BB, Ridley AJ. Crossing the endothelial barrier during metastasis. Nature Reviews Cancer. 2013;13(12):858-70.
110. Aceto N, Bardia A, Miyamoto DT, Donaldson MC, Wittner BS, Spencer JA, et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell. 2014;158(5):1110-22.
111. Au SH, Storey BD, Moore JC, Tang Q, Chen Y-L, Javaid S, et al. Clusters of circulating tumor cells traverse capillary-sized vessels. Proceedings of the National Academy of Sciences. 2016;113(18):4947-52.
112. Schumacher D, Strilic B, Sivaraj KK, Wettschureck N, Offermanns S. Platelet-derived nucleotides promote tumor-cell transendothelial migration and metastasis via P2Y2 receptor. Cancer cell. 2013;24(1):130-7.
113. Labelle M, Begum S, Hynes RO. Direct signaling between platelets and cancer cells induces an epithelial-mesenchymal-like transition and promotes metastasis. Cancer cell. 2011;20(5):576-90.
114. Strilic B, Yang L, Albarrán-Juárez J, Wachsmuth L, Han K, Müller UC, et al. Tumour-cell-induced endothelial cell necroptosis via death receptor 6 promotes metastasis. Nature. 2016;536(7615):215.
115. Paget S. The distribution of secondary growths in cancer of the breast. The Lancet. 1889;133(3421):571-3.
116. Luzzi KJ, MacDonald IC, Schmidt EE, Kerkvliet N, Morris VL, Chambers AF, et al. Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. The American journal of pathology. 1998;153(3):865-73.
117. Crick FH, editor On protein synthesis. Symp Soc Exp Biol; 1958.
118. Brody L. Gene Expression: National Human Geneome Research Institute 2022 [updated December 19. 2022.
119. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. 1953;171(4356):737-8.
120. Kornberg RD, Thomas JO. Chromatin Structure: Oligomers of the Histones: The histones comprise an (F2A1) 2 (F3) 2 tetramer, a different oligomer of F2A2 and F2B, and monomer of F1. Science. 1974;184(4139):865-8.
121. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature. 1997;389(6648):251-60.

122. Choudhuri S. Chapter 1 - Fundamentals of Genes and Genomes**The opinions expressed in this chapter are the author's own and they do not necessarily reflect the opinions of the FDA, the DHHS, or the Federal Government. In: Choudhuri S, editor. Bioinformatics for Beginners. Oxford: Academic Press; 2014. p. 1-25.
123. Vanzan L, Sklias A, Boskovic M, Herceg Z, Murr R, Suter DM. Chapter 3 - Mechanisms of Histone Modifications. In: Tollefsbol TO, editor. Handbook of Epigenetics (Third Edition): Academic Press; 2023. p. 27-54.
124. Smith BC, Denu JM. Chemical mechanisms of histone lysine and arginine modifications. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 2009;1789(1):45-57.
125. Biswas S, Rao CM. Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. Eur J Pharmacol. 2018;837:8-24.
126. Milazzo G, Mercatelli D, Di Muzio G, Triboli L, De Rosa P, Perini G, et al. Histone Deacetylases (HDACs): Evolution, Specificity, Role in Transcriptional Complexes, and Pharmacological Actionability. Genes (Basel). 2020;11(5).
127. Shilatifard A. Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annu Rev Biochem. 2006;75:243-69.
128. Pekowska A, Benoukraf T, Ferrier P, Spicuglia S. A unique H3K4me2 profile marks tissue-specific gene regulation. Genome Res. 2010;20(11):1493-502.
129. Ninova M, Fejes Tóth K, Aravin AA. The control of gene expression and cell identity by H3K9 trimethylation. Development (Cambridge, England). 2019;146(19).
130. Tang L, Nogales E, Ciferri C. Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. Progress in biophysics and molecular biology. 2010;102(2-3):122-8.
131. Zaret KS. Pioneer Transcription Factors Initiating Gene Network Changes. Annual Review of Genetics. 2020;54(1):367-85.
132. Cirillo LA, McPherson CE, Bossard P, Stevens K, Cherian S, Shim EY, et al. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. The EMBO Journal. 1998;17(1):244-54.
133. Luger K, Hansen JC. Nucleosome and chromatin fiber dynamics. Current opinion in structural biology. 2005;15(2):188-96.
134. Orphanides G, Lagrange T, Reinberg D. The general transcription factors of RNA polymerase II. Genes & development. 1996;10(21):2657-83.
135. Zawel L, Reinberg D. Initiation of transcription by RNA polymerase II: a multi-step process. Progress in nucleic acid research and molecular biology. 1993;44:67-108.
136. Novina CD, Roy AL. Core promoters and transcriptional control. Trends in genetics : TIG. 1996;12(9):351-5.
137. Pedersen AG, Baldi P, Chauvin Y, Brunak S. The biology of eukaryotic promoter prediction—a review. Computers & Chemistry. 1999;23(3):191-207.
138. Wobbe CR, Struhl K. Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro. Molecular and Cellular Biology. 1990;10(8):3859-67.

139. Hahn S, Buratowski S, Sharp PA, Guarente L. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. Proceedings of the National Academy of Sciences. 1989;86(15):5718-22.
140. Smith AD, Sumazin P, Zhang MQ. Tissue-specific regulatory elements in mammalian promoters. Mol Syst Biol. 2007;3:73.
141. Smale ST. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. Biochimica et biophysica acta. 1997;1351(1-2):73-88.
142. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The human transcription factors. Cell. 2018;172(4):650-65.
143. Buratowski S, Hahn S, Guarente L, Sharp PA. Five intermediate complexes in transcription initiation by RNA polymerase II. Cell. 1989;56(4):549-61.
144. Torchia J, Glass C, Rosenfeld MG. Co-activators and co-repressors in the integration of transcriptional responses. Curr Opin Cell Biol. 1998;10(3):373-83.
145. Deaton AM, Bird A. CpG islands and the regulation of transcription. Genes Dev. 2011;25(10):1010-22.
146. Helbo AS, Lay FD, Jones PA, Liang G, Grønbæk K. Nucleosome Positioning and NDR Structure at RNA Polymerase III Promoters. Sci Rep. 2017;7:41947.
147. Fuda NJ, Ardehali MB, Lis JT. Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature. 2009;461(7261):186-92.
148. Proudfoot NJ. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science. 2016;352(6291):aad9926.
149. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. Cell. 2013;152(6):1237-51.
150. Nebert DW. Transcription factors and cancer: an overview. Toxicology. 2002;181-182:131-41.
151. Jia Q, Chen S, Tan Y, Li Y, Tang F. Oncogenic super-enhancer formation in tumorigenesis and its molecular mechanisms. Experimental & Molecular Medicine. 2020;52(5):713-23.
152. Strazzabosco M, Fabris L. Development of the bile ducts: essentials for the clinical hepatologist. Journal of hepatology. 2012;56(5):1159-70.
153. Sadler TW, Langman J. Langman's Medical Embryology (13th ed). 12th ed ed. Philadelphia, Pa

London: Lippincott Williams & Wilkins/Wolters Kluwer Health; 2015.
154. Sia D, Villanueva A, Friedman SL, Llovet JM. Liver Cancer Cell of Origin, Molecular Class, and Effects on Patient Prognosis. Gastroenterology. 2017;152(4):745-61.
155. Ahn KS, Kang KJ. Molecular heterogeneity in intrahepatic cholangiocarcinoma. World J Hepatol. 2020;12(12):1148-57.
156. Bogert PT, LaRusso NF. Cholangiocyte biology. Curr Opin Gastroenterol. 2007;23(3):299-305.
157. Yoo K-S, Lim WT, Choi HS. Biology of Cholangiocytes: From Bench to Bedside. Gut Liver. 2016;10(5):687-98.
158. Banales JM, Cardinale V, Carpino G, Marzioni M, Andersen JB, Invernizzi P, et al. Cholangiocarcinoma: current knowledge and future perspectives consensus

statement from the European Network for the Study of Cholangiocarcinoma (ENS-CCA). Nature Reviews Gastroenterology & Hepatology. 2016;13(5):261-80.
159. Krasinskas AM. Cholangiocarcinoma. Surg Pathol Clin. 2018;11(2):403-29.
160. Lowe RC, Anderson CD. Epidemiology, pathogenesis, and classification of cholangiocarcinoma: UpToDate Inc.; 2017 [Available from: https://www.uptodate.com/contents/epidemiology-pathogenesis-and-classification-of-cholangiocarcinoma?source=bookmarks_widget#H2.
161. Klatskin G. Adenocarcinoma of the hepatic duct at its bifurcation within the porta hepatis: An unusual tumor with distinctive clinical and pathological features. The American Journal of Medicine. 1965;38(2):241-56.
162. DeOliveira ML, Cunningham SC, Cameron JL, Kamangar F, Winter JM, Lillemoe KD, et al. Cholangiocarcinoma: thirty-one-year experience with 564 patients at a single institution. Annals of surgery. 2007;245(5):755-62.
163. Kendall T, Verheij J, Gaudio E, Evert M, Guido M, Goeppert B, et al. Anatomical, histomorphological and molecular classification of cholangiocarcinoma. Liver Int. 2019;39 Suppl 1:7-18.
164. Shin HR, Lee CU, Park HJ, Seol SY, Chung JM, Choi HC, et al. Hepatitis B and C virus, Clonorchis sinensis for the risk of liver cancer: a case-control study in Pusan, Korea. International journal of epidemiology. 1996;25(5):933-40.
165. Watanapa P. Cholangiocarcinoma in patients with opisthorchiasis. The British journal of surgery. 1996;83(8):1062-64.
166. Watanapa P, Watanapa WB. Liver fluke-associated cholangiocarcinoma. The British journal of surgery. 2002;89(8):962-70.
167. Schistosomes, liver flukes and Helicobacter pylori. IARC Monogr Eval Carcinog Risks Hum. 1994;61:1-241.
168. SADUN EH. Studies on Opisthorchis viverrini in Thailand. American Journal of Hygiene. 1955.
169. Claessen MM, Vleggaar FP, Tytgat KM, Siersema PD, van Buuren HR. High lifetime risk of cancer in primary sclerosing cholangitis. Journal of hepatology. 2009;50(1):158-64.
170. LaRusso NF, Shneider BL, Black D, Gores GJ, James SP, Doo E, et al. Primary sclerosing cholangitis: summary of a workshop. Hepatology (Baltimore, Md). 2006;44(3):746-64.
171. Kowdley KV. Primary sclerosing cholangitis: Epidemiology and pathogenesis: UpToDate Inc.; 2017 [Available from: https://www.uptodate.com/contents/primary-sclerosing-cholangitis-epidemiology-and-pathogenesis?search=primary%20sclerosing%20cholangitis&source=search_result&selectedTitle=3~103&usage_type=default&display_rank=3#H5.
172. Hsing AW, Gao YT, Han TQ, Rashid A, Sakoda LC, Wang BS, et al. Gallstones and the risk of biliary tract cancer: a population-based study in China. British journal of cancer. 2007;97(11):1577-82.
173. Lee PC, Hu YW, Hu LY, Chen SC, Chien SH, Shen CC, et al. Risk of cancer in patients with cholecystitis: a nationwide population-based study. Am J Med. 2015;128(2):185-91.

