Profiling and Functional Interactomics of Circular RNAs across Human Stem Cell Differentiation

By

Stuart Timothy Webb, B. Sci. (Hons)

A thesis submitted to Flinders University for the degree of

Doctor of Philosophy

June 2022

College of Medicine and Public Health

Flinders University

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and the research within will not be submitted for any other future degree or diploma without the permission of Flinders University; and to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

Stuart Timothy Webb

19/11/2021

Contents

Declaration ii

List of Figures ix

List of Tables xiii

Acknowledgments xiv

Abbreviations xvii

Thesis Summary xx

Chapter 1 Literature Review 1

- 1.1 Introduction 2
- 1.2 CircRNA Biogenesis, Subcellular Localisation 3
- 1.3 Circular RNA functions and interactions 6
  - 1.3.1 Interactions with DNA 7
  - 1.3.2 Interactions with RNA 8
  - 1.3.3 Interactions with protein 9
  - 1.3.4 Protein coding circRNAs 10
- 1.4 Pathological implications of circRNAs 11
- 1.5 Epithelial-Mesenchymal Transition (EMT) – Pathway for motility 13
- 1.6 Human Embryonic Stem Cell Differentiation and EMT 15
- 1.7 EMT in Cancer 18
- 1.8 Partial/Hybrid EMT 19
- 1.9 Alternative splicing 22
- 1.10 Profiling and function of circRNAs across EMT 23
- 1.11 Cell Cycle 27
- 1.12 Pathways for pluripotency and differentiation 30
  - 1.12.1 TGFβ in hESC differentiation 32
  - 1.12.2 TGFB and the cell cycle 37
  - 1.12.3 Cell cycle and circRNAs 39
- 1.13 Methods for the reduction of RNA transcript abundances 39
  - 1.13.1 Small hairpin RNAs 39
  - 1.13.2 Small interfering RNAs 40
  - 1.13.3 Antisense Oligonucleotides 41
  - 1.13.4 Conventional RNA interference comparison 41
  - 1.13.5 CRISPR 42
- 1.14 Summary 44
- 1.15 General hypothesis and aims 44
  - 1.15.1 Aim 1 45
  - 1.15.2 Aim 2 45

1.15.3 Aim 3 45
1.16 Experiments/data in thesis completed/given by others 45
1.16.1 CircRNA-seq samples and next generation sequencing 45
1.16.2 H9 Karyotype analysis 46
1.16.3 qPCR analysis of neurons 46
1.16.4 Immunofluorescence staining and imaging 46
1.16.5 Biotinylated Probe pulldown 46
1.16.6 CRISPR Cas13b 47
Chapter 2 Materials and Methods 48
2.1 Chemicals 49
2.2 RNA samples for hESC RNA sequencing 49
2.3 Next Generation Sequencing 49
2.3.1 Circ-RNA seq library preparation and high-throughput sequencing and bioinformatics analysis 49
2.3.2 MRNA seq library prep and high-throughput sequencing and bioinformatic analysis 50
2.3.3 Total-RNA Sequencing 51
2.4 Cell Culture 52
2.4.1 Cell lines 52
2.4.2 Growth Medium 53
2.4.3 Extracellular Matrices 53
2.4.3.1 Geltrex 54
2.4.3.2 Poly-L lysine 54
2.4.4 Passaging cells 54
2.4.5 Freezing/Thawing Cells 54
2.4.6 Embryoid Body formation 55
2.4.7 Antibiotics, Chemicals and solutions used during Cell Culture 55
2.4.7.1 Puromycin 55
2.4.7.2 G418 56
2.4.7.3 Blasticidin 56
2.4.7.4 Polybrene 56
2.4.7.5 Ethanol 56
2.4.7.6 Propidium Iodide 56
2.4.7.7 Hoechst 34580 57
2.4.7.8 HumanKine™ Transforming Growth Factor-β1 57
2.4.7.9 Paraformaldehyde 57
2.4.7.10 Phosphate-buffered saline 57

        - 2.4.7.11 TrypLE Express...57
    - 2.4.8 Cell cycle analysis..58
        - 2.4.8.1 Fixed cell cycle analysis..58
        - 2.4.8.2 Live cell cycle analysis..58
        - 2.4.8.3 Cell Synchronisation by Serum starvation 58
        - 2.4.8.4 Cell Synchronisation by Nocodazole treatment 59
    - 2.4.9 Production of lentiviral particles..59
    - 2.4.10 Viral transduction of hESCs 60
    - 2.4.11 Overexpression of circRNAs 60
- 2.5 Bacterial Culture and plasmid preparation 61
    - 2.5.1 Bacterial strain and culture 61
    - 2.5.2 Plasmids 62
    - 2.5.3 Design of CRISPR Cas13b gRNA for pLKO.1 plasmid 62
    - 2.5.4 Design of shRNA cassette for pLKO.1 plasmid 63
    - 2.5.5 Miniprep and Midiprep of plasmids...63
    - 2.5.6 Nucleic acid concentration..64
    - 2.5.7 Digestion of DNA 64
    - 2.5.8 Gel purification 64
    - 2.5.9 Column purification 65
    - 2.5.10 Ligation 65
    - 2.5.11 Gateway Cloning..65
    - 2.5.12 Screening for positive colonies - Colony PCR 65
    - 2.5.13 Agarose gel electrophoresis 66
- 2.6 Polymerase Chain Reaction (PCR)...66
    - 2.6.1 Genomic DNA Isolation 66
    - 2.6.2 Total RNA isolation...66
    - 2.6.3 Reverse transcription of total RNA to cDNA 67
    - 2.6.4 Oligonucleotide design and provision...67
    - 2.6.5 PCR reactions and running conditions..68
    - 2.6.6 qPCR 69
        - 2.6.6.1 Selection of house-keeping genes 69
        - 2.6.6.2 Sanger Sequencing 69
- 2.7 CircRNA pulldown..70
    - 2.7.1 CircRNA pulldown cell transfection...70
    - 2.7.2 CircRNA pulldown probes..70
    - 2.7.3 *CircACVR2A* pulldown probes 70
    - 2.7.4 Preparation of Streptavidin C1 magnetic beads...................................71

2.7.5 *circACVR2A* circRNA-protein pulldown 71
2.8 Mass spectrometry 72
2.8.1 MS sample preparation 72
2.8.2 MS data analysis 72
2.9 CellNet analysis 73
2.10 Immunofluorescence processing 73
2.11 RNA immunoprecipitation 75
2.11.1 RNA immunoprecipitation cell transfection 75
2.11.2 RNA immunoprecipitation of Cyclin H FLAG protein 75
2.12 Western blotting 76
2.12.1 Protein preparation 76
2.12.2 SDS-PAGE electrophoresis and transfer 77
2.12.3 Antibody probing 78
2.12.4 Antibodies 79
Chapter 3 Transcriptomics of Human Embryonic Stem Cell Differentiation 80
3.1 Introduction 81
3.2 Results 86
3.2.1 CircRNA Sequencing 86
3.2.2 CircRNA Differential Expression Analysis using FDR-adjusted p-values 89
3.2.3 CircRNA differential expression is only calculated based on backsplice reads 92
3.2.4 CircRNA Differential Expression Analysis using non FDR-adjusted p-values 94
3.2.5 Candidate circRNA selection 99
3.2.6 Poly-A mRNA sequencing 101
3.2.6.1 Correlation between circRNA and parental gene expression 106
3.2.7 Sequencing Validation 114
3.2.8 Selection of House-keeping Genes for qRT-PCR during Stem Cell Differentiation 115
3.3 Discussion 119
3.4 Chapter 3 Conclusion 125
Chapter 4 Stable Knockdown of Circular RNAs in Human Embryonic Stem Cells 126
4.1 Introduction 127
4.2 Results 128
4.2.1 Design of gRNAs targeting circRNA backsplice hESCs 128

4.2.2 Design of shRNAs targeting circRNA backsplice junction 130
4.2.3 Stable knockdown of circRNAs in H9 hESCs by targeting backsplice junction..131
4.2.4 Knockdown of *circFAT3* 133
4.2.5 Knockdown of *circCDYL*...134
4.2.5.1 CircCDYL drives alternative splicing of CDYL 135
4.2.6 Knockdown of *circPIF1* 138
4.2.7 Knockdown of *circZNF609* 139
4.2.8 Knockdown of *circACVR2A* 140
4.2.8.1 Knockdown of ACVR2A does not induce EB formation 148
4.2.9 Knockdown of *circACVR2A* with shRNAs targeting non-backsplice sequences 150
4.3 Discussion..153
4.3.1 CircRNAs regulating parental gene transcripts 158
4.4 Chapter 4 Conclusion 161
Chapter 5 Differentiation of Human Embryonic Stem Cells Induced by Knockdown of *CircACVR2A*...162
5.1 Introduction..163
5.2 Results..164
5.2.1 *CircACVR2A* EBs have altered adherence properties and neural differentiation...164
5.2.1.1 CircACVR2A knockdown embryoid bodies are non-adherent...164
5.2.1.2 CircACVR2A knockdown embryoid body differentiation on different ECMs..165
5.2.1.3 Neural-like cell production by circACVR2A shRNA EBs 168
5.2.1.4 QRT-PCR analysis of EBs and outgrown neural-like cells 171
5.2.1.5 Embryoid body immunofluorescent analysis 174
5.2.2 RNA-seq profiling indicates *circACVR2A* knockdown embryoid bodies have undergone neural differentiation 177
5.2.2.1 Total-RNA sequencing of EBs...177
5.2.2.2 RNA-seq Analysis of H9 circACVR2A knockdown EBs 178
5.2.2.3 Gene Ontology Enrichment analysis..181
5.2.2.4 CellNet analysis of H9 circACVR2A shRNA EBs.....................185
5.2.2.5 Neuronal markers by RNA-seq 190
5.2.2.6 Adhesion markers by RNA-seq..192
5.2.2.7 Changes to hESC maintenance and differentiation signalling pathways...194
5.2.2.8 CircACVR2A does not appear to act as a microRNA sponge in hESCs 199

5.2.3 RNA-seq analysis for *circFAT3* and *circCDYL* 201
5.3 Discussion 203
5.3.1 Immunofluorescence 205
5.3.2 RNA-seq analysis 206
5.4 Chapter 5 Conclusion 210
Chapter 6 Profiling *circACVR2A*-Protein Interactions and Functional Consequences in A549 Cells 212
6.1 Introduction 213
6.2 Results 214
6.2.1 *CircACVR2A* is not regulated during hTGFβ1-induced EMT in A549 cells. 214
6.2.2 A549 cells with decreased *circACVR2A* 549 cells show changes to EMT markers in response to hTGFβ1 218
6.2.3 Knockdown of *circACVR2A* does not affect SMAD levels 220
6.2.4 Knockdown of *circACVR2A* negatively affects proliferation of A549 cells. 221
6.2.5 Knockdown of *circACVR2A* negatively affects migration of A549 cells and nuclear size 222
6.2.6 *CircACVR2A* pull-down and Mass Spectrometry 225
6.2.7 RNA immunoprecipitation of Cyclin H 228
6.2.8 Cell cycle analysis of *CircACVR2A* knockdown A549s 233
6.2.9 Regulation of *circACVR2A* during the cell cycle 234
6.2.10 Decreased CDK1/2 activity during *circACVR2A* KD 239
6.3 Discussion 240
6.4 Chapter 6 Conclusion 245
Chapter 7 Major Project Findings 247
7.1 Thesis summary 248
Appendix 1 251
Appendix 2 252
Appendix 3 255
Appendix 4 268
Appendix 5 269
Appendix 6 270
References 271

List of Figures

Figure 1-1 Linear versus circular RNA splicing 4
Figure 1-2 Interactions of circRNAs 6
Figure 1-3 Phenotypic and Molecular changes in EMT.......................................14
Figure 1-4 Different subtypes of EMT...21
Figure 1-5 Partial EMT in cancer...22
Figure 1-6 Alternative splicing drives RNA diversity 23
Figure 1-7 Quaking drives circRNA biogenesis during EMT 25
Figure 1-8 CAK complex in the cell cycle...29
Figure 1-9 Differentiation pathways for pluripotent stem cells 31
Figure 1-10 TGFβ signalling pathway 33
Figure 1-11 Mitogenic signalling for stem cell differentiation 34
Figure 1-12 Differentiation of EBs produces cells from all three lineages...........36
Figure 1-13 Propensity for differentiation of hESCs in early and late G1............38
Figure 1-14 Current technologies for targeted cleavage of RNAs........................40
Figure 2-1 Oligonucleotide sequence cloned into pLKO.1 vector for Cas13b 63
Figure 2-2 Design of circRNA qPCR primers 68
Figure 3-1 Alignment of RNA-seq reads 81
Figure 3-2 Overview of RNA-seq library preparation..82
Figure 3-3 Backsplice junction of circRNAs 84
Figure 3-4 Germ layers and resulting terminal cell lines 86
Figure 3-5 Venn diagram showing the number of detected circRNAs with backsplice junction reads overlapping from each lineage.....................................87
Figure 3-6 Analysis of circRNA-seq data 89
Figure 3-7 Venn diagram showing the number of overlapping statistically significantly differentially expressed circRNAs (padj<0.05) from each lineage..91
Figure 3-8 Distribution of circRNA backsplice reads in circRNA-seq data.........93
Figure 3-9 Venn diagram showing the number of overlapping statistically significantly expressed circRNAs (p<0.05) from each lineage.............................95
Figure 3-10 Abundance of commonly regulated circRNAs in undifferentiated H9 hESCs..99
Figure 3-11 circRNA candidate Log2Fold changes in differentiated cell lines..101
Figure 3-12 hESC and mRNA lineage markers in hESCs and differentiated cells 102
Figure 3-13 PCA plots of mRNA-seq data 104
Figure 3-14 Changes to markers of pluripotency and EMT................................105
Figure 3-15 Differences between number of significantly differentially expressed genes during differentiation...106
Figure 3-16 Correlation analysis of most abundant circRNAs and parental transcripts in hESCs 107
Figure 3-17 Correlation analysis of most abundant and statistically differentially expressed circRNAs and parental transcripts in each differentiated lineage 108
Figure 3-18 List of Log2Fold changes of candidate circRNAs and their parental transcripts 109
Figure 3-19 Correlation between expression of some candidate circRNAs and parental transcripts 110
Figure 3-20 Correlation between circRNA and parental transcripts of commonly statistically differentially expressed circRNAs (p<0.05) 111

Figure 3-21 Heatmap of Log2Fold changes of 90 commonly statistically differentially expressed circRNAs (p<0.05) 112
Figure 3-22 Correlation between circRNA and parental transcripts of commonly differentially expressed circRNAs (padj<0.05)...113
Figure 3-23 Heatmap of Log2Fold changes of 90 commonly differentially expressed circRNAs (padj<0.05) 113
Figure 3-24 Changes in expression of RNA Binding Proteins implicated in circRNA biogenesis with circRNA abundance after differentiation..................114
Figure 3-25 Design of qPCR primers and sequence confirmation of circRNA PCR amplificon...115
Figure 3-26 Changes to housekeeping genes during differentiation in RNA-seq data 116
Figure 3-27 *YWHAZ* standard curve..117
Figure 3-28 Sequencing Validation of *YWHAZ* amplicon 117
Figure 3-29 Validation of lineage-specific circRNAs in the circRNA-seq data .118
Figure 4-1 Knockdown of *circACVR2A* by CRISPR Cas13b is not effective in H9 hESCs 129
Figure 4-2 shRNA design for targeting circRNA backsplice junction 130
Figure 4-3 Empty Vector and scrambled control shRNA plasmids show no altered phenotype in H9 hESCs 133
Figure 4-4 Reduction of *circFAT3* by shRNA induces the loss of compact colonies in H9 hESCs 134
Figure 4-5 Reduction of *circCDYL* by shRNA induces the loss of compact colonies in H9 hESCs 135
Figure 4-6 Changing the levels of *circCDYL* abundance alters the splicing *CDYL* mRNA isoforms in H9 hESCs..137
Figure 4-7 Reduction of *circPIF1* by shRNA induces adherent EB formation in H9 hESCs 139
Figure 4-8 Reduction of *circZNF609* by shRNA induces altered cell morphology in H9 hESCs 140
Figure 4-9 *circACVR2A* primetime qPCR standard curve 141
Figure 4-10 Reduction of *circACVR2A* by shRNA induces EB formation in H9 hESCs 142
Figure 4-11 Reduction of *circACVR2A* by a third shRNA induces EB formation in H9 hESCs 143
Figure 4-12 Knockdown of *circACVR2A* in hESCs cultured on vitronectin 144
Figure 4-13 Increased hESC EB proportion using half *circACVR2A* shRNA 1 lentivirus...145
Figure 4-14 Neural rosette formation in hESC colony from *circACVR2A* shRNA 2 Knockdown on geltrex 145
Figure 4-15 Neural rosette structures observed in *circACVR2A* knockdown induced EBs 147
Figure 4-16 Morphological classification of EBs following *circACVR2A* knockdown 148
Figure 4-17 *ACVR2A* mRNA shRNAs were designed downstream of *circACVR2A* 149
Figure 4-18 Knockdown of *ACVR2A* mRNA by shRNA does not alter hESC morphology 150
Figure 4-19 Image of *circACVR2A* and shRNA locations.................................151

Figure 4-20 Knockdown of *circACVR2A* with non-backsplice targeting shRNAs 152
Figure 5-1 Embryoid Body attachment and outgrowth phenotypes on different binding matrices 167
Figure 5-2 Neural-like outgrowth of *circACVR2A* shRNA 2 EBs......................169
Figure 5-3 Neural-like outgrowth of *circACVR2A* shRNA 2 EBs......................170
Figure 5-4 Differentiation of neural cells from Neurosphere.............................171
Figure 5-5 Pluripotency and neuronal specific QRT-PCR analysis of *circACVR2A* shRNA induced EBs...173
Figure 5-6 EV EBs and *circACVR2A* shRNA 2 EBs are similar sizes for Immunofluorescence comparisons...174
Figure 5-7 NANOG pluripotency factor is decreased in early *circACVR2A* shRNA 2 EBs 175
Figure 5-8 POU5F1 pluripotency factor is retained in late *circACVR2A* shRNA 2 EBs 176
Figure 5-9 FOXA2 midbrain floorplate marker is expressed in early *circACVR2A* shRNA 2 EBs 176
Figure 5-10 Changes to pluripotency and EMT transcripts through RNA-seq with *circACVR2A* shRNA knockdown in hESCs..179
Figure 5-11 Changes to lineage specific transcripts through RNA-seq with *circACVR2A* shRNA knockdown in hESCs..181
Figure 5-12 GO analysis of differentially expressed RNA transcripts common to both shRNA treatments compared to H9 EV hESCs 183
Figure 5-13 GO analysis of differentially expressed RNA transcripts of EV control EBs, and early and late *circACVR2A* shRNA 2 EBs..185
Figure 5-14 CellNet analysis of differentiate cells used in mRNA-seq.............187
Figure 5-15 CellNet analysis of *circACVR2A* shRNA knockdown EBs............188
Figure 5-16 CellNet analysis of GRN status of *circACVR2A* shRNA knockdown EBs 189
Figure 5-17 Changes to neural RNA transcripts in late *circACVR2A* shRNA EBs 191
Figure 5-18 Changes to Adhesion RNA transcripts in *circACVR2A* shRNA EBs. 193
Figure 5-19 Changes to TGFβ signalling pathway transcripts in *circACVR2A* shRNA EBs by RNA-seq 196
Figure 5-20 Changes to WNT signalling pathway transcripts in *circACVR2A* shRNA EBs by RNA-seq 197
Figure 5-21 Changes to FGF signalling pathway transcripts in *circACVR2A* shRNA EBs by RNA-seq 198
Figure 5-22 Putative miRNA binding sites on *circACVR2A* 200
Figure 5-23 No changes in both *circACVR2A* shRNA treatments for miRNA target genes...200
Figure 5-24 Pluripotency and EMT markers altered in *circFAT3* and *circCDYL* shRNA knockdown H9 hESCs...202
Figure 5-25 GO analysis of differentially expressed RNA transcripts from *circFAT3* and *circCDYL* knockdown hESCs compared to H9 EV hESCs 203
Figure 6-1 Analysis of *circACVR2A* localisation in bladder cancer 214
Figure 6-2 EMT factor response to hTGF β1 induced EMT in A549 cells 215
Figure 6-3 A549 *circACVR2A* shRNA knockdown efficiency and sensitivity to TGFβ1 216

Figure 6-4 A549 *circACVR2A* shRNA knockdown cells responsive to TGFβ1 . 217
Figure 6-5 EMT marker changes with *circACVR2A* shRNA knockdown in A549s 218
Figure 6-6 Sensitivity of EMT markers during TGFB induced EMT of *circACVR2A* knockdown A549s 220
Figure 6-7 No changes to SMAD2/3 protein levels with *circACVR2A* shRNA knockdown in A549s 221
Figure 6-8 Proliferation assay of *circACVR2A* shRNA knockdown in A549s ... 222
Figure 6-9 Migration assay of *circACVR2A* shRNA knockdown A549s 224
Figure 6-10 Migration assay of *circACVR2A* shRNA knockdown A549s 225
Figure 6-11 String analysis of Cyclin H 228
Figure 6-12 Western blot verifying pcDNA 3.1 CCNH expression with C-terminal FLAG tag 229
Figure 6-13 QRT-PCR of *circACVR2A* overexpression during CCNH FLAG RIP 230
Figure 6-14 Western blog of RNA immunoprecipitation of CCNH by FLAG antibody bound to Dynabeads 231
Figure 6-15 Raw fluorescence from qRT-PCR of CCNH FLAG RIP detecting *circACVR2A, U1* snRNA, and *GAPDH*. 232
Figure 6-16 Cell cycle analysis of *circACVR2A* shRNA knockdown A549s 234
Figure 6-17 *circACVR2A* cell cycle profiling after serum starvation 236
Figure 6-18 *circACVR2A* cell cycle profiling after Nocodazole treatment 238
Figure 6-19 *circACVR2A* abundance in serum starved and nocodazole treated cells 239
Figure 6-20 Western blot of phosphor-CDK substrates in *circACVR2A* shRNA knockdown A549s 240

List of Tables

Table 2-1 Antibiotic concentration for LB Broth and Agar 61
Table 2-2 Plasmid information used for mammalian cell culture 62
Table 2-3 Antibodies used in this study 79
Table 3-1 Number of unique circRNAs detected in each circRNA-seq sample ... 87
Table 3-2 Statistically differentially expressed circRNAs (padj<0.05) from each lineage 91
Table 3-3 List of commonly regulated significantly differentially expressed circRNAs in all cell lineages 91
Table 3-4 Statistically significantly differentially expressed circRNAs from each lineage 94
Table 3-5 Commonly statistically significantly regulated circRNAs in all differentiated lineages 95
Table 3-6 Statistically significantly differentially expressed circRNAs that are exclusively differentially expressed in each lineage 97
Table 3-7 Proportion of commonly statistically differentially expressed circRNAs of the total number of statistically differentially expressed circRNAs for each lineage 98
Table 3-8 Criteria for circRNA candidate selection 100
Table 3-9 Statistically differentially expressed genes during hESC differentiation in each lineage 106
Table 4-1 Candidate circRNAs with cellular affect after knockdown 132
Table 5-1 Adherence of H9 *circACVR2A* shRNA 2 EBs and EV EBs 165
Table 5-2 Samples for qRT-PCR analysis by SAHMRI 172
Table 5-3 Maturation of control and *circACVR2A* knockdown EBs used for Immunofluorescence 174
Table 5-4 List of samples for total-RNA sequencing 177
Table 5-5 Number of statistically differentiated genes from H9 EV control (p<0.05) 178
Table 5-6 Statistically differentially expressed RNA transcripts for *circFAT3* and *circCDYL* knockdown in H9 hESCs 201
Table 6-1 Doubling time of control and *circACVR2A* shRNA knockdown A549s 222
Table 6-2 Proteins identified from biotinylated probe pulldown and mass spec 227

Acknowledgments

I would first like to thank my fiancé Ayaka Miyazawa. You have been instrumental for helping me maintain my sanity over the course of my PhD, with talking everyday, travelling continents to be with me, and then supporting me after moving in with me. I can't wait for our lifelong journey to start when we can travel the world together and eat many different types of foods (let's hope we don't get food poisoning too often).

I wish to thank my family, Mum, Dad, Steph, Steve and little Jackson. Mum and Dad, you have given me the roof over my head, the food in my belly and for the most part, the clothes on my back. But much more than that, you have supported all my life choices and made me the person I am today.

An incredibly big thank you to the now Professor Simon Conn (congratulations on losing the Associate), for taking me on before my PhD, and for your constant support throughout the 3.5 years of my candidature. I have learnt so much during my PhD, and not just from the project within this thesis, but also with being a good scientist and overcoming setbacks and challenges. A further thank you to my secondary supervisor Professor Ross McKinnon who would come to check up on me, be present at all my candidature reviews, and also for purchasing vital equipment for my project. A big thank you to the rest of the circRNA laboratory at Flinders including (in order of length of time in the lab) Dr Vanessa Conn, Kirsty Kirk, Dr Marta Gabryelska and Dr Brett Stringer. Everyone has helped with discussions, presentation practises, coffee breaks or physically helping me with the techniques or generating data/figures. A special thank you to Kirsty for allowing myself and Ayaka look after your beautiful dog Lucy on a number of occasions.

A further thank you to the previous level 3 FCIC honours students Jordyn Binelli, Timothy Penn and Aasha Purling. It was fun helping with your projects, having discussions with other students, and also going to the amazing socially distant Halloween party at Aasha's house. Although our overlapping time at Flinders was brief, I would like to say congratulations and thank you to Dr Saira Ali and Dr Ayla Orang for completing your PhDs, and for your help when I started in the building as a fresh PhD student.

Another thank you to all the members of our neighbouring laboratory in Luke Selth's lab. Whether it was for help with techniques, borrowing reagents, having lunch together or grabbing coffee on level 4, you have been amazing. This also extends to the other laboratories in FCIC including all members from Associate Professor Michael Michael's lab and Professor Janni Petersen's lab. Within the hospital, I would like to thank proteomics for answering questions about Mass Spectrometry, and completing said Mass Spectrometry, as well at Pat Vilimas in Flinders microscopy with teaching me how to not remove my fingers when using a microtome or cryostat.

I would also like to thank Associate Professor Dan Thomas and his laboratory members located at SAHMRI, as they have taken me on in their lab for my 20-day Australian Commonwealth Scholarship placement. Watching another lab troubleshooting, overcoming challenges, writing publications, and being involved in lab meetings and journal clubs has given me a broader knowledge of the field and also cellular metabolism that I was not expecting to learn within my last 6 months of my PhD, but definitely worth it.

Finally, I would like to wish a big thank you to the Flinders College of Medicine and Public Health and the Australian Government Research Training Scholarship for my RTP stipend that allowed me to concentrate on my PhD. A further acknowledgement to the people at ScopeGlobal and the Commonwealth Scholarships Program for South Australia that has provided additional support during my PhD and organising my placement within Associate Professor Dan Thomas's laboratory.

Abbreviations

µl; ml; l: microlitre; millilitre; litre

ρM; µM; mM; M: picomolar; micromolar; millimolar; molar

ActB: Actin Beta

AS: alternative splicing

ASO: antisense oligonucleotides

bp: base pairs

Cardio: cardiomyocytes

CCNH: Cyclin H

CDK: cyclin-dependent kinases

cDNA: complementary dna

circRNA: circular rna

circRNA-seq: circular rna sequencing

CSC: cancer stem cell

CTC: circulating tumour cell

CV: coefficient of variation

DE: definitive endoderm

DESeq: differential expression analysis

DPBS: dulbecco's phosphate-buffered saline

DMEM: dulbecco's modified eagle medium

DN or Dopa: dopaminergic neurons

DNA: deoxyribonucleic acid

DTT: diothiothreitol

dNTPs: dinucleotide triphosphates

EB: embryoid body

ECM: extracellular matrix

EDTA: ethylenediaminetetraacetic acid

EMT: epithelial-mesenchymal transition

ESC: embryonic stem cells

EV: empty vector control

FBS: fetal bovine serum

FDR: false discovery rate

FP: floor plate precursors

G1: gap1 of growth 1

G2: gap2 or growth 2

GO: gene ontology

Gapdh: Glyceraldehyde 3-phosphate dehydrogenase

H9: female h9 hesc

H9 EV: female h9 hesc empty vector control cell line

H9 EV EBs: Embryoid bodies of female h9 hesc empty vector control cells

hESC: human embryonic stem cells

hr: hour

hr: hours

HRP: horseradish peroxidase

IF: immunofluorescence

iPSC: induced pluripotent stem cell

IRES: internal ribosomal entry sites

KD: knockdown

kDa: kilodaltons

LB: luria broth

Liver: hepatocytes

MET: mesenchymal-epithelial transition

miRNA: microrna

mRNA: messenger rna

MS: mass spectrometry

mW: molecular weight

ng; µg; mg; kg: nanograms; micrograms; milligrams; kilograms

NGS: next generation sequencing

padj: fdr adjusted p value

PBS: phosphate buffered saline

PCR: polymerase chain reaction

Pol: polymerase

QKI: quaking

qRT-PCR: quantitative real time polymerase chain reaction

Rb: retinoblastoma protein

RBPs: rna binding proteins

RNA: ribonucleic acid

Rnase R: ribonuclease r

RNA-seq: rna sequencing

RPKM: fragments per kilobase of transcript per million mapped reads

RT: room temperature

S: synthesis

SDS: sodium dodecyl sulphate

secs: seconds

SEM: standard error of the mean

shRNA: small hairpin rna

siRNA: small interfering rna

Std. Dev.: standard deviation

TBE: tris-borate EDTA

TGFβ: tumour growth factor beta

TF: transcription factor

UV: ultraviolet

VTN-N: vitronectin

WB: western blot

YWHAZ: tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta

Thesis Summary

Human embryonic stem cells (hESC) are a powerful tool to examine many aspects of cell development, differentiation and disease. Directed differentiation of ESCs into specific lineages can be achieved with high fidelity and reproducibility using specific growth factors invoking dramatic phenotypic and transcriptomic responses. Each of these differentiation processes incorporate a key process called epithelial-mesenchymal transition (EMT), which has a significant prognostic impact in cancer. EMT is the conversion of stationary cells into cells with greater motility and is indispensable for processes including cellular differentiation, wound repair and organ development. Therefore, as EMT is required for cellular differentiation, transcriptomic profiling was performed to investigate circular RNA (circRNA) expression changes and molecular functions during hESC differentiation into cell types of the three germ cell lineages.

Alternative Splicing (AS) is a regulated process that generates a variety of RNA molecules from the same pre-mRNA transcript. AS is observed in ~95% of multi-exonic human genes resulting in unique mRNA transcripts that may function synonymously, in opposition to, or in an unrelated manner. Importantly, AS is regulated during cell differentiation leading to changes in both coding and non-coding RNAs.

CircRNAs are a highly abundant class of RNA, ubiquitous among eukaryotes and usually located within protein coding genes. Circular RNAs are produced during a non-canonical AS event, termed back-spicing, and therefore changes to AS have a significant impact on circRNA biosynthesis which leads to the production of differentiation and EMT-associated circRNAs. Currently, over 400,000 circRNAs have been identified and despite a growing repertoire of functions being reported, this is only for comparatively few circRNAs. Therefore, the vast proportion (>99%) remain functionally enigmatic.

Performing high-throughput RNA-sequencing of H9 hESCs differentiated into 5 different cell types, lead to the identification of 27 candidate circRNAs that are either (i) commonly regulated during each cell-specific differentiation event, or (ii) showed lineage-specific changes in expression. Furthermore, unique characteristics of individual circRNAs, including, differential splicing of exons and co-regulation with their parental gene's mRNA transcripts provides insight into additional regulation of circRNAs. To study their effect on differentiation, and associated programs including EMT, circRNA abundance was manipulated through knockdown (KD) experiments. Reduction to *circACVR2A* levels was found to induce consistent and rapid neuronal differentiation in hESCs and partially block the cell cycle in A549 cells. Undertaking a *circACVR2A*-protein interactome analysis, Cyclin H was identified as a binding partner. Cyclin H, together with CDK7 and MAT1, forms the CDK-activating kinase (CAK) trimeric complex, necessary for regulating proliferation and cell cycle progression. Altering *circACVR2A* levels altered CDK activity and phosphorylation of targets, providing a functional explanation for altered passage through the cell cycle and for cellular phenotypes observed in both cell lines, altering cell fate and resulting in neuronal differentiation of hESCs.