174. Shaib YH, El-Serag HB, Davila JA, Morgan R, McGlynn KA. Risk factors of intrahepatic cholangiocarcinoma in the United States: a case-control study. Gastroenterology. 2005;128(3):620-6.
175. Donato F, Gelatti U, Tagger A, Favret M, Ribero ML, Callea F, et al. Intrahepatic cholangiocarcinoma and hepatitis C and B virus infection, alcohol intake, and hepatolithiasis: a case-control study in Italy. Cancer causes & control : CCC. 2001;12(10):959-64.
176. Kobayashi M, Ikeda K, Saitoh S, Suzuki F, Tsubota A, Suzuki Y, et al. Incidence of primary cholangiocellular carcinoma of the liver in japanese patients with hepatitis C virus-related cirrhosis. Cancer. 2000;88(11):2471-7.
177. Scott J, Shousha S, Thomas HC, Sherlock S. Bile duct carcinoma: a late complication of congenital hepatic fibrosis. Case report and review of literature. The American journal of gastroenterology. 1980;73(2):113-9.
178. Lipsett PA, Pitt HA, Colombani PM, Boitnott JK, Cameron JL. Choledochal cyst disease. A changing pattern of presentation. Annals of surgery. 1994;220(5):644-52.
179. Yamada A, Komaki Y, Komaki F, Micic D, Zullow S, Sakuraba A. Risk of gastrointestinal cancers in patients with cystic fibrosis: a systematic review and meta-analysis. The Lancet Oncology. 2018;19(6):758-67.
180. Welzel TM, Graubard BI, Zeuzem S, El-Serag HB, Davila JA, McGlynn KA. Metabolic syndrome increases the risk of primary liver cancer in the United States: a study in the SEER-Medicare database. Hepatology (Baltimore, Md). 2011;54(2):463-71.
181. Welzel TM, Graubard BI, El-Serag HB, Shaib YH, Hsing AW, Davila JA, et al. Risk factors for intrahepatic and extrahepatic cholangiocarcinoma in the United States: a population-based case-control study. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2007;5(10):1221-8.
182. Jing W, Jin G, Zhou X, Zhou Y, Zhang Y, Shao C, et al. Diabetes mellitus and increased risk of cholangiocarcinoma: a meta-analysis. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP). 2012;21(1):24-31.
183. Pinlaor S, Hiraku Y, Ma N, Yongvanit P, Semba R, Oikawa S, et al. Mechanism of NO-mediated oxidative and nitrative DNA damage in hamsters infected with Opisthorchis viverrini: a model of inflammation-mediated carcinogenesis. Nitric Oxide. 2004;11(2):175-83.
184. Wehbe H, Henson R, Meng F, Mize-Berge J, Patel T. Interleukin-6 contributes to growth in cholangiocarcinoma cells by aberrant promoter methylation and gene expression. Cancer Res. 2006;66(21):10517-24.
185. Nakeeb A, Pitt HA, Sohn TA, Coleman J, Abrams RA, Piantadosi S, et al. Cholangiocarcinoma. A spectrum of intrahepatic, perihilar, and distal tumors. Annals of surgery. 1996;224(4):463-73; discussion 73-5.
186. Blechacz B, Komuta M, Roskams T, Gores GJ. Clinical diagnosis and staging of cholangiocarcinoma. Nature reviews Gastroenterology & hepatology. 2011;8(9):512-22.

187. Lowe RC, Anderson CD, Kowdley KV. Clinical manifestations and diagnosis of cholangiocarcinoma: UpToDate Inc.; [Available from: https://www.uptodate.com/contents/clinical-manifestations-and-diagnosis-of-cholangiocarcinoma/contributors.
188. Cardinale V, Bragazzi MC, Carpino G, Matteo SD, Overi D, Nevi L, et al. Intrahepatic cholangiocarcinoma: review and update. Hepatoma Research. 2018;4:20.
189. Rimola J, Forner A, Reig M, Vilana R, de Lope CR, Ayuso C, et al. Cholangiocarcinoma in cirrhosis: absence of contrast washout in delayed phases by magnetic resonance imaging avoids misdiagnosis of hepatocellular carcinoma. Hepatology (Baltimore, Md). 2009;50(3):791-8.
190. Rushbrook SM, Kendall TJ, Zen Y, Albazaz R, Manoharan P, Pereira SP, et al. British Society of Gastroenterology guidelines for the diagnosis and management of cholangiocarcinoma. Gut. 2023.
191. Hong SM, Pawlik TM, Cho H, Aggarwal B, Goggins M, Hruban RH, et al. Depth of tumor invasion better predicts prognosis than the current American Joint Committee on Cancer T classification for distal bile duct carcinoma. Surgery. 2009;146(2):250-7.
192. Spolverato G, Bagante F, Weiss M, Alexandrescu S, Marques HP, Aldrighetti L, et al. Comparative performances of the 7th and the 8th editions of the American Joint Committee on Cancer staging systems for intrahepatic cholangiocarcinoma. Journal of surgical oncology. 2017;115(6):696-703.
193. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. Journal of hepatology. 2018;69(1):182-236.
194. Komaya K, Ebata T, Yokoyama Y, Igami T, Sugawara G, Mizuno T, et al. Recurrence after curative-intent resection of perihilar cholangiocarcinoma: analysis of a large cohort with a close postoperative follow-up approach. Surgery. 2018;163(4):732-8.
195. Fung S, Syed YY. Durvalumab: A Review in Advanced Biliary Tract Cancer. Targeted Oncology. 2023;18(6):965-72.
196. Soufi A, Smith C, Clarke AR, Gaston K, Jayaraman P-S. Oligomerisation of the Developmental Regulator Proline Rich Homeodomain (PRH/Hex) is Mediated by a Novel Proline-rich Dimerisation Domain. Journal of Molecular Biology. 2006;358(4):943-62.
197. Williams H, Jayaraman PS, Gaston K. DNA wrapping and distortion by an oligomeric homeodomain protein. J Mol Biol. 2008;383(1):10-23.
198. Soufi A, Sawasdichai A, Shukla A, Noy P, Dafforn T, Smith C, et al. DNA compaction by the higher-order assembly of PRH/Hex homeodomain protein oligomers. Nucleic acids research. 2010;38(21):7513-25.
199. Crompton MR, Bartlett TJ, MacGregor AD, Manfioletti G, Buratti E, Giancotti V, et al. Identification of a novel vertebrate homeobox gene expressed in haematopoietic cells. Nucleic acids research. 1992;20(21):5661-7.
200. Bedford FK, Ashworth A, Enver T, Wiedemann LM. HEX: a novel homeobox gene expressed during haematopoiesis and conserved between mouse and human. Nucleic acids research. 1993;21(5):1245-9.

201. Hromas R, Radich J, Collins S. PCR cloning of an orphan homeobox gene (PRH) preferentially expressed in myeloid and liver cells. Biochemical and biophysical research communications. 1993;195(2):976-83.
202. Ghosh B, Jacobs HC, Wiedemann LM, Brown A, Bedford FK, Nimmakayalu MA, et al. Genomic structure, cDNA mapping, and chromosomal localization of the mouse homeobox gene, Hex. Mammalian genome. 1999;10:1023-5.
203. Morgutti M, Demori E, Pecile V, Amoroso A, Rustighi A, Manfioletti G. Genomic organization and chromosome mapping of the human homeobox gene HHEX. Cytogenetics and Cell Genetics. 2001;94(1-2):30-2.
204. Zorn AM, Butler K, Gurdon JB. Anterior endomesoderm specification in Xenopus by Wnt/beta-catenin and TGF-beta signalling pathways. Developmental biology. 1999;209(2):282-97.
205. McLin VA, Rankin SA, Zorn AM. Repression of Wnt/beta-catenin signaling in the anterior endoderm is essential for liver and pancreas development. Development (Cambridge, England). 2007;134(12):2207-17.
206. Zhang W, Yatskievych TA, Cao X, Antin PB. Regulation of Hex gene expression by a Smads-dependent signaling pathway. The Journal of biological chemistry. 2002;277(47):45435-41.
207. Zhang W, Yatskievych TA, Baker RK, Antin PB. Regulation of Hex gene expression and initial stages of avian hepatogenesis by Bmp and Fgf signaling. Developmental biology. 2004;268(2):312-26.
208. Shin D, Shin CH, Tucker J, Ober EA, Rentzsch F, Poss KD, et al. Bmp and Fgf signaling are essential for liver specification in zebrafish. Development (Cambridge, England). 2007;134(11):2041-50.
209. Guo X, Wang XF. Signaling cross-talk between TGF-beta/BMP and other pathways. Cell Res. 2009;19(1):71-88.
210. Sato A, Keng VW, Yamamoto T, Kasamatsu S, Ban T, Tanaka H, et al. Identification and characterization of the hematopoietic cell-specific enhancer-like element of the mouse hex gene. Journal of biochemistry. 2004;135(2):259-68.
211. Kikkawa E, Hinata M, Keng VW, Myint Z, Sato A, Yamada K, et al. Sp family members stimulate transcription of the hex gene via interactions with GC boxes. Journal of biochemistry. 2001;130(6):885-91.
212. Puppin C, D'Elia AV, Pellizzari L, Russo D, Arturi F, Presta I, et al. Thyroid-specific transcription factors control Hex promoter activity. Nucleic acids research. 2003;31(7):1845-52.
213. Puppin C, Presta I, D'Elia AV, Tell G, Arturi F, Russo D, et al. Functional interaction among thyroid-specific transcription factors: Pax8 regulates the activity of Hex promoter. Molecular and Cellular Endocrinology. 2004;214(1):117-25.
214. Puppin C, Puglisi F, Pellizzari L, Manfioletti G, Pestrin M, Pandolfi M, et al. HEX expression and localization in normal mammary gland and breast carcinoma. BMC Cancer. 2006;6(1):192.
215. Kitchen P, Lee KY, Clark D, Lau N, Lertsuwan J, Sawasdichai A, et al. A Runaway PRH/HHEX-Notch3-Positive Feedback Loop Drives Cholangiocarcinoma and Determines Response to CDK4/6 Inhibition. Cancer Res. 2020;80(4):757-70.

216. Denson LA, McClure MH, Bogue CW, Karpen SJ, Jacobs HC. HNF3beta and GATA-4 transactivate the liver-enriched homeobox gene, Hex. Gene. 2000;246(1-2):311-20.
217. Yatskievych TA, Pascoe S, Antin PB. Expression of the homeobox gene Hex during early stages of chick embryo development. Mechanisms of development. 1999;80(1):107-9.
218. Keng VW, Fujimori KE, Myint Z, Tamamaki N, Nojyo Y, Noguchi T. Expression of Hex mRNA in early murine postimplantation embryo development. FEBS Lett. 1998;426(2):183-6.
219. Martinez Barbera JP, Clements M, Thomas P, Rodriguez T, Meloy D, Kioussis D, et al. The homeobox gene Hex is required in definitive endodermal tissues for normal forebrain, liver and thyroid formation. Development (Cambridge, England). 2000;127(11):2433-45.
220. Smithers LE, Jones CM. Xhex-expressing endodermal tissues are essential for anterior patterning in Xenopus. Mechanisms of development. 2002;119(2):191-200.
221. Martinez Barbera JP, Clements M, Thomas P, Rodriguez T, Meloy D, Kioussis D, et al. The homeobox gene Hex is required in definitive endodermal tissues for normal forebrain, liver and thyroid formation. Development (Cambridge, England). 2000;127(11):2433.
222. Bogue CW, Ganea GR, Sturm E, Ianucci R, Jacobs HC. Hex expression suggests a role in the development and function of organs derived from foregut endoderm. Developmental dynamics : an official publication of the American Association of Anatomists. 2000;219(1):84-9.
223. Hallaq H, Pinter E, Enciso J, McGrath J, Zeiss C, Brueckner M, et al. A null mutation of Hhex results in abnormal cardiac development, defective vasculogenesis and elevated Vegfa levels. Development (Cambridge, England). 2004;131(20):5197-209.
224. Hunter MP, Wilson CM, Jiang X, Cong R, Vasavada H, Kaestner KH, et al. The homeobox gene Hhex is essential for proper hepatoblast differentiation and bile duct morphogenesis. Developmental biology. 2007;308(2):355-67.
225. Goodings C, Smith E, Mathias E, Elliott N, Cleveland SM, Tripathi RM, et al. Hhex is Required at Multiple Stages of Adult Hematopoietic Stem and Progenitor Cell Differentiation. Stem cells (Dayton, Ohio). 2015;33(8):2628-41.
226. Newman CS, Chia F, Krieg PA. The XHex homeobox gene is expressed during development of the vascular endothelium: overexpression leads to an increase in vascular endothelial cell number. Mechanisms of development. 1997;66(1-2):83-93.
227. Soufi A, Jayaraman PS. PRH/Hex: an oligomeric transcription factor and multifunctional regulator of cell fate. The Biochemical journal. 2008;412(3):399-413.
228. Soufi A, Noy P, Buckle M, Sawasdichai A, Gaston K, Jayaraman PS. CK2 phosphorylation of the PRH/Hex homeodomain functions as a reversible switch for DNA binding. Nucleic acids research. 2009;37(10):3288-300.
229. Billeter M. Homeodomain-type DNA recognition. Progress in biophysics and molecular biology. 1996;66(3):211-25.

230. Kasamatsu S, Sato A, Yamamoto T, Keng VW, Yoshida H, Yamazaki Y, et al. Identification of the transactivating region of the homeodomain protein, hex. Journal of biochemistry. 2004;135(2):217-23.
231. Gaston K, Tsitsilianos M-A, Wadey K, Jayaraman P-S. Misregulation of the proline rich homeodomain (PRH/HHEX) protein in cancer cells and its consequences for tumour growth and invasion. Cell & bioscience. 2016;6:12-.
232. Guiral M, Bess K, Goodwin G, Jayaraman PS. PRH represses transcription in hematopoietic cells by at least two independent mechanisms. The Journal of biological chemistry. 2001;276(4):2961-70.
233. Topisirovic I, Culjkovic B, Cohen N, Perez JM, Skrabanek L. The proline-rich homeodomain protein, PRH, is a tissue-specific inhibitor of eIF4E-dependent cyclin D1 mRNA transport and growth. 2003;22(3):689-703.
234. Noy P, Sawasdichai A, Jayaraman PS, Gaston K. Protein kinase CK2 inactivates PRH/Hhex using multiple mechanisms to de-repress VEGF-signalling genes and promote cell survival. Nucleic acids research. 2012;40(18):9008-20.
235. BESS KL, SWINGLER TE, RIVETT AJ, GASTON K, JAYARAMAN P-S. The transcriptional repressor protein PRH interacts with the proteasome. Biochemical Journal. 2003;374(3):667-75.
236. Soufi A, Gaston K, Jayaraman P-S. Purification and characterisation of the PRH homeodomain: removal of the N-terminal domain of PRH increases the PRH homeodomain–DNA interaction. International journal of biological macromolecules. 2006;39(1-3):45-50.
237. Pellizzari L, D'Elia A, Rustighi A, Manfioletti G, Tell G, Damante G. Expression and function of the homeodomain-containing protein Hex in thyroid cells. Nucleic acids research. 2000;28(13):2503-11.
238. Schaefer LK, Wang S, Schaefer TS. Functional interaction of Jun and homeodomain proteins. The Journal of biological chemistry. 2001;276(46):43074-82.
239. Tanaka H, Yamamoto T, Ban T, Satoh S-i, Tanaka T, Shimoda M, et al. Hex stimulates the hepatocyte nuclear factor 1α-mediated activation of transcription. Archives of Biochemistry and Biophysics. 2005;442(1):117-24.
240. Oyama Y, Kawai-Kowase K, Sekiguchi K, Sato M, Sato H, Yamazaki M, et al. Homeobox Protein Hex Facilitates Serum Responsive Factor–Mediated Activation of the SM22α Gene Transcription in Embryonic Fibroblasts. Arteriosclerosis, thrombosis, and vascular biology. 2004;24(9):1602-7.
241. Shukla A, Burton NM, Jayaraman PS, Gaston K. The proline rich homeodomain protein PRH/Hhex forms stable oligomers that are highly resistant to denaturation. PLoS One. 2012;7(4):e35984.
242. Litchfield DW. Protein kinase CK2: structure, regulation and role in cellular decisions of life and death. The Biochemical journal. 2003;369(Pt 1):1-15.
243. Guo Y, Zhu Z, Huang Z, Cui L, Yu W, Hong W, et al. CK2-induced cooperation of HHEX with the YAP-TEAD4 complex promotes colorectal tumorigenesis. Nature Communications. 2022;13(1):4995.

244. Binder JX, Pletscher-Frankild S, Tsafou K, Stolte C, O'Donoghue SI, Schneider R, et al. COMPARTMENTS: unification and visualization of protein subcellular localization evidence. Database (Oxford). 2014;2014:bau012.
245. D'Elia AV, Tell G, Russo D, Arturi F, Puglisi F, Manfioletti G, et al. Expression and localization of the homeodomain-containing protein HEX in human thyroid tumors. The Journal of clinical endocrinology and metabolism. 2002;87(3):1376-83.
246. Sawasdichai A, Chen H-T, Abdul Hamid N, Jayaraman P-S, Gaston K. In situ subcellular fractionation of adherent and non-adherent mammalian cells. J Vis Exp. 2010(41):1958.
247. Jackson JT, Nutt SL, McCormack MP. The Haematopoietically-expressed homeobox transcription factor: roles in development, physiology and disease. Front Immunol. 2023;14:1197490.
248. Ghosh B, Ganea GR, Denson LA, Iannucci R, Jacobs HC, Bogue CW. Immunocytochemical Characterization of Murine Hex, a Homeobox-Containing Protein. Pediatric Research. 2000;48(5):634-8.
249. Kershaw RM, Roberts D, Wragg J, Shaaban AM, Humphreys E, Halsall J, et al. Proline-Rich Homeodomain protein (PRH/HHEX) is a suppressor of breast tumour growth. Oncogenesis. 2017;6(6):e346.
250. Topisirovic I, Guzman ML, McConnell MJ, Licht JD, Culjkovic B, Neering SJ, et al. Aberrant eukaryotic translation initiation factor 4E-dependent mRNA transport impedes hematopoietic differentiation and contributes to leukemogenesis. Molecular and Cellular Biology. 2003;23(24):8992-9002.
251. Ploski JE, Topisirovic I, Park KW, Borden KL, Radu A. A mechanism of nucleocytoplasmic trafficking for the homeodomain protein PRH. Mol Cell Biochem. 2009;332(1-2):173-81.
252. Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, et al. A single-cell type transcriptomics map of human tissues. Sci Adv. 2021;7(31).
253. Cong R, Jiang X, Wilson CM, Hunter MP, Vasavada H, Bogue CW. Hhex is a direct repressor of endothelial cell-specific molecule 1 (ESM-1). Biochemical and biophysical research communications. 2006;346(2):535-45.
254. Brickman JM, Jones CM, Clements M, Smith JC, Beddington RS. Hex is a transcriptional repressor that contributes to anterior identity and suppresses Spemann organiser function. Development (Cambridge, England). 2000;127(11):2303-15.
255. Noy P, Williams H, Sawasdichai A, Gaston K, Jayaraman PS. PRH/Hhex Controls Cell Survival through Coordinate Transcriptional Regulation of Vascular Endothelial Growth Factor Signaling. Molecular and Cellular Biology. 2010;30(9):2120-34.
256. Swingler TE, Bess KL, Yao J, Stifani S, Jayaraman PS. The proline-rich homeodomain protein recruits members of the Groucho/Transducin-like enhancer of split protein family to co-repress transcription in hematopoietic cells. The Journal of biological chemistry. 2004;279(33):34938-47.
257. Minami T, Murakami T, Horiuchi K, Miura M, Noguchi T, Miyazaki J, et al. Interaction between hex and GATA transcription factors in vascular endothelial cells