The outcome of this project identified that specific circRNAs play critical functional roles within the context of regulating parental mRNA levels, cell cycle control, human stem cell differentiation and EMT.

Chapter 1 Literature Review

CHAPTER 1

1.1 Introduction

The central dogma of genetics proposes that DNA is transcribed into RNA, then translated into protein. However, this is a vastly over-simplified view. Discovery of different functional RNA molecules that arise from the same DNA locus as protein coding mRNAs, yet encode a different variant of the protein, or no protein all together (so-called non-coding RNA), further advanced the notion of complex gene regulation (Tomizawa *et al.*, 1981). Many such RNA species have been discovered and have been shown to have roles in protein synthesis (Hoagland, Keller and Zamecnik, 1956), RNA splicing (Matera and Wang, 2014) and gene regulation (Lee, Feinbaum and Ambros, 1993). Among the RNAs identified, circular RNAs (circRNAs), which can be derived from exonic, intronic, intergenic or exonic-intronic sequences (Meng *et al.*, 2017), remain poorly characterised as they have been difficult to detect by traditional RNA sequencing methods (Jeck *et al.*, 2013). CircRNAs are covalently-closed single stranded, largely non-coding RNA transcripts and, due to their circular form, do not possess a 5’ cap or 3’ poly A tail (Memczak *et al.*, 2013). Due to these factors, circRNAs are extremely stable as, unlike messenger RNA (mRNA), they are immune from exoribonuclease attack which target unpaired RNA termini to cleave phosphodiester bonds (Salzman *et al.*, 2013). Despite being able to be targeted by endoribonucleases, this circular form confers a longer half-life, exceeding 48 hours (Jeck *et al.,* 2013), with one study demonstrating the median half-lives of 60 circRNAs being 2.5 times longer than their linear counterparts (Enuka *et al.*, 2016).

1.2 CircRNA Biogenesis, Subcellular Localisation

Due to their head-to-tail splicing pattern, circRNAs were initially thought to be splicing artefacts (Nigro *et al.*, 1991; Cocquerelle *et al.*, 1993). However, recent studies have highlighted that circRNAs are more widespread than first believed yet are not pervasive, but cell-specific (Salzman *et al.*, 2013), gene-specific (Jeck *et al.*, 2013) and expressed across many eukaryotic (Wang *et al.*, 2014; Sun *et al.*, 2016), and certain prokaryotic species (Danan *et al.*, 2012; LaRoche-Johnston *et al.*, 2018). Recently, circRNA species have been reported to be transcribed from viruses (Huang *et al.*, 2019) with 72774 circRNA molecules encoded by viruses identified by computational tools (Cai *et al.*, 2021). Investigations into the biogenesis of circRNAs revealed that flanking reverse complementary sequences, including *Alu* repeats (Jeck *et al.*, 2013), complementary sequences (Zhang *et al.*, 2014), and RNA-binding protein recognition sites (Conn *et al.*, 2015; Aktaş *et al.*, 2017; Errichelli *et al.*, 2017; Fei *et al.*, 2017) are able to drive circularisation (Figure 1-1) whereby a downstream 3' splice site is ligated to an upstream 5' splice site (Geng *et al.*, 2020). Remarkably, *circMBL*, expressed in both humans and *Drosophila Melanogaster*, generated from the second exon of the *muscleblind* gene, is circularised by the encoded *muscleblind* (MBL) protein (Ashwal-Fluss *et al.*, 2014). MBL binds within the introns adjacent to the second exon at conserved muscleblind binding sites, promoting circularisation of *circMBL* by bridging the two flanking intronic sequences. Furthermore, *circMBL* also has conserved MBL binding sites and is able to sponge MBL protein when in excess, regulating MBL levels by competing with linear splicing.

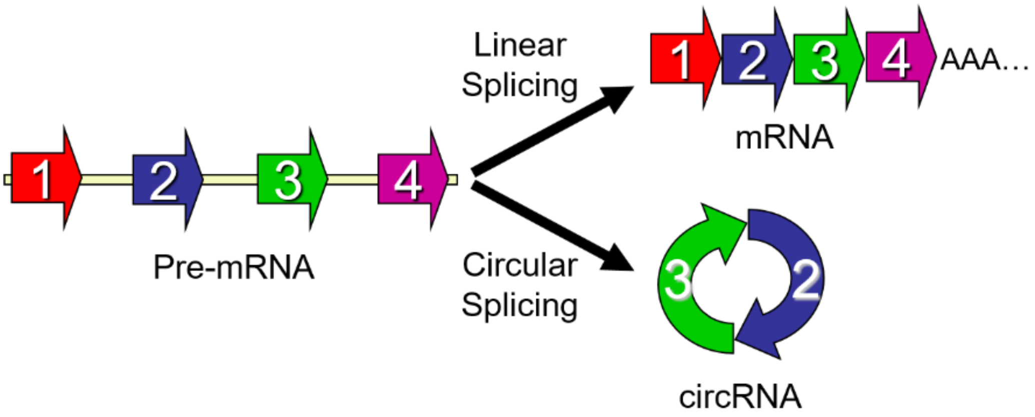


Figure 1-1 Linear versus circular RNA splicing

Canonical (linear) splicing of pre-mRNA leads to the sequential ligation of exons to form a mature mRNA. CircRNAs are largely non-coding RNAs that form from a 3’ end of a downstream exon binding to a 5’ end of an upstream exon in a process known as backsplicing. This leads to a covalently-closed circular RNA transcript that lacks 5’ or 3’ termini. This figure displays a two-exon circRNA comprising exons 2 and 3 of the pre-mRNA.

Notably, our laboratory demonstrated that over one-third of EMT related circRNAs are regulated by alternative splicing factor Quaking (QKI), and *de novo* exonic circRNAs could be generated by introduction of QKI splicing sites adjacent to exon boundaries (Conn *et al.*, 2015), thereby converting linear splicing to circular splicing. Negative regulation of circRNA expression is also observed as double-strand RNA-editing enzyme ADAR1, known to modify *Alu* elements (Athanasiadis, Rich and Maas, 2004; Levanon *et al.*, 2004), alters the complementary stem structures that promote circularisation (Ivanov *et al.*, 2015).

RNA binding proteins (RBPs) may not only contribute to circularisation but also to subcellular distribution as Zhang *et al.* (2019) revealed a wide subcellular distribution of circRNAs within the nucleus, cytoplasm, ribosome, cytosol, and exosome. Some circRNAs are predominantly localised in the nucleus and recruit proteins to chromatin (Wang *et al.*, 2019, p. 1) as well as promote protein nuclear retention (Q. Yang *et al.*, 2017). Hundreds of mitochondrial-encoded circRNAs have also been identified, with evidence suggesting they act as molecular chaperones and bind to protein and RNA importation factors including TOM40 and PNPASE (X. Liu *et al.*, 2020; Wu *et al.*, 2020; Zhao *et al.*, 2020). Although circRNAs are overwhelmingly localised in the cytoplasm (Salzman *et al.*, 2012), RNAi depletion of URH49 and UAP56 in *Drosophila Melanogaster* were found to enrich the levels of short and long circRNAs, respectively, in the nucleus (C. Huang *et al.*, 2018). However, despite orthologues present in human and mouse, this phenomenon has yet to be demonstrated in mammalian cells. Nuclear export by recruitment of YTHDC1 to N6-methyladenosine (m6A) sites on *circNSUN2* provides further evidence of an active system for the translocation of circRNAs (R.-X. Chen *et al.*, 2019). These systems imply an active circRNA transport process mediated by RNA binding proteins that bind RNA sequence motifs (J. Zhang *et al.*, 2019, p. 2), bind to RNA modifications or control the shutting of circRNAs in a length-dependent manner.

While most circRNAs are present at lower abundance compared to their cognate linear mRNA, there are exceptions such as highly expressed circRNA *CDR1as* (Salzman *et al.*, 2012, 2013; Jeck *et al.*, 2013). Guo *et al.* compiled a list of 57 circRNAs, including *CDR1as*, with which the circular RNA was detected higher than other transcript isoforms in upwards of 39 different cell lines (Guo *et al.*, 2014). The diversity of unique circRNAs, combined with their long half-lives, cell-and-gene specific expression, evidence of active biosynthesis of circRNAs by RBPs (Ashwal-Fluss *et al.*, 2014; Conn *et al.*, 2015; Khan *et al.*, 2016; Errichelli *et al.*, 2017; Fei *et al.*, 2017), and detectable RNA modifications (Zhou *et al.*, 2017; Park *et al.*, 2019; Di Timoteo *et al.*, 2020), advocates functional roles for circRNAs.

1.3 Circular RNA functions and interactions

Recent publications have described the ability of circRNAs to bind a range of molecular targets including miRNAs (Hansen *et al.*, 2013; Memczak *et al.*, 2013; Ashwal-Fluss *et al.*, 2014; Zheng *et al.*, 2016; Piwecka *et al.*, 2017; Zhang *et al.*, 2017; Song and Li, 2018; C. Yang *et al.*, 2018; S. Chen *et al.*, 2019), proteins (Du *et al.*, 2016; Abdelmohsen *et al.*, 2017; Q. Yang *et al.*, 2017; N. Chen *et al.*, 2018; Wu *et al.*, 2019), as well as DNA forming R-loops (Conn *et al.*, 2017; Y. Liu *et al.*, 2020; Liu *et al.*, 2021), and interact with ribosomes to encode for proteins, if they possess internal ribosome entry site (IRES)-like motifs (Legnini *et al.*, 2017; Pamudurti *et al.*, 2017; Begum *et al.*, 2018; Y. Yang *et al.*, 2018; M. Zhang *et al.*, 2018) (Figure 1-2).

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 1-2 Interactions of circRNAs

CircRNAs can; (a) bind to DNA forming R-loops and causing transcriptional pausing or altering splicing of genes; (b) enhance parental gene transcription by binding to U1 snRNP which interacts with Pol II; (c) bind to and sponge miRNA; (d) interact with proteins; (e) encode for protein and can be transcribed if they contain IRES sites; (f) be methylated which recruits translation factors inducing protein expression. (Image: (Zhou et al., 2020)).

MiRNAs typically function by silencing RNA through base-pairing to complementary RNA sequences leading to decreased translation efficiency or degradation of the RNA transcript (Fabian, Sonenberg and Filipowicz, 2010). A small number of circRNAs contain a high density of miRNA binding sites – *CDR1as*, *Sry* etc. (Hansen *et al.*, 2013; Memczak *et al.*, 2013), and are highly expressed in comparison to the target miRNA; two requirements that are necessary to be an effective miRNA sponge (Guo *et al.*, 2014; Z. Li *et al.*, 2015; Ebbesen, Hansen and Kjems, 2017). However, few circRNAs satisfy these requirements and thus the miRNA sponge hypothesis is not a stoichiometrically favourable phenomenon for the majority of reported circRNA:miRNA interactions (Stagsted *et al.*, 2019; Chaudhary *et al.*, no date). Studies have also concluded that the vast majority of circRNAs do not show an enrichment for miRNA target sites, except for *CDR1as* and *circZNF91* as they harbour more miRNA binding sites than expected by chance for their size, with *CDR1as* possessing 74 *miR-7* binding sites across its 1485 nucleotide length (Guo *et al.*, 2014; Stagsted *et al.*, 2019). Overexpression of circRNAs far above average physiological levels allows for enrichment of circRNAs for downstream applications, however, this may give misleading evidence for the ability of a normally lowly abundant circRNA to effectively sponge miRNAs. While miRNA sponging is the most commonly reported function for circRNAs, the majority of circRNAs are stoichiometrically unlikely to function as effective miRNA sponges with only a few examples being physiologically relevant.

1.3.1 Interactions with DNA

A miRNA or protein sponge provides a practical regulatory function for highly abundant circRNAs, however the low abundance (0.1%-10% expression compared to parental mRNA expression) (Lasda and Parker, 2014) and small size of the majority of circRNAs (~500bp) (Zheng *et al.*, 2016) would not allow for optimum sponging. Therefore, a circRNA-DNA interaction would be a stoichiometrically favourable option for low copy number circRNAs. A seminal paper from my laboratory demonstrated that a circRNA in *Arabidopsis thaliana* could bind to its cognate locus and drive alternative splicing through the formation of R-loops (DNA-RNA hybrid), leading to exon skipping in the parental mRNA (Conn *et al.*, 2017). Following this work, circRNAs forming R-loops in other species of plants were discovered affecting chromatin structure (Y. Liu *et al.*, 2020) and altering splicing via intron retention of parental gene transcripts (Liu *et al.*, 2021).

1.3.2 Interactions with RNA

Production of a circRNA from a gene has the capacity to impact splicing of the cognate mRNA as circRNA splicing competes with canonical splicing as shown by *circMBL* in *Drosophila* and humans (Ashwal-Fluss *et al.*, 2014). Interestingly, circRNAs can also affect cognate RNA levels by RNA-RNA interactions as exemplified by a class of circRNAs that have retained introns between exons, termed exon-intron circRNAs (EIciRNA). Evidence suggests that EIciRNAs interact with the RNA component of the small nuclear Ribonucleoprotein (snRNP) *U1* by RNA-RNA interaction which further interacts with DNA polymerase II (Pol II) at parental gene promotors promoting transcription (Zhang *et al.*, 2013; Z. Li *et al.*, 2015). Similarly, intronic RNA sequences that circularise are also able to affect parental genes, with an intronic circRNA from *ANKRD52* co-localising at the parent gene loci, associating with elongation RNA Pol II machinery and acting as a positive transcription regulator (Zhang *et al.*, 2013). Bioinformatical analysis by

Guo *et al.* (2014) found that approximately 20% of mammalian circRNAs had reads that mapped to introns, suggesting this effect may be even more widely spread. As EIciRNAs and intronic circRNAs can affect transcription of their parental gene, a high number of molecules may not be required for an effect on parental RNA levels and offers a more favourable stochiometric interaction than miRNA sponging.

1.3.3 Interactions with protein

Multiple circRNAs have been identified as binding to proteins resulting in the dissociation or interaction or protein complexes (Zhang *et al.*, 2013; Ashwal-Fluss *et al.*, 2014; Z. Li *et al.*, 2015; Abdelmohsen *et al.*, 2017; F. Yang *et al.*, 2018). *CircCCNB1* binds to both CCNB1 and CDK1, dissociating their interaction and decreasing proliferation by arresting cell cycle progression (Fang *et al.*, 2019, p. 1). Conversely, *circFOXO3* binds to both p53 and MDM2, promoting their interaction, thereby enhancing p53 degradation by ubiquitination (William W Du *et al.*, 2017). CircRNAs can also block complex formation or interaction of proteins involved with binding DNA and RNA. *CircHuR* sequesters CNBP, a nucleic acid binding protein, from binding to the *HuR* promotor, thereby downreguating HuR protein and repressing tumour progression (F. Yang *et al.*, 2019). HuR is also affected by *circPABPN1*, which binds to and supresses the RNA binding protein binding to the *PABPN1* mRNA, negatively affecting the mRNA stability decreasing translation of PABPN1 (Abdelmohsen *et al.*, 2017). Transcription is further affected by circRNAs by recruitment of transcription factors (TFs), modifying enzymes or chromatin remodelers. *CircRHOT1* recruits TIP60 to *NR2F6* gene promotor to initiate transcription with KD of *circRHOT1* suppressing hepatocellular carcinoma proliferation, invasion and migration (Wang *et al.*, 2019, p. 1). Expanding on the epigenetic roles for circRNAs, a study by Chen *et al.* (2018,

p. 1) discovered that an exonic circRNA from the *FLI1* gene coordinates the methylation status of its parental gene promoter by interactions with methylating and demethylating enzymes. Furthermore, a circRNA from the *MRPS35* gene, a tumour suppressor gene in gastric cancer, was found to act as a molecular scaffold and recruit histone acetyltransferase KAT7 to gene promotors thereby eliciting histone acetylation, activating transcription of target genes (Jie *et al.*, 2020). It is clear that the protein binding capacity of circRNAs have significant biological and, in some cases, pathological roles in humans.

1.3.4 Protein coding circRNAs

Since the first discovery of protein coding potential of *circMBL* from the muscleblind (*MBL*) gene in *Drosophila Melanogaster* (Pamudurti *et al.*, 2017), a number of other circRNAs have been identified that produce peptide sequences (Legnini *et al.*, 2017; Y. Yang *et al.*, 2017, 2018, p. 7; M. Zhang *et al.*, 2018; Liang *et al.*, 2019; Zheng *et al.*, 2019; Weigelt *et al.*, 2020). Of note, m^6A modifications found on circRNAs, confer protein-coding potential as hundreds of circRNA encoded peptides were detected by mass spectrometry in human cells (Y. Yang *et al.*, 2017) and polypeptides encoded by the backsplice junction of circRNAs in developing and mature germ cells during murine spermatogenesis (Tang *et al.*, 2020). Interestingly, circRNA *HPV16-circE7* encoded in human papillomaviruses is detectable in HPV16-transformed cells and is m6A modified, leading to the translation of an E7 oncoprotein (Zhao *et al.*, 2019).

1.4 Pathological implications of circRNAs

Although the functions for the majority of circRNAs have not yet been determined, specific circRNAs have been found with important roles in the fields of cancer progression (Geng *et al.*, 2020), cell proliferation (Dang, Liu and Li, 2017; X. Wang *et al.*, 2018; Yu *et al.*, 2020), cellular mobility (Lin *et al.*, 2018; Yu *et al.*, 2020), apoptosis (Geng *et al.*, 2016; Song and Li, 2018) angiogenesis (He *et al.*, 2018, p. 1; Garikipati *et al.*, 2019), and other cellular alterations. This includes cancers from tissues and organs of the lung (Xu *et al.*, 2018; Zhou *et al.*, 2019), breast (D. Gao *et al.*, 2017; Yan *et al.*, 2017; Y. Liu *et al.*, 2018, p. 1; Zeng *et al.*, 2018; Sang *et al.*, 2019), bladder (C. Yang *et al.*, 2018; X. Chen *et al.*, 2018; Dong *et al.*, 2019; Li *et al.*, 2019), pancreas (Shao *et al.*, 2018; Xu, Yu and Ding, 2018), blood (Pan *et al.*, 2018; J. Shang *et al.*, 2019), and other tissues (Lee *et al.*, 2019). CircRNAs have also been associated with multiple diseases and are promising diagnostic and prognostic markers for disease onset and progression (Vausort *et al.*, 2016; Verduci *et al.*, 2021). Diseases include cardiovascular disease (Geng *et al.*, 2016), diabetes (Xu *et al.*, 2015; Zhao *et al.*, 2017), neurological diseases (Zhao *et al.*, 2016; Shi *et al.*, 2017), autoimmune diseases (Holdt *et al.*, 2016; Chen *et al.*, 2017; C.-X. Liu *et al.*, 2019) and viral infections (Mo *et al.*, 2021) as well as bacterial infections (Z.-K. Huang *et al.*, 2018).

Certain circRNAs are detectable in extracellular fluids, including blood (Memczak *et al.*, 2015), saliva (Jafari Ghods, 2018), seminal plasma (Zhang, Yang and Xiao, 2018) and within exosomes (Y. Li *et al.*, 2015). The ease of sample collection and hyperstability of circRNAs makes them clinically attractive and promising prognostic and diagnostic biomarkers for diseases as well as possible intracellular drug targets facilitating diagnosis and treatment (Ahmed *et al.*, 2016;

Görlach and Holdenrieder, 2017; Yang, Fu and Zhou, 2018). Recent studies have hypothesised roles for circRNAs in contributing to tumour metastasis by activating the EMT pathway by binding and perturbing miRNA functions (Q. Chen *et al.*, 2018; X. Zhang *et al.*, 2018, p. 2; Zeng *et al.*, 2018; Zhou *et al.*, 2018; Li *et al.*, 2019; Z. Yang *et al.*, 2019).

It is known that circRNAs are hyper stable compared to their parental gene mRNA, however it is not well known how circRNAs are degraded by cells to regulate their expression independent from mRNA degradation. Park *et al.* (2019) showed that circRNAs containing m6A modifications are subject to cleavage by endoribonucleases RNase P/MRP complexed with YTHDF2, a m6A reader protein, and HRSP12 adapter protein. Recent studies have found that RNase L non-specifically degrades all circRNA and mRNAs within cells during viral infection. The production of mRNA is more efficient than circRNA production at maintaining mRNA levels during RNase L degradation (Y. Zhang *et al.*, 2016), whereas circRNA levels are heavily diminished from RNase L degradation (C.-X. Liu *et al.*, 2019).

Interestingly, genome wide association studies identified that the majority of disease associated single nucleotide polymorphisms (SNP) lie within non-coding regions where some non-coding RNAs are produced, including long non-coding RNAs and intergenic circRNAs (Maurano *et al.*, 2012; Giral, Landmesser and Kratzer, 2018; Zhao and Qu, 2020). This suggests that changes to non-coding RNAs in the absence of changes to parental coding sequences, can contribute to disease pathogenesis, highlighting a functional importance for circRNA coding sequences as well as other non-coding RNAs.

1.5 Epithelial-Mesenchymal Transition (EMT) – Pathway for motility

When required, cells undergo EMT to gain more motile properties, indispensable for processes including embryogenesis, tissue remodelling, wound healing, stem cell differentiation and cancer metastasis (Lamouille, Xu and Derynck, 2014; Neumann, Goodall and Gregory, 2018). EMT allows cells to undergo a myriad of biochemical and structural changes (Figure 1-3) and is dependent on the tissue and signalling context (Lamouille, Xu and Derynck, 2014). Induction of EMT decreases physical restrictions (cell-to-cell connections) allowing cells to acquire invasive and migratory properties (Acloque *et al.*, 2009). This generally requires the repression of epithelial markers, leading to alterations in cytoskeletal architecture, the loss of cell-cell adhesion, apical-basal polarity, and increase in appearance of mesenchymal properties (Gibbons *et al.*, 2009; Lamouille, Xu and Derynck, 2014). Several factors are involved with the activation of EMT including the mesenchymal markers Vimentin, N-Cadherin, Fibronectin, Slug and Snail. During EMT, a reciprocal loss of epithelial markers including, E-Cadherin, Syndecan-1 and occludins is observed (Thiery *et al.*, 2009; Lamouille, Xu and Derynck, 2014; Dongre and Weinberg, 2018). Mesenchymal-Epithelial Transition (MET) is the reverse process that allows motile cells to become more stationary and polarised, and is involved in stem cell differentiation and the establishment of circulating cancer cells in forming secondary tumour locations (Yao, Dai and Peng, 2011). The induction of EMT has been found to dissociate cancer cells and is believed to be a significant factor for dissemination of cancers cells (Thiery *et al.*, 2009), with MET driving re-epithelisation in secondary locations around the body (Nieto, 2013).

EMT is not a highly conserved process, as not all EMT-associated markers are observed to be affected in individual EMT events (Huang *et al.*, 2013) and can depend on the context and tissue type (Tan *et al.*, 2014). Most studies have only determined the extent to which EMT is induced based on dramatic phenotypic changes (Nieto *et al.*, 2016), however changes to the definition of EMT have included the possibly of changes to epithelial and mesenchymal states without observable cellular changes to classic EMT phenotypes (Huang *et al.*, 2013; Yu *et al.*, 2013; Nieto *et al.*, 2016). A multitude of different EMT factors exist including a network of TFs, epigenetic regulators and non-coding RNA that can contribute to unique or partial EMT events (Li and Kang, 2016),.

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 1-3 Phenotypic and Molecular changes in EMT

Major EMT TFs including ZEB, Twist, SnaiI and/or Slug (Snai2), control the activation of mesenchymal genes while repressing epithelial genes. This drives the conversion of stationary epithelial cells into more mesenchymal/motile cells that lose tight junctions, adherens, and desmosomes leading to an increase in invasive properties. The process of MET describes the reverse process whereby cells re-acquire an epithelial state(Image: (Dongre and Weinberg, 2018)).

Although the study of TFs have been the focus for EMT studies, EMT is also modulated by miRNAs, epigenetic modifications, alternative splicing, protein stability and localisation of proteins (Nieto *et al.*, 2016). The miRNAs *hsa-mir-200* and *hsa-mir-34* are classical miRNAs that control the expression of ZEB1 and Snail (SNAI1) TFs respectively by negative feedback loops (Gregory *et al.*, 2008; Brabletz and Brabletz, 2010; Siemens *et al.*, 2011). ZEB1 and SNAI1 function to control E-cadherins, occludins and claudins (Puisieux, Brabletz and Caramel, 2014). GRHL2 and OVOL1/2 proteins also from negative feedback loops with *ZEB1* (Hong *et al.*, 2015) with GRHL2 regulating the expression of *hsa-mir-200* upstream of *ZEB1* (Cieply *et al.*, 2012; Chung *et al.*, 2016). SNAI1 binds to promotor regions and can recruit HDAC1/2 (Peinado *et al.*, 2004), LSD1 (Lin *et al.*, 2010) and PRC2 (Herranz *et al.*, 2008) triggering transient and long-lasting modifications that can activate or repress transcription of target genes (Javaid *et al.*, 2013). The miRNA *hsa-mir-200* is also regulated by epigenetic modifications as histone demethylase KDM5B controls the expression of *hsa-mir-200* by the addition of histone H3 methylation of miRNA regulatory regions (Enkhbaatar *et al.*, 2013). At the protein level, protein modifications can also mediate EMT with SNAI1 phosphorylation leading to ubiquitin-mediated degradation by GSK3β (Zhou *et al.*, 2004).

1.6 Human Embryonic Stem Cell Differentiation and EMT

EMT was first described following observations by Greenburg and Hay (Greenburg, 1982) that epithelial cells and embryonic epithelia were able to lose polarity and invade the adjacent extracellular matrix. It was reported that these ‘transformations’ occur at specific times in particular epithelia. During embryonic development, specifically at gastrulation, cells undergo EMT and is essential for the transformation from single-layered blastula into the three germ layers (mesoderm, ectoderm and endoderm) (Nakaya and Sheng, 2008; Acloque *et al.*, 2009). Specific cells then undergo several rounds of EMT and MET, allowing the cells to migrate and subsequently generate cellular structures within the embryo, which facilitates the production of distinct cell types (Acloque *et al.*, 2009; Lamouille, Xu and Derynck, 2014; Li *et al.*, 2017).

Human embryonic stem cells (hESCs) are pluripotent and self-renewing cells derived from the inner cell mass of a blastocyte and have the capacity to differentiate into cells from all three germ lineages (Hoffman and Carpenter, 2005), giving rise to germ layer-specific differentiated cells and tissues. Stem cell differentiation is controlled and influenced by external factors like cell-cell contacts and secretions of chemicals or growth factors, activating or inhibiting internal signalling pathways, essential for cell specification during differentiation (Zakrzewski *et al.*, 2019). Understanding the stem cell environment and external molecular factors, stem cells can be cultured *in vitro* indefinitely and maintained in an undifferentiated state (Vazin and Freed, 2010). HESCs were initially cultured on Mouse Embryonic Fibroblasts (MEF) feeder layers. Studies of excreted factors from MEFs, and supplementation of these in culture media, including FGF, TGFβ and WNTs, eliminated the need for MEF feeder layers, while still maintaining hESC pluripotency during routine culturing of these cells (James *et al.*, 2005; Lu *et al.*, 2006; Brabletz and Brabletz, 2010). Targeted manipulation of these conditions can enrich for specific precursors and terminal cell types during controlled differentiation protocols.

Human and mouse embryonic stem cells, have been used extensively to study early development and provide pathways for possible therapies. Notably, somatic cells can be induced into stem cells by the introduction and expression of key TFs OCT3/4, SOX2, c-MYC and KLF4 (Okita, Ichisaka and Yamanaka, 2007). These factors as well as others including master regulator of pluripotency NANOG, control the pluripotent state of stem cells and reduction of any of these factors causes the loss of pluripotency and induces stem cell differentiation (Zafarana *et al.*, 2009; Fischer *et al.*, 2010; De *et al.*, 2014; Liao *et al.*, 2018, p. 4). One of the earliest notable cellular changes that occur during reprograming of mesenchymal cells into induced pluripotent stem cells is the acquisition of epithelial characteristics after addition of reprograming factors (Li *et al.*, 2010, 2017; Samavarchi-Tehrani *et al.*, 2010). A link between stemness and EMT is further supported by the detection of a rise in SOX2, NANOG, OCT3/4, LIN28B and NOTCH1 during platelet derived growth factor induced EMT in prostate cells (Kong *et al.*, 2010).

The differentiation of induced, mouse or human ESCs are useful *in vitro* model systems to study EMT and may be useful to model shared molecular and biochemical events that occur during differentiation and cancer metastasis (Eastham *et al.*, 2007).

1.7 EMT in Cancer

Cancer cells, like stem cells, also utilise EMT. By adopting a mesenchymal phenotype they become motile and can disseminate from the primary tumour location and spread to adjacent tissues or around the body (Thiery, 2002; Nieto *et al.*, 2016). Some circulating cells are able to reverse this process and undergo MET, as seen in stem cells during differentiation, and colonise in secondary locations exhibiting epithelial phenotypes (Zeisberg, Shah and Kalluri, 2005; Brabletz, 2012; Ocaña *et al.*, 2012; Tsai *et al.*, 2012; Davis *et al.*, 2014). EMT also grants cancer cells enhanced survivability including the repression of apoptosis (Mani *et al.*, 2008), as well as migratory and invasive properties, which leads to an increased resistance to a variety of cancer therapies (Thomson *et al.*, 2008; Shintani *et al.*, 2011; Jiang *et al.*, 2012; Fischer *et al.*, 2015; Zheng *et al.*, 2015).

Tumours are a heterogenous mass of cells, as cells at the tumour front can present with more mesenchymal features and the interior component show more epithelial characteristics (Brabletz *et al.*, 2001; Nieto *et al.*, 2016). This feature makes the study of EMT difficult in cell culture as the tumour microenvironment and heterogeneity are not replicated in cell culture models (Nieto *et al.*, 2016). The presence of circulating tumour cells (CTC) with mesenchymal markers are associated with disease progression and poorer prognosis (Cohen *et al.*, 2009; Krebs *et al.*, 2011; Yu *et al.*, 2013; Caixeiro *et al.*, 2014), however CTCs can present with both epithelial and mesenchymal markers (Yu *et al.*, 2013; Khoo *et al.*, 2015). Evidence suggests that CTCs express extracellular matrix (ECM) components which could be useful for the reattachment of these cells to organs of the body (Ting *et al.*, 2014). Interestingly, clusters of CTCs that present with mesenchymal and epithelial markers are more effective at establishing in secondary locations in the body than single mesenchymal CTC (Aceto *et al.*, 2014). It is proposed that the mesenchymal cells allow dissociation of a cluster of cells, while the epithelial cells within the cluster, have an increased propensity to colonise secondary sites.

Therefore, understanding and targeting EMT has become more challenging due to heterogeneity of cancer cells. However, EMT still remains of great interest in the treatment of cancer or as an adjuvant treatment alongside conventional therapies (Davis *et al.*, 2014).

1.8 Partial/Hybrid EMT

The role of EMT in promoting cancer metastasis has come under debate in recent years as two separate studies in 2015 found that inhibition of EMT did not lead to decreased levels of cancer metastasis but did contribute to conferring chemoresistance (Fischer *et al.*, 2015; Zheng *et al.*, 2015). These two studies were quickly rebutted suggesting that the genetic manipulations by these groups failed to completely suppress versions of the EMT program (Aiello *et al.*, 2017; Ye *et al.*, 2017). Supporting these rebuttals, several studies have since observed multiple EMT states or partial/hybrid/intermediate EMT states whereby cells can display both epithelial and mesenchymal phenotypes, giving an explanation for the incomplete inhibition of EMT in the previous studies (Blanco *et al.*, 2007; Leroy and Mostov, 2007; Arnoux *et al.*, 2008; Strauss *et al.*, 2011; Huang *et al.*, 2013; Grande *et al.*, 2015; Lovisa *et al.*, 2015; Schliekelman *et al.*, 2015; Grigore *et al.*, 2016; Aiello *et al.*, 2018; Jolly, Mani and Levine, 2018; Pastushenko *et al.*, 2018). In fact, “metastable” cells have been identified that are able to induce and also revert the EMT process (Lee *et al.*, 2006; Tam and Weinberg, 2013). Although loss of some EMT factors show little or no changes to cancer metastasis (Fischer *et al.*, 2015; Zheng *et al.*, 2015), EMT cannot be ruled out as a major contributor to cancer progression (Maheswaran and Haber, 2015), as cooperation of multiple EMT TFs suggest that loss of any individual factor may be insufficient (Nieto *et al.*, 2016) as compensation or redundancy may occur. Of note, although EMT can be viewed as deleterious for cancer metastasis (Fischer *et al.*, 2015; Zheng *et al.*, 2015), EMT induced during wound healing is beneficial, and therefore investigating EMT has the capacity to offer, not only cancer treatment or diagnosis improvements, but the possibility for improved therapies for healing and wound repairs (Leopold, Vincent and Wang, 2012).