inhibits flk-1/KDR-mediated vascular endothelial growth factor signaling. The Journal of biological chemistry. 2004;279(20):20626-35.
258. Denson LA, Karpen SJ, Bogue CW, Jacobs HC. Divergent homeobox gene hex regulates promoter of the Na(+)-dependent bile acid cotransporter. American journal of physiology Gastrointestinal and liver physiology. 2000;279(2):G347-55.
259. Su J, You P, Zhao JP, Zhang SL, Song SH, Fu ZR, et al. A potential role for the homeoprotein Hhex in hepatocellular carcinoma progression. Medical oncology (Northwood, London, England). 2012;29(2):1059-67.
260. Kershaw RM, Siddiqui YH, Roberts D, Jayaraman PS, Gaston K. PRH/HHex inhibits the migration of breast and prostate epithelial cells through direct transcriptional regulation of Endoglin. Oncogene. 2014;33(49):5592-600.
261. Nakagawa T, Abe M, Yamazaki T, Miyashita H, Niwa H, Kokubun S, et al. HEX acts as a negative regulator of angiogenesis by modulating the expression of angiogenesis-related gene in endothelial cells in vitro. Arteriosclerosis, thrombosis, and vascular biology. 2003;23(2):231-7.
262. Levy L, Hill CS. Alterations in components of the TGF-β superfamily signaling pathways in human cancer. Cytokine & growth factor reviews. 2006;17(1-2):41-58.
263. Massagué J. TGFβ in cancer. Cell. 2008;134(2):215-30.
264. Hartwell KA, Muir B, Reinhardt F, Carpenter AE, Sgroi DC, Weinberg RA. The Spemann organizer gene, Goosecoid, promotes tumor metastasis. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(50):18969-74.
265. Sarrazin S, Adam E, Lyon M, Depontieu F, Motte V, Landolfi C, et al. Endocan or endothelial cell specific molecule-1 (ESM-1): a potential novel endothelial cell marker and a new target for cancer therapy. Biochimica et biophysica acta. 2006;1765(1):25-37.
266. Smith S, Tripathi R, Goodings C, Cleveland S, Mathias E, Hardaway JA, et al. LIM domain only-2 (LMO2) induces T-cell leukemia by two distinct pathways. PLoS One. 2014;9(1):e85883.
267. Puppin C, Puglisi F, Pellizzari L, Manfioletti G, Pestrin M, Pandolfi M, et al. HEX expression and localization in normal mammary gland and breast carcinoma. BMC Cancer. 2006;6:192.
268. George A, Morse HC, Justice MJ. The homeobox gene Hex induces T-cell-derived lymphomas when overexpressed in hematopoietic precursor cells. Oncogene. 2003;22(43):6764-73.
269. Lee K. THE ROLE OF THE PROLINE RICH HOMEODOMAIN TRANSCRIPTION FACTOR (PRH/HHex) IN CHOLANGIOCARCINOMA: The University of Birmingham; 2020.
270. Ferrari S, Cannizzaro LA, Battini R, Huebner K, Baserga R. The gene encoding human vimentin is located on the short arm of chromosome 10. Am J Hum Genet. 1987;41(4):616-26.
271. Quax W, Meera Khan P, Quax-Jeuken Y, Bloemendal H. The human desmin and vimentin genes are located on different chromosomes. Gene. 1985;38(1-3):189-96.

272. Franke WW, Schmid E, Osborn M, Weber K. Different intermediate-sized filaments distinguished by immunofluorescence microscopy. Proc Natl Acad Sci U S A. 1978;75(10):5034-8.
273. Coulombe PA, Wong P. Cytoplasmic intermediate filaments revealed as dynamic and multipurpose scaffolds. Nat Cell Biol. 2004;6(8):699-706.
274. Herrmann H, Aebi U. Intermediate Filaments: Structure and Assembly. Cold Spring Harb Perspect Biol. 2016;8(11).
275. Rittling SR, Baserga R. Functional analysis and growth factor regulation of the human vimentin promoter. Molecular and Cellular Biology. 1987;7(11):3908-15.
276. Sommers CL, Skerker JM, Chrysogelos SA, Bosseler M, Gelmann EP. Regulation of vimentin gene transcription in human breast cancer cell lines. Cell Growth Differ. 1994;5(8):839-46.
277. Izmailova ES, Wieczorek E, Perkins EB, Zehner ZE. A GC-box is required for expression of the human vimentin gene. Gene. 1999;235(1):69-75.
278. Li L, Davie JR. The role of Sp1 and Sp3 in normal and cancer cell biology. Annals of Anatomy-Anatomischer Anzeiger. 2010;192(5):275-83.
279. Angel P, Karin M. The role of Jun, Fos and the AP-1 complex in cell-proliferation and transformation. Biochimica et biophysica acta. 1991;1072(2-3):129-57.
280. Wang Q, Zhu G, Lin C, Lin P, Chen H, He R, et al. Vimentin affects colorectal cancer proliferation, invasion, and migration via regulated by activator protein 1. Journal of Cellular Physiology. 2021;236(11):7591-604.
281. Lilienbaum A, Paulin D. Activation of the human vimentin gene by the Tax human T-cell leukemia virus. I. Mechanisms of regulation by the NF-kappa B transcription factor. The Journal of biological chemistry. 1993;268(3):2180-8.
282. Pantuck AJ, An J, Liu H, Rettig MB. NF-kappaB-dependent plasticity of the epithelial to mesenchymal transition induced by Von Hippel-Lindau inactivation in renal cell carcinomas. Cancer research. 2010;70(2):752-61.
283. Chen JH, Vercamer C, Li Z, Paulin D, Vandenbunder B, Stehelin D. PEA3 transactivates vimentin promoter in mammary epithelial and tumor cells. Oncogene. 1996;13(8):1667-75.
284. Sergio S, Coluccia AML, Lemma ED, Spagnolo B, Vergara D, Maffia M, et al. 3D-microenvironments initiate TCF4 expression rescuing nuclear β-catenin activity in MCF-7 breast cancer cells. Acta Biomaterialia. 2020;103:153-64.
285. Wu Y, Diab I, Zhang X, Izmailova ES, Zehner ZE. Stat3 enhances vimentin gene expression by binding to the antisilencer element and interacting with the repressor protein, ZBP-89. Oncogene. 2004;23(1):168-78.
286. Kuwahara J, Yonezawa A, Futamura M, Sugiura Y. Binding of transcription factor Sp1 to GC box DNA revealed by footprinting analysis: different contact of three zinc fingers and sequence recognition mode. Biochemistry. 1993;32(23):5994-6001.
287. Wu Y, Zhang X, Salmon M, Lin X, Zehner ZE. TGFbeta1 regulation of vimentin gene expression during differentiation of the C2C12 skeletal myogenic cell line requires Smads, AP-1 and Sp1 family members. Biochimica et biophysica acta. 2007;1773(3):427-39.

288. Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. Science. 1975;187(4173):226-32.
289. Jung S, Yi L, Kim J, Jeong D, Oh T, Kim CH, et al. The role of vimentin as a methylation biomarker for early diagnosis of cervical cancer. Mol Cells. 2011;31(5):405-11.
290. Zou H, Harrington JJ, Shire AM, Rego RL, Wang L, Campbell ME, et al. Highly Methylated Genes in Colorectal Neoplasia: Implications for Screening. Cancer Epidemiology, Biomarkers & Prevention. 2007;16(12):2686-96.
291. Wu Y, Zhang X, Salmon M, Zehner ZE. The zinc finger repressor, ZBP-89, recruits histone deacetylase 1 to repress vimentin gene expression. Genes to Cells. 2007;12(8):905-18.
292. Esteller M. Non-coding RNAs in human disease. Nature Reviews Genetics. 2011;12(12):861-74.
293. Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet. 2008;9(2):102-14.
294. Lin S-L, Lin Y-H, Chi H-C, Lin T-K, Chen W-J, Yeh C-T, et al. A Novel Long Non-Coding RNA-01488 Suppressed Metastasis and Tumorigenesis by Inducing miRNAs That Reduce Vimentin Expression and Ubiquitination of Cyclin E. Cells. 2020;9(6):1504.
295. The Human Protein Atlas 2022 [https://www.proteinatlas.org/ENSG00000026025-VIM/tissue].
296. Danielsson F, Peterson MK, Caldeira Araújo H, Lautenschläger F, Gad AKB. Vimentin Diversity in Health and Disease. Cells. 2018;7(10).
297. Franke WW, Grund C, Kuhn C, Jackson BW, Illmensee K. Formation of cytoskeletal elements during mouse embryogenesis. III. Primary mesenchymal cells and the first appearance of vimentin filaments. Differentiation. 1982;23(1):43-59.
298. Larsson A, Wilhelmsson U, Pekna M, Pekny M. Increased cell proliferation and neurogenesis in the hippocampal dentate gyrus of old GFAP(-/-)Vim(-/-) mice. Neurochem Res. 2004;29(11):2069-73.
299. Perlson E, Hanz S, Ben-Yaakov K, Segal-Ruder Y, Seger R, Fainzilber M. Vimentin-dependent spatial translocation of an activated MAP kinase in injured nerve. Neuron. 2005;45(5):715-26.
300. Eckes B, Colucci-Guyon E, Smola H, Nodder S, Babinet C, Krieg T, et al. Impaired wound healing in embryonic and adult mice lacking vimentin. Journal of Cell Science. 2000;113(13):2455-62.
301. van Engeland NCA, Suarez Rodriguez F, Rivero-Müller A, Ristori T, Duran CL, Stassen O, et al. Vimentin regulates Notch signaling strength and arterial remodeling in response to hemodynamic stress. Sci Rep. 2019;9(1):12415.
302. Taylor AC. Microtubules in the microspikes and cortical cytoplasm of isolated cells. J Cell Biol. 1966;28(2):155-68.
303. Goldman RD, Follett EA. The structure of the major cell processes of isolated BHK21 fibroblasts. Exp Cell Res. 1969;57(2):263-76.
304. Goldman RD. The role of three cytoplasmic fibers in BHK-21 cell motility. I. Microtubules and the effects of colchicine. J Cell Biol. 1971;51(3):752-62.