These observations are in concordance with earlier studies that observed different EMT types and classified them into three subtypes (Figure 1-4) (Kalluri and Weinberg, 2009; Zeisberg and Neilson, 2009), illustrating the dynamism of EMT rather than a canonical, rigid process, however this is most likely now viewed as a simplified model as EMT can be viewed as a spectrum (Huang *et al.*, 2013). The existence of cancer cells that are able to migrate as a collective, displaying both mesenchymal and epithelial markers further supports the notion of a dynamic EMT whereby cells retain context-specific cell-to-cell contact (Bronsert *et al.*, 2014).

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 1-4 Different subtypes of EMT

Type 1 EMT is associated with developmental processing such as organ development and embryo formation, giving rise to multiple cell types. Type 2 is the acquisition of EMT properties during repair and restructure due to healing, tissue regeneration and fibrosis. Type 3 EMT is associated with cancer metastasis with cancer cells increasing their invasive capabilities (Image: (Kalluri and Weinberg, 2009)**).**

In cancer, a partial EMT is often observed, giving rise to the high plasticity of cancer cells (Figure 1-5) (Brabletz, 2012; Nieto, 2013; Pastushenko *et al.*, 2018; Stemmler *et al.*, 2019). EMT plasticity complicates the analysis of individual EMT models for cancer metastasis, as no single *in vitro* EMT model represents the entirety of all EMT/metastatic events. Although cells can exist in-between the epithelial and mesenchymal states, some molecular and structural changes are commonly observed amongst the different subtypes (Kalluri and Weinberg, 2009; Nieto *et al.*, 2016). Furthermore, complete epithelial or mesenchymal states are not always the terminal result as cells can remain with a partial EMT phenotype (Grande *et al.*, 2015).

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 1-5 Partial EMT in cancer

Activation of the EMT program induces the mesenchymal state, however partial activation of EMT is observed in cancer cells, leading to an increased motility and resistance to current cancer therapies (Image: (Stemmler *et al.*, 2019)**).**

1.9 Alternative splicing

Alternative splicing (AS) of RNA has been observed in over 90% of human transcripts and is one of the main sources of expansion of the functional proteome in plants and animals (Wang *et al.*, 2008; Nilsen and Graveley, 2010; Syed *et al.*, 2012; Reddy *et al.*, 2013). AS is responsible for the formation of circRNAs, but also generates other RNA transcripts that either includes or excludes exons and also sequences adjacent to exon boundaries (Figure 1-6). Large changes in AS is observed during EMT with mesenchymal cancer cells showing distinct splicing patterns (Li *et al.*, 2018). Epithelial Splicing Regulatory Protein 1 and 2 (ESRP1 and ESRP2) are well studied RBPs shown to control the splicing patterns of cells within the epithelial state (Warzecha *et al.*, 2009, 2010). These splicing factors regulate cell motility, cell-cell adhesion and cell-matrix adhesion by altering splicing of p120-catenin, FGFR2, MENA and CD44. Other splicing factors including Quaking (QKI) and RNA Binding Fox-1 Homolog 2 (RBFOX2) and

SRSF2 , have been shown to promote a mesenchymal state driving alternative splicing of RNA during EMT (Bonomi *et al.*, 2013; Braeutigam *et al.*, 2014; Conn *et al.*, 2015; Yang *et al.*, 2016; Li *et al.*, 2018; Pillman *et al.*, 2018). These splicing changes can lead to altered functional proteins and increase the variety of RNA transcripts from a single gene which can function independently (i.e mRNAs, circRNAs, noncoding RNAs) .

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 1-6 Alternative splicing drives RNA diversity

Illustration of different outcomes of alternative splicing of pre-mRNA. Dashed lines indicate the alternative splicing patterns that can occur between include sequences (blue bars) and alternatively splices sequences (red and green bars). The black solid lines represent introns (Image: (Warzecha and Carstens, 2012)).

1.10 Profiling and function of circRNAs across EMT

CircRNAs are highly regulated during EMT (Conn *et al.*, 2015), with the biosynthesis of approximately one-third of circRNAs found to be regulated during this process by QKI (Conn *et al.*, 2015) in immortalized human mammary epithelial (HMLE) cells. Interestingly, over 50% of circRNAs (2537 of 4667 circRNAs) detected exclusively in mesenchymal HMLE cells were regulated independently from their parental mRNA transcript after TGFβ-induced EMT, highlighting that increased backsplicing is responsible for the majority of circRNA changes during EMT. QKI, acting as a dimer (Teplova *et al.*, 2013), binds to motifs in the introns flanking circRNA exons, within a pre-mRNA, and bring the boundaries of the circRNA into close proximity, thereby encouraging the circularisation of RNA by the spliceosome (Figure 1-7). To test this hypothesis, Conn *et al.* were able to generate *de novo* circRNAs by inserting QKI binding sites to DNA that had not been found to generate circRNA (Conn *et al.*, 2015). Furthermore, QKI is sufficient to influence cell plasticity, invasion and tumour growth by directing AS changes leading to an increase in circRNA number and abundance, without substantially affecting the cognate mRNA levels (Figure 1-7) (Pillman *et al.*, 2018). Since QKI is able to drive AS changes, and is able to regulate circRNA expression during EMT, it is possible that circRNAs are effectors of EMT. In fact, this has been demonstrated for a cohort of circRNAs which have been shown to either promote or inhibit EMT postulated to be occur from the sponging of miRNAs (L. Wang *et al.*, 2018; Meng *et al.*, 2018; X. Chen *et al.*, 2018; Zhou *et al.*, 2018; Li *et al.*, 2019; X. Zhang *et al.*, 2019).

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 1-7 Quaking drives circRNA biogenesis during EMT

QKI, acts as a dimer to bring exon boundaries of per-mRNA into close proximity and generates circularisation of RNA. An increase to the number and abundance of circRNAs during EMT is observed and corelates with an increase of QKI during EMT (Image: (Conn *et al.*, 2015)).

As most circRNAs arise from within genes, either exonic sequences, intronic sequences or a combination of both, circRNAs are regulated by similar factors as their linear counterparts. This suggests that some circRNAs and linear RNAs from the same gene may have synonymous functions, either functioning in the same pathway, or even acting as positive or negative regulators. In fact, this has been observed as a circRNA from the *PPARα* tumour suppressor gene was also shown to inhibit tumorigenesis and metastasis (N. Zhang *et al.*, 2018). However, this may not be the case for all circRNAs as changes in either circular or cognate linear forms do not always correlate. This is also further exemplified by vast changes to circRNA expression, with little change to linear RNA levels, by the KD of the RNA binding protein QKI (Conn *et al.,* 2015).

A review by Shang *et al.* (2019) gives many examples of circRNAs found to be involved in EMT by modulating EMT TFs, signalling (TGFβ, WNT), and adhesion and cytoskeletal molecules. These circRNAs are commonly found to effect EMT by sponging miRNAs that regulate these EMT related genes, and the relevance of these interactions at physiological levels is contentious as many of these studies overexpress circRNAs far above cellular levels. As explained earlier, circRNAs are generally found to be lowly expressed within cells and therefore a sponging mechanism is not stoichiometrically favourable.

In regards to hESCs, two circRNAs *circBIRC6 and circCORO1C* were discovered to be functionally related to the maintenance of pluripotency as KD by small hairpin RNA (shRNA) resulted in differentiation and loss of pluripotency in hESCs (Yu *et al.*, 2017). Conversely, overexpression of these circRNAs upregulated expression of pluripotency-associated TFs *POU5F1*, *NANOG* and SOX2, while decreasing lineage-related TFs. One study employed the use of high-throughput sequencing from Breast Cancer Stem Cells and found 27 differentially regulated circRNAs. Gene Ontology (GO) Analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis indicated processes of stem cell proliferation and differentiation as well enrichment for stem cell related pathways (Hippo, Notch and Wnt). KD of one circRNA, *circVRK1*, was able to enhance breast cancer stem cell characteristics (Yan *et al.*, 2017). However, whether these circRNAs are involved within the context of EMT in these cells was not investigated. Additionally, *circZKSCAN1* has recently been demonstrated to supress hepatocellular carcinoma cell (HCC) stemness by physically binding to the RBP fragile X mental retardation protein (FMRP) thereby decreasing the activity of WNT signalling. Interestingly, a reduced expression of the EMT-regulated RBP QKI was found to be responsible for the decreased *circZKSCAN1* levels (Zhu *et al.*, 2019). These studies demonstrate that altered levels of circRNAs and RBPs can affect stem cell characteristics, prompting the need for further investigation into the effects of these molecules.

1.11 Cell Cycle

The mitotic cell cycle is a highly regulated process that involves the replication of DNA and splitting of one cell into two. This is broken down into 4 phases, Gap 1 (G1), DNA Synthesis (S), Gap 2 (G2) and Mitosis (M) (L. Liu *et al.*, 2019). Each phase includes checkpoints, tightly regulated processes that search for damaging faults and errors obtained during cell growth and DNA synthesis, allowing for correction before division in mitosis. If these errors cannot be fixed, the cell will enter programmed cell death via apoptosis to ensure that mutated or damaged cells do not continue replicating (Malumbres and Barbacid, 2009). These checkpoints are heavily deregulated in cancer cells, imparting increased proliferative abilities and an increase to the number of mutation events in checkpoint-regulating proteins (Molinari, 2000; Leal-Esteban and Fajas, 2020). G1 is the phase of a newly divided cell that is sensitive to growth signals, allowing cells to increase in size before the commitment to another round of mitosis. Cells in S replicate their DNA to achieve a tetraploid state and G2 phase allows for more growth and preparation for mitosis. M phase encompasses multiple individual steps, that after completion, give rise to two independent daughter cells.

Cell cycle regulation is controlled by the sequential expression and destruction of cyclin proteins which bind to their phase-depended CDK partners to activate and repress cell cycle regulators and allow for successful progression through each cell cycle checkpoint (Malumbres and Barbacid, 2009). Cyclin D, bound with CDK4 and CDK6 acting as kinases, are most active during G1 and increase in response to mitogenic stimuli (Mens and Ghanbari, 2018). Cyclin D/CDK4 and CDK6 begin to partially phosphorylate proteins that control cell cycle progression including Retinoblastoma (Rb) and other Rb family members including p107 and p130 (Johnson and Walker, 1999; Conklin, Baker and Sage, 2012; Mens and Ghanbari, 2018; Leal-Esteban and Fajas, 2020). Phosphorylated Rb inhibits the binding of histone deacetylase to E2F TFs which starts to transcribe downstream E2F target genes including Cyclin E (Lim and Kaldis, 2013; L. Liu *et al.*, 2019). During G1/S transition, Cyclin E with its CDK partner CDK2, continue to hyper phosphorylate Rb, leading to higher expression of Cyclin E and A to support the progression into S phase by their involvement in nucleotide synthesis and DNA replication (Lim and Kaldis, 2013). Progression to S phase requires the activation of CDK2/cyclin E, by CDK-activating kinase (CAK), which goes on to phosphorylate many proteins including Rb (Lim and Kaldis, 2013) and p53 (KO *et al.*, 1998, p. 7; Schneider, Montenarh and Wagner, 1998). CAK is a three-subunit complex comprising of Cyclin H, CDK7 and ring finger protein MAT1 and is detected in the nucleus and cytoplasm in mouse ESCs (Patel and Simon, 2010). This complex also activates a range of different cyclin-associated kinases (CDK1, CDK2, CDK4 and CDK6) complexes by threonine phosphorylation (Figure 1-8) throughout the cell cycle.

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 1-8 CAK complex in the cell cycle

Role of CAK complex (CDK7, Mat1, Cyclin H) in progression of cell cycle by activation of multiple CDK/cyclin complexes (Wang *et al.*, 2020).

Cyclin A/CDK2 are active in S phase leading to the phosphorylation and activation of important mitotic genes required in G2 including A and B cyclins and CDK1 (Leal-Esteban and Fajas, 2020). During G2, critical genes, including FOXM1, become active by cyclin A/B/CDK1 phosphorylation, and target genes for activation involved with mitosis initiation and chromosomal segregation regulators (Major, Lepe and Costa, 2004). Cyclin-dependent kinase inhibitors function to slow progression of the cell cycle by binding to and inhibiting cyclin-CDK complexes (Molinari, 2000; Vermeulen, Van Bockstaele and Berneman, 2003).

CDKs abundance during the cell cycle remains constant however their activity is regulated by their phosphorylation status and whether they are bound to their cyclin counterpart, which can oscillate in abundance throughout the cell cycle (Arellano and Moreno, 1997). CDKs are activated (phosphorylated) by the CAK complex. This complex also has little variability during the cell cycle and its ability to activate different CDKs throughout the cell cycle in a sequential manner is highly regulated (Larochelle, 2001). Interestingly, CCNH/CDK7 is sufficient to phosphorylate CDK2, however MAT1 is required for efficient phosphorylation of p53, with MAT1 acting as a CAK substrate specificity determining factor (KO *et al.*, 1998). Therefore, the addition of proteins to the CAK complex can change the substrate specificity. Of note, the CAK complex is also a component of the general transcription factor TFIIH, which associates with RNA pol II (Peissert *et al.*, 2020). CDK7 of the CAK complex, phosphorylates the C-terminal domain of the largest pol II subunit, which contributes to promotor escape (Compe *et al.*, 2019). The switch from cell cycle activation to RNA pol II activation is also regulated by MAT1 addition to the CAK complex (Larochelle, 2001).

1.12 Pathways for pluripotency and differentiation

The expression of the master stem cell transcription factors (OCT3/4, SOX2 and NANOG) are controlled by several signalling pathways such as WNT, FGF, TGFβ and BMP signalling. Activation or inhibition of these pathways can maintain pluripotency and/or induce differentiation into cells from each germ layer depending on the concentration, duration, or whether multiple pathways are manipulated (Figure 1-9). Several studies indicate that WNT signalling activation leads to loss of pluripotency and differentiation toward mesodermal and endodermal lineages (Bone *et al.*, 2011; Davidson *et al.*, 2012; Martin and Kimelman, 2012), however conflicting reports indicate the importance for WNT signalling for maintenance of self-renewal of ESCs (Sato *et al.*, 2004; Ogawa *et al.*, 2006; Anton, Kestler and Kühl, 2007; Sokol, 2011). Additionally, FGF2 has been routinely been added into hESC culture media and are thought to be required for the continuous undifferentiated culture of stem cells (Eiselleova *et al.*, 2009; Ding *et al.*, 2010; Amita *et al.*, 2013; Mossahebi-Mohammadi *et al.*, 2020). A recent study contradicted this by showing that hESCs were unaffected by inhibition of FGF signalling, however IGF signalling was required for hESC maintenance (Wamaitha *et al.*, 2020). This highlights the complexity of the signalling pathways to maintain pluripotency or induce differentiation into cells of the three germ layers.

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 1-9 Differentiation pathways for pluripotent stem cells

Major signalling molecules that contribute to the directed differentiation of stem cells into the three germ layers of ectoderm, mesoderm and endoderm. These include molecules and inhibitors primarily from the TGFβ, WNT, FGF signalling pathways (*Embryonic and Induced Pluripotent Stem Cell Differentiation Pathways & Lineage-specific Markers*, no date).

1.12.1 TGFβ in hESC differentiation

TGFβ signalling plays roles in cell differentiation and development, inflammation, and EMT, while also being a signal for cell growth and cell death, depending on the cell type and context (Hao, Baker and ten Dijke, 2019). TGFβ signalling is crucial to maintain the pluripotent status of hESCs (Vallier, Alexander and Pedersen, 2005). Perturbation of TGFβ signalling via the interaction of multiple ligands (Activin, Nodal, BMPs, FGF, GDPs) or inhibitors of TGFβ cell-surface receptors are exploited to differentiate hESCs and induced pluripotent stem cells (iPSC) into a multitude of cell types. TGFβ type I and type II receptors function as dimers, with ligands binding to type II receptors on the cell surface activating type I receptors. Active type I receptors phosphorylate the downstream effectors of TGFβ signalling, the R-SMAD complexes (SMAD2/3 complex or SMAD1/5/8), which then associate with Smad4 and translocates into the nucleus (Figure 1-11). Translocated R-Smads act as TFs and can induce differentiation and lineage specification.

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 1-10 TGFβ signalling pathway

Receptor activated Smads (R-Smads) are the downstream effector molecules for TGFβ signalling which bind to Smad4 and translocate to the nucleus to act as TFs. (Image: (Pauklin and Vallier, 2015)).

The Activin/Nodal/Smad2/3 signalling branch of TGFβ signalling is necessary to activate Nanog expression, and thus, for maintaining self-renewal properties and supporting both naïve and primed pluripotency. Inhibition of this signalling pathway, by overexpression of Nodal antagonists Lefty or Cerberus, or by inhibitors, SB431542 or follistatin, leads to neuroectoderm differentiation of hESCs; whereas treatment with ligands of this pathway, Activin and Nodal, when FGF is depleted, results in mesendoderm (precursor of endoderm and mesoderm) differentiation. This is in part controlled by a downstream effector of SMAD2/3 signalling, a key pluripotency factor, Nanog, which blocks differentiation into neuroectoderm (Figure 1-12). Activin/Nodal signalling also maintains positive H3K4me3 histone markers on “bivalent genes” known to be master regulators of mesendoderm specification. Conversely, inhibition of Activin/Nodal and Nanog KD are both associated with a decrease of H3K4me3 at these loci, impairing the capacity of these cells to differentiate toward mesendoderm (Bertero *et al.*, 2015). Activin/Nodal signalling also regulates the deposition of m6A modification to a number of RNA transcripts including *Nanog* mRNA. A decrease of *Nodal* mRNA stability by m6A deposition during inhibition of activin/nodal signalling facilitates exit from pluripotency as well as specification into the neuroectoderm lineage (Gordeeva, 2019).

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 1-11 Mitogenic signalling for stem cell differentiation

Model illustrating that Nanog prevents differentiation into neuroectoderm differentiation, induced by FGF signalling, and limits endoderm differentiation by limiting the Smad2/3 signalling cascade (Image: (Vallier *et al.*, 2009)).

Sustained activation of Activin signalling stimulates the differentiation of the primitive streak population, with high and low levels inducing definitive endoderm and mesoderm, respectively (Watabe and Miyazono, 2009).

BMP/Smad1/5/8 signalling decreases the expression of pluripotency factors, NANOG and SOX2, and, in cooperation with Activin/Nodal signalling, leads to expansion of the endodermal lineage (Teo *et al.*, 2012). However, in the absence of Activin/Nodal and FGF signalling activity, BMP signalling promotes the differentiation of trophoblast and extraembryonic endoderm lineages (D'Amour *et al.*, 2005; Amita *et al.*, 2013).

Finally, the synergistic action of inhibitors for both branches of the TGFβ family signalling pathways, such as Noggin (BMP inhibitor) and SB431542 (TGFβ/Activin/Nodal inhibitor), facilitate the induction of neuroectodermal lineages in hESCs and iPSC in both adherent and EB-mediated differentiation protocols (Chambers *et al.*, 2009; Mak *et al.*, 2012). Use of only one of the above inhibitors does not induce neural differentiation from adherent hESCs or iPSCs, but can differentiate cells towards neural lineages during more extensive differentiation protocols (Gordeeva, 2019), or during high bFGF conditions (Francis and Wei, 2010).

Differentiation of ESCs or iPSCs into Embryoid Bodies (EBs) is a common and widely used strategy to direct further differentiation into specific cell types. EBs are three-dimensional aggregates of ESCs or iPSCs that are formed by suspension cells in non-adherent conditions, and can differentiate into cells that comprise of the three germ cell lineages (Itskovitz-Eldor *et al.*, 2000) (Figure 1-13). As EBs differentiate they activate well-conserved genes that control early events during gastrulation and germ layer formation (Rust, Sadasivam and Dunn, 2006).

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 1-12 Differentiation of EBs produces cells from all three lineages

Embryoid bodies generated from ESCs and IPSCs can be differentiate into cells of mesoderm, endoderm and ectoderm (Image: (*Differentiation of Stem Cells | LSR | Bio-Rad*, no date))

1.12.2 TGFB and the cell cycle

G1 phase is generally the longest phase of the cell cycle and can be further split into early G1 (post-mitotic) and late G1 (pre-synthesis) with hESC cells sensitive to mitogenic signals during early G1. HESCs in early G1 can only initiate differentiation into endoderm and mesoderm, while cells in late G1 are limited to neuroectoderm lineages (Pauklin and Vallier, 2013), and little differentiation potential within S and G2/M. Not only does early and late G1 affect differentiation but also cell migration as early G1 phase cells are more sensitive to migratory signals. Largely, migration is reduced in G2/M phases of the cell cycle (Bonneton, Sibarita and Thiery, 1999).

During early G1, signalling pathways such as the TGFβ signalling pathway is active as downstream effectors SMAD2/3 are able to enter the nucleus and act as TFs (Pauklin and Vallier, 2013). Differentiation cues during early G1 (low Cyclin D/CDK4/6 and active SMAD2/3) promote endoderm and mesoderm lineages. As G1 progresses into late G1, Cyclin D/CDK4/6 inactivate TGFβ signalling by phosphorylation of SMAD2/3 linker regions as well as progressing cell cycle regulation by inactivating Rb via phosphorylation. In late G1, and in the absence of TGFβ signalling, the cells are limited to the “neuroectoderm default pathway” (Vallier, Reynolds and Pedersen, 2004; Pauklin and Vallier, 2013) (Figure 1-14). Interestingly, the total length of G1, modulated by the WNT pathway, can also change the differentiation capacity of stem cells with a longer G1 phase in single hESC cells increasing the bias towards neuro-ectoderm differentiation (Jang *et al.*, 2019). Furthermore, a constitutive increase to cyclin D1, D2 and D3 can increase neuronal differentiation independently of SMAD2/3 CDK4/6 cross-talk (Pauklin and Vallier, 2013, 2015). Additionally, perturbation of the cell cycle by stalling cells in G1 by a CDK2 inhibitor is enough to prolong phospho-Rb activation promoting the loss of master stem cell regulator POU5F1 (Filipczyk *et al.*, 2007). Therefore, any changes to the cell cycle, G1 length or a cells early/late G1 status can influence cell differentiation, proliferation and migratory abilities. Of note, inhibition of the TGFβ pathway by chemical inhibitors is a common method for the differentiation of stem cells into neural lineages (Kelava and Lancaster, 2016).

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 1-13 Propensity for differentiation of hESCs in early and late G1

During Early G1 Smad2/3 is unphosphorylated and can translocate to the nucleus and drive transcription on endoderm and mesodermal genes. Smad2/3 become phosphorylated in late G1 and G1/S and can no longer translocate leaving the neuroectoderm as the only pathway available for differentiation (Image: (Pauklin and Vallier, 2013)).

1.12.3 Cell cycle and circRNAs

Manipulation of multiple circRNAs has shown to affect the cell cycle (Y. Yang *et al.*, 2018; Karedath *et al.*, 2019; P. Chen *et al.*, 2020, p. 1; Li, Li and Yu, 2020, p. 609; Yu *et al.*, 2020, p. 1; Chaudhary *et al.*, no date). Interestingly, and unlike the majority of groups that report miRNA sponge mechanisms, *circFOXO3* was found to form a ternary complex with CDK2 and p21 that inhibits their interactions, and therefore, halts cell cycle progression (Du *et al.*, 2016). Furthermore, *circFBXW7* contains an internal ribosome entry site and produces a protein that inhibits proliferation an cell cycle acceleration by antagonising the stabilisation of c-Myc (Y. Yang *et al.*, 2018).

1.13 Methods for the reduction of RNA transcript abundances

Multiple methods exist for reducing RNA transcripts, including linear RNAs and circRNAs, with the choice of method determined by cost, method of action/delivery, cell type or cellular location of the target RNA transcript (McIntyre and Fanning, 2006). These methods include short hairpin RNAs (shRNAs), small interfering RNAs (siRNAs), antisense oligonucleotides (ASOs), and more recently, CRISPR-Cas13.

1.13.1 Small hairpin RNAs

RNA interference (RNAi) can be achieved by small hairpin RNAs (shRNAs) and are the most cost-effective strategy for reducing a panel of RNA targets. These can be utilised to transiently reduce target RNA levels or can be used to generate stable shRNA-expressing cell lines, if the plasmid encoding the shRNA cassette is first used to generate lentivirus that is capable of infecting and integrating into a cell’s genome (Brummelkamp, Bernards and Agami, 2002; Moore *et al.*, 2010). The shRNA cassette encoded on the plasmid is expressed by the cell to produce an shRNA hairpin molecule which is processed via endogenous Drosha to form the pre-shRNA in the nucleus containing a loop, sense and antisense strand. The pre-shRNA is exported into the cytoplasm by endogenous Exportin 5 and further processed by endogenous Dicer, removing the hairpin, and loaded into the RNA-induced silencing complex (RISC) (Sheng, Flood and Xie, 2020). This complex can be loaded with either the antisense or potentially the sense strand, providing specificity for RISC-mediated degradation of RNA transcripts (perfect base pairing) or blockage of translation (imperfect base pairing) (Mockenhaupt *et al.*, 2015) (Figure 1-15).

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 1-14 Current technologies for targeted cleavage of RNAs

Current technologies for reduction of RNA by cleavage and degradation include (A) shRNA and siRNA loaded into RISC, (B) ASO mediated RNase H cleavage and (C) targeting and cleavage of RNA by CRISPR CAS variants (Image adapted from: (Liu and Lim, 2018)).

1.13.2 Small interfering RNAs

Similar to shRNAs, small interfering RNAs (siRNAs) are delivered into cells, loaded into a RISC complex in the cytoplasm, and subsequently degrade RNA transcripts. Although shRNAs offer greater flexibility, as it offers stable or transient expression, the use of the sense and antisense strand for RISC specificity can be a limiting factor when the shRNA target location is a limiting factor for shRNA design. SiRNAs offer strand-specific KD and can be titrated to reduce RNA targets to varying degrees potentially decreasing unwanted off-target effects (Sharma and Rao, 2009).

1.13.3 Antisense Oligonucleotides

Antisense Oligonucleotides (ASOs) are single stranded DNA molecules that also target specific RNA molecules based on Watson-Crick base pairing, but does not use RISC-mediated cleavage to degrade RNA levels. ASOs bind to their target and recruit RNase H that cleaves the double stranded RNA-DNA hybrid thereby decreasing RNA levels (Crooke, 2017). ASOs and siRNAs degrade RNA transcripts via different mechanisms, however they are similar in that they do not require the processing machinery that shRNAs require to make the mature shRNA molecule. Similarly to siRNA transfection, this mechanism is transient only and can occur in the cytoplasm and nucleus (Crooke, 2017), while shRNA and siRNA-mediated RNA degradation occurs mainly in the cytoplasm, however some literature details that siRNA machinery can migrate to the nucleus and target RNA (Haussecker and Proudfoot, 2005; Langlois *et al.*, 2005) as well as affecting DNA methylation (Morris *et al.*, 2004).

1.13.4 Conventional RNA interference comparison

All the above techniques have been heavily reviewed in regards to their benefits and drawbacks. ShRNAs are the most cost effective and provide the ability of researchers to stably target many RNA transcripts to probe for molecular or cellular consequences due to a reduction to RNA targets. Drawbacks include the possible production of both a sense and antisense RNA that can be loaded into a RISC complex which can target different unique RNA (off-target effects) (Wei *et* *al.*, 2009; Mockenhaupt *et al.*, 2015). Another is that siRNAs and shRNAs can have identical nucleotide sequences, however differences in expression profiles can be observed, highlighting that differential processing occurs for siRNAs and shRNAs than can confound analysis (Rao *et al.*, 2009). Furthermore, multiple shRNA hairpins of different sequences or sizes may be transcribed as RNA pol III does not always start transcription at the intended transcription start site, however it can be improved with introducing a purine at the +1 position for U6 promotors (Z. Gao *et al.*, 2017), however, when designing shRNAs across backsplice junctions of circRNAs, this design criteria may not be met. siRNAs or ASOs also have the potential to bind off-target RNA transcripts however production of alternate RNA-silencing molecules is not a factor. Use of these RISC or RNase H-mediated RNA degradation pathways requires the use of multiple targeting molecules to increase confidence of resulting phenotypes being a result of the reduction of the target gene and not off-target RNAs. A large advantage with using siRNAs or ASOs over shRNAs is that reduction of the target RNA can occur more efficiently when transfecting a small oligonucleotide instead of transfecting a shRNA-expression plasmid. Generation of stably producing shRNA cell lines can also take multiple days and requires selection of cells, either by flow cytometry or antibiotic selection. During this time, cellular effects like cellular differentiation, can prove disadvantageous in trying to probe for the initial cellular repercussion of reducing RNA KD.

1.13.5 CRISPR

Clustered regularly interspaced short palindromic repeats (CRISPR) is a novel and widely adopted technology used for DNA and gene editing, and recently for degrading RNA transcripts (Abudayyeh *et al.*, 2017). Discovered in prokaryotic organisms, CRISPR was initially identified as a family of DNA sequences residing inside cells that are left over from previous bacteriophage infections and used to detect and destroy foreign DNA in subsequent infections (Barrangou *et al.*, 2007). Investigation into the enzymes involved with the cleavage of DNA in other prokaryotes, revealed that Cascade proteins (Cas) utilise CRISPR RNA sequences to target viral DNA, imparting protection for the host bacteria (Brouns *et al.*, 2008). This system has been extensively investigated and has uncovered multiple Cas enzymes with different functional advantages and affinities. Cas13 was discovered to be an RNA-guided RNA cleaving enzyme and has been adapted to function in mammalian cells to achieve RNAi with the use of guide RNAs (gRNAs) that imparts specificity for Cas13 RNA degradation (Abudayyeh *et al.*, 2016, 2017; Xu *et al.*, 2020, p. 13; Zhang *et al.*, 2021). The gRNA provides similar base-pairing targeting system as other RNA targeting systems, however has been shown to have significantly lower off-target effects than shRNAs (Abudayyeh *et al.*, 2017; Xu *et al.*, 2020; Zhang *et al.*, 2021). All Cas13 subtypes (Cas13a-d) have shown RNase activity however can vary in different model systems (Xu *et al.*, 2021). Utilising CRISPR technology for RNA KD still requires multiple gRNAs for a confident phenotypic effect of targeting a specific RNA however off-target effects are able to be reduced. Recently announced, a new Cas7-11 CRISPR enzyme, derived from a fusion between a putative Cas11 domain and multiple Cas7 subunits, also has effective RNAi functionality and with potentially less off-target effects and cell toxicity than current Cas13 subtypes and shRNAs (Özcan *et al.*, 2021). Advancements in CRISPR-mediated RNAi are providing a new method for accomplishing RNA degradation with comparable KD efficiencies in many cells lines, and with far fewer off-target effects (Abudayyeh *et al.*, 2017; Özcan *et al.*, 2021).

1.14 Summary

The prevalence of circRNAs throughout the genomes of eukaryotes, prokaryotes and viruses, their ability to interact with other RNA species, DNA and proteins, encode for protein, as well as affect epigenetic modifications to DNA, demonstrate the broad regulatory roles for these diverse molecules. Recent studies of circRNAs during EMT have revealed extensive regulation by RBPs leading to an increase in number and abundance of circRNAs in mesenchymal cells.

Shared stemness properties of pluripotent stem cells and cancer stem cells, along with their EMT capabilities, allows the use of stem cell differentiation to mimic EMT observed during cancer metastasis as well as other shared programs. The plasticity of stem cells undergoing differentiation into different lineages, allows for a more diverse transcriptional landscape to be investigated, to discover critical EMT effector molecules that are shared between different EMT programs. Examining the roles of regulated circRNA across EMT and differentiation will give insight into the complex nature of EMT and provide further evidence of circRNA function, specifically in hESCs.

1.15 General hypothesis and aims

Stem cell differentiation is accompanied by waves of EMT, imparting structural plasticity, permitting cells to undergo the required changes to become intermediate and terminal cell types. CircRNAs are regulated during EMT and differentiation, with activation of EMT imbuing cancer cells with stem cell properties, possibly linking aspects of the two programs. Profiling of circRNAs across stem cell differentiation will be used to identify candidate circRNAs that may play a role in differentiation and/or EMT.

1.15.1 Aim 1

To transcriptionally profile circRNAs during stem cell differentiation, and to select circRNA candidates that show defined expression properties across differentiation.

1.15.2 Aim 2

To knockdown circRNA candidates and characterise the transcriptional and phenotypic changes that may indicate consequences to stem cell differentiation or changes in Epithelial or Mesenchymal characteristics that form a basis for further profiling.

1.15.3 Aim 3

To identify the mechanisms underlying circRNA function in stem cell differentiation.

1.16 Experiments/data in thesis completed/given by others

1.16.1 CircRNA-seq samples and next generation sequencing

The circRNA-seq of H9 hESCs and 5 differentiated cell lines was completed in 2018 before the commencement of my PhD. The differentiation of H9 hESCs was performed by Mr Michael Kulik, Mr Thomas Colunga and Professor Stephen Dalton from the University of Georgia, USA (Kriks *et al.*, 2011; Gasimli *et al.*, 2014; Colunga *et al.*, 2019). Cell pellets of these lines were kindly donated and RNA extraction was completed by Dr Vanessa Conn from my laboratory. Refer to chapter 2 for methods on the processing of RNA for enrichment for circRNAs, and for subsequent library preparation, circRNA sequencing and sequencing analysis. Bioinformatical analysis of circRNA-seq was completed by John Toubia (Centre for Cancer Biology, South Australia, Australia).

1.16.2 H9 Karyotype analysis

Karyotype analysis of H9 hESCs was generously performed by Kasey Peachey and Rhonda Hutchinson at SA Pathology at Women’s and Children’s Hospital, Australia.

1.16.3 qPCR analysis of neurons

Analysis by qRT-PCR of neural genes from embryoid body RNA samples was generously performed by PhD student Jenne Tran within Associate Professor Cedric Bardy’s Laboratory at SAHMRI, South Australia.

1.16.4 Immunofluorescence staining and imaging

Microtome sectioning of fixed embedded embryoid bodies, immunofluorescent staining of sections and fluorescence microscopy was completed by lab member and lab technician Kirsty Kirk.

1.16.5 Biotinylated Probe pulldown

Biotinylated probes for the RNA pulldown of *circACVR2A* were designed, and the pulldown was performed, by lab member Dr Marta Gabryelska.

1.16.6 CRISPR Cas13b

The Cas13b (WT) in pLX311 was generously donated by Associate Professor Joseph Rosenbluh from Monash University.

Chapter 2 Materials and Methods

2.1 Chemicals

All chemicals were obtained from Sigma-Aldrich (Sydney, Australia) unless otherwise specified. Water used for resuspension of chemicals was MilliQ grade, water for Injection.

2.2 RNA samples for hESC RNA sequencing

RNA samples for hESCs and all 5 lineages (Dopaminergic Neurons, Hepatocytes, Cardiomyocytes, Definitive Endoderm, Floorplate Precursors) were kindly provided by Professor Stephen Dalton.

2.3 Next Generation Sequencing

2.3.1 Circ-RNA seq library preparation and high-throughput sequencing and bioinformatics analysis

RNA was treated with DNase I (Roche) at 37°C for 15 mins, and purified with RNeasy RNA minelute cleanup kit (Qiagen, Venlo, Netherlands). The purified RNA (2.5ug) was digested with 5 Units of RNase R (2 U/µg RNA), then purified with RNeasy micro kit (Qiagen, Venlo, Netherlands). RNA was then depleted for rRNA using the NEBNext® rRNA Depletion Kit (Human/Mouse/Rat) (New England Biolabs, Ipswich, MA, USA) before libraries were constructed with the KAPA RNA HyperPrep Kit for illumina® (KAPA Biosystems). For qRT-PCR, RNA was reverse transcribed with QuantiTect® reverse transcription kit (Qiagen, Venlo, Netherlands).

RNase R-treated RNA-seq libraries from 3 biological replicates; of pluripotent stem cells (H9) and 5 differentiated states (cardiomyocytes, definitive endoderm, dopaminergic neurons, floor plate precursors and hepatocytes), were multiplexed and sequenced on the Illumina NextSeq 500 platform using the stranded, single end protocol with a read length of 150. Raw reads were adapter trimmed and filtered for short sequences using cutadapt v1.8.1(Martin, 2011) setting minimum-length option to 18, error-rate 0.2, quality cut-off 28 and overlap 5. The resulting FASTQ files averaging 35 million reads per sample were analysed and quality checked using the FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Reads were mapped against the human reference genome (hg19) using the STAR spliced alignment algorithm (Dobin *et al.*, 2013) (version 2.5.3a with default parameters and --chimSegmentMin 20, --quantMode GeneCounts) returning an average unique alignment rate of 54%. Abundance estimations of circRNAs were conducted utilising the chimeric junction file produced by the STAR aligner as processed by the CIRCexplorer2 (version 2.3.2 using default parameters) annotate function (X.-O. Zhang *et al.*, 2016). Differential expression analysis was evaluated from TMM normalized circRNA counts using R (version 3.2.3) and edgeR (Robinson, McCarthy and Smyth, 2010) (version 3.3) following protocols as described (Lun, Chen and Smyth, 2016). Alignments were visualised and interrogated using the Integrative Genomics Viewer v2.3.80 (Thorvaldsdóttir, Robinson and Mesirov, 2013).

2.3.2 MRNA seq library prep and high-throughput sequencing and bioinformatic analysis

Libraries were prepared using the NEBNext Ultra II Directional RNA-Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA). PolyA mRNA was isolated using the Poly(A) mRNA Magnetic Isolation Module (NEB#E7490) (New England Biolabs, Ipswich, MA, USA) using the same DNase I treated RNA from the circ-RNA seq library prep. Prepared libraries were sequenced at GENEWIZ (Suzhou, China) using a HiSeq 2500. Platform using the pair-end protocol with read length of 150. Raw were trimmed and filtered for short sequences using cutadapt v.1.9.1(Martin, 2011) setting minimum-length option to 18, error-rate 0.2 and overlap 5. The resulting FASTQ files averaging 39 million reads per sample were analysed and quality checked using FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Reads were mapped against the human reference genome (hg38) using STAR spliced alignment algorithm (Dobin *et al.*, 2013) (version 2.7.0 with default parameters) returning an average of unique alignment rate of 88%. Using featureCounts (Liao, Smyth and Shi, 2014) reads assigned to each exonic region and derived the counts. Differential expression analysis was evaluated from filtered RPKM normalised counts using DESeq2 statistical tool (Love, Huber and Anders, 2014). Alignments were visualised and interrogated using the Integrative Genomics Viewer v2.3.80 (Thorvaldsdóttir, Robinson and Mesirov, 2013).

2.3.3 Total-RNA Sequencing

Libraries were prepared using the TruSeq Stranded Total Library Prep kit for Illumina by South Australian Genomics Centre – Flinders Node. Prepared libraries were sequenced at GENEWIZ (Suzhou, China) using HiSeq 2500 (Illumina) platform using the paired-end protocol with read length of 150. Raw reads were trimmed and filtered for short sequences using cutadapt v.1.9.1 (Martin, 2011) setting minimum-length option to 18, error-rate 0.2 and overlap 5. The resulting FASTQ files averaging 124 million reads per sample were analysed and quality checked using FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Reads were mapped against the human reference genome (hg38) using STAR spliced alignment algorithm (Dobin et al., 2013) (version 2.7.0 with default parameters and --chimSegmentMin 20, --quantMode GeneCounts) returning an average of unique alignment rate of 88%. Differential expression analysis was evaluated from filtered RPKM normalised gene counts using DESeq2 statistical tool (Love et al., 2014). For abundance estimations of circRNAs were conducted utilising the chimeric junction file produced by the STAR aligner as processed by the CIRCexplorer2 (version 2.3.3 using default parameters) annotate function (Zhang et al., 2016). Alignments were visualised and interrogated using the Integrative Genomics Viewer v2.3.80 (Thorvaldsdottir et al., 2013).

2.4 Cell Culture

2.4.1 Cell lines

The cell lines used in this study are H9 hESCs, HEK293Ts, HEK293FTs and A549 cells and were confirmed by STR analysis, performed by AGRF. H9 hESCs, were confirmed to possess a normal XX karyotype with 22 pairs of autosomes (Appendix 1). The H9 line was derived from the inner cell mass of embryos in the blastocyst stage by Thomson *et al.* (1998) (Thomson *et al.*, 1998). This cell line is capable of forming derivatives of all three embryonic germ layers including endoderm, mesoderm and ectoderm. The H9 hESC cell along with the H1 hESC line, with a normal XY karyotype, was a gift from Lachlan Jolly (University of Adelaide, Australia). HEK-293T cells, were utilised for routine cell work and producing lentivirus particles. This cell line is a derivative of HEK 293 cells (Graham *et al.*, 1977), into which the SV40 T-antigen mutant tsA1609 has been inserted to replicate vectors carrying the SV40 region of replication. This derivative was generated by DuBridge *et al.* 1987 (DuBridge *et al.*, 1987). HEK293FT cells were used for lentiviral production, and are also a derivative of the HEK293 cell line. This line was established as a fast growing variant of HEK293Ts and therefore still contain the SV40 T-antigen (Oka *et al.*, 2010). A549 cells, a human non-small cell lung Carcinoma cell line, were utilised as a TGFβ sensitive cell line. This cell line consist of hypotriploid alveolar basal epithelial cells that were developed by D. J. Giard *et al.* 1973 by culturing pulmonary carcinoma tissue (Giard *et al.*, 1973). All cells were maintained at 37°C and 5% CO2.

2.4.2 Growth Medium

Media for all cells was supplied by Thermo Fisher. Both hESC cells lines (H9 and H1) were grown in Gibco™ StemFlex™ Medium, optimized to enable long-term feeder-free culture of PSCs. HEK-293T and A549 cells were grown in complete media of Gibco Dulbecco’s Modified Eagle Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS) without antibiotics.

2.4.3 Extracellular Matrices

All hESC cell lines were cultured on Gibco™ Vitronectin Recombinant Human Protein, Truncated (VTN-N) coated plates. Coating was applied at a concentration of 0.5 µg/cm² diluted in DPBS and using manufacturer’s recommended volume for each culture dish/plate. VTN-N provides a defined surface for feeder-free culture of human pluripotent stem cells while maintaining pluripotency and normal growth.

2.4.3.1 Geltrex

To promote adherence and differentiation, embryoid bodies were cultured on Gibco™ Geltrex® LDEV-Free hESC-qualified Reduced Growth Factor Basement Membrane Matrix at a concentration of 0.12–0.18 mg/ml.

2.4.3.2 Poly-L lysine

To promote adherence during transfections of HEK293FT Poly-L-lysine solution (0.1% (w/v) in H_2O) (Cat: P8920-100ML) was diluted in water 1:10 and transferred to culture dish at 1 ml per 25 cm^2.

2.4.4 Passaging cells

To passage and subculture hESCs, a 0.5mM EDTA solution in Dulbecco's phosphate-buffered saline (DPBS) was added after a single wash with DPBS. After a 5 min incubation, the solution was removed and cells were dislodged with an appropriate volume of StemFlex Medium. For general passaging, a 1:6 split was performed with an average passage time of 3-4 days.

A549, HEK-293T and HEK293FT cells were passaged using TrypLE Express (ThermoFisher, Waltham, Massachusetts, USA) and subcultured using standard cell culture methods when cells reached ~85% confluency.

2.4.5 Freezing/Thawing Cells

Human ESCs were cryopreserved using the Gibco™ Cryopreservation Kit, a ready to use xeno-free solution, following the manufacturers guidelines. To thaw hESCs, the cryovials were removed from liquid nitrogen vapour storage and immersed in a 37°C water bath, centrifuged at 200 g for 4 mins, and resuspended in StemFlex medium containing RevitaCell™ Supplement at the recommended concentration and cells were seeded on vitronectin coated plates. The RevitaCell supplement is a post-thaw recovery solution (100X concentrate) used to improve cell viability and was used for single cell passaging applications.

HEK293FT and A549 cells were cryopreserved in 50% complete media, 40% FBS and 10% DMSO in cryovials at a density of $1x10^6$ cells/ml. The vials were placed into a CoolCell® LX at, -80°C for short term storage, and in liquid nitrogen vapour for long term storage. To thaw HEK293FT and A549 cells, the vials were warmed in a 37°C water bath and seeded in a 25 cm^2 flask (Sarstedt).

2.4.6 Embryoid Body formation

To generate Embryoid Bodies (EBs) hESCs at 60-80% confluency were incubated for 2 hours in the presence of 1X RevitaCell before passaging. hESCs were single cell passaged following the StemFlex User guide and plated in a low adherence culture dish without further dilution. EBs were removed from the low adherence culture dish the next day and allowed to continue culturing in TC-treated dishes.

2.4.7 Antibiotics, Chemicals and solutions used during Cell Culture

2.4.7.1 Puromycin

Puromycin (25 mg/ml) was prepared by mixing 25 mg Puromycin in 1 ml of milliQ grade water and passed through a sterile µm filter. This stock was stored at -20°C and, prior to use, further diluted in water to 500 ng/ul for use in cell culture at required concentrations for cell lines.

2.4.7.2 G418

G418 solution (50 mg/ml) was obtained from Roche (Catalogue number 4727878001) and diluted directly in cell culture media at the required concentrations for cell lines.

2.4.7.3 Blasticidin

Blasticidin was reconstituted at 10 mg/ml and diluted directly into cell culture media at the required concentrations for cell lines.

2.4.7.4 Polybrene

Polybrene (10 mg/ml) was prepared by mixing 10mg polybrene in 1 ml of milliQ grade, water and passed through a 0.22nm filter.

2.4.7.5 Ethanol

Ethanol 100% denatured was obtained from Chem-supply and diluted to 70% with MilliQ grade water for sterilisation of equipment in cell culture. Ethanol 100% was also used for the fixation of cells prior to cell cycle analysis with Propidium Iodide or Hoechst 34580.

2.4.7.6 Propidium Iodide

Propidium Iodide (1 mg/ml) was obtained via the Flinders Flow Cytometry Facility and used to stain cells at 20 µg/ml in PBS for cell cycle analysis.

2.4.7.7 Hoechst 34580

Hoechst 34580 (1 mg/ml) (Catalogue number 63493) was used to stain cells at 0.2 µg/ml in PBS for cell cycle analysis.

2.4.7.8 HumanKine™ Transforming Growth Factor-β1

Human Transforming Growth Factor Beta 1 (hTGFβ1) was reconstituted in 4 nM HCL with 0.1% BSA to 5 µg/ml and used to induce EMT in A549 cells.

2.4.7.9 Paraformaldehyde

Paraformaldehyde (16% solution) was obtained from ProSciTech and diluted to 4% with milliQ grade water for fixation of cells.

2.4.7.10 Phosphate-buffered saline

Phosphate buffered saline (PBS) (10X) was obtained from ThermoFisher (Waltham, Massachusetts, USA) and diluted in water for irrigation for use in cell culture.

Dulbecco's phosphate-buffered saline (DPBS) (1X) was used in cell culture and to dilute vitronectin for hESC ECM.

2.4.7.11 TrypLE Express

TrypLE express was obtained from ThermoFisher (Waltham, Massachusetts, USA) and used for routine passaging of cells and single cell passage of hESCs following above protocols.

2.4.8 Cell cycle analysis

2.4.8.1 Fixed cell cycle analysis

Cells were harvested by trypsin and the cell pellet was washed with PBS and centrifuged at 300 g and resuspended at 0.5-1 x 10^6 cells in 500 µLs PBS. While vortexing, 500 µLs of ice-cold 100% ethanol was added dropwise to the cells in PBS. Cells were stored at -20°C for a minimum of 2 hours. To stain cells for analysis, cells were centrifuged at 300 g for 5 minutes. The Ethanol solution was decanted and the cell pellet was washed with 1 ml PBS for 5 minutes at room temperature. Cells were centrifuged at 300 x g for 5 mins and resuspended in 1 mL staining buffer (20 µg/ml Propidium Iodide, 200 µg/ml RNase A, 0.1% (v/v) Triton X-100 in PBS, pH 7.4) and incubated at room temperature for 30 mins in the dark. After staining, cells were used directly for cell cycle analysis on the CytoFLEX Flow Cytometer Platform (Beckman Coulter).

2.4.8.2 Live cell cycle analysis

Cells were harvested by trypsin and the cell pellet was resuspended in 10 µg/ml Hoechst 34580 in complete medium. Cells were incubated at 37°C for 30-60 mins. Cells were centrifuged at 300 g for 5 mins and resuspended in 1 x PBS and used directly for cell cycle analysis on the FACSAria™Fusion (Becton Dickinson).

2.4.8.3 Cell Synchronisation by Serum starvation

To synchronise cells in G1/G0 cell cycle phase, cells were seeded so that they were at ~30% confluent the next day at the time of serum starvation. Complete media was removed from the cells and replaced with DMEM containing no FBS.

Cells were incubated for 48 hours, DMEM with no FBS was removed and compete media was added back to the cells to release serum starvation and cell cycle blockage. 0.5 – 1 x 10^6 cells were used for cell cycle analysis via Propidium Iodide staining and the remaining cells were harvested by Trizol for RNA.

2.4.8.4 Cell Synchronisation by Nocodazole treatment

To synchronise cells in G2/M cell cycle phase, cells were seeded so that they were at ~30% confluent the next day, at the time of nocodazole treatment. Complete media was removed from the cells and replaced with DMEM containing 0.2 µg/ml Nocodazole for 24 hours. To release cell cycle blockage, media containing Nocodazole was removed an replaced with complete media without Nocodazole. 0.5 – 1 x 10^6 cells were used for cell cycle analysis via Hoechst 34580 staining and the remaining cells were harvested by Trizol for RNA.

2.4.9 Production of lentiviral particles

HEK293FT cells were seeded in T25 flasks at 30% confluency so that the cells would be 60-80% confluent. The following day, molar ratios of the expression vector (pLKO.1 and pLX311) along with psPAX2 packaging plasmid and VSVg envelope plasmid were mixed in 20 µLs of Opti-MEM reduced serum media. 6 µLs of Lipofectamine2000 was dilute to 80 µLs of Opti-MEM and incubated at room temperature for 10 mins. The DNA/Opti-MEM mixture was mixed with the Lipofectamine2000/Opti-MEM mixture and incubated for 20 mins at room temperature. The DNA/Lipofectamine/Opti-MEM mixture was added to the T25 flask containing HEK293T or HEK293FT cells with DMEM media, and allowed to incubate for 12-15 hrs at 37°C at 5% CO_2. The following day, the media containing the lipofectamine complexes was removed and replaced with fresh

DMEM media. Lentiviral supernatants were harvested at 48 hrs post media change and stored at -80°C.

2.4.10 Viral transduction of hESCs

hESCs were maintained on vitronectin and passaged once cell reached 60-80% confluency using 0.5mM EDTA solution for 5 mins to passage cells as aggregates instead of single cells. Cells were seeded at a high density (1:3 ratio) into a 24-well culture plate (Sarstedt). The following day, the media was change to 225 µls of StemFlex media containing 3 µg/ml polybrene (hexadimethrine bromide). 25 µls of viral supernatant was added to each well and centrifuged for 1 hour at 800 g. After centrifugation, cells were washed twice with DPBS before replacing with fresh StemFlex media containing no polybrene. Cells were allowed to recover for 24 hours before the addition of puromycin at 3 mg/ml to select for successfully transduced cells. Puromycin treatment lasted 3 days and cells that remained were expanded in media without puromycin. Cells were allowed to recover for 48 hours before the addition of blasticidin at 1 µg/ml to select for successfully transduced cells with blasticidin. Blasticidin treatment lasted until all control cells (wells with no viruses added) were dead.

2.4.11 Overexpression of circRNAs

Overexpression of circRNAs was completed by two methods depending on the cell line. For HEK293T cells, cells were transfected following standard lipofectamine2000 protocols using pcDNA 3.1 vector (Invitrogen, Waltham, Massachusetts, USA) following the protocol from Liu *et al* 2018 on the design of circRNA overexpressing vectors (D. Liu *et al.*, 2018).

To overexpress circRNAs in hESCs, the same design as above, for overexpressing circRNAs was utilised, however the pAAVS1-P-CAG-DEST (Addgene plasmid #80490) plasmid was used in combination with pTALdNC-AAVS1_T1 (Addgene plasmid #80495) and pTALdNC-AAVS1_T1 (Addgene plasmid #80496) (Oceguera-Yanez *et al.*, 2016). This results in integration of the circRNA overexpressing cassette into the AAVS1 sate site. Transfection was completed with Lipofectamine Stem Reagent following manufacturers protocol. Puromycin selection was performed 48 hours post transfection for 3-5 days at a concentration of 500ng/ml.

2.5 Bacterial Culture and plasmid preparation

2.5.1 Bacterial strain and culture

One Shot® TOP10 Competent Cells (Invitrogen, Waltham, Massachusetts, USA) were stored at -80°C and were transformed with plasmids using the heat shock method following the manufacturers standard transformation protocol. Cells were grown in LB Broth (Amyl Media) or plated on LB Agar (Amyl Media) supplemented with the appropriate antibiotic resistance (Table 1).

Table 2-1 Antibiotic concentration for LB Broth and Agar

Antibiotic	Concentration
Ampicillin	100 µg/ml
Kanamycin	50 µg/ml

2.5.2 Plasmids

pLKO.1 puro was obtained from Addgene (Plasmid #8453) and was ligated with annealed oligos to target specific candidate RNAs (circRNAs and mRNAs). Restriction enzyme digestion of the vector was completed sequentially with *Eco*RI-HF® (New England Biolabs, Ipswich, MA, USA) and *Age*I-HF® (New England Biolabs, Ipswich, MA, USA) in 1X CutSmart buffer at 37°C for 1 hr and gel purified for subsequent ligation of shRNA designed oligos. Table 2-2 contains information of other plasmids used in this study.

Table 2-2 Plasmid information used for mammalian cell culture

Plasmid name	Mammalian Antibiotic Resistance	Mammalian Antibiotic Concentration	Application	Obtained
pLKO.1	Puro	3 µg/ml	Expression of shRNAs and Cas13b gRNAs	Addgene (plasmid #8453)
pcDNA3.1	G418	N/A	Expression of proteins and circRNAs	Invitrogen (V79020)
pENTR2B	N/A	N/A	Transfer of circRNAs into pAAVS1-P-CAG-DEST	Invitrogen (A10463)
pAAVS1-P-CAG-DEST	Puro	3 µg/ml	Expression of circRNAs	Addgene (plasmid #80490)
pLX311	Blast	1 µg/ml	Expression of CRISPR Cas13b	Rosenbluh lab (Monash University)

2.5.3 Design of CRISPR Cas13b gRNA for pLKO.1 plasmid

GBlocks were ordered from Integrated DNA technologies (IDT) containing the required sequences for cloning the 3 individual gRNAs into pLKO.1 plasmid (Figure 2-1). A 30bp sequence (NX30) specific to the circRNA backsplice junction was included in the gblock (NX30) which gives the targeting specificity or Cas13b. The pLKO.1 vector was digested the same as for shRNA ligation (2.5.2 Plasmids), however the gblock was inserted into the vector by NEBuilder® HiFi DNA Assembly (New England Biolabs, Ipswich, MA, USA) following manufacturer's instructions.

```
5′ TGGAAAGGACGAAACACCGG(NX30)gttgtggaaggtccagttttgaggggctattacaacTTTTTTctcgacctcgagacaaat 3′
```

Figure 2-1 Oligonucleotide sequence cloned into pLKO.1 vector for Cas13b

Sequence that was ordered as an oligonucleotide containing a 30-nucleotide sequence (NX30) giving specificity to Cas13b to target *circACVR2A*.

2.5.4 Design of shRNA cassette for pLKO.1 plasmid

Top and bottom oligos were ordered separately from IDT and resuspended at 100 µM in annealing buffer (10 mM Tris pH 7.5, 1 mM EDTA, 50 mM NaCl). Oligos were annealed together by adding 2 µL of each top and bottom strand oligos into a 0.2 ml PCR tube with 16 µL of annealing buffer. The tubes placed in a thermocycler and were heated to 95°C for 2 mins. The thermocycler was set to cool to 20°C over 30 mins.

Top: 5' CCGG – 21bp sense -CTGCAG- 21bp antisense-TTTTTG 3'

Bottom: 5' AATTCAAAAA – 21bp sense -CTGCAG- 21bp antisense

2.5.5 Miniprep and Midiprep of plasmids

Plasmids were transformed into competent cells as neat plasmid or after ligation of DNA into the plasmid and spread on a agar plate with appropriate antibiotic. After overnight incubation, individual colonies were grown in 3-5 mls of LB broth with appropriate antibiotic. After 16-24 hrs, plasmids were harvested using the Bioline ISOLATE II Plasmid Mini Kit following manufacturer's instructions and eluted in supplied elution buffer (Bioline, London, UK). For Midiprep, the 2 mls of overnight culture was expanded in 200 mls LB broth for another overnight incubation. Plasmids were harvested using the Qiagen Plasmid Midi Kit following manufacturer's instructions and eluted in supplied elution buffer (Qiagen, Venlo, Netherlands).

2.5.6 Nucleic acid concentration

Measuring DNA or RNA concentration was performed on a Thermo Scientific™ Nanodrop™ One following manufacturer's instructions using the relevant elution buffer to blank before analysis. Purity ratios were assessed to obtain a 260/280 ratio of ~1.8 and 260/230 between the ranges of 1.8-2.2.

2.5.7 Digestion of DNA

DNA (plasmid and linear DNA) was digested following recommended conditions with Cutsmart buffer (New England Biolabs, Ipswich, MA, USA) or other buffers. Digested DNA was either purified by gel electrophoresis and purification or direct column purification.

2.5.8 Gel purification

Digested DNA or PCR amplified DNA was separated on a TAE agarose gel (1-2%) containing SYBR Safe DNA Gel Stain. Bands were viewed briefly on a Vilber E BOX using low UV light and removed using sterile Gel Extraction Tips. The DNA was eluted using QIAquick Gel Extraction Kit according to manufacturer's instructions (Victoria, Australia) and resuspended in provided elution buffer.

2.5.9 Column purification

Digested DNA or PCR amplified DNA was purified using the QIAquick PCR Purification Kit according to manufacturer's instructions (Qiagen, Victoria, Australia) and resuspended in provided elution buffer.

2.5.10 Ligation

Ligation of PCR products, annealed oligos or after Gateway cloning into plasmids was performed using T4 DNA Ligase (New England Biolabs) according to manufacturer's instructions. Ligation was performed with using a 3 to 5-fold molar excess of insert to vector in a 10-20 µl reaction volume at 4°C for a minimum of 20 hrs before bacterial transformation.

2.5.11 Gateway Cloning

Gateway™ LR Clonase™ II (Invitrogen, Waltham, Massachusetts, USA) was performed to manufacturer's instructions, with the reaction size reduced to 5 µls instead of 10 µls. Destination vector (75 ng) was mixed with entry vector (25ng) and made up to 4 µls with TE buffer pH 8.0. LR Clonase™ II enzyme mix (1 µl) was added to the reaction, and the tube was briefly vortexed before an incubation of 25°C for 1 hr. Proteinase K (1 µl) was added and the mixture was incubated at 37°C for 10 mins to terminate the reaction. 1 µl of the Gateway™ reaction was transformed into competent cells and screened by PCR for successful transformants.

2.5.12 Screening for positive colonies - Colony PCR

Colony PCR was performed by stabbing a single colony from an LB agar plate using a sterile pipette tip and dipping into a single PCR mixture with plasmid-specific primers to detect the amplification of the inserted DNA fragment. PCR was performed with the MyTaq Protocol (Bioline, London, UK) according to manufacturer's instructions.

2.5.13 Agarose gel electrophoresis

PCR or plasmid samples were separated based on size by agarose gel electrophoresis. Samples were run for 30-60 mins at 80-100V on a 1-2% (w/v) agarose gel in 1x TAE (40mM Tris-Base, 20mM Glacial Acetic Acid, 1mM EDTA, pH8) using a 1:20,000 dilution of SYBR Safe DNA Gel Stain (Invitrogen, Waltham, Massachusetts, USA). Gels were photographed on a Vilber E-BOX CX5 UV-Transilluminator (Eberhardzell, Germany).

2.6 Polymerase Chain Reaction (PCR)

2.6.1 Genomic DNA Isolation

Genomic DNA was extracted from cells using the AllPrep DNA/RNA Mini Kit according to the manufacturer's instructions (Qiagen, Victoria, Australia).

2.6.2 Total RNA isolation

Total RNA was extracted from cells using Trizol (Invitrogen, Waltham, Massachusetts, USA), according to the manufacturer's instructions. 200 µL of chloroform was added for each 1 ml of sample and shaken vigorously for 15 secs. The sample was incubated at room temperature for 2 mins and then centrifuged at 12,000 g for 15 mins at 4°C. The upper aqueous layer was transferred to a fresh 1.5 ml Eppendorf with 500 µL isopropanol for 1 ml Trizol used. Samples were mixed by inversion and incubated at room temperature for 5 mins followed by centrifugation at 12,000 g for 15 mins at 4°C. The supernatant was discarded and the RNA pellet was washed with 1 ml 75% ice-cold ethanol and vortexed briefly. The pellets were centrifuged at 7500 g for 5 mins at 4°C. The ethanol was discarded and the pellet was air dried in a fume hood for 5-10 mins. The dried RNA pellet was resuspended in 20 µLs RNase free water.

Total RNA was also purified from Trizol (Invitrogen, Waltham, Massachusetts, USA) using Zymo Research Direct-zol RNA Miniprep kit and following manufacturer’s instructions.

2.6.3 Reverse transcription of total RNA to cDNA

Reverse transcription was carried out with 1 µg of RNA using QuantiTect Reverse Transcription Kit (Qiagen, Venlo, Netherlands) according to the manufacturer’s instructions. cDNA was diluted 1/5 in sterile water and stored at -20°C.

2.6.4 Oligonucleotide design and provision

All primers and oligos were manufactured by Integrated DNA technologies (Iowa, USA), and suspended in either primer resuspension buffer (10 nM Tris-HCL pH 7.5, 0.1 mM EDTA pH 8) or oligo annealing buffer (10 nM Tris pH 7.5, 0.1 µM EDTA pH 8, 50 nM NaCl) at 100 µM concentration and stored at -20°C. Primer working concentrations (10µ) were prepared and stored at 4°C for several months. All qRT-PCR primers and oligonucleotides used in this study are listed in Appendix 2.

Primers used for qPCR were designed using the web-based tool Primer 3 (http://bioinfo.ut.ee/primer3-0.4.0/) to satisfy generic primer design rules and cross evaluated in NetPrimer (http://www.premierbiosoft.com/netprimer/). For SYBR based-qPCR, amplicons were designed to be 100-300bps in length. To amplify circRNA backsplice regions, divergent primers were designed so that the amplicon would cross over the backsplice site (Figure 2-2). Alternatively, primers could be designed so that the backsplice region sits within the primer sequence. For PrimeTime assays, IDT PrimerQuest® Tool was used to generate a probe sequence that contained the backsplice region and two primers on each side with an amplicon length of 75-150bps.

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 2-2 Design of circRNA qPCR primers

Designing divergent PCR primers for unique amplicon comprising of backsplice junction (Image: (Panda and Gorospe, 2018)).

2.6.5 PCR reactions and running conditions

For routine PCRs, 2X HotStart MyTaq RedMix (Bioline, London, UK) was used according to manufacturer’s instructions. PCR products were loaded directly onto an agarose gel and gel purified, then ligated into the pLKO.1 vector. All vectors underwent Sanger-sequencing, carried out at Flinders Medical Centre Sequencing Facility.

2.6.6 qPCR

For qPCR, Qiagen SYBR Green PCR Kit (Qiagen, Venlo, Netherlands) was used according to manufacturer’s instructions. For probe-based qPCR, PrimeTime® Gene Expression Master Mix (Integrated DNA technologies, Iowa, USA) was used according to manufacturer’s instructions with a probe to primer ratio of 1:2. qPCR was performed using the Rotorgene Q according to manufacturer’s instructions.