305. Ishikawa H, Bischoff R, Holtzer H. Formation of arrowhead complexes with heavy meromyosin in a variety of cell types. J Cell Biol. 1969;43(2):312-28.
306. Herrmann H, Häner M, Brettel M, Ku N-O, Aebi U. Characterization of distinct early assembly units of different intermediate filament proteins11Edited by W. Baumeister. Journal of Molecular Biology. 1999;286(5):1403-20.
307. Goldie KN, Wedig T, Mitra AK, Aebi U, Herrmann H, Hoenger A. Dissecting the 3-D structure of vimentin intermediate filaments by cryo-electron tomography. J Struct Biol. 2007;158(3):378-85.
308. Interaction In Vitro of Type III Intermediate Filament Proteins with Higher Order Structures of Single-Stranded DNA, Particularly with G-Quadruplex DNA. DNA and Cell Biology. 2005;24(2):85-110.
309. Ceschi S, Berselli M, Cozzaglio M, Giantin M, Toppo S, Spolaore B, et al. Vimentin binds to G-quadruplex repeats found at telomeres and gene promoters. Nucleic acids research. 2022;50(3):1370-81.
310. Herrmann H, Häner M, Brettel M, Müller SA, Goldie KN, Fedtke B, et al. Structure and assembly properties of the intermediate filament protein vimentin: the role of its head, rod and tail domains. Journal of molecular biology. 1996;264(5):933-53.
311. Eckelt A, Herrmann H, Franke WW. Assembly of a tail-less mutant of the intermediate filament protein, vimentin, in vitro and in vivo. Eur J Cell Biol. 1992;58(2):319-30.
312. Duarte S, Viedma-Poyatos Á, Navarro-Carrasco E, Martínez AE, Pajares MA, Pérez-Sala D. Vimentin filaments interact with the actin cortex in mitosis allowing normal cell division. Nature Communications. 2019;10(1):4200.
313. Danielsson F, Peterson MK, Caldeira Araújo H, Lautenschläger F, Gad AKB. Vimentin diversity in health and disease. Cells. 2018;7(10):147.
314. Lopez CG, Saldanha O, Huber K, Köster S. Lateral association and elongation of vimentin intermediate filament proteins: A time-resolved light-scattering study. Proceedings of the National Academy of Sciences. 2016;113(40):11152.
315. Herrmann H, Aebi U. Intermediate filaments: molecular structure, assembly mechanism, and integration into functionally distinct intracellular Scaffolds. Annu Rev Biochem. 2004;73:749-89.
316. Aziz A, Hess JF, Budamagunta MS, Voss JC, Kuzin AP, Huang YJ, et al. The structure of vimentin linker 1 and rod 1B domains characterized by site-directed spin-labeling electron paramagnetic resonance (SDSL-EPR) and X-ray crystallography. The Journal of biological chemistry. 2012;287(34):28349-61.
317. Prahlad V, Yoon M, Moir RD, Vale RD, Goldman RD. Rapid movements of vimentin on microtubule tracks: kinesin-dependent assembly of intermediate filament networks. J Cell Biol. 1998;143(1):159-70.
318. Meier M, Padilla GP, Herrmann H, Wedig T, Hergt M, Patel TR, et al. Vimentin coil 1A-A molecular switch involved in the initiation of filament elongation. J Mol Biol. 2009;390(2):245-61.
319. Svitkina TM, Verkhovsky AB, Borisy GG. Plectin sidearms mediate interaction of intermediate filaments with microtubules and other components of the cytoskeleton. J Cell Biol. 1996;135(4):991-1007.

320. Herrmann H, Aebi U. Intermediate filaments: molecular structure, assembly mechanism, and integration into functionally distinct intracellular scaffolds. Annual review of biochemistry. 2004;73(1):749-89.
321. Ku N-O, Liao J, Chou C-F, Omary MB. Implications of intermediate filament protein phosphorylation. Cancer and Metastasis Reviews. 1996;15(4):429-44.
322. Zackroff RV, Goldman RD. In vitro assembly of intermediate filaments from baby hamster kidney (BHK-21) cells. Proceedings of the National Academy of Sciences. 1979;76(12):6226-30.
323. Robert A, Rossow MJ, Hookway C, Adam SA, Gelfand VI. Vimentin filament precursors exchange subunits in an ATP-dependent manner. Proceedings of the National Academy of Sciences. 2015;112(27):E3505-E14.
324. Byun Y, Chen F, Chang R, Trivedi M, Green KJ, Cryns VL. Caspase cleavage of vimentin disrupts intermediate filaments and promotes apoptosis. Cell Death & Differentiation. 2001;8(5):443-50.
325. Shoeman RL, Höner B, Stoller TJ, Kesselmeier C, Miedel MC, Traub P, et al. Human immunodeficiency virus type 1 protease cleaves the intermediate filament proteins vimentin, desmin, and glial fibrillary acidic protein. Proc Natl Acad Sci U S A. 1990;87(16):6336-40.
326. Ostrowska-Podhorodecka Z, Ding I, Norouzi M, McCulloch CA. Impact of Vimentin on Regulation of Cell Signaling and Matrix Remodeling. Front Cell Dev Biol. 2022;10:869069.
327. Kalyanasundaram A, Li N, Gardner ML, Artiga EJ, Hansen BJ, Webb A, et al. Fibroblast-Specific Proteotranscriptomes Reveal Distinct Fibrotic Signatures of Human Sinoatrial Node in Nonfailing and Failing Hearts. Circulation. 2021;144(2):126-43.
328. Hyder CL, Pallari HM, Kochin V, Eriksson JE. Providing cellular signposts--post-translational modifications of intermediate filaments. FEBS Lett. 2008;582(14):2140-8.
329. Starger JM, Goldman RD. Isolation and preliminary characterization of 10-nm filaments from baby hamster kidney (BHK-21) cells. Proc Natl Acad Sci U S A. 1977;74(6):2422-6.
330. Da Q, Behymer M, Correa JI, Vijayan KV, Cruz MA. Platelet adhesion involves a novel interaction between vimentin and von Willebrand factor under high shear stress. Blood. 2014;123(17):2715-21.
331. Mor-Vaknin N, Punturieri A, Sitwala K, Markovitz DM. Vimentin is secreted by activated macrophages. Nat Cell Biol. 2003;5(1):59-63.
332. Pantel K, Brakenhoff RH, Brandt B. Detection, clinical relevance and specific biological properties of disseminating tumour cells. Nat Rev Cancer. 2008;8(5):329-40.
333. Luo W, Fang W, Li S, Yao K. Aberrant expression of nuclear vimentin and related epithelial-mesenchymal transition markers in nasopharyngeal carcinoma. Int J Cancer. 2012;131(8):1863-73.
334. Janmey PA, Euteneuer U, Traub P, Schliwa M. Viscoelastic properties of vimentin compared with other filamentous biopolymer networks. J Cell Biol. 1991;113(1):155-60.

335. Liu CY, Lin HH, Tang MJ, Wang YK. Vimentin contributes to epithelial-mesenchymal transition cancer cell mechanics by mediating cytoskeletal organization and focal adhesion maturation. Oncotarget. 2015;6(18):15966-83.
336. Berr AL, Wiese K, dos Santos G, Koch CM, Anekalla KR, Kidd M, et al. Vimentin is required for tumor progression and metastasis in a mouse model of non–small cell lung cancer. Oncogene. 2023;42(25):2074-87.
337. Horkovics-Kovats S, Traub P. Specific interaction of the intermediate filament protein vimentin and its isolated N-terminus with negatively charged phospholipids as determined by vesicle aggregation, fusion, and leakage measurements. Biochemistry. 1990;29(37):8652-7.
338. Evans DL, Harris DT, Leary JH, 3rd, St John AL, Jaso-Friedman L. Identification of a vimentin-like function associated molecule (FAM) on rat NK cells: evidence for receptor function. Scand J Immunol. 1993;37(2):131-42.
339. Perides G, Harter C, Traub P. Electrostatic and hydrophobic interactions of the intermediate filament protein vimentin and its amino terminus with lipid bilayers. The Journal of biological chemistry. 1987;262(28):13742-9.
340. Alevriadou BR, Moake JL, Turner NA, Ruggeri ZM, Folie BJ, Phillips MD, et al. Real-time analysis of shear-dependent thrombus formation and its blockade by inhibitors of von Willebrand factor binding to platelets. Blood. 1993;81(5):1263-76.
341. Fasipe TA, Hong SH, Da Q, Valladolid C, Lahey MT, Richards LM, et al. Extracellular Vimentin/VWF (von Willebrand Factor) Interaction Contributes to VWF String Formation and Stroke Pathology. Stroke. 2018;49(10):2536-40.
342. Riegger J, Brenner RE. Increase of cell surface vimentin is associated with vimentin network disruption and subsequent stress-induced premature senescence in human chondrocytes. eLife. 2023;12:e91453.
343. Xu B, deWaal RM, Mor-Vaknin N, Hibbard C, Markovitz DM, Kahn ML. The endothelial cell-specific antibody PAL-E identifies a secreted form of vimentin in the blood vasculature. Molecular and Cellular Biology. 2004;24(20):9198-206.
344. Huet D, Bagot M, Loyaux D, Capdevielle J, Conraux L, Ferrara P, et al. SC5 mAb represents a unique tool for the detection of extracellular vimentin as a specific marker of Sezary cells. J Immunol. 2006;176(1):652-9.
345. Thalla DG, Rajwar AC, Laurent AM, Becher JE, Kainka L, Lautenschläger F. Extracellular vimentin is expressed at the rear of activated macrophage-like cells: Potential role in enhancement of migration and phagocytosis. Front Cell Dev Biol. 2022;10:891281.
346. Georgatos SD, Blobel G. Lamin B constitutes an intermediate filament attachment site at the nuclear envelope. J Cell Biol. 1987;105(1):117-25.
347. Satelli A, Li S. Vimentin in cancer and its potential as a molecular target for cancer therapy. Cell Mol Life Sci. 2011;68(18):3033-46.
348. Mergui X, Puiffe ML, Valteau-Couanet D, Lipinski M, Bénard J, Amor-Guéret M. p21Waf1 expression is regulated by nuclear intermediate filament vimentin in neuroblastoma. BMC Cancer. 2010;10:473.
349. Hartig R, Shoeman RL, Janetzko A, Tolstonog G, Traub P. DNA-mediated transport of the intermediate filament protein vimentin into the nucleus of cultured cells. Journal of Cell Science. 1998;111(24):3573-84.

350. Hartig R, Shoeman RL, Janetzko A, Grüb S, Traub P. Active nuclear import of single-stranded oligonucleotides and their complexes with non-karyophilic macromolecules. Biol Cell. 1998;90(5):407-26.
351. Fan L, Zheng M, Zhou X, Yu Y, Ning Y, Fu W, et al. Molecular mechanism of vimentin nuclear localization associated with the migration and invasion of daughter cells derived from polyploid giant cancer cells. Journal of Translational Medicine. 2023;21(1):719.
352. Rodríguez JA. Interplay between nuclear transport and ubiquitin/SUMO modifications in the regulation of cancer-related proteins. Seminars in Cancer Biology. 2014;27:11-9.
353. Janmey PA, Euteneuer U, Traub P, Schliwa M. Viscoelastic properties of vimentin compared with other filamentous biopolymer networks. Journal of Cell Biology. 1991;113(1):155-60.
354. Wang N, Stamenović D. Contribution of intermediate filaments to cell stiffness, stiffening, and growth. American Journal of Physiology-Cell Physiology. 2000;279(1):C188-C94.
355. Goldman RD, Khuon S, Chou YH, Opal P, Steinert PM. The function of intermediate filaments in cell shape and cytoskeletal integrity. J Cell Biol. 1996;134(4):971-83.
356. Patteson AE, Vahabikashi A, Pogoda K, Adam SA, Mandal K, Kittisopikul M, et al. Vimentin protects cells against nuclear rupture and DNA damage during migration. J Cell Biol. 2019;218(12):4079-92.
357. Mendez MG, Restle D, Janmey PA. Vimentin enhances cell elastic behavior and protects against compressive stress. Biophys J. 2014;107(2):314-23.
358. Langlois B, Belozertseva E, Parlakian A, Bourhim M, Gao-Li J, Blanc J, et al. Vimentin knockout results in increased expression of sub-endothelial basement membrane components and carotid stiffness in mice. Sci Rep. 2017;7(1):11628.
359. Usman S, Waseem NH, Nguyen TKN, Mohsin S, Jamal A, Teh MT, et al. Vimentin Is at the Heart of Epithelial Mesenchymal Transition (EMT) Mediated Metastasis. Cancers (Basel). 2021;13(19).
360. Tata PR, Rajagopal J. Cellular plasticity: 1712 to the present day. Current Opinion in Cell Biology. 2016;43:46-54.
361. Tata PR, Rajagopal J. Cellular plasticity: 1712 to the present day. Curr Opin Cell Biol. 2016;43:46-54.
362. Hayakawa Y, Fox JG, Wang TC. Isthmus Stem Cells Are the Origins of Metaplasia in the Gastric Corpus. Cell Mol Gastroenterol Hepatol. 2017;4(1):89-94.
363. Chen M, Puschmann TB, Marasek P, Inagaki M, Pekna M, Wilhelmsson U, et al. Increased Neuronal Differentiation of Neural Progenitor Cells Derived from Phosphovimentin-Deficient Mice. Mol Neurobiol. 2018;55(7):5478-89.
364. Petrie RJ, Koo H, Yamada KM. Generation of compartmentalized pressure by a nuclear piston governs cell motility in a 3D matrix. Science. 2014;345(6200):1062-5.
365. Emad A, Ray T, Jensen TW, Parat M, Natrajan R, Sinha S, et al. An epithelial-mesenchymal-amoeboid transition gene signature reveals subtypes of breast cancer progression and metastasis. bioRxiv. 2018:219410.

366. Bhattacharya R, Gonzalez AM, DeBiase PJ, Trejo HE, Goldman RD, Flitney FW, et al. Recruitment of vimentin to the cell surface by β3 integrin and plectin mediates adhesion strength. Journal of Cell Science. 2009;122(9):1390-400.
367. Hynes RO. Integrins: bidirectional, allosteric signaling machines. Cell. 2002;110(6):673-87.
368. Kim J, Yang C, Kim EJ, Jang J, Kim S-J, Kang SM, et al. Vimentin filaments regulate integrin–ligand interactions by binding to the cytoplasmic tail of integrin β3. Journal of Cell Science. 2016;129(10):2030-42.
369. Kim J, Jang J, Yang C, Kim EJ, Jung H, Kim C. Vimentin filament controls integrin α5β1-mediated cell adhesion by binding to integrin through its Ser38 residue. FEBS Letters. 2016;590(20):3517-25.
370. Wiche G, Baker MA. Cytoplasmic network arrays demonstrated by immunolocalization using antibodies to a high molecular weight protein present in cytoskeletal preparations from cultured cells. Exp Cell Res. 1982;138(1):15-29.
371. Foisner R, Leichtfried FE, Herrmann H, Small JV, Lawson D, Wiche G. Cytoskeleton-associated plectin: in situ localization, in vitro reconstitution, and binding to immobilized intermediate filament proteins. J Cell Biol. 1988;106(3):723-33.
372. Mendez MG, Kojima S, Goldman RD. Vimentin induces changes in cell shape, motility, and adhesion during the epithelial to mesenchymal transition. Faseb j. 2010;24(6):1838-51.
373. Lauffenburger DA, Horwitz AF. Cell migration: a physically integrated molecular process. Cell. 1996;84(3):359-69.
374. De Pascalis C, Pérez-González C, Seetharaman S, Boëda B, Vianay B, Burute M, et al. Intermediate filaments control collective migration by restricting traction forces and sustaining cell-cell contacts. J Cell Biol. 2018;217(9):3031-44.
375. Costigliola N, Ding L, Burckhardt CJ, Han SJ, Gutierrez E, Mota A, et al. Vimentin fibers orient traction stress. Proc Natl Acad Sci U S A. 2017;114(20):5195-200.
376. Schaedel L, Lorenz C, Schepers AV, Klumpp S, Köster S. Vimentin intermediate filaments stabilize dynamic microtubules by direct interactions. Nature Communications. 2021;12(1):3799.
377. Esue O, Carson AA, Tseng Y, Wirtz D. A Direct Interaction between Actin and Vimentin Filaments Mediated by the Tail Domain of Vimentin*. Journal of Biological Chemistry. 2006;281(41):30393-9.
378. Eckes B, Dogic D, Colucci-Guyon E, Wang N, Maniotis A, Ingber D, et al. Impaired mechanical stability, migration and contractile capacity in vimentin-deficient fibroblasts. Journal of Cell Science. 1998;111(13):1897-907.
379. Zhu Y, Zhang Y, Sui Z, Zhang Y, Liu M, Tang H. USP14 de-ubiquitinates vimentin and miR-320a modulates USP14 and vimentin to contribute to malignancy in gastric cancer cells. Oncotarget. 2017;8(30):48725-36.
380. Cheng F, Shen Y, Mohanasundaram P, Lindström M, Ivaska J, Ny T, et al. Vimentin coordinates fibroblast proliferation and keratinocyte differentiation in wound healing via TGF-β-Slug signaling. Proc Natl Acad Sci U S A. 2016;113(30):E4320-7.

381. Garde A, Sherwood DR. Fueling Cell Invasion through Extracellular Matrix. Trends in Cell Biology. 2021;31(6):445-56.
382. Berr AL, Wiese K, dos Santos G, Davis JM, Koch CM, Anekalla KR, et al. Vimentin is Required for Tumor Progression and Metastasis in a Mouse Model of Non-Small Cell Lung Cancer. bioRxiv. 2020:2020.06.04.130963.
383. Rathje L-SZ, Nordgren N, Pettersson T, Rönnlund D, Widengren J, Aspenström P, et al. Oncogenes induce a vimentin filament collapse mediated by HDAC6 that is linked to cell stiffness. Proceedings of the National Academy of Sciences. 2014;111(4):1515-20.
384. Ding Y, Lv C, Zhou Y, Zhang H, Zhao L, Xu Y, et al. Vimentin loss promotes cancer proliferation through up-regulating Rictor/AKT/β-catenin signaling pathway. Experimental Cell Research. 2021;405(1):112666.
385. Salvador J, Hernandez GE, Ma F, Abrahamson CW, Pellegrini M, Goldman R, et al. Transcriptional Evaluation of the Ductus Arteriosus at the Single-Cell Level Uncovers a Requirement for Vim (Vimentin) for Complete Closure. Arteriosclerosis, thrombosis, and vascular biology. 2022;42(6):732-42.
386. Adair TH, Montani JP. Integrated Systems Physiology: from Molecule to Function to Disease. Angiogenesis. San Rafael (CA): Morgan & Claypool Life Sciences