2.6.6.1 Selection of house-keeping genes

Analysis of RPKM values from the mRNA-seq data for these hESCs and differentiated cell lines was used to find genes with a low coefficient of variation (CV) that are also known housekeeping genes. *YWHAZ* was found to be the most stably expressed common housekeeping gene and was utilised in all hESC qRT-PCR studies. This was previously utilised by Martin *et al.* (2020).

For qRT-PCR analysis of A549 during cell cycle studies, *GAPDH* and *ACTB* mRNAs were used as house-keeping genes for normalisation (Jain *et al.*, 2012; Aviner *et al.*, 2015; Woźniak *et al.*, 2016, p. 53; Amatori, Persico and Fanelli, 2017).

2.6.6.2 Sanger Sequencing

PCR products or plasmids were purified and underwent dye termination and sequencing at the Flinders Medical Centre Sequencing Facility using forward and/or reverse primers for the PCR amplicon or primers anchored in the plasmid.

2.7 CircRNA pulldown

2.7.1 CircRNA pulldown cell transfection

HEK293 cells were grown on 15cm dish in standard conditions up to 60-70% confluency. Cells were transfected with 40 µg of pcDNA 3.1 plasmid harbouring the *circACVR2A* overexpression construct in 2 ml OptiMEM media mixed with 70 µl of Lipofectamine 2000 (in 2 mls OptiMEM media). This was incubated at room temperature for 5 mins and added, dropwise to cells already grown in DMEM (without FBS and antibiotics). RNA pulldown was performed 40 hrs post transfection.

2.7.2 CircRNA pulldown probes

5' biotinylated (5'TEG-biotin) DNA oligonucleotides reverse complementary to the circRNA backsplice junction were designed to be 30 nts long and overlapped *circACVR2A* backsplice site equally with 15 nts from either side of the backsplice junction. Using blast analysis for analysing potential off-target effects, and secondary structure folding analysis RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi), the best nucleotide length with no strong higher-dimensional folding was 30 nts.

2.7.3 *CircACVR2A* pulldown probes

circACVR2A – negative control (sense DNA strand)

5' – CTTGTTCCAACTCAAGTGCTATACTTGGTA – 3'

circACVR2A – (antisense DNA strand)

5' – TACCAAGTATAGCACTTGAGTTGGAACAAG – 3'

2.7.4 Preparation of Streptavidin C1 magnetic beads

Prior to circRNA pulldown, 150 µl beads were washed 3 times in B&W 2x wash buffer (10 mM Tris-HCL, pH 7.5, 2 M NaCl, 1 mM EDTA) and twice in Co-IP buffer (20 mM Tris-HCL, pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.5 % NP-40, and protease inhibitor (Complete™ ULTRA Tablets, Mini, EDTA-free, *EASYpack* Protease Inhibitor Cocktail, Merck)) and resuspended in 150 µl Co-Ip buffer.

2.7.5 *circACVR2A* circRNA-protein pulldown

RNA pull-down assays were carried out as described (William W. Du *et al.*, 2017, p. 3; William W Du *et al.*, 2017, p. 3). Briefly, 40 hr post transfection confluent 1 x 10^7 HEK293 cells were washed in ice-cold PBS and centrifuged for 5 mins at 200 g. Cells were resuspended and lysed in 500 µl co-IP buffer for 30 mins on ice, and incubated with 4 µls of 100 uM biotinylated DNA oligo probes against over expressed *circACVR2A* or negative control probes (sense strand DNA) at room temperature for 2 hrs. 150 µl washed Streptavidin C1 magnetic beads (Invitrogen, Waltham, Massachusetts, USA) and 20 µg of BSA (to prevent unspecific binding) were added to each binding reaction and further incubated at room temperature for another 2 hrs. Beads were washed briefly with co-IP buffer four times, followed by one wash with water. Elution was performed in 150 µl of 5 mM tris pH 8 with 1 µg of RNase A and incubated overnight at 37°C on a rotation wheel. The bound proteins in the pull-down material were analysed by MS. Three biological replicates were prepared for the positive and negative probes.

2.8 Mass spectrometry

2.8.1 MS sample preparation

Pull-down material was concentrated, and protein level measured using standard BCA protocol (Bio-Rad Protein Assay Dye reagent concentrate) (Biorad, Hercules, California, USA). One microgram of total protein was subjected to in-solution trypsin digestion. For In-solution trypsin digestion, protein samples were incubated for 30 minutes at 56°C in the following conditions: 111 mM Tris PH8, 1.1 mM Cacl2 (Sigma-Aldrich, Sydney, Australia), 5.6 mM DTT (ThermoFisher, Waltham, Massachusetts) in a total volume of 18 µLs. Subsequently, 1 µL of 200 mM 2-Chloroacetamide (Sigma-Aldrich, Sydney, Australia) was added to above reaction after it has cooled to room temperature and incubated in the dark at room temperature for 30 mins. Protein samples were digested in a 1:50 enzyme:protein ratio, containing 19ul of the above reaction mixture and Pierce™ Trypsin Protease, MS Grade (ThermoFisher, Waltham, Massachusetts) solution to the final concentration of 1 ng/µl. Sample was vortexed and incubated at 37°C overnight, transferred to mass spectrometry vials (Axygen) (Corning, Corning, New York, USA) and analysed by MS within 24 hours. Details of the mass spectrometry hardware, software and setting are in Appendix 3.

2.8.2 MS data analysis

Proteome data for negative and positive probes, separately, were analysed as follows. Results were filtered, rejecting protein hits that did not receive a high level of confidence or was identified with only a single peptide. Proteins found in any negative control were subtracted from positive results. Three replicates were compared and proteins found in at least two positive replicas were included in the final list of proteins interactions.

2.9 CellNet analysis

CellNet analysis of mRNA-seq FASTQ files and total RNA-seq FASTQ files was performed by Dr Marta Gabryelska following standard protocol (Radley *et al.*, 2017).

2.10 Immunofluorescence processing

The EBs were moved to a 1.5 ml microfuge tube and the supernatant was removed. The EBs were washed with 2 times with DPBS then resuspended in 1 ml 4% paraformaldehyde overnight at 4°C. Paraformaldehyde was removed and the fixed EBs were washed 3 times with DPBS before resuspending the EBs in 200-400 µls in molten low melting point agarose and incubated at 42°C for 2 hours in a heating block to allow EBs to settle to the bottom of the agarose. The tube was taken out of the heating block and allowed to cool upright in a rack at 16°C to solidify the agarose. One the agarose had solidified the agarose embedded EBs were taken out of the tube and excess agarose was cut off and discarded. The agarose embedded EBs were placed into a cassette and dehydrated in a series of ethanol solutions for 15 minutes each (70%, 80%, 90% (v/v)). The agarose embedded EBs were further dehydrated with 100% ethanol, 3 times at 6 mins each. After dehydration, the agarose embedded EBs were placed in chloroform overnight. After overnight incubation, the agarose embedded EBs were placed directly into a beaker of paraffin wax for 15 mins then moved to a second beaker of fresh paraffin wax for a further 15 mins. The agarose embedded EBs were taken out of the cassette and placed into a paraffin mould with the EB side of the agarose faced down, and molten paraffin was added and left to solidify on the paraffin embedding station cold plate. Once the paraffin had cooled the block was loaded into a microtome and sections were cut at 5µm increments.

Slides were placed inside antigen retrieval buffer (10mM Sodium Citrate, ph6) and boiled in a microwave at 100% power for 1m 40 seconds. Slides were further heated at 20% power for 10 minutes. After allowing the slides to cool, the slides were washed 3 times with deionized water for 2 minutes each. A PAP pen was used to draw a hydrophobic barrier around the sections. The sections were incubated for 30 mins with 5% BSA 0.1% PBS-T at room temperature. After blocking the sections were rinsed 3 times with PBS for 5 minutes each. The sections were incubated overnight with the primary antibody in 0.1% PBS-T with 1% BSA. After the overnight incubation, the sections were rinsed 3 times with PBS or 5 minutes each. The sections were incubated at room temperature in a dark chamber for 1 hour with a fluorophore-conjugated secondary antibody diluted in PBS-T with 1% BSA. The sections were rinsed 3 times with PBS for 5 minutes each. DAPI stained for 10 minutes diluted in 1% BSA PBS-T, 1/1000. The sections were washed 2 times in PBS or 5 minutes each. A coverslip was sealed on the top of the section by adding a drop of buffered glycerol, pH 8.6 and laying the coverslip carefully to not introduce bubbles. Sections were images on an Olympus AX70 at Flinders Medical Centre Microscopy with 40X magnification.

2.11 RNA immunoprecipitation

2.11.1 RNA immunoprecipitation cell transfection

HEK293FT cells were grown in a 6 well plate in standard conditions up to 60% confluency. One well was transfected with 2 µg of pcDNA 3.1 plasmid harbouring the *circACVR2A* overexpression construct and 2 µg of pcDNA 3.1 plasmid harbouring the Cyclin H FLAG overexpression construct in 250 µls OptiMEM media mixed with 7.5 µls of Lipofectamine 2000 also in 250 µls OptiMEM media. This was incubated at room temperature for 5 mins and added, dropwise to cells already grown in 2.5 mls DMEM without antibiotics. RNA pulldown was performed 40 hrs post transfection.

2.11.2 RNA immunoprecipitation of Cyclin H FLAG protein

Forty hours post transfection, single cells were collected by trypsin treatment and centrifuged at 1000 x *g* for 5 mins. Cells were washed twice with 1 x ice-cold PBS before being resuspended in 1 ml of RIPA buffer (150 mM NaCl, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% SDS, 50mM Tris pH 7.4, 200 units/ml RNase OUT, 1 x cOmplete Tablets, Mini EDTA-free (Roche, Basel, Switzerland), 1mM sodium orthovanadate, 1mM sodium fluoride). The lysate was rotated at 4°C for 30 mins. 25µl was kept for input protein, 50 µl for input RNA, then remaining lysate split into 2 x 500 µl samples – one for IgG IP and one for FLAG IP.

Dynabeads™ magnetic beads (ThermoFisher, Waltham, Massachusetts) were rotated for 5 minutes to resuspend the beads. 30 µls of Dynabeads™ magnetic beads were added to a 1.5ml tube. The tube was added to a magnet to remove separate the beads from the solution and the supernatant was removed. Antibodies (5 µg anti-FLAG and Mouse IgG) were diluted in 200 µls PBS with Tween™ 20 and added to the tube containing the Dynabeads™ magnetic beads. The bead-antibody mixture was incubated for 20 mins at room temperature. The bead-antibody mixture was put on the magnet and the supernatant was removed leaving the antibody conjugated beads. The conjugated beads were washed with 200 µls PBS with Tween™ 20 and the mixture was washed by gentle pipetting. The antibody-conjugated beads tube was put on the magnet to remove the supernatant and the prepared thawed lysate (500 µls) was added directly to the beads. The sample was incubated for 4 hrs at 4°C with rotation to allow beads to bind to the antigen. The tube was placed on the magnet at the supernatant was removed. The antigen-antibody-bead complex was washed 3 times by gentle pipetting adding 200 µls RIPA buffer then placing the tube on the magnet and removing the supernatant. After the last wash the beads were separated into two tubes for RNA and protein. For RNA, the 25% of the beads were resuspended in 500 µl Trizol and processed by an RNA extraction method. For protein, the antigen was eluted off the beads by boiling 75% of the beads in 1 X SDS loading dye (50 mM Tris-HCl, pH 6.8, 6% (w/v) glycerol, 2% SDS, 0.004% bromophenol blue, 1% 2-mercaptoethanol) at 95°C for 5 mins.

2.12 Western blotting

2.12.1 Protein preparation

Cells were lysed directly on the culture dish using RIPA Buffer containing protease inhibitor (cOmplete Tablets, Mini EDTA-free, EASYpack (Roche, Basel, Switzerland)) and phosphatase inhibitor (Pierce Phosphatase Inhibitor Mini Tablets (ThermoFisher, Waltham, Massachusetts)) where appropriate. The lysate was maintained under constant agitation for 30 mins at 4°C. The lysate was centrifuged for 16,000g for 20 mins at 4 degrees and the supernatant was transferred to a fresh tube on ice and mixed with 2x Laemmli (65.8 mM Tris-HCl, pH 6.8, 26.3% (w/v) glycerol, 2.1% SDS, 0.01% bromophenol blue, 5% 2-mercaptoethanol) for SDS-PAGE.

2.12.2 SDS-PAGE electrophoresis and transfer

SDS-PAGE was performed using precast Mini-PROTEAN TGX Stain-Free Gels (Biorad, #4568124) as per manufacturer's instructions (Biorad, Hercules, California, USA) in Running Buffer (25mM Tris, 190mM Glycine, 0.1% (w/v) SDS) at 200V. The gel was imaged on a ChemiDoc XRS+ (Biorad, Hercules, California, USA) for total protein.

SDS-PAGE was performed as per Laemmli (1970). The separating gel (375mM Tris-Cl, pH 8.8; 0.1% SDS; 6-15% Bis/Acrylamide (1:37.5); 0.05% Ammonium Persulphate; 0.2% TEMED) was poured and overlaid with isopropanol for 30 mins. After removal of isopropanol, the stacking gel (125mM Tris-Cl, pH. 6.8; 0.1% SDS; 4% Bis/Acrylamide; 0.05% Ammonium Persulphate; 0.2% TEMED) was poured on top of the separating gel and the comb inserted for 30 mins. Before the addition of the samples the comb was removed and the wells rinsed with running buffer (25 mM Tris, 190 mM Glycine, 0.1% (w/v) SDS), the gel was placed in the transfer tank with running buffer (25 mM Tris, 190 mM Glycine, 0.1% (w/v) SDS) added so that it submerged the wells. Protein samples (20 µls) were loaded into each well and the gel run at 200V until the dye front reached the bottom of the gel or until desired.

The transfer of proteins onto a Nitrocellulose membrane (Merck Millipore, Burlington, Massachusetts, USA) was performed by tank transfer as per manufacturer’s instructions (Biorad, Hercules, California, USA) in Transfer Buffer (25mM Tris, 190mM Glycine, 20% Methanol) at 80V for 50 minutes at 4°C.

2.12.3 Antibody probing

After transfer, the membrane was blocked with skim milk (3-5%) in TBST (20mM Tris, 150mM NaCl, 0.1% Tween 20) for 1 hr. The membrane was incubated with the primary antibody in fresh blocking solution overnight at 4°C with gentle agitation on a rocking platform. The membrane was washed 3 x 10 minutes with TBST, followed with secondary antibody incubation at room temperature or 1 hr with gentle agitation. After secondary antibody incubation the membrane was washed with TBST 3 x 10 mins. Chemiluminescence was detected using SuperSignal West Pico PLUS Chemiluminescent Substrate following manufacturer’s instructions (ThermoFisher, Waltham, Massachusetts) and imaged using the Biorad ChemiDox XRS+ (Biorad, Hercules, California, USA). Normalisation was performed using the Biorad imaging software and using total protein for each blot.

2.12.4 Antibodies

Table 2-3 Antibodies used in this study

Antibodies	**Application**	**Dilution**	**Manufacturer**	**Source**	**Lot.no**
Anti-SMAD2/3	WB	1:400 WB	R&D systems (AF3797)	Goat	N/A
Anti-FLAG	WB, IP	1:1000 WB 5ug IP	Sigma-Aldrich (F1804)	Mouse	SLCD3524
Anti-Cyclin H	WB	1:1000 WB	Cell Signalling (2927)	Rabbit	12/2019
p-CDK Substrates (pTPXK)	WB	1:1000 WB	Cell Signalling (14371)	Rabbit	08/2020
Normal mouse IgG	IP	5ug IP	Santa Cruz (sc-2025)	Mouse	G2020
Anti-Mouse-HRP	WB	1:10,000 WB	ThermoFisher (A16078)	Goat	mLA16078
Anti-Rabbit-HRP	WB	1:10,000 WB	ThermoFisher (A16110)	Goat	mLA16110
Veriblot	WB	1:500 WB	Abcam (ab131366)	N/A	GR3343276-1
Anti-OCT3/4	IF	1:500	Santa Cruz (sc-5279)	Mouse	A0219
Anti-NANOG	IF	1:500	Abcam (ab109250)	Rabbit	n/a
Anti-FOXA2	IF	1:300	Abcam (ab108422)	Rabbit	n/a
Anti-Rabbit Alexa Fluor® 488	IF	1:100	ThermoFisher (A11008)	Goat	1735088
Anti-Mouse Alexa Fluor® 488	IF	1:100	ThermoFisher (A11001)	Goat	1008801
Anti-Rabbit Cy5®	IF	1:500	ThermoFisher (A10523)	Goat	1270183

Chapter 3 Transcriptomics of Human Embryonic Stem Cell Differentiation

3.1 Introduction

Next-generation sequencing (NGS), or more specifically, RNA sequencing (RNA-seq) is a group of techniques which provides a high resolution RNA transcriptional landscape of particular cellular material (Kukurba and Montgomery, 2015). Analysis of RNA-seq data can inform on expression changes between samples or treatments as well as other transcriptomic features (Figure 3-1) including exon boundaries, alternatively spliced transcripts, single nucleotide polymorphisms, mutations and abundances of different species of RNA (mRNA, miRNA, circRNA etc.) (Zhao *et al.*, 2014).

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 3-1 Alignment of RNA-seq reads

Aligning RNA-seq reads to reference genome, providing information regarding exon positions, alternatively spliced variants and intron retention. (Image: (Marco-Puche *et al.*, 2019))

Total RNA-seq entails sequencing all coding and non-coding RNA and, through random priming, is useful for determining the global transcriptomic profile.

To investigate an individual RNA species more deeply, enrichment techniques are employed, which involves either enriching for the target transcript(s) and/or the removal of unwanted RNA transcripts (Kukurba and Montgomery, 2015), before converting the RNA into cDNA libraries required for sequencing (Figure 3-2). After completion of sequencing of the short cDNA fragments, the reads (sequencing fragments) are aligned to a reference genome and differential expression can be performed to determine the expression levels of genes (Kukurba and Montgomery, 2015).

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 3-2 Overview of RNA-seq library preparation

Total RNA is extracted from cells including coding and non-coding transcripts. Specific kits for selection or depletion allow for the enrichment for RNA of interest. The RNA is fragmented to shorter lengths and converted to cDNA. Sequencing adapters are ligated to the ends of the cDNA to allow for PCR amplification and sequencing. (Image: (Kukurba and Montgomery, 2015)

Poly-A enriched RNA-seq is a common method for transcriptomic profiling of genes which removes all RNAs that do not contain a large length of adenosine sequences. This enrichment removes the ability to detect changes to circRNAs and other RNA transcripts that do not contain a poly-A sequence. CircRNAs have largely been ignored for many years as a possible consequence, or cause, for multiple diseases.

CircRNAs contain a unique backsplice sequence (Figure 3-3), as a downstream terminus of the RNA transcript joins to its own upstream terminus. This unique sequence allows for the detection of circRNAs even in the presence of other overlapping RNA species, as any reads that overlap the backsplice junction, can be effectively used to quantify circRNA levels. However, properties including exon exclusion or intron retention within circRNA transcripts may not be able to be distinguished. To quantify and illuminate on circRNA attributes, RNA isolation techniques are employed to enrich for circRNAs. This enrichment process includes digesting total RNA with RNase R, an exoribonuclease that cleaves all linear RNA, and completing a ribosomal RNA depletion (Zhang *et al.*, 2020). To further enrich for circRNAs, other treatments are possible that includes polyadenylation of any remaining linear and small RNAs after RNase R digestions, coupled with a depletion of polyadenylated RNAs (Pandey *et al.*, 2019). This effectively isolates circRNAs for downstream RNA-seq applications and can help to identify circRNAs that have not been previously discovered by other enrichments methods. Furthermore, addition of these enrichment steps can help to reveal full length circRNA sequences as the reads coming from linear RNAs is vastly depleted (Pandey *et al.*, 2019). As the majority of mapped reads from total RNA are linear transcripts, and typically less than 0.1% are circRNAs (Gruner *et al.*, 2016), removal of all other RNAs is crucial to allow for deeper sequencing of circRNAs so changes in expression can be confidently quantified.

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 3-3 Backsplice junction of circRNAs

CircRNAs are identified by detecting the backsplice sequence. This unique sequence is used for detecting circRNAs by RNA-seq, targeting the circRNA by RNAi and in designing and validating qRT-PCR (Image adapted from: (Cooper, Cortés-López and Miura, 2018)).

CircRNA regulation during stem cell differentiation, and particularly into multiple lineages, has not been thoroughly tracked and investigated. To determine which circRNAs are regulated during human stem cell differentiation, circRNA-seq was performed on RNA collected from undifferentiated H9 hESCs and cells that were differentiated into five separate cell types (Definitive Endoderm, Hepatocytes, Cardiomyocytes, Floor Plate Precursors and Dopaminergic Neurons) that comprise the three germ cell lineages (Endoderm, Mesoderm and Ectoderm) (Figure 3-4). Poly-A RNA-seq (mRNA-seq) was also performed to give insight on the regulation of circRNAs along with their parental gene transcripts. Analysis of these RNA-seq data sets will inform on the expression changes of circRNAs during differentiation, during the loss of pluripotency and stemness characteristics, and after the cells undergo waves of EMT to reach both precursor and terminal cell types. Utilising circRNA-seq and mRNA-seq for these cells will also inform on the relationship between candidate circRNAs and parental mRNA expression. The study of multiple differentiation events from a single cell type will enhance the confidence of identified correlations between expression changes of unique circRNAs and their parental mRNAs. The identified regulated and de-regulated expression patterns of circRNAs compared to their parental mRNAs, will inform on the link between these so-expressed transcripts (arising from the same gene), and may reveal circRNAS that are required for differentiation into specific lineages. Furthermore, as EMT is a required process during stem cell differentiation, identification of commonly regulated circRNAs during all differentiation events will provide a list of candidate circRNAs that may be involved with an EMT program shared during differentiation into multiple cell types. The benefit of these multiple differentiation events, including 2 precursor cell lines (Definitive Endoderm and Floor Plate Precursors), and 3 mature cell lines (Hepatocytes, Cardiomyocytes, and Dopaminergic Neurons), is that it captures multiple stages of differentiation. This strategy allows for the identification of EMT related circRNAs, while lineage-specific and independently regulated circRNAs can be eliminated. Although other strategies exist to profile circRNAs during EMT, including TGFβ treatment of HMLE cells that has been utilised in our lab, cells that have undergone differentiation represent cells of varying epithelial and mesenchymal characteristics, providing an alternative approach for the study of EMT traits as well as investigating circRNA regulation during differentiation.

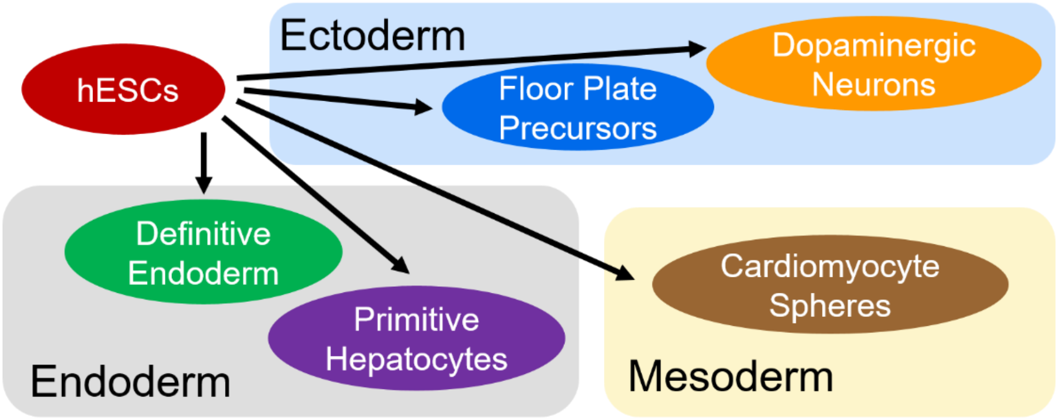

Ectoderm
hESCs
Dopaminergic Neurons
Floor Plate Precursors
Definitive Endoderm
Primitive Hepatocytes
Endoderm
Cardiomyocyte Spheres
Mesoderm

Figure 3-4 Germ layers and resulting terminal cell lines

Graphical representation of which cell types arise from which of the three germ cell layers.

3.2 Results

3.2.1 CircRNA Sequencing

H9 hESCs were differentiated into 5 cell types comprising of the three germ layers, ectoderm, mesoderm and endoderm. RNA from biological triplicates were harvested and underwent RNase R digestion and rRNA depletion. Stranded libraries were sequenced on Illumina NextSeq (150SE) and circRNAs identified using circExplorer (Zhang *et al.*, 2014). Differential expression analysis of mapped reads overlapping circRNA backsplice junctions was performed using edgeR. Between 5000 to 14000 circRNAs were detected in each sample with a minimum of 2 backsplice reads (Table 3-1). Depending on the cell type, 30-50% of these circRNAs were detected in all 3 replicates. 2373 circRNAs were found to be commonly expressed in all cell types, using a cut-off value of 2 average backsplice reads across the 3 biological replicates (Figure 3-5). Furthermore, each cell type has uniquely detected circRNAs, with Dopaminergic Neurons (Dopa) and Floor Plate Precursors (FP) having the most at 2214 and 2122 circRNAs, respectively.

Table 3-1 Number of unique circRNAs detected in each circRNA-seq sample

Number of unique circRNAs that are detected within each circRNA-seq sample with a minimum of 2 reads overlapping the backsplice junction.

Cell type	Replicate 1	Replicate 2	Replicate 3	Average	Common between replicates	Percent of Ave
H9 hESCs	5328	5335	5835	5499.33	2215	40.28%
Definitive Endoderm	6677	7509	6694	6960.00	3019	43.38%
Floor Plate Precursors	11522	11507	13731	12253.33	6175	50.39%
Dopaminergic neurons	13717	12910	7394	11340.33	3773	33.27%
Primitive Hepatocytes	5713	8530	7610	7284.33	3133	43.01%
Cardiomyocyte Spheres	8588	10187	7990	8921.67	4092	45.87%

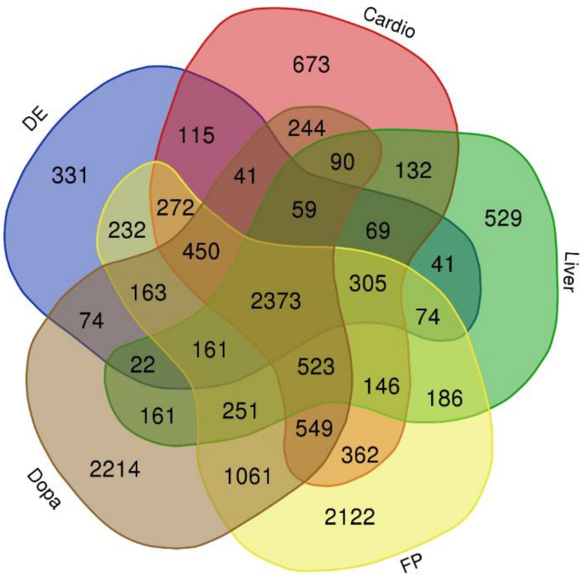


Figure 3-5 Venn diagram showing the number of detected circRNAs with backsplice junction reads overlapping from each lineage

Venn diagram showing the number of detected circRNAs, with a minimum of 2 overlapping backsplice reads on average from the three replicates of each differentiated cell type. Vann diagram could not include H9 hESCs as this includes too many samples to complete the analysis. DE = Definitive Endoderm, Cardio = Cardiomyocytes, Liver = Primitive Hepatocytes, FP = Floorplate Precursors, Dopa = Dopaminergic Neurons.

Analysis of circRNA-seq using an MDS plot, a method for visualisation of sample dissimilarity, revealed that each cell line has a unique circRNA transcriptional profile that is not stochastic, as the three replicates for each cell lineage cluster together, while the different cell lineages are distinct from each other (Figure 3-6A). CircRNA-seq differential expression, a common analysis performed to identify RNA transcripts that have altered expression, is only calculated for circRNAs based on reads that overlap the backsplice junction. Sequencing reads that do not overlap the backsplice junction can be visualised to uncover unique characteristics of circRNAs. This includes the differential splicing of *circCORO1C* (Figure 3-6B), as the second exon is incorporated into the final circRNA at varying levels between the cell types. Furthermore, reads that map to sequences in between exons, as is the case for *circCAMSAP1* (Figure 3-6C), identity circRNAs that have retained introns, yet for this circRNA, changes to the abundance of the circRNA throughout differentiation is minimal. Finally, differential expression analysis of backsplice reads and visualisation of reads between the backsplice junction inform on the regulation of circRNA expression as demonstrated by *circSMARCA5* (Figure 3-6D).

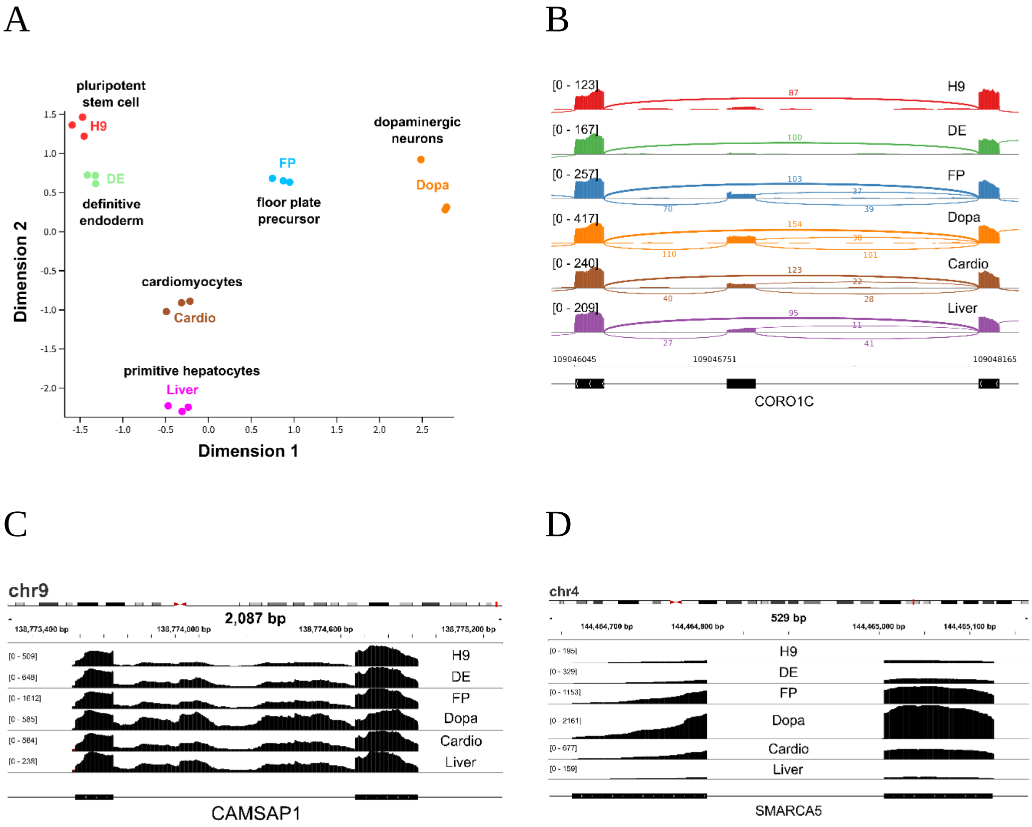


Figure 3-6 Analysis of circRNA-seq data

Analysis of circRNA-seq data identifies unique characteristics for each cellular lineage. (A) MDS plot of high-confidence circRNAs for each biological replicate identifying sample similarity and lineage-specificity based on circRNA expression clustering. (B) Sashimi plot illustrating differential exon inclusion of *circCORO1C*. (C) Coverage plot illustrating differential intron exclusion of *circCAMSAP1*. (D) Coverage plot illustrating differential expression of *circSMARCA5* in each lineage. DE = Definitive Endoderm, Cardio = Cardiomyocytes, Liver = Primitive Hepatocytes, FP = Floorplate Precursors, Dopa = Dopaminergic Neurons.

3.2.2 CircRNA Differential Expression Analysis using FDR-adjusted p-values

To determine significantly regulated circRNAs, differential expression analysis was performed. To initially screen for regulated circRNAs in this data, statistically significant Log2Fold changes were used to identify circRNAs that were commonly regulated during each differentiation/EMT event compared to H9 hESCs. Using a FDR-corrected p-adjusted value (padj) of 0.05 circRNAs from each cell type were found to be statistically differentially expressed (Table 3-2).