387. Antfolk D, Sjöqvist M, Cheng F, Isoniemi K, Duran CL, Rivero-Muller A, et al. Selective regulation of Notch ligands during angiogenesis is mediated by vimentin. Proc Natl Acad Sci U S A. 2017;114(23):E4574-e81.
388. Kumar N, Robidoux J, Daniel KW, Guzman G, Floering LM, Collins S. Requirement of Vimentin Filament Assembly for β3-Adrenergic Receptor Activation of ERK MAP Kinase and Lipolysis*. Journal of Biological Chemistry. 2007;282(12):9244-50.
389. Yang CY, Chang PW, Hsu WH, Chang HC, Chen CL, Lai CC, et al. Src and SHP2 coordinately regulate the dynamics and organization of vimentin filaments during cell migration. Oncogene. 2019;38(21):4075-94.
390. Wilhelmsson U, Stillemark-Billton P, Borén J, Pekny M. Vimentin is required for normal accumulation of body fat. Biol Chem. 2019;400(9):1157-62.
391. Liu H, Ye G, Liu X, Xue M, Zhou Q, Zhang L, et al. Vimentin inhibits type I interferon production by disrupting the TBK1-IKKε-IRF3 axis. Cell Rep. 2022;41(2):111469.
392. Su LX, Pan P, Wang XT, Long Y, Liu DW, Zhou X. Vimentin modulates apoptosis and inflammatory cytokine release by a human monocytic cell line (THP-1) in response to lipopolysaccharides in vitro. Chin Med J (Engl). 2019;132(11):1336-43.
393. Olson EN, Nordheim A. Linking actin dynamics and gene transcription to drive cellular motile functions. Nat Rev Mol Cell Biol. 2010;11(5):353-65.
394. Posern G, Treisman R. Actin'together: serum response factor, its cofactors and the link to signal transduction. Trends in cell biology. 2006;16(11):588-96.
395. Cen B, Selvaraj A, Burgess RC, Hitzler JK, Ma Z, Morris SW, et al. Megakaryoblastic leukemia 1, a potent transcriptional coactivator for serum

response factor (SRF), is required for serum induction of SRF target genes. Molecular and Cellular Biology. 2003;23(18):6597-608.
396. GILLES C, POLETTE M, PIETTE J, DELVIGNE A-C, THOMPSON EW, FOIDART J-M, et al. VIMENTIN EXPRESSION IN CERVICAL CARCINOMAS: ASSOCIATION WITH INVASIVE AND MIGRATORY POTENTIAL. The Journal of Pathology. 1996;180(2):175-80.
397. Williams AA, Higgins JPT, Zhao H, Ljungberg B, Brooks JD. CD 9 and vimentin distinguish clear cell from chromophobe renal cell carcinoma. BMC Clinical Pathology. 2009;9(1):9.
398. KÖditz B, Stog A, GÖbel H, Heidegger I, Fries J, Heidenreich A, et al. Vimentin 3 Expression in Prostate Cancer Cells. Anticancer Res. 2021;41(1):169-74.
399. Lehtinen L, Ketola K, Mäkelä R, Mpindi JP, Viitala M, Kallioniemi O, et al. High-throughput RNAi screening for novel modulators of vimentin expression identifies MTHFD2 as a regulator of breast cancer cell migration and invasion. Oncotarget. 2013;4(1):48-63.
400. Richardson AM, Havel LS, Koyen AE, Konen JM, Shupe J, Wiles WG, et al. Vimentin Is Required for Lung Adenocarcinoma Metastasis via Heterotypic Tumor Cell-Cancer-Associated Fibroblast Interactions during Collective Invasion. Clin Cancer Res. 2018;24(2):420-32.
401. Mogre S, Makani V, Pradhan S, Devre P, More S, Vaidya M, et al. Biomarker Potential of Vimentin in Oral Cancers. Life (Basel). 2022;12(2).
402. Otsuki S, Inokuchi M, Enjoji M, Ishikawa T, Takagi Y, Kato K, et al. Vimentin expression is associated with decreased survival in gastric cancer. Oncol Rep. 2011;25(5):1235-42.
403. Al-Maghrabi J. Vimentin immunoexpression is associated with higher tumor grade, metastasis, and shorter survival in colorectal cancer. Int J Clin Exp Pathol [Internet]. 2020 2020; 13(3):[493-500 pp.]. Available from: http://europepmc.org/abstract/MED/32269687

https://europepmc.org/articles/PMC7137029

https://europepmc.org/articles/PMC7137029?pdf=render.
404. Ngan CY, Yamamoto H, Seshimo I, Tsujino T, Man-i M, Ikeda JI, et al. Quantitative evaluation of vimentin expression in tumour stroma of colorectal cancer. British journal of cancer. 2007;96(6):986-92.
405. Sun S, Poon RT, Lee NP, Yeung C, Chan KL, Ng IO, et al. Proteomics of hepatocellular carcinoma: serum vimentin as a surrogate marker for small tumors (<or=2 cm). J Proteome Res. 2010;9(4):1923-30.
406. Nussler AK, Vergani G, Gollin SM, Dorko K, Gansauge S, Morris SM, et al. Isolation and characterization of a human hepatic epithelial-like cell line (AKN-1) from a normal liver. In Vitro Cellular & Developmental Biology-Animal. 1999;35(4):190-7.
407. Shimizu Y, Demetris AJ, Gollin SM, Storto PD, Bedford HM, Altarac S, et al. Two new human cholangiocarcinoma cell lines and their cytogenetics and responses to growth factors, hormones, cytokines or immunologic effector cells. International journal of cancer. 1992;52(2):252-60.

408. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a cancer dependency map. Cell. 2017;170(3):564-76. e16.
409. Galbraith DW, Anderson MT, Herzenberg LA. Chapter 19: Flow Cytometric Analysis and FACS Sorting of Cells Based on GFP Accumulation. In: Sullivan KF, Kay SA, editors. Methods in Cell Biology. 58: Academic Press; 1998. p. 315-41.
410. Lefkovits I, Waldmann H. Limiting dilution analysis of cells in the immune system: Cambridge University Press; 1979.
411. Greenfield EA. Single-Cell Cloning of Hybridoma Cells by Limiting Dilution. Cold Spring Harbor Protocols. 2019;2019(11).
412. Pageon SV, Nicovich PR, Mollazade M, Tabarin T, Gaus K. Clus-DoC: a combined cluster detection and colocalization analysis for single-molecule localization microscopy data. Mol Biol Cell. 2016;27(22):3627-36.
413. Hari-Gupta Y, Fili N, Dos Santos Á, Cook AW, Gough RE, Reed HCW, et al. Myosin VI regulates the spatial organisation of mammalian transcription initiation. Nat Commun. 2022;13(1):1346.
414. Elson EL, Magde D. Fluorescence correlation spectroscopy. I. Conceptual basis and theory. Biopolymers: Original Research on Biomolecules. 1974;13(1):1-27.
415. Goulding J, Kondrashov A, Mistry SJ, Melarangi T, Vo NTN, Hoang DM, et al. The use of fluorescence correlation spectroscopy to monitor cell surface β2-adrenoceptors at low expression levels in human embryonic stem cell-derived cardiomyocytes and fibroblasts. Faseb j. 2021;35(4):e21398.
416. Kruger NJ. The Bradford method for protein quantitation. The protein protocols handbook: Springer; 2009. p. 17-24.
417. Wang T, Antonacci-Fulton L, Howe K, Lawson HA, Lucas JK, Phillippy AM, et al. The Human Pangenome Project: a global resource to map genomic diversity. Nature. 2022;604(7906):437-46.
418. Nassar LR, Barber GP, Benet-Pagès A, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2023 update. Nucleic acids research. 2022;51(D1):D1188-D95.
419. Rodríguez A, Rodríguez M, Córdoba JJ, Andrade MJ. Design of primers and probes for quantitative real-time PCR methods. Methods Mol Biol. 2015;1275:31-56.
420. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic acids research. 2001;29(9):e45.
421. Justus CR, Leffler N, Ruiz-Echevarria M, Yang LV. In vitro cell migration and invasion assays. J Vis Exp. 2014(88).
422. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. Nucleic acids research. 2019;47(W1):W556-W60.
423. Community TG. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. Nucleic acids research. 2022;50(W1):W345-W51.
424. Andrews S. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom; 2010.

425. Krueger F. Trim galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. 2015;516(517).
426. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome research. 1998;8(3):186-94.
427. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105-11.
428. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-30.
429. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(12):550.
430. Martin FJ, Amode MR, Aneja A, Austine-Orimoloye O, Azov Andrey G, Barnes I, et al. Ensembl 2023. Nucleic acids research. 2022;51(D1):D933-D41.
431. Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, et al. FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics. 2015;15(15):2597-601.
432. Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. Nucleic acids research. 2015;43(W1):W566-W70.
433. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005;102(43):15545-50.
434. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature Genetics. 2003;34(3):267-73.
435. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14:128.
436. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic acids research. 2016;44(W1):W90-7.
437. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology. 2009;10(3):R25.
438. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012;9(4):357-9.
439. Institute, B. (n.d.). Picard. In Broad Institute, GitHub repository. GitHub. http://broadinstitute.github.io/picard/ [
440. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biology. 2008;9(9):R137.
441. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. Nature Protocols. 2012;7(9):1728-40.

442. Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics. 2010;26(17):2204-7.
443. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2.
444. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature Communications. 2019;10(1):1523.
445. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576-89.
446. Watanabe R, Hirano Y, Hara M, Hiraoka Y, Fukagawa T. Mobility of kinetochore proteins measured by FRAP analysis in living cells. Chromosome Research. 2022;30(1):43-57.
447. Hochreiter B, Pardo-Garcia A, Schmid JA. Fluorescent Proteins as Genetically Encoded FRET Biosensors in Life Sciences. Sensors. 2015;15(10):26281-314.
448. Herzenberg LA, Tung J, Moore WA, Herzenberg LA, Parks DR. Interpreting flow cytometry data: a guide for the perplexed. Nature Immunology. 2006;7(7):681-5.
449. Ramboer E, De Craene B, De Kock J, Berx G, Rogiers V, Vanhaecke T, et al. Development and characterization of a new human hepatic cell line. Excli j. 2015;14:875-89.
450. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nature Methods. 2012;9(7):676-82.
451. Arkill KP, Qvortrup K, Starborg T, Mantell JM, Knupp C, Michel CC, et al. Resolution of the three dimensional structure of components of the glomerular filtration barrier. BMC Nephrology. 2014;15(1):24.
452. Kaufman WL, Kocman I, Agrawal V, Rahn H-P, Besser D, Gossen M. Homogeneity and persistence of transgene expression by omitting antibiotic selection in cell line isolation. Nucleic acids research. 2008;36(17):e111-e.
453. Westermann L, Li Y, Göcmen B, Niedermoser M, Rhein K, Jahn J, et al. Wildtype heterogeneity contributes to clonal variability in genome edited cells. Sci Rep. 2022;12(1):18211.
454. Monberg ME, Geiger H, Lee JJ, Sharma R, Semaan A, Bernard V, et al. Occult Polyclonality of Preclinical Pancreatic Cancer Models Drives in Vitro Evolution. Nature Communications. 2022;13(1).
455. Llufrio EM, Wang L, Naser FJ, Patti GJ. Sorting cells alters their redox state and cellular metabolome. Redox Biol. 2018;16:381-7.
456. Marcolino E, Siddiqui YH, van den Bosch M, Poole AW, Jayaraman P-S, Gaston K. Blood platelets stimulate cancer extravasation through TGFβ-mediated downregulation of PRH/HHEX. Oncogenesis. 2020;9(2):10.
457. Harvey AR, Ehlert E, de Wit J, Drummond ES, Pollett MA, Ruitenberg M, et al. Use of GFP to analyze morphology, connectivity, and function of cells in the

central nervous system. Viral Applications of Green Fluorescent Protein: Methods and Protocols. 2009:63-95.
458. Wang T, Larcher LM, Ma L, Veedu RN. Systematic Screening of Commonly Used Commercial Transfection Reagents Towards Efficient Transfection of Single-Stranded Oligonucleotides. Molecules. 2018;23(10):2564.
459. Ansari AM, Ahmed AK, Matsangos AE, Lay F, Born LJ, Marti G, et al. Cellular GFP Toxicity and Immunogenicity: Potential Confounders in in Vivo Cell Tracking Experiments. Stem Cell Rev Rep. 2016;12(5):553-9.
460. Franke WW, Schmid E, Osborn M, Weber K. Different intermediate-sized filaments distinguished by immunofluorescence microscopy. Proceedings of the National Academy of Sciences. 1978;75(10):5034-8.
461. Ostrowska-Podhorodecka Z, Ding I, Lee W, Tanic J, Abbasi S, Arora PD, et al. Vimentin tunes cell migration on collagen by controlling β1 integrin activation and clustering. Journal of Cell Science. 2021;134(6).
462. Miller DM, Thomas SD, Islam A, Muench D, Sedoris K. c-Myc and cancer metabolism. Clin Cancer Res. 2012;18(20):5546-53.
463. Mergui X, Puiffe M-L, Valteau-Couanet D, Lipinski M, Bénard J, Amor-Guéret M. p21 Waf1 expression is regulated by nuclear intermediate filament vimentin in neuroblastoma. BMC cancer. 2010;10(1):473.
464. Banday S, Farooq Z, Rashid R, Abdullah E, Altaf M. Role of Inner Nuclear Membrane Protein Complex Lem2-Nur1 in Heterochromatic Gene Silencing. The Journal of biological chemistry. 2016;291(38):20021-9.
465. Glushonkov O, Réal E, Boutant E, Mély Y, Didier P. Optimized protocol for combined PALM-dSTORM imaging. Scientific Reports. 2018;8(1):8749.
466. Schwille P, Haustein E. Fluorescence correlation spectroscopy. A tutorial for the Biophysics Textbook Online (BTOL) Biophysical Society, Rockville, MD. 2002.
467. Shin CJ, Wong S, Davis MJ, Ragan MA. Protein-protein interaction as a predictor of subcellular location. BMC Syst Biol. 2009;3:28.
468. Ivanov AA, Khuri FR, Fu H. Targeting protein-protein interactions as an anticancer strategy. Trends Pharmacol Sci. 2013;34(7):393-400.
469. Sivagurunathan S, Vahabikashi A, Yang H, Zhang J, Vazquez K, Rajasundaram D, et al. Expression of vimentin alters cell mechanics, cell-cell adhesion, and gene expression profiles suggesting the induction of a hybrid EMT in human mammary epithelial cells. Frontiers in cell and developmental biology. 2022:1839.
470. Aaron JS, Taylor AB, Chew T-L. Image co-localization – co-occurrence versus correlation. Journal of Cell Science. 2018;131(3).
471. Li H, Liu J, Lai J, Su X, Wang X, Cao J, et al. The HHEX-ABI2/SLC17A9 axis induces cancer stem cell-like properties and tumorigenesis in HCC. Journal of Translational Medicine. 2024;22(1):537.
472. Kagami T, Chen S, Memar P, Choi MK, Foster LJ, Numata M. Identification and Biochemical Characterization of the SLC9A7 Interactome. Molecular Membrane Biology. 2008;25(5):436-47.
473. Bill A, Espinola S, Guthy D, Haling JR, Lanter M, Lu M, et al. EndoBind detects endogenous protein-protein interactions in real time. Communications Biology. 2021;4(1):1085.

474. Isolation of SDS-Stable Complexes of the Intermediate Filament Protein Vimentin with Repetitive, Mobile, Nuclear Matrix Attachment Region, and Mitochondrial DNA Sequence Elements from Cultured Mouse and Human Fibroblasts. DNA and Cell Biology. 2001;20(9):531-54.
475. Herrick-Davis K, Grinde E, Lindsley T, Cowan A, Mazurkiewicz JE. Oligomer size of the serotonin 5-hydroxytryptamine 2C (5-HT2C) receptor revealed by fluorescence correlation spectroscopy with photon counting histogram analysis: evidence for homodimers without monomers or tetramers. Journal of Biological Chemistry. 2012;287(28):23604-14.
476. Wei J, Gibbs JS, Hickman HD, Cush SS, Bennink JR, Yewdell JW. Ubiquitous Autofragmentation of Fluorescent Proteins Creates Abundant Defective Ribosomal Products (DRiPs) for Immunosurveillance. The Journal of biological chemistry. 2015;290(26):16431-9.
477. Cassany A, Gerace L. Reconstitution of nuclear import in permeabilized cells. Methods Mol Biol. 2009;464:181-205.
478. Hegazy M, Cohen-Barak E, Koetsier JL, Najor NA, Arvanitis C, Sprecher E, et al. Proximity Ligation Assay for Detecting Protein-Protein Interactions and Protein Modifications in Cells and Tissues in Situ. Curr Protoc Cell Biol. 2020;89(1):e115.
479. Porter KA, Xia B, Beglov D, Bohnuud T, Alam N, Schueler-Furman O, et al. ClusPro PeptiDock: efficient global docking of peptide recognition motifs using FFT. Bioinformatics. 2017;33(20):3299-301.
480. Tang G, Cho M, Wang X. OncoDB: an interactive online database for analysis of gene expression and viral infection in cancer. Nucleic acids research. 2021;50(D1):D1334-D9.
481. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. Nature genetics. 2000;25(1):25-9.
482. Paccione RJ, Miyazaki H, Patel V, Waseem A, Gutkind JS, Zehner ZE, et al. Keratin down-regulation in vimentin-positive cancer cells is reversible by vimentin RNA interference, which inhibits growth and motility. Molecular Cancer Therapeutics. 2008;7(9):2894-903.
483. Saentaweesuk W, Araki N, Vaeteewoottacharn K, Silsirivanit A, Seubwai W, Talabnin C, et al. Activation of Vimentin Is Critical to Promote a Metastatic Potential of Cholangiocarcinoma Cells. Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics. 2017;26.
484. Salic A, Mitchison TJ. A chemical method for fast and sensitive detection of DNA synthesis in vivo. Proceedings of the National Academy of Sciences. 2008;105(7):2415-20.
485. Pagano M. Application of electrophoresis and related methods, such as western blotting and zymography to the study of some proteins and enzymes. Analytica Chimica Acta. 1999;383(1–2):119-25.
486. Jonkman J, Brown CM, Wright GD, Anderson KI, North AJ. Tutorial: guidance for quantitative confocal microscopy. Nature Protocols. 2020;15(5):1585-611.
487. Parker AL, Teo WS, McCarroll JA, Kavallaris M. An Emerging Role for Tubulin Isotypes in Modulating Cancer Biology and Chemotherapy Resistance. Int J Mol Sci. 2017;18(7).

488. Sun ZG, Yadav V, Amiri S, Cao W, De La Cruz EM, Murrell M. Cofilin-mediated actin filament network flexibility facilitates 2D to 3D actomyosin shape change. European Journal of Cell Biology. 2024;103(1):151379.
489. Thompson EW, Newgreen DF, Tarin D. Carcinoma invasion and metastasis: a role for epithelial-mesenchymal transition? Cancer Res. 2005;65(14):5991-5; discussion 5.
490. KORITA PV, WAKAI T, AJIOKA Y, INOUE M, TAKAMURA M, SHIRAI Y, et al. Aberrant Expression of Vimentin Correlates with Dedifferentiation and Poor Prognosis in Patients with Intrahepatic Cholangiocarcinoma. Anticancer Research. 2010;30(6):2279-85.
491. Mao X, Chen D, Wu J, Li J, Zhou H, Wu Y, et al. Differential expression of fascin, E-cadherin and vimentin: Proteins associated with survival of cholangiocarcinoma patients. The American journal of the medical sciences. 2013;346(4):261-8.
492. Rappaport N, Fishilevich S, Nudel R, Twik M, Belinky F, Plaschkes I, et al. Rational confederation of genes and diseases: NGS interpretation via GeneCards, MalaCards and VarElect. Biomedical Engineering Online. 2017;16(1):1-14.
493. Zhao X, Yu C, Zheng M, Sun J. Prognostic value of the mRNA expression of gap junction α members in patients with gastric cancer. Oncol Lett. 2019;18(2):1669-78.
494. Wagenblast E, Soto M, Gutiérrez-Ángel S, Hartl CA, Gable AL, Maceli AR, et al. A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis. Nature. 2015;520(7547):358-62.
495. Fayard B, Bianchi F, Dey J, Moreno E, Djaffer S, Hynes NE, et al. The serine protease inhibitor protease nexin-1 controls mammary cancer metastasis through LRP-1–mediated MMP-9 expression. Cancer research. 2009;69(14):5690-8.
496. Prieto TG, Machado-Rugolo J, Baldavira CM, Velosa APP, Teodoro WR, Saber AMA, et al. The Fibrosis-Targeted Collagen/Integrins Gene Profile Predicts Risk of Metastasis in Pulmonary Neuroendocrine Neoplasms. Front Oncol. 2021;11.
497. Tan Z, Zhang Z, Yu K, Yang H, Liang H, Lu T, et al. Integrin subunit alpha V is a potent prognostic biomarker associated with immune infiltration in lower-grade glioma. Frontiers in Neurology. 2022;13.
498. Wu Y, Zheng J, Yan Y, Liu J, Zhou Y. Gelsolin can be a prognostic biomarker and correlated with immune infiltrates in gastric cancer. International Journal of General Medicine. 2022:927-36.