Definitive Endoderm (DE) cells have the least number of circRNAs that are statistically differentially expressed (padj<0.05) from H9 hESCs, however this cell population clustered closer to H9 hESCs than any other cell type using the MDS plot analysis, implying the circRNA expression profile of DE cells is more comparable to hESCs than other cell types. Conversely, Dopa and Hepatocytes (Liver), which clustered furthest away from H9 hESCs in the MDS analysis, had the greatest number of statistically differentially expressed (padj<0.05) circRNAs, with 1256 and 698 respectively (Table 3-2). The use of a Venn diagram displays the number of statistically differentially expressed circRNAs for each cell type and also if these circRNAs share their expression changes with other cell types (Figure 3-7). Only 1 circRNA (*circCDYL*) was found to be commonly statistically increased in all lineages from H9 hESCs and 11 circRNAs were found to be commonly decreased in expression (Figure 3-7 and Table 3-3). Additionally, a larger proportion of circRNAs were found to overlap between Dopa and FP and also between Cardiomyocyte spheres (Cardio) and Primitive Hepatocytes (Liver) than with other cell types, in both increased and decreased circRNAs Venn diagrams. This is in line with the MDS plot analysis as these cell lineages are clustered closer (Figure 3-6A) than with the other cell types.

Table 3-2 Statistically differentially expressed circRNAs (padj<0.05) from each lineage

Number of statistically significantly differentially expressed circRNAs in each lineage ($p<0.05$) compared with undifferentiated H9 hESCs, detected by RNA-seq.

Cell lineage	Differentially expressed circRNAs that increase	Differentially expressed circRNAs that decrease	Total differentially expressed circRNAs
Definitive Endoderm	23	33	56
Cardiomyocyte Spheres	214	228	442
Primitive Hepatocytes	314	384	698
Floor Plate Precursors	244	259	503
Dopaminergic neurons	659	597	1256

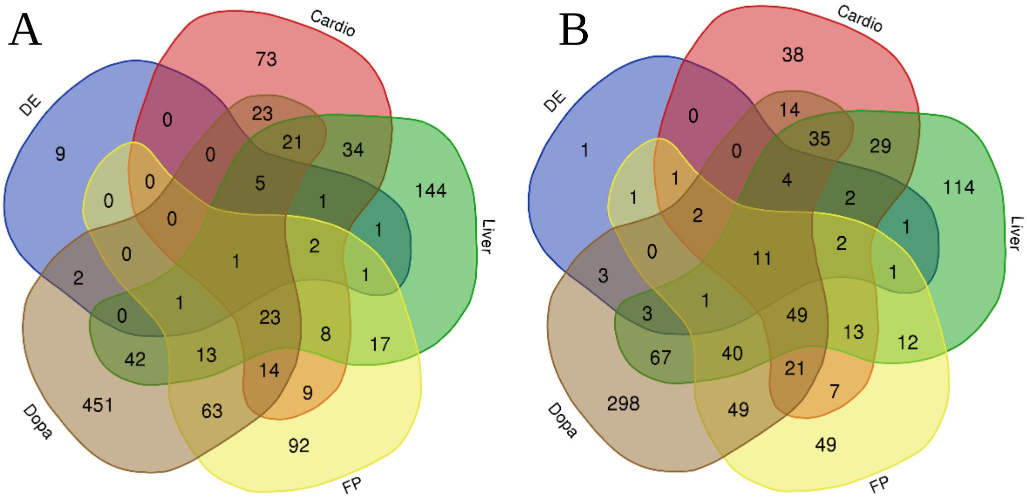


Figure 3-7 Venn diagram showing the number of overlapping statistically significantly differentially expressed circRNAs (padj<0.05) from each lineage

Venn diagram showing the number overlapping statistically significantly expressed circRNAs (padj<0.05) that are (A) upregulated circRNAs from each lineage and (B) downregulated circRNAs from each lineage. DE = Definitive Endoderm, Cardio = Cardiomyocytes, Liver = Primitive Hepatocytes, FP = Floorplate Precursors, Dopa = Dopaminergic Neurons.

Table 3-3 List of commonly regulated significantly differentially expressed circRNAs in all cell lineages

List of commonly regulated significantly differentially expressed circRNAs in all cell lineages (padj $p<0.05$) including their genome location in hg19 genome, detected by RNA-seq. One upregulated circRNA and eleven downregulated circRNAs.

Upregulated circRNAs
chr6:4891946-4892613_+_CDYL

Downregulated circRNAs
chr8:102570646-102571040_+_GRHL2
chr7:148851036-148851432_+_ZNF398
chr19:11230767-11231198_+_LDLR
chr7:121753171-121773795_-_AASS
chr5:7414685-7521012_+_ADCY2
chr2:64083439-64085070_+_UGP2
chr6:34059659-34101636_-_GRM4
chr12:111855922-111856681_+_SH2B3
chr1:85331067-85331821_-_LPAR3
chr20:31379406-31380576_+_DNMT3B
chr9:133907562-133911531_+_LAMC3

Although this list provides strong candidates that may be involved with an EMT event, the cells with which RNA was extracted may no longer show differential expression of EMT-related circRNAs as EMT and MET can occur in succession during differentiation (Li *et al.*, 2017). Furthermore, this list may only provide circRNAs that are regulated similarly during differentiation and not provide a comprehensive list of differentially regulated circRNAs. To this end, using less stringent statistical analyses would broaden the number of circRNAs that could be functionally important during EMT and differentiation.

3.2.3 CircRNA differential expression is only calculated based on backsplice reads

To further support using a less stringent statistical analysis, of the 18.9 million aligned reads on average per sample, only 6.97% of these aligned over backsplice junctions and could be used for differential expression. Therefore 93.03% of aligned reads on average could not be used for determining statistical differential expression changes of circRNAs. Furthermore, 90.91% of unique circRNAs were identified with only 1 backsplice read in only 1 of the 16 samples, and this accounted for 23.45% of backsplice reads (Figure 3-8A) and do not give any significant statistical data. The most expressed unique circRNAs with at least over 100 reads on average per sample, only accounted for 0.02% of unique circRNAs (Figure 3-8B) and comprised 7.65% of all backsplice reads, and 0.53% of total aligned reads (Figure 3-8A). Due to the number of unique circRNAs with low copy numbers per cell, which make up the majority of backsplice sequencing reads, and because only back-splicing reads can be used for detection of unique circRNAs, statistical significance was determined using non-adjusted p-values ($p<0.05$) for the remainder of the chapter unless otherwise stated.

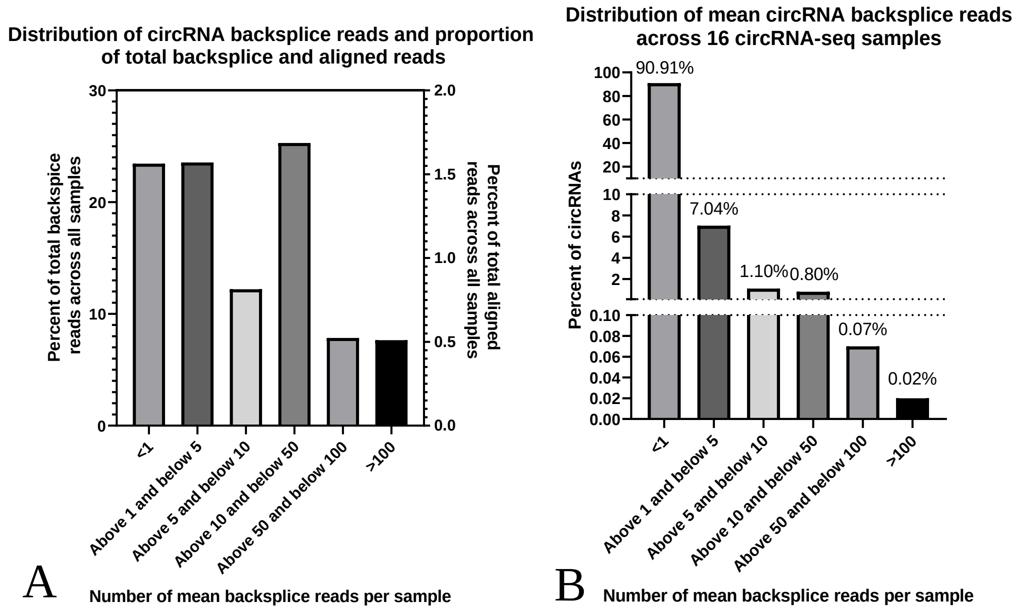


Figure 3-8 Distribution of circRNA backsplice reads in circRNA-seq data

(A) Distribution of circRNA backsplice reads and proportion of total backsplice and aligned reads. (B) Distribution of mean circRNA backsplice reads across the 16 circRNA-seq samples.

3.2.4 CircRNA Differential Expression Analysis using non FDR-adjusted p-values

Statistically differentially expressed ($p<0.05$) circRNAs from H9 hESCs were identified from each lineage (Table 3-4) using non FDR-adjusted p-values. Again, DE cells have the lowest number of statistically differentially expressed ($p<0.05$) circRNAs, at 631 circRNAs, while Dopa and Liver cells have the most, at 2549 and 1729 respectively. Using a Venn diagram, 27 circRNAs are now found to be commonly increased during differentiation, while 63 circRNAs are decreased (Figure 3-9 and Table 3-5) providing a larger list than with using FDR-adjusted p-values.

Table 3-4 Statistically significantly differentially expressed circRNAs from each lineage

Number of statistically significantly differentially expressed circRNAs in each lineage ($p<0.05$) compared with undifferentiated H9 hESCs, detected by RNA-seq.

Cell lineage	Differentially expressed circRNAs that increase	Differentially expressed circRNAs that decrease	Total differentially expressed circRNAs
Definitive Endoderm	288	343	631
Cardiomyocyte Spheres	675	717	1392
Primitive Hepatocytes	824	905	1729
Floor Plate Precursors	773	724	1497
Dopaminergic neurons	1380	1169	2549

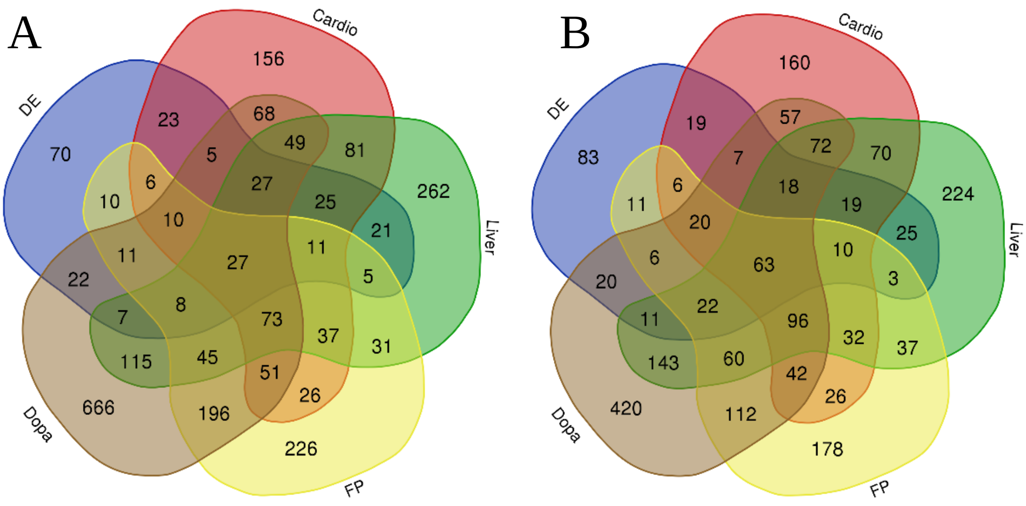


Figure 3-9 Venn diagram showing the number of overlapping statistically significantly expressed circRNAs (p<0.05) from each lineage

Venn diagram showing the number overlapping statistically significantly expressed circRNAs (p<0.05) that are (A) upregulated circRNAs from each lineage and (B) downregulated circRNAs from each lineage. DE = Definitive Endoderm, Cardio = Cardiomyocytes, Liver = Primitive Hepatocytes, FP = Floorplate Precursors, Dopa = Dopaminergic Neurons.

Table 3-5 Commonly statistically significantly regulated circRNAs in all differentiated lineages

Commonly statistically significantly upregulated and downregulated circRNAs in all differentiated lineages (p<0.05) compared with undifferentiated H9 hESCs, detected by RNA-seq.

Commonly upregulated circRNAs	Commonly Downregulated circRNAs
chr2:40655612-40657444_-_SLC8A1	chr17:37864573-37866734_+_ERBB2
chr8:68018139-68028357_+_CSPP1	chrX:130883333-130928494_-_FIRRE
chr5:64747301-64769779_-_ADAMTS6	chr10:126631025-126631876_+_ZRANB1
chr12:31595708-31600703_-_DENND5B	chr4:146767107-146770713_-_ZNF827
chrX:79544404-79565732_-_CHMP1B2P	chr16:84773914-84779279_+_USP10
chr8:17847367-17851128_+_PCM1	chr7:73100965-73101425_+_WBSCR22
chr21:40578033-40601362_-_BRWD1	chr5:128861976-128864370_+_ADAMTS19
chr21:38559349-38568337_+_TTC3	chr18:8718421-8720494_+_SOGA2
chr7:16298014-16317851_-_ISPD	chr20:31379406-31380576_+_DNMT3B
chr18:9182379-9221997_+_ANKRD12	chr22:47188416-47193517_+_TBC1D22A
chr1:65830317-65831879_+_DNAJC6	chr5:113740134-113740553_+_KCNN2
chr3:57276883-57282379_+_APPL1	chr2:64083439-64085070_+_UGP2
chr11:102233626-102239279_+_BIRC2	chr14:51710573-51716483_+_TMX1
chr10:122273422-122280607_+_PPAPDC1A	chr15:101775286-101775782_-_CHSY1
chr1:117944807-117963271_+_MAN1A2	chr8:128902834-128903244_+_PVT1
chr7:23650789-23651172_+_CCDC126	chr9:96277948-96278551_+_FAM120A
chr6:13632601-13644961_-_RANBP9	chr11:6962801-6977013_+_ZNF215
chr11:120345268-120348235_+_ARHGEF12	chr2:64083439-64090062_+_UGP2
chrX:107083899-107097934_+_MID2	chr15:65110144-65111422_-_PIF1
chr12:42604156-42604482_-_YAF2	chr20:31376659-31380576_+_DNMT3B
chr11:46098304-46113774_-_PHF21A	chr8:102570646-102571040_+_GRHL2
chr11:120916382-120930794_+_TBCEL	chr6:34059659-34101636_-_GRM4
chr6:4891946-4892613_+_CDYL	chr6:10703637-10705077_+_PAK1IP1
chr8:109462051-109468159_+_EMC2	chr15:65040617-65043837_-_RBPMS2

chr17:57430575-57430887_+_YPEL2	chr6:34100754-34101636_-_GRM4
chr20:57014000-57016139_+_VAPB	chr1:233334684-233344435_-_PCNXL2
chr14:66028054-66028484_+_FUT8	chr19:11230767-11231198_+_LDLR
	chr12:53670403-53671011_+_ESPL1
	chr20:31379406-31381401_+_DNMT3B
	chr7:148851036-148851432_+_ZNF398
	chr1:85331067-85331821_-_LPAR3
	chr20:31383214-31385105_+_DNMT3B
	chr15:92459222-92459688_+_SLCO3A1
	chr12:95914772-95918564_-_USP44
	chr22:36914809-36915584_-_EIF3D
	chr5:7414685-7521012_+_ADCY2
	chr4:16587544-16597498_-_LDB2
	chr3:52028167-52029057_-_RPL29
	chrX:96603116-96639043_+_DIAPH2
	chr16:77850817-77859358_+_VAT1L
	chr17:80714040-80772810_+_TBCD
	chr17:20149238-20163607_+_SPECC1
	chr8:129021835-129022690_+_PVT1
	chr6:99955303-99958106_-_USP45
	chr8:12583235-12600791_-_LONRF1
	chr1:6659106-6659512_-_KLHL21
	chr22:33402336-33402809_-_SYN3
	chr19:4430545-4432204_+_CHAF1A
	chr7:121753171-121773795_-_AASS
	chr5:80419460-80422982_+_RASGRF2
	chr1:233334684-233372726_-_PCNXL2
	chr12:111855922-111856681_+_SH2B3
	chr5:89781341-89802491_+_POLR3G
	chr9:133907562-133911531_+_LAMC3
	chr13:41133645-41134997_-_FOXO1
	chr1:233296028-233363117_-_PCNXL2
	chr13:114157810-114175048_+_TMCO3
	chr20:31376659-31381401_+_DNMT3B
	chrX:8553307-8565297_-_KAL1
	chrX:96603116-96684744_+_DIAPH2
	chr6:29575020-29576377_-_GABBR1
	chr15:25212175-25213229_+_SNRPN
	chr3:185307901-185318643_+_SENP2

The Venn diagrams also illustrates that, for each lineage, a portion of the total number of circRNAs that are statistically differentially expressed ($p<0.05$) from H9 hESCs, is differentially expressed uniquely in that lineage. Dopa cells have the largest number of uniquely statistically differentially expressed circRNAs (1086 circRNAs) which represents 42.6% of the total number of circRNAs differentially expressed in Dopa cells (Table 3-6). Furthermore, Cardio cells only have 486 uniquely statistically differentially expressed circRNAs, representing 22.7% of the total number of circRNAs differentially expressed in Cardio cells. Again, a larger overlap is observed between the populations of Dopa with FP and

DE with Liver (Figure 3-9), as similarly observed with the FDR-adjusted statistically differentially expressed circRNA Venn diagram.

Table 3-6 Statistically significantly differentially expressed circRNAs that are exclusively differentially expressed in each lineage

Number of statistically significantly differentially expressed circRNAs that are exclusively regulated from a single lineage ($p<0.05$) compared with undifferentiated H9 hESCs. Included is the percentage of the total number of statistically differentially expressed circRNAs in each lineage ($p<0.05$) compared with undifferentiated H9 hESCs, detected by RNA-seq.

Cell lineage	Differentially expressed circRNAs that increase	Differentially expressed circRNAs that decrease	Total differentially expressed circRNAs
Definitive Endoderm	70 (24.3%)	83 (24.2%)	153 (24.2%)
Cardiomyocyte Spheres	156 (23.1%)	160 (22.3%)	316 (22.7%)
Primitive Hepatocytes	262 (31.8%)	224 (24.8%)	486 (28.1%)
Floor Plate Precursors	226 (29.2%)	178 (24.6%)	404 (27.0%)
Dopaminergic neurons	666 (48.3%)	420 (35.6%)	1086 (42.6%)

Overall, 90 circRNAs were commonly significantly differentially expressed ($p<0.05$) from undifferentiated hESCs to all of the 5 cell types (Table 3-5) with 27 circRNAs increasing in expression and 63 circRNAs decreasing in expression. As Dopa cells have the highest number of statistically differentially expressed circRNAs, common differentially expressed circRNAs makes up only 3.5% (Table 3-7) in Dopa cells. Comparatively, 14.2% of all statistically significantly differentially expressed circRNAs in DE cells are commonly differentially expressed in all other lineages.

Table 3-7 Proportion of commonly statistically differentially expressed circRNAs of the total number of statistically differentially expressed circRNAs for each lineage

Percentage of the common statistically significantly differentially expressed circRNAs of the total statistically differentially expressed circRNAs for each lineage ($p<0.05$) compared with undifferentiated H9 hESCs, detected by RNA-seq.

Cell lineage	Differentially expressed circRNAs that increase	Differentially expressed circRNAs that decrease	Total differentially expressed circRNAs
Definitive Endoderm	27 (9.4%)	63 (18.4%)	90 (14.2%)
Cardiomyocyte Spheres	27 (4.0%)	63 (8.8%)	90 (6.5%)
Primitive Hepatocytes	27 (3.3%)	63 (7.0%)	90 (5.2%)
Floor Plate Precursors	27 (3.5%)	63 (8.7%)	90 (6.0%)
Dopaminergic neurons	27 (2.0%)	63 (5.4%)	90 (3.5%)

Interestingly, there are more common statistically differentially expressed circRNAs that were decreased than increased (Table 3-7), however, this may be due to all cell lineages arising from H9 hESCs and therefore have a common ancestral cell line. In fact, commonly decreased circRNAs are statistically more abundant in H9 hESCs than the commonly increased circRNAs (Figure 3-10). A higher number of commonly regulated circRNAs are lost from the precursor cell line than circRNAs that are gained in all differentiation events. Downregulation can be more difficult to detect for lowly expressed circRNAs, and therefore, need a higher initial backsplice count to be statistically differentially expressed, leading to this difference in upregulated and downregulated circRNAs.

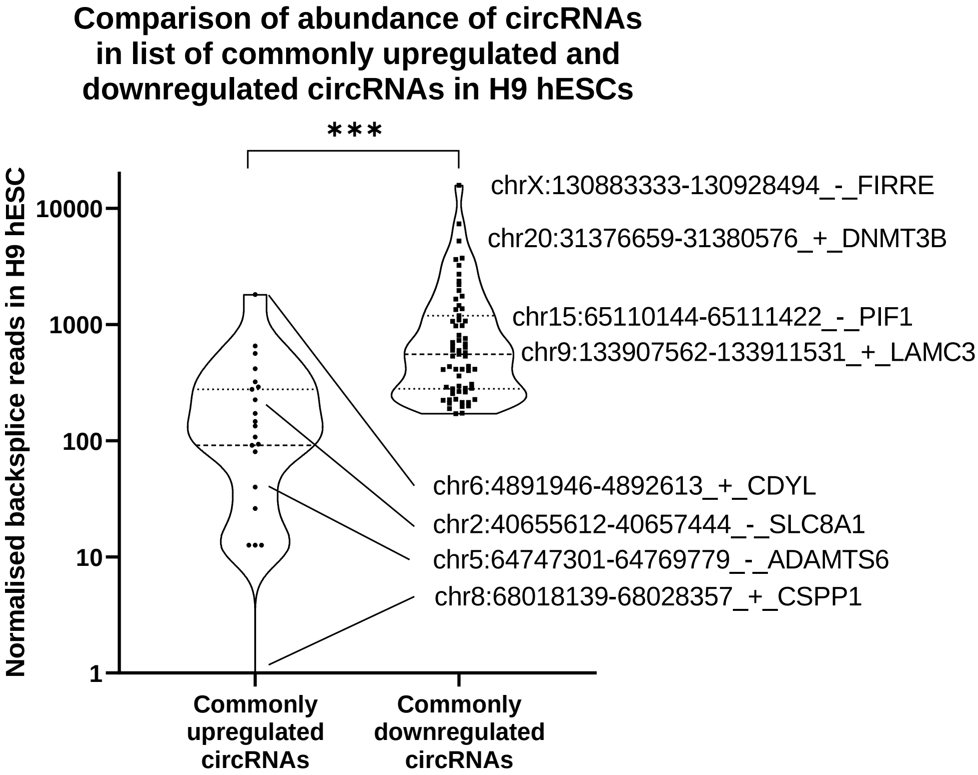


Figure 3-10 Abundance of commonly regulated circRNAs in undifferentiated H9 hESCs

Two-tailed unpaired t-test with Welch’s correction for not assuming equal standard deviation, comparing the abundance of the commonly statistically significantly upregulated and downregulated circRNAs in H9 hESCs. Y axis is the average normalised backsplice reads in H9 hESCs across the 3 RNA-seq libraries. Y axis is Log format. **** = p<0.0001

3.2.5 Candidate circRNA selection

To maximize the ability to manipulate and quantify expression changes of circRNAs that might be linked to EMT, circRNA candidates that show the most significant changes in expression across differentiation and, ideally, are highly abundant, and found to be commonly regulated circRNAs (Table 3-5) were prioritised. This aligns with the theory that a highly expressed and highly regulated circRNA has a greater likelihood of being a driver of EMT than a circRNA possessing only one of these characteristics.

CircRNA candidates were chosen based on several factors including their abundance within undifferentiated H9 hESCs including *circZNF608* (2nd highest), *circFAT3* (12th highest) and *circZNF609* (58th highest), as well as large fold changes from H9 hESCs after differentiation into individual cell types (*circSLC8A1*, *circCSPP1*, *circLAMC3*) (Table 3-8 and Figure 3-11). CircRNAs that are observed to increase in some cell types and decrease in others are also included as positive or negative regulation during EMT/differentiation is a possibility. EMT-related circRNAs may not present with transcriptional regulation in all differentiation events so other candidates have also been included that may not be regulated in all cell types (i.e. 4 out of 5 lineages).

Table 3-8 Criteria for circRNA candidate selection

Criteria
Commonly increase in abundance
Commonly decrease in abundance
Lineage-specific abundance changes
Correlation with parental transcripts
No Correlation with parental transcripts
Abundant circRNAs

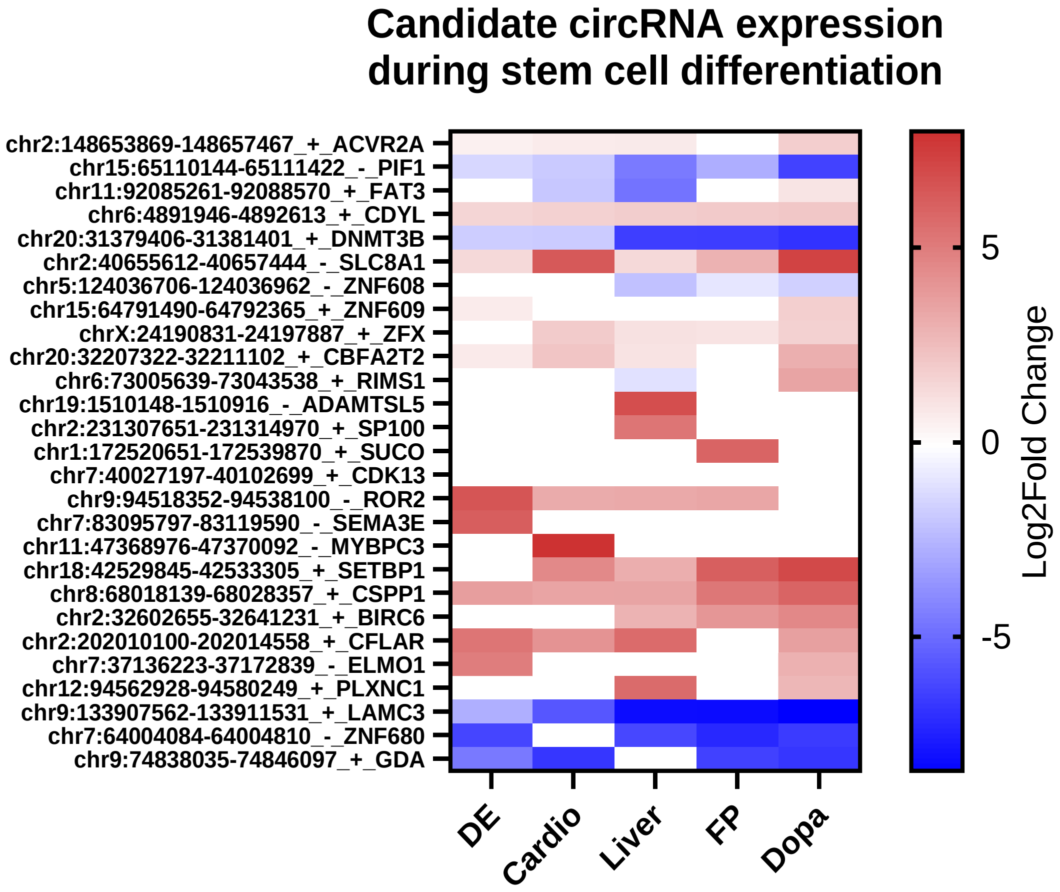


Figure 3-11 circRNA candidate Log2Fold changes in differentiated cell lines

Heatmap of circRNA candidates including Log2Fold change from undifferentiated H9 stem cells. Data is only given if the change from H9 hESCs is statistically significant ($p<0.05$). DE=Definitive Endoderm. Cardio=Cardiomyocytes. Liver=Hepatocytes. FP=Floor Plate Precursors. Dopa=Dopaminergic Neurons.

3.2.6 Poly-A mRNA sequencing

To identify the regulatory relationship between circRNAs and their parental genes, Poly-A mRNA sequencing was performed using the remaining RNA that was used to generate the circRNA-seq library. Differential expression analysis of trimmed and mapped reads was completed by the DEseq2 program. Analysis of pluripotency and differentiation markers for all replicates reveals that the five *in vitro* differentiation protocols produced cells with an increase in the expected lineage markers by RNA-seq and confirmed the successful differentiation into the different cell types (Figure 3-12). This analysis also highlights discrepancies for replicate 2 from hESCs, FP and Dopa as these did not correlate with the other replicates, as the normalised read counts did not match, or entire genes were missing read counts after bioinformatical analysis of reads.

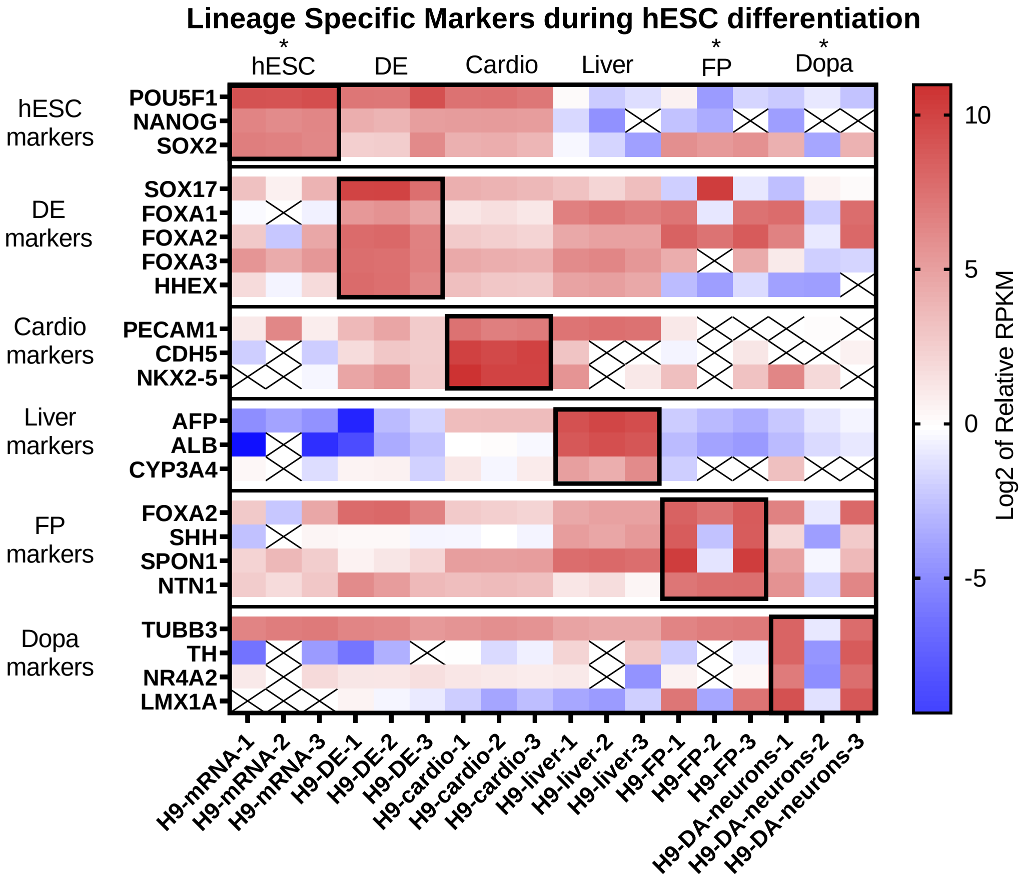


Figure 3-12 hESC and mRNA lineage markers in hESCs and differentiated cells

Heat map analysis of mRNA seq data for hESC and differentiation markers during hESC differentiation. Matrix coordinates represent the Log2 of the normalised RPKM (DESeq2) for each cell line and replicate. The data is only relative to the other coordinates in each row and does not represent expression values between different genes. Each row has been manipulated so that expression changes of highly and lowly expressed genes can be represented. Black boxes outline the lineage markers for the representative cell type. The replicates that do not match closely with the other replicates are marked with an asterisk (*).

Principal component analysis of all biological replicates of the mRNA-seq similarly revealed that the same samples from the PCA plot do not cluster with the other replicates from the same cell type, while the other replicates clustered very closely (Figure 3-13A). The remaining RNA from these samples after circRNA-seq analysis was low, and so these samples went through more amplification cycles during the RNA-seq library preparation, however, did not perform well as a biological replicate. These samples were removed to generate another PCA plot (Figure 3-13B) and was also removed from the differential expression analysis as they greatly altered the standard deviation of mRNA transcripts which impedes statistical analysis during differential expression analysis. The circRNA-seq and mRNA-seq PCA plots were both able to stratify the cell populations to comparable levels, however, the circRNA plot did stratify DE cells as a distinct population marginally better than the mRNA-seq plot. This may be due to the tightness of the replicates or due to having only two replicates for H9 hESC.

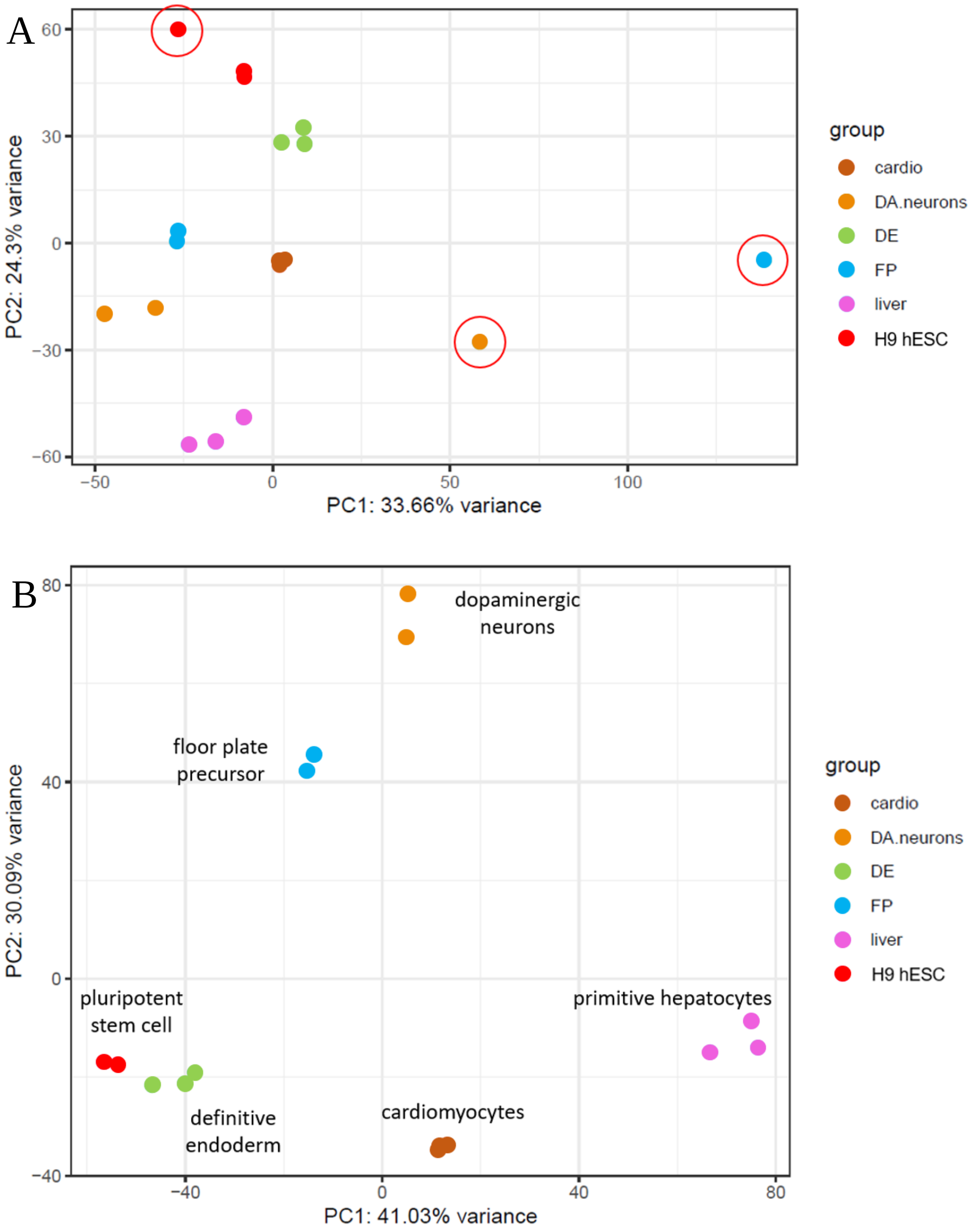


Figure 3-13 PCA plots of mRNA-seq data

PCA plot of mRNA-seq data for each biological replicate identifying sample similarity and lineage-specificity based on mRNA expression clustering. (A) PCA plots of all replicates from samples. Replicates in Red circles do not cluster well with other replicates from the same cell type. (B) exclusion of replicates that do not match the clustering with the other replicates from the same cell type.

Each differentiation was found to generally decrease the expression of pluripotency genes (Figure 3-14A), increase mesenchymal gene expression (Figure 3-14B), and show altered epithelial gene profiles (Figure 3-14C). Reepithelialisation of cells can occur after initial differentiation into some cell lines, and is observed to occur in the differentiation from ESCs to hepatic cells as the majority of these cells co-express epithelial and mesenchymal markers (Li, Pei and Zheng, 2014), further supported in this data for Liver cells (Figure 3-13C).

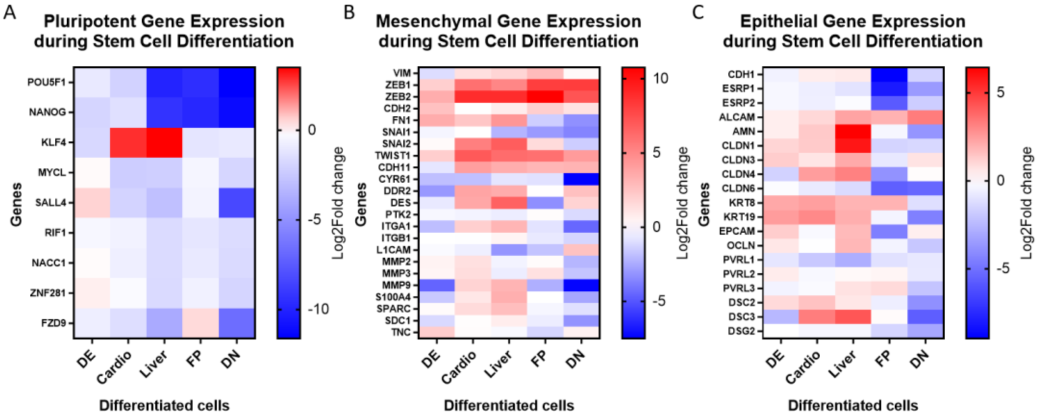


Figure 3-14 Changes to markers of pluripotency and EMT

Heat maps summarising changes in expression of (A) Pluripotent genes, (B) Mesenchymal genes and (C) Epithelial genes during differentiation of H9 hESCs into multiple cell lineages. Matrix coordinate is only presented if the change from H9 hESCs is statistically significant using RNA-seq differential expression analysis (DESeq2) ($p<0.05$).

Thousands of genes were found to be significantly differentially expressed (padj<0.05) from stem cells to each cell type (Table 3-9). Interestingly, Cardio differentiation resulted in more detectable significantly differentiated gene transcripts than other cell types, however, Cardio cells showed the second lowest number of differentially expressed circRNAs (Figure 3-15). Dopa cells were found to have the highest number of differentially expressed circRNAs compared to other cell types, however, no comparable increase to mRNA transcripts was observed during differentiation from H9 hESCs.

Table 3-9 Statistically differentially expressed genes during hESC differentiation in each lineage

Number of statistically differentially expressed genes during hESC differentiation in each lineage (padj<0.05), detected by RNA-seq. Statistical significance is quantified using RNA-seq differential expression analysis (DESeq2).

Cell lineage	Increased Differentially Expressed genes	Decreased Differentially Expressed genes
Definitive Endoderm	2799	2522
Cardiomyocytes	6608	6065
Hepatocytes	4479	3361
Floor Plate Precursors	3612	4465
Dopaminergic Neurons	3577	3206

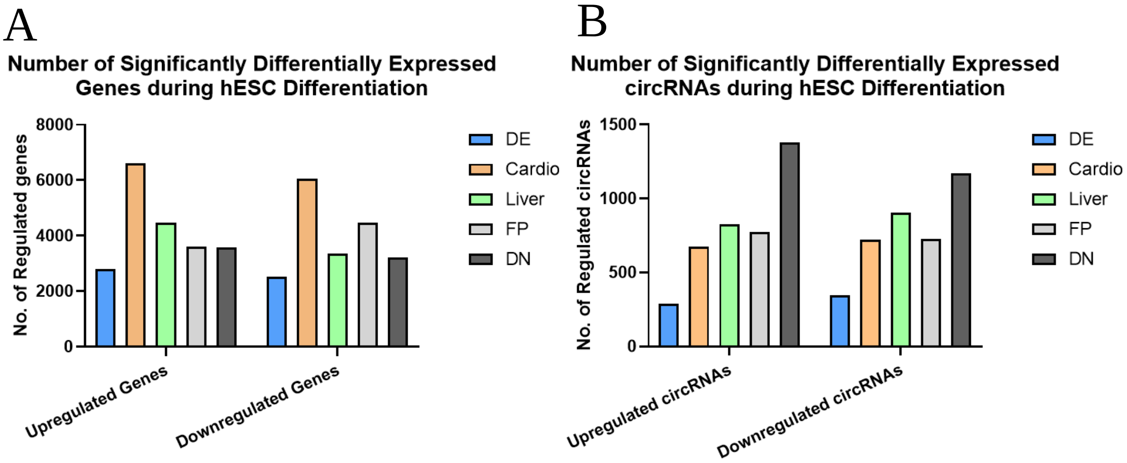


Figure 3-15 Differences between number of significantly differentially expressed genes during differentiation

Comparative differences between the number of (A) genes and (B) circRNAs identified as significantly differentially expressed between each lineage compared to H9 hESC, detected by RNA-seq. DE=Definitive Endoderm. Cardio=Cardiomyocytes. Liver=Hepatocytes. FP=Floor Plate Precursors. DN=Dopaminergic Neurons.

3.2.6.1 Correlation between circRNA and parental gene expression

Log2Fold changes for the 100 most abundant circRNAs in H9 hESCs, and their parental gene transcripts were compared to determine if changes to parental gene expression impacts circRNA expression during differentiation. Correlation analysis revealed a positive correlation between circRNA expression and parental gene expression (Figure 3-16). The changes observed in parental mRNA expression explains some, but not all of the variation of circRNA levels as r-squared values range between 0.26 to 0.61. This positive correlation is less appreciable in DE with an r-squared value of 0.26, but prominent for the other cell types. A positive correlation is also observed when comparing not only abundant, but also abundant and statistically differentially expressed circRNAs from each lineage (Figure 3-17).

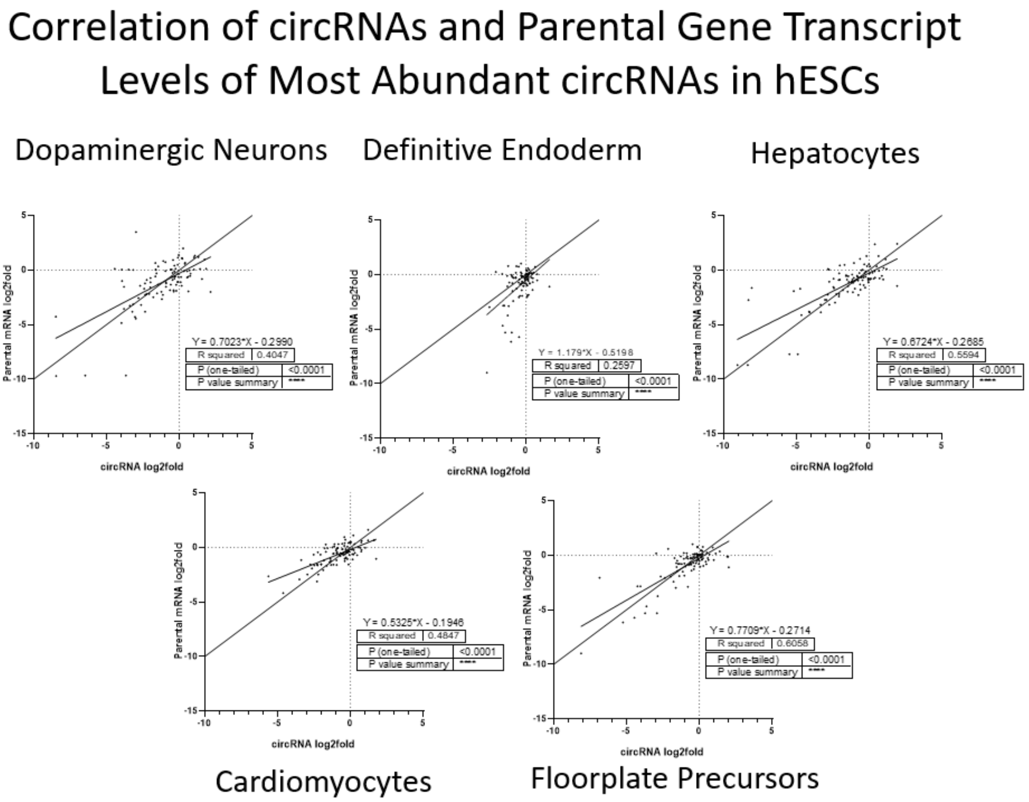


Figure 3-16 Correlation analysis of most abundant circRNAs and parental transcripts in hESCs

Correlation of the top 100 most abundant circRNAs with their parental gene. One-tailed Pearson's correlation analysis of circRNA and parental gene log2fold changes in each lineage, detected by RNA-seq. Graphs include line of identity to observe the slope when x and y coordinates are equal.

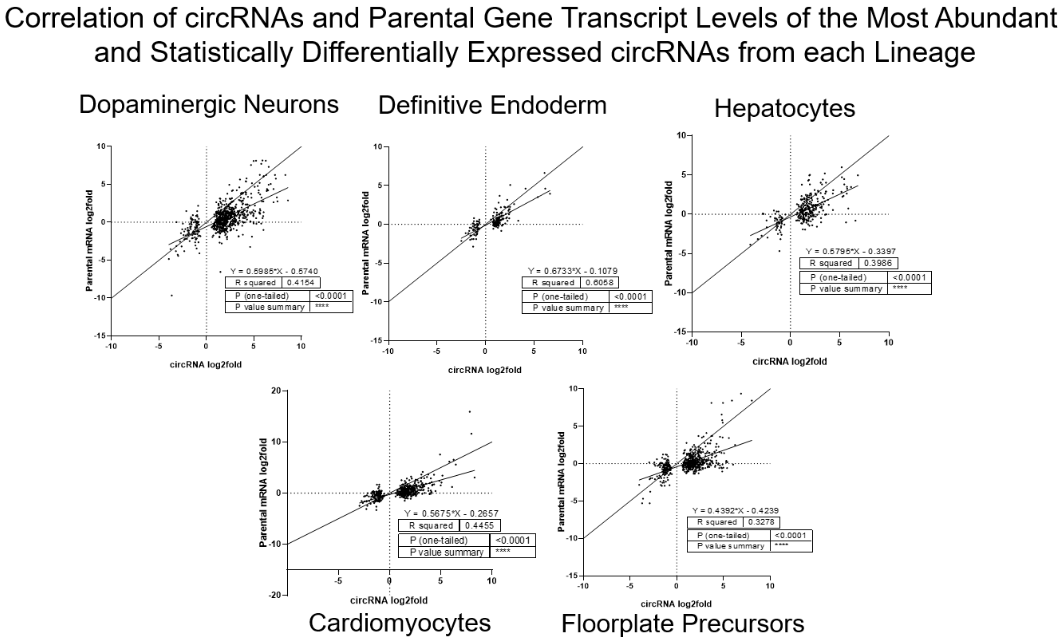


Figure 3-17 Correlation analysis of most abundant and statistically differentially expressed circRNAs and parental transcripts in each differentiated lineage

RNA-seq analysis of the most abundant (minimum of 10 reads on average from 3 replicates) and statistically differentially expressed ($p<0.05$) circRNAs and parental gene expression shows correlation during stem cell differentiation. One-tailed Pearson's correlation analysis of circRNA and parental gene log2fold changes in each lineage, detected by RNA-seq. Graphs include line of identity to observe the slope when x and y coordinates are equal.

Comparing the Log2Fold change for the circRNA candidates (Figure 3-18A) with their parental genes (Figure 3-18B), both co-expression and independent expression is observed visually using heatmaps. A one-tailed Pearson's Correlation analysis identifies *circACVR2A*, *circPIF1* and *circFAT3* as somewhat correlated with their parental gene transcript (p=<0.05) during each differentiation event (Figure 3-19). In contrast, *circCSPP1* and *circBIRC6* transcripts are regulated independently from their parental gene with as no statistically significant correlation observed.

Some circRNAs that increased in expression only in one or two celltypes (*circADAMTSL5* and *circPLXNC1*) follow their parental gene trend while *circSUCO* is observed to increase only in FP, however this is not reflected with an increase to parental gene expression (Figure 3-18). *CircCDYL* is unique in this subset of circRNAs in that it shows a statistically significant inverse correlation with its parental gene as the circRNA abundance increased during each differentiation event, while its parental mRNA decreased (Figure 3-18, Figure 3-19). Interestingly, this is the only circRNA found to have an inverse correlation with its parental gene throughout each unique differentiation event.

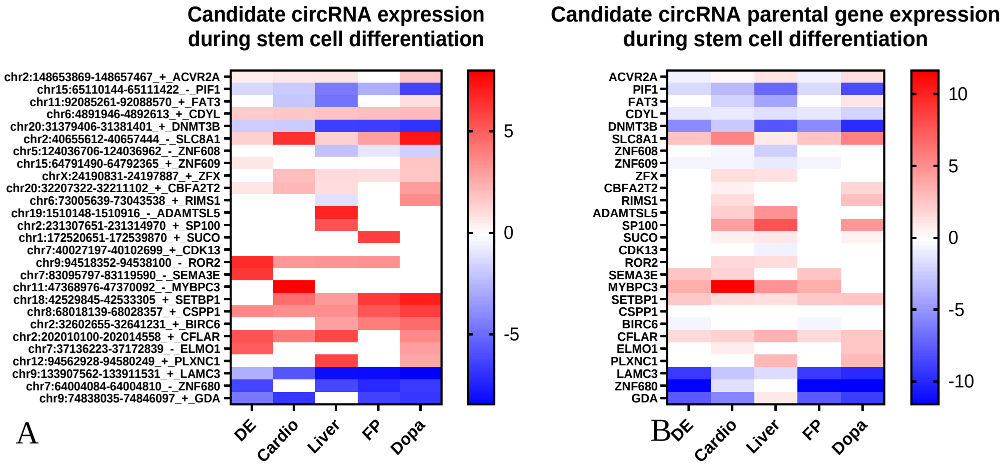


Figure 3-18 List of Log2Fold changes of candidate circRNAs and their parental transcripts

List of (A) circRNA candidates and (B) their parental gene Log2Fold change from undifferentiated H9 stem cells and their expression level before differentiation. Data is only given if the change from H9 hESCs is statistically significant ($p<0.05$). DE=Definitive Endoderm. Cardio=Cardiomyocytes. Liver=Hepatocytes. FP=Floor Plate Precursors. DN=Dopaminergic Neurons.

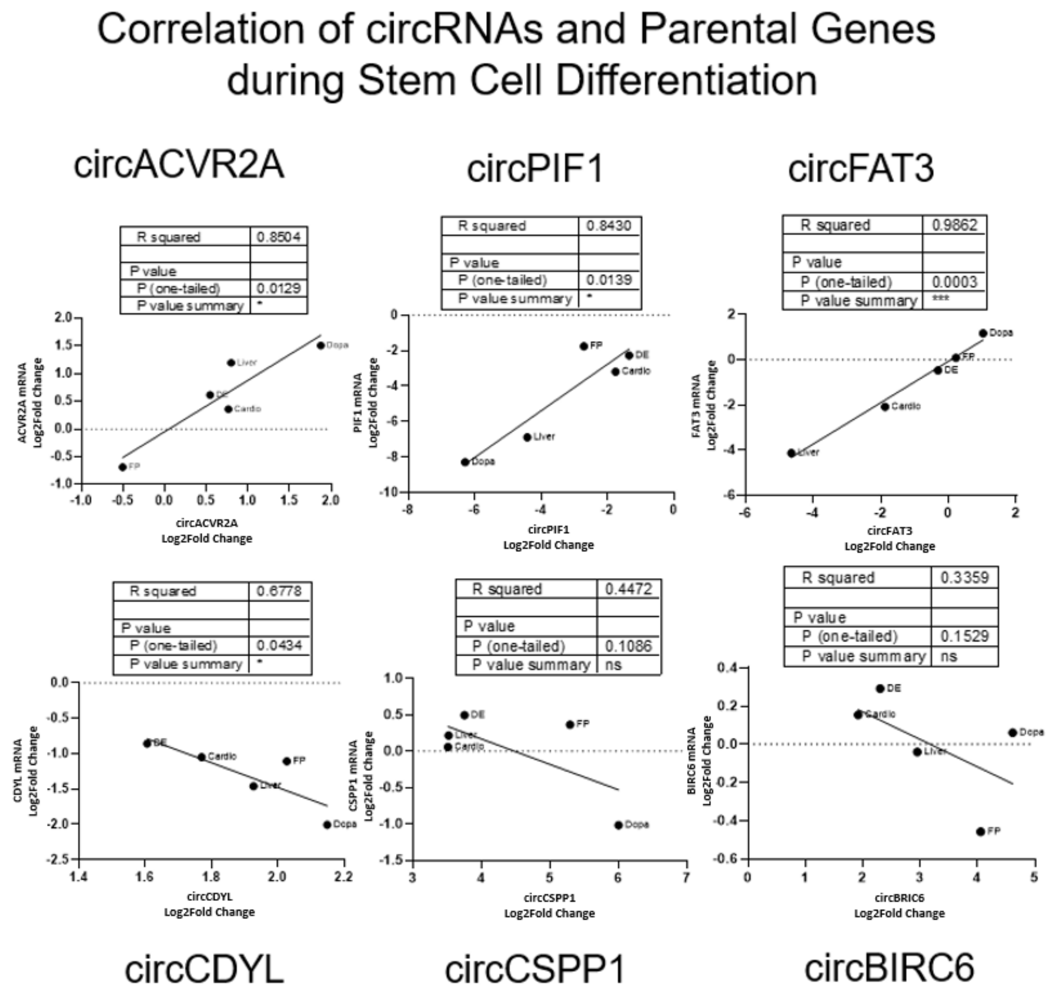


Figure 3-19 Correlation between expression of some candidate circRNAs and parental transcripts

One-tailed Pearson's correlation analysis of circRNA and parental gene log2fold changes of normalised TMM (circRNA seq) and RPKM (mRNA seq) in each lineage, detected by RNA-seq. DE=Definitive Endoderm. Cardio=Cardiomyocytes. Liver=Hepatocytes. FP=Floor Plate Precursors. DN=Dopaminergic Neurons.

Of the list of 90 commonly statistically differentially expressed circRNAs ($p<0.05$) (Table 3-4), parental mRNA expression changes accounted for some of the changes to circRNA levels (Figure 3-20, Figure 3-21), with positive correlations observed in all lineages and r-squared values between 0.47 and 0.62. This further demonstrates that parental transcript expression is a significant contributing factor for dictating circRNA abundance. Even so, individual circRNAs within this list can show unique increases or decreases in expression while no change, or the opposite direction of expression, is found for parental gene expression (Figure 3-20). An example of this is *circUSP10* that decreases during each differentiation, however, the parental transcript shows both an increase and decrease depending on the differentiated cell type (Figure 3-21).

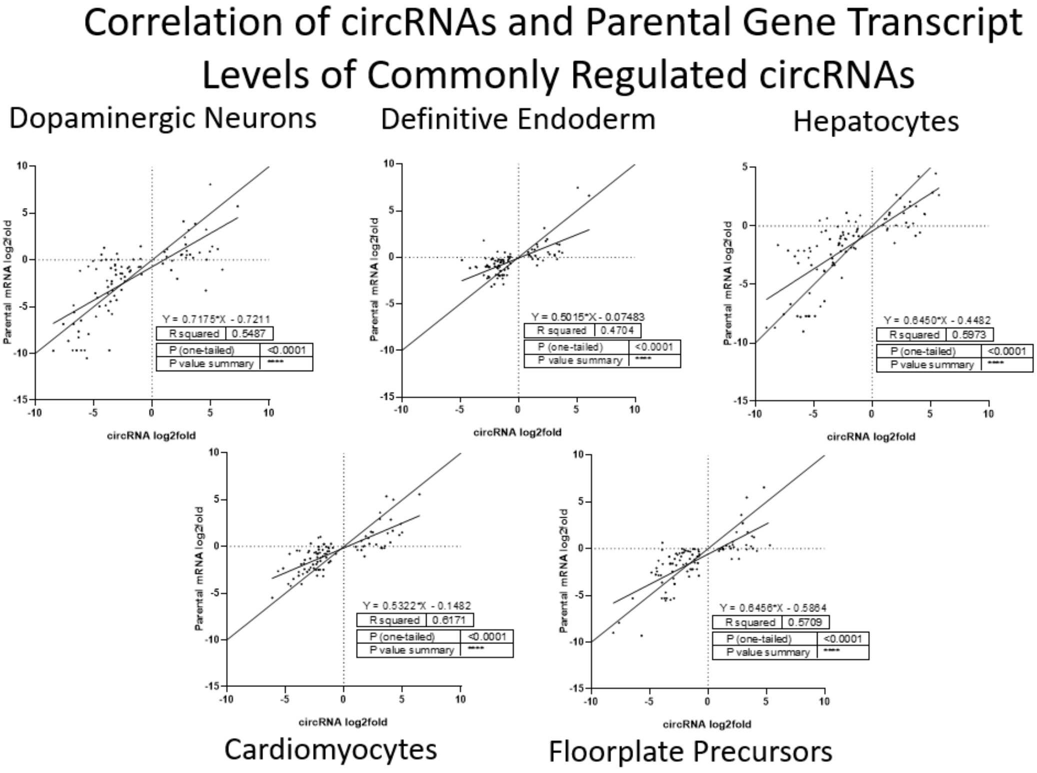


Figure 3-20 Correlation between circRNA and parental transcripts of commonly statistically differentially expressed circRNAs (p<0.05)

One-tailed Pearson's correlation analysis of the 90 commonly statistically differentially expressed (non-FDR adjusted p-value<0.05) circRNAs and parental gene log2fold changes in each lineage, detected by RNA-seq. Graphs include line of identity to observe the slope when x and y coordinates are equal.

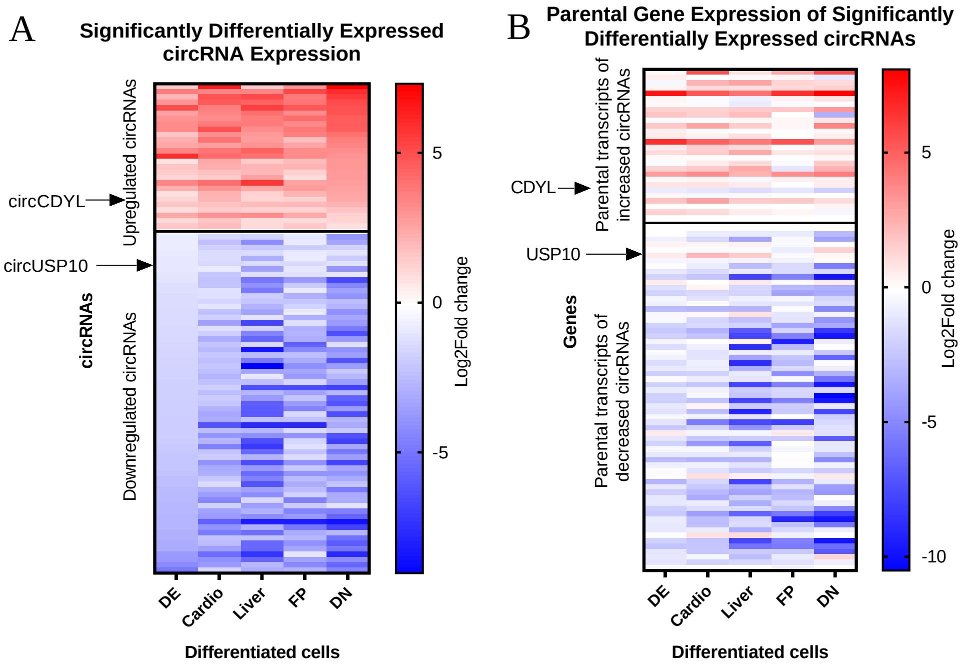


Figure 3-21 Heatmap of Log2Fold changes of 90 commonly statistically differentially expressed circRNAs ($p<0.05$)

Visual representation of (A) statistically significantly differentially expressed (non-FDR adjusted p-value<0.05) circRNA expression changes alongside (B) parental transcript changes during hESC differentiation into five different lineages.

The FDR-adjusted commonly statistically differentially expressed circRNAs also show a positive correlation with their parental gene in each lineage, except during DE differentiation (Figure 3-22). *CircCDYL* is removed from this analysis as it is the one outlier and only circRNA throughout the entire dataset that shows an inverse relationship with its parental gene during each differentiation event (Figure 3-23). These circRNA changes can be explained mainly by parental gene expression, with high R squared values observed (between 0.5 and 0.7) and significant grouping around a linear regression analysis. This gives further support for looking at non FDR-adjusted statistically differentially expressed circRNAs, so that circRNAs that show different regulatory characteristics can be investigated.

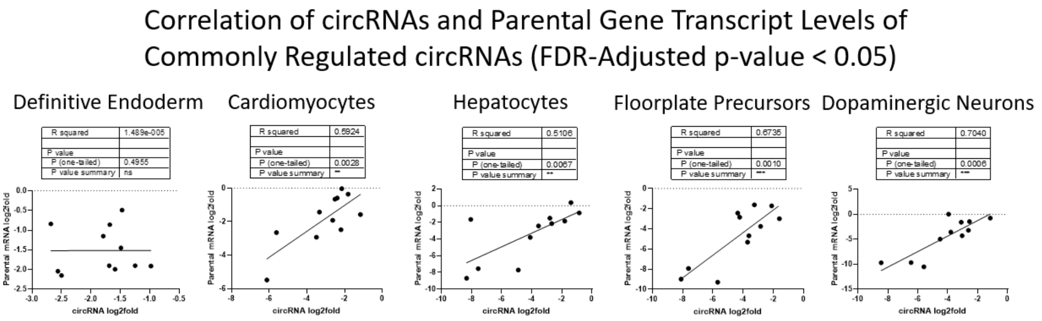


Figure 3-22 Correlation between circRNA and parental transcripts of commonly differentially expressed circRNAs (padj<0.05)

One-tailed Pearson's correlation analysis of commonly statistically differentially expressed (padj <0.05) circRNAs and parental gene log2fold changes in each lineage, detected by RNA-seq. Removed *circCDYL* from analysis.

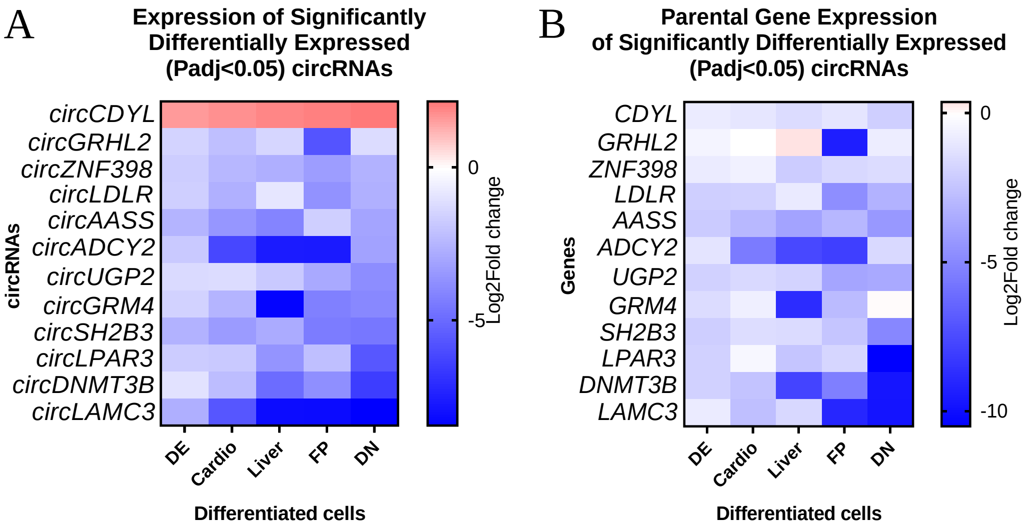


Figure 3-23 Heatmap of Log2Fold changes of 90 commonly differentially expressed circRNAs (padj<0.05)

Visual representation of (A) statistically significant differentially expressed (padj <0.05) circRNA expression changes alongside (B) parental transcript changes during hESC differentiation into five different lineages. Data is only given if the change from H9 hESCs is statistically significant (p<0.05)

To further understand the transcriptional changes to molecular determinants of circRNA expression, splicing factors and other proteins known to promote circRNA circularisation (Huang and Zhu, 2021) were investigated for transcriptional changes during hESC differentiation. Although an increase in circRNAs is observed in each cell type compared to undifferentiated hESCs, no increase to circularisation factor proteins was observed, and in fact, many factors were decreased (Figure 3-24).

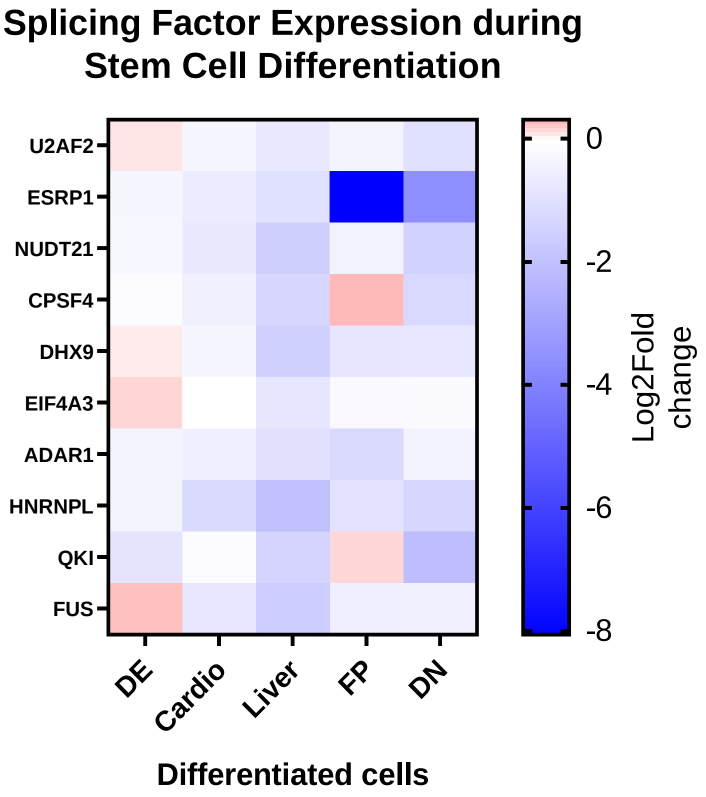


Figure 3-24 Changes in expression of RNA Binding Proteins implicated in circRNA biogenesis with circRNA abundance after differentiation

(A) Visual representation of circRNA circularisation proteins expression changes through differentiation into the five lineages determined by DEseq2.

3.2.7 Sequencing Validation

To confirm the existence of the candidate circRNAs, divergent primers were designed to amplify from cDNA containing circRNA transcripts (Figure 3-25A). CircRNAs were amplified from H9 hESC or HEK293T cDNA and Sanger sequenced to confirm the expected backsplice junction (Figure 3-25B).

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

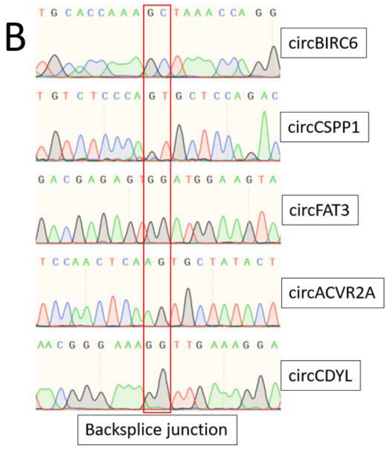


Figure 3-25 Design of qPCR primers and sequence confirmation of circRNA PCR amplificon

(A) Design of circRNA Divergent primers for PCR. (Image adapted from: (Shanmugapriya *et al.*, 2018)). (B) Chromatogram of successful detection of circRNA backsplice junctions (Boxed in red).

3.2.8 Selection of House-keeping Genes for qRT-PCR during Stem Cell Differentiation

Stem cell differentiation is a major cellular change and can affect cell properties including cytoskeletal organisation, cell-to-cell interactions and metabolism (Panina *et al.*, 2018). To combat transcriptional changes and to normalise data during qRT-PCR analysis, house-keeping genes that maintain steady levels within cells are required to normalise for input between different samples that are being tested, allowing for direct comparison of target RNA levels. Even the most widely utilised house-keeping genes can vary greatly depending on the model system. Determination of which gene is best for a given cell line or treatment is required before comparisons of RNA levels can be accurately and confidently accessed. Panina *et al.* (2018) analysed twelve common house-keeping genes during induced pluripotent stem cell reprogramming and found the most stable transcripts to be *ATP5F1*, *PGK1* and *GAPDH*.

To identify genes that can be used as references genes during qRT-PCR analysis of differentiating hESCs, the same common housekeeping genes were analysed within the mRNA-seq data. Analysis of the Fragments Per Kilobase of transcript per million mapped reads (RPKM) for each gene revealed *YWHAZ* as the most stable gene to use for qRT-PCR analysis (Figure 3-26) with the smallest Coefficient of Variation (CV).

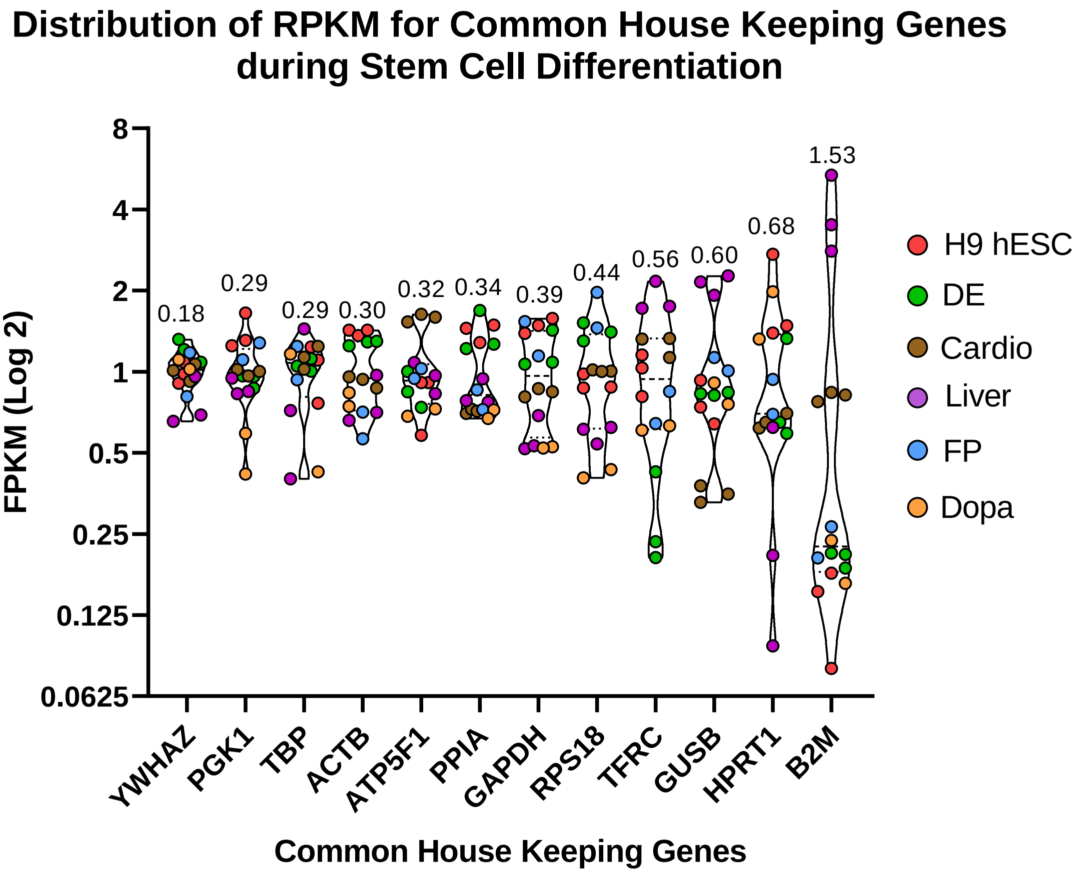


Figure 3-26 Changes to housekeeping genes during differentiation in RNA-seq data

Violin plots of the distribution of Common Housekeeping Genes during Stem Cell Differentiation RNA-seq data. Number above each plot is the Coefficient of Variation calculated by the Standard Deviation divide by the Average RPKM value.

Primers were designed to target all *YWHAZ* RNA variants by qRT-PCR as RPKM values are calculated based on all transcripts from the gene. Performing a qRT-PCR standard curve at 56 degrees celsius revealed an efficiency of 98% with an R squared value of 0.99914 (Figure 3-27). The amplicon was Sanger sequenced (Figure 3-28A) and analysed by Basic Local Alignment Search Tool (BLAST) to confirm the sequence that was amplified are the *YWHAZ* transcripts (Figure 3-28B).

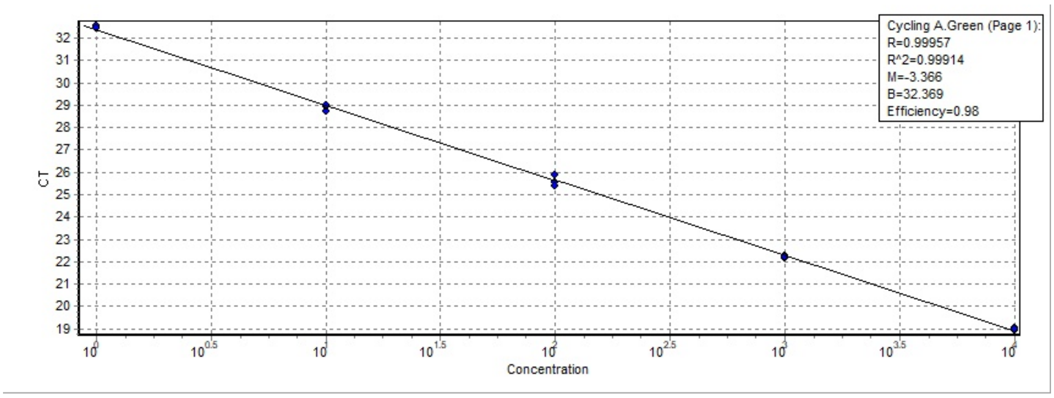
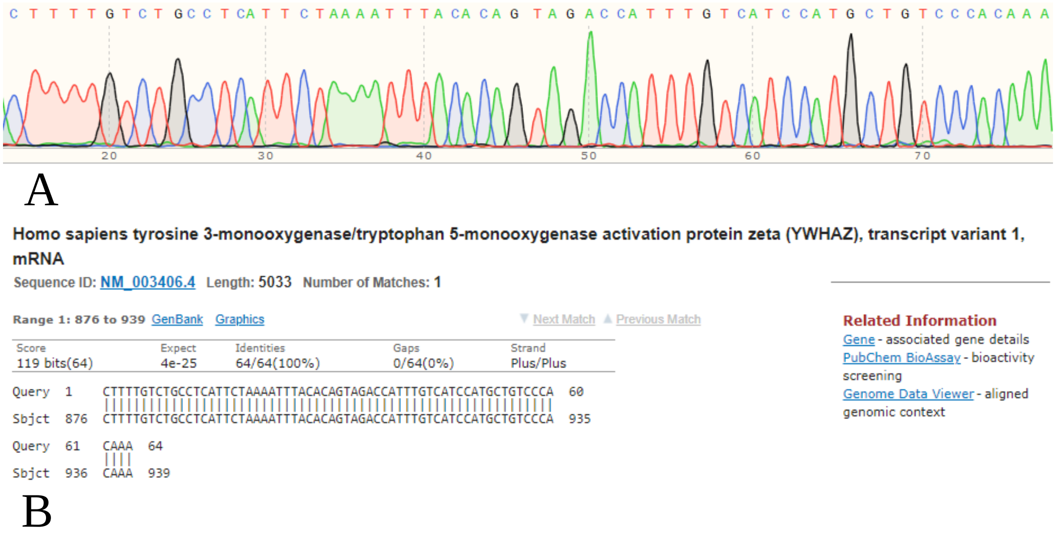


Figure 3-27 *YWHAZ* standard curve

YWHAZ House Keeping Gene qPCR standard curve by qPCR at 56 degrees Celsius with SYBR 2X mix.


CTTTTGTCTGCCTCATTCTAAAATTTACACAGTAGACCATTTGTCATCCATGCTGTCCCACAAA
20
30
40
50
60
70
A
B

Homo sapiens tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta (YWHAZ), transcript variant 1, mRNA

Sequence ID: NM_003406.4 Length: 5033 Number of Matches: 1

Range 1: 876 to 939 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
119 bits(64)	4e-25	64/64(100%)	0/64(0%)	Plus/Plus

```
Query  1    CTTTTGTCTGCCTCATTCTAAAATTTACACAGTAGACCATTTGTCATCCATGCTGTCCCA  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  876  CTTTTGTCTGCCTCATTCTAAAATTTACACAGTAGACCATTTGTCATCCATGCTGTCCCA  935

Query  61   CAAA  64
            ||||
Sbjct  936  CAAA  939
```

Related Information

Gene - associated gene details

PubChem BioAssay - bioactivity screening

Genome Data Viewer - aligned genomic context

Figure 3-28 Sequencing Validation of *YWHAZ* amplicon

(A) Sanger Sequencing and (B) blast alignment (B) of *YWHAZ* primer pair PCR amplicon.

3.2.9 Validation of lineage-specific circRNAs in circRNA-seq data

Insufficient RNA remained to generate cDNA to validate each cell lineage and replicate of the circRNA-seq data for all candidate circRNAs. However, in support of these findings, circRNAs from the circRNA-seq data were found in the literature to be upregulated in specific cell lines. *CircCACNA1D* is found to be upregulated in Cardiomyocytes (Lei *et al.*, 2018), *circRIMS1* in the brain (Xia *et al.*, 2017) and *circGSE1* in the liver (Xu *et al.*, 2017). In validation of the circRNA-seq dataset, these circRNAs were found to be statistically and specifically upregulated after the differentiation into these cell types respectively (Figure 3-29). DE and FP are precursor cells, and due to the study of circRNAs being novel, compared to other non-coding RNAs, this report is the first to perform circRNA expression profiling in these cell lines.

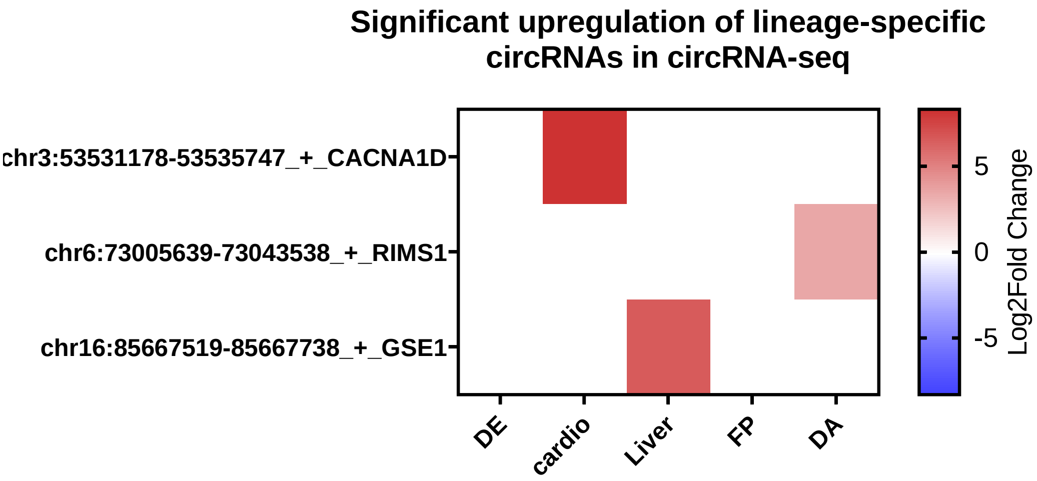


Figure 3-29 Validation of lineage-specific circRNAs in the circRNA-seq data

Visual representation of statistically significant differentially expressed (padj <0.05) Data is only given if the change from H9 hESCs is statistically significant (padj<0.05).

3.3 Discussion

RNA-seq is a powerful tool that, when sequenced at a sufficient depth, can quantify changes to most expressed mRNA transcripts, inform on splicing changes and also identify novel non-poly-A RNA transcripts. Although enrichment for circRNAs by RNase R has significantly improved the identification of circRNAs, the ability to define expression changes is still restricted to identifying changes to the number of reads that span over the backsplice junction of an individual circRNA, as some circRNAs share the same sequences with others from the same gene. This major limitation, along with the large number of detected circRNAs, hinders the statistical power to determine significant differential expression.

If a circRNA is to play a functional role in activating or repressing differentiation and/or EMT, it would require at least one interacting partner, being RNA, protein, or DNA. Therefore, large transcriptional changes to circRNAs may not be the ideal measure for prioritising circRNA candidates. Furthermore, these expression changes may be impacted by parental mRNA expression and/or by its transport, or the role individual circRNAs play in other processes that occur during differentiation. Using only FDR-adjusted p-values has grouped together commonly differentially expressed circRNAs that are strongly correlated with their parental gene expression, excluding circRNAs with different regulatory mechanisms during differentiation. Therefore, using non-adjusted p-values was chosen and has highlighted that circRNA expression, although correlated with parental gene expression, is not solely dependent on parental gene expression as circRNAs are regulated by other factors (e.g. QKI).

Each unique cell-specific differentiation may also have undergone phenotypically similar, yet transcriptionally unique, EMT events. Many EMT-related circRNAs may not be captured if all differentiation events are thought to use the exact same and complete EMT pathway. In fact, these cells undergo waves of EMT and MET, most likely using a transcriptionally unique EMT during each wave leading to the generation of specific cell types. This also supports the use of less stringent statistical analysis, as EMT is not an all or nothing process, and circRNA expression may be unique during differentiation into distinct cell types.

Analysis of circRNA-seq for hESCs, and resulting differentiated cells, has revealed many circRNAs that are commonly and uniquely expressed in all lineages. CircRNA expression alone is able to stratify the cell types as effectively as using mRNA seq data, supporting that circRNA regulation is not stochastic but highly regulated. hESCs were found to have the lowest number of circRNAs detected whereas Dopa and FP cells were found to have the most circRNAs detected, with Dopa cells having the most differentially expressed circRNAs from H9 hESCs. This is not surprising as others have also found that circRNAs are upregulated during neuronal differentiation specifically, in primary neuron culture (Rybak-Wolf *et al.*, 2015; D'Ambra, Capauto and Morlando, 2019) and also in the mammalian brain (Sekar and Liang, 2019). Interestingly, circRNAs, and not their parental gene mRNA transcripts, are highly enriched in synapses, with some groups suggesting that circRNAs might act as scaffolds for RNAs or proteins (D'Ambra, Capauto and Morlando, 2019).

Several candidate circRNAs that are commonly differentially expressed were discovered ($padj<0.05$) and these circRNAs may be related to EMT occurring during differentiation. Previously published work by our lab investigated circRNA expression during TGFβ1-induced EMT in HMLE cells (Conn *et al.*, 2015). Only 8 of the commonly downregulated circRNAs (padj<0.05) from the hESC data, were also identified in HMLE cells, and 3 of these were shown to have reduced expression during TGFβ1-induced EMT. This includes *circGRHL2*, *circZNF398* and *circSH2B3*. This suggests that EMT regulated circRNAs are also regulated during EMT throughout differentiation into multiple cell types.

The addition of mRNA-sequencing has informed on the regulation of these circRNAs, identifying circRNAs whose expression correlates with their parental gene transcripts. Unfortunately, 3 mRNA-seq samples did not match with the other replicates of the same cell types, when looking at lineage markers, and bioinformatically using a PCA plot, and were excluded from other analysis. This was likely due to a low RNA input after circRNA-seq library preparation, therefore, more amplification cycles were required during mRNA-seq library preparation, resulting in low concentration libraries and disproportionate amplification for these replicates. Additionally, differentiation could have been more variable for these replicates, however, these individual libraries were reamplified as the concentration after library preparation was low. Furthermore, the replicates in the circRNA-seq MDS plot, generated from the same RNA as the mRNA-seq libraries, shows close grouping, therefore, variable differentiation was likely not the cause of the grouping inconsistencies of the mRNA-seq library replicates.

To identify a correlation between parental gene mRNAs and circRNA expression, whole mRNA reads were used to calculate Log2Fold changes instead of individual mRNA isoform reads or reads adjacent to back splice junctions. Many mRNA isoforms and non-coding sequences are comprised of the same sequences from a gene and so a using a particular sequence adjacent to the circRNA would remove some, but not all of the isoforms generated from the parental gene. Therefore, all parental gene mRNA reads were used for each circRNA-mRNA correlation, so that a consistent approach is utilised. However, this approach will not identify if certain isoforms have a negative correlation with circRNAs but will only identify if circRNA levels are different from the overall gene transcript changes.

CircFAT3 expression correlates strongly with its parent gene during differentiation, while, in contrast, *circCDYL* has an inverse correlation with its parental gene. Furthermore, *circACVR2A* has shown both up and downregulation of its expression during different differentiation events while still being correlated with its parental mRNA. Overall, a weak to moderate positive correlation was observed between circRNA and mRNA expression during each cell differentiation. This data highlights that circRNA regulation during stem cell differentiation may be controlled by similar factors, including positive or negative co-regulation with their parental gene. This model only partially explains the variance between circRNA and mRNA expression as independent and cell lineage-specific factors, such as alternative splicing, may also regulate circRNA expression, in situations when the parental mRNA may be affected to a lesser degree or not at all. Whether looking at abundant statistically differentially expressed circRNAs, or just abundant circRNAs, a moderate correlation is observed and highlights that transcription of parental genes is a major driver of circRNA abundance. Interestingly, the slope of most of the correlation analyses above were between 0.3 to 0.7 indicating that circRNAs do not increase or decrease as much than their parental gene, highlighting that although parental gene expression is a strong factor for circRNA expression, it is not the only factor that can affect circRNA levels.

Additionally, transcriptional profiling of known circRNA splicing/circularisation factors reveals a decrease for these RNA transcripts, yet the number of unique circRNAs increases during differentiation. This illustrates that an increase in circRNA levels is not due to a general increase to circRNA splicing/circularisation factor levels, however these factors may require recruitment to RNA before effective circularisation can be performed. An example of this is the recruitment of the spliceosome protein U2 snRNP by the factor U2AF65 which binds to the polypyrimidine tract adjacent to a 3' splice site (Black, 2003). This further supports that parental gene transcription is a major regulation factor for circRNA biogenesis during stem cell differentiation.

As our understanding of circRNA biogenesis is still expanding, more factors that promote circRNA formation, or the destruction of circRNA transcripts, may be uncovered and explain for the regulatory mechanisms that drive apparent cell type--specific circRNA regulation. Interestingly, each cell type did have uniquely regulated circRNAs. These circRNAs may be involved with cell type-specific functions or possibly could drive lineage specificity during differentiation, however, a exhaustive catalogue of functions that circRNAs can perform are still yet to be elicited.

Examining only the 90 commonly significantly regulated circRNAs from all cell types, correlation analysis confirms that parental gene expression is a strong determinant of circRNA expression in this group of circRNAs. Looking at individual circRNAs within this list gives insight into the cell type-specific regulation of circRNA expression as parental transcripts can show opposite expression changes. This list gives a good number of circRNAs that may prove to be involved with EMT or differentiation processes.

Differential expression changes of a circRNA transcript, with no change to the parental transcript, as seen for *circCSPP1* and *circBIRC6*, could be explained by a circRNA being produced as a by-product of alternative splicing of the parental transcript giving rise to alternative mRNA isoforms. Both *CSPP1* and *BIRC6* have multiple mRNA isoforms, however, whether these isoforms change in abundance was not determined. Alternatively, for genes that have no recorded splice variants, a circRNA may be independently regulated whereby splicing to produce different alternative parental isoforms isn't a consequence of circRNA formation as resulting mRNA may be degraded after splicing. Either mechanism may not show any significant changes to the total read count for the parental gene, while changes to a circRNA transcript may be observed. A theory of increased back-splicing, and not other forms of splicing, may allow for higher circRNA levels regardless of an increase or decrease of parental mRNA transcription. As circRNAs generally have low copy numbers per cell, an increase of back splicing could have a significant impact on the Log2Fold change of a circRNA with a minimal effect on parental mRNA levels. Both independent regulation and co-regulation with parental mRNA, gives insight into the nuances of cellular control of circRNA levels.

To broaden the potential for identifying circRNAs with an ability to affect EMT or differentiation, different circRNA properties were chosen (cell type-specific, commonly regulated, dysregulation to parental gene etc.). As there may be many shared cellular programs during stem cell differentiation, including EMT and cell cycle regulation, circRNAs that can impact a variety of cellular processes may be discovered. Identification of circRNA as effectors of EMT or differentiation from these candidates will inform on the diversity of regulatory mechanisms that control these enigmatic RNA molecules.

3.4 Chapter 3 Conclusion

Profiling circRNAs across human stem cell differentiation has revealed that the majority of circRNAs are co-regulated with their parental gene expression. Individual circRNAs can show independent regulation and inverse correlation with their parental gene transcripts and these circRNAs are of great interest as they are uniquely regulated during differentiation.

Circ-RNA seq analysis has identified unique characteristics of circRNAs as overlapping mRNA isoforms are removed during library preparation before RNA sequencing. Analysis of circRNAs contained within this data will give insight into the functional relationships of circRNAs and how they may affect EMT and differentiation.

Chapter 4 Stable Knockdown of Circular RNAs in Human Embryonic Stem Cells

4.1 Introduction

The bulk of circRNA publications focus on the ability of circRNAs to bind to and sponge miRNAs through complementary base-pairing (Zheng *et al.*, 2016; Sun *et al.*, 2017; Song and Li, 2018). These conclusions of a circRNA-miRNA network are largely drawn by overexpression of circRNAs and the analysis of downstream mRNAs that are a target by putative miRNAs (W. Chen *et al.*, 2020). Although this binding may occur, overexpression of a normally lowly abundant circRNAs, far above normal cellular levels, may not accurately recapitulate the function of circRNAs within cells and give misleading evidence for effective sponging of miRNAs whose abundance is orders of magnitude higher than the circRNA.

To begin investigating functional roles for possible EMT-related circRNAs, a simple approach that perturbs the level of these molecules and assaying a simple phenotypic response was chosen (i.e. ability to undergo EMT/differentiation, increase in migration or invasion, loss of epithelial traits). Decreasing the abundance of individual circRNAs was chosen as the best initial approach over circRNA overexpression, due to the simplicity, as consequences of expressing the circRNA at far above normal biological levels, lead to an increased chance of non-canonical localisation or off-target circRNA functions. Introducing an siRNA or shRNA molecule does introduce potential for off-target binding and silencing by RISC mediated degradation, however, this is addressed by utilising multiple individual molecules that target the circRNA backsplice junction, increasing confidence if consistent phenotypic responses are observed. Stable integration of either shRNAs or CRISPR Cas13b, along with a gRNA, is the preferred method to generate cell lines that do not require multiple transfections to continue reducing

RNA levels. This is especially important in hESCs as stem cells undergoing EMT and differentiation may show altering phenotypes depending if targeted RNAs are reduced only within a 72 hour window or constitutively. Differentiation protocols into certain lineages can take weeks to perform and so a sustained KD is preferable, if probing for altered differentiation profiles is to be completed.

To this end, attempts to probe circRNA functions were undertaken by knocking down circRNAs with CRISPR Cas13b, rather than overexpressing circRNAs. If a phenotypic response was seen from multiple gRNAs, this would increase confidence in this result arising from direct KD of the circRNA rather than potential off-target effects. As RNAi or CRISPR Cas13b target the backsplice region of the circRNA, there is potential that the RNA sequence that is also present in the parental mRNA transcript could be targeted and should be identified (Zhang *et al.*, 2021). Furthermore, production of a circRNA and mRNA transcript from the same gene may suggest that they have synonymous or opposing functions (Ashwal-Fluss *et al.*, 2014), as individual circRNAs may be co-regulated with its parental gene, therefore any off-target effects on the parental gene should be determined.

4.2 Results

4.2.1 Design of gRNAs targeting circRNA backsplice hESCs

To design guide RNAs for CRISPR Cas13b KD of *circACVR2A*, identified from circRNA-seq in chapter 3, 3 independent oligonucleotides were ordered containing the required sequences and 30 nucleotides that give Cas13b specificity for *circACVR2A* (NX30). These oligonucleotides were directly cloned into the pLKO.1 vector by Gibson cloning as outlined in Materials and Methods.

Lentivirus was generated from the pLKO.1 plasmids, and also the Cas13b plasmid, by co-transfecting psPAX2 and pCMV-VSV-G, into HEK293FT cells and collecting the supernatant. H9 hESCs were transduced with lentiviral supernatant of both Cas13b and pLKO.1 lentivirus to generate cells expressing Cas13b and the required gRNA. Successful integration of the lentiviral cassette (inclusive of the DNA sequence between the 5’ and 3’ viral long terminal repeats) randomly into the target cells genome was selected for by the use of puromycin resistance (pLKO.1) and blasticidin resistance (pLX311). After successful integration of the lentiviral cassette, antibiotic resistance is conferred to the hESCs. After selection of successfully transduced cells using puromycin and blasticidin, RNA was harvested, and qRT-PCR was performed to ascertain the efficiency of *circACVR2A* KD with Cas13b in hESCs. Although H9 hESCs were successfully transduced with both viruses and survive antibiotic selection, no significant KD was achieved (Figure 4-1). CRISPR Cas13b was not further employed for KD of circRNAs in H9 hESCs and shRNAs were designed to target circRNAs.

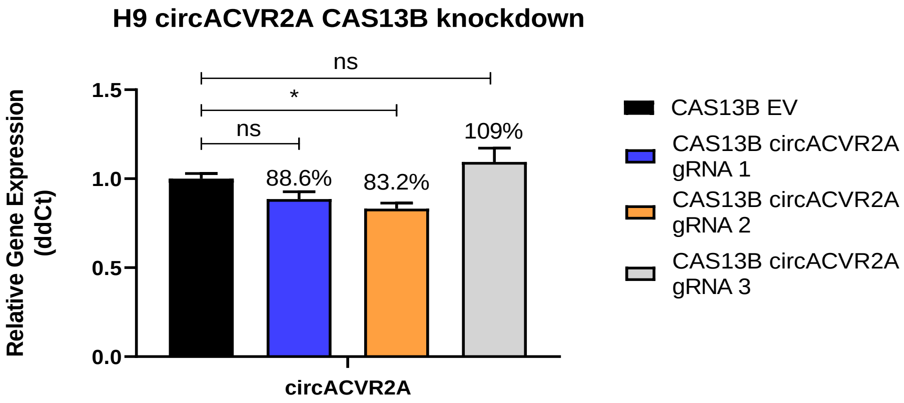


Figure 4-1 Knockdown of *circACVR2A* by CRISPR Cas13b is not effective in H9 hESCs

qRT-PCR validation of *circACVR2A* KD in H9 hESCs. Relative to *YWHAZ* expression. Statistical analysis was performed with an Ordinary one-way ANOVA with Fisher’s LSD test. (**** = $p<0.0001$) (n=3; Mean+SEM)

4.2.2 Design of shRNAs targeting circRNA backsplice junction

To evaluate whether shRNAs are appropriate in hESCs for reducing circRNAs, 2 independent shRNAs were designed that bind to the unique backsplice junctions of circRNAs and should silence the transcript by RISC-mediated RNA cleavage. As the shRNA overlaps the backsplice junction, the shRNAs were not completely complementary to the parental mRNA, and therefore were specific only for the circRNA. The DNA sequence, and the corresponding antisense sequence, were used to design oligos (Figure 4-2) that could be annealed together and directly ligated into the pLKO.1 vector as outlined in Materials and Methods.

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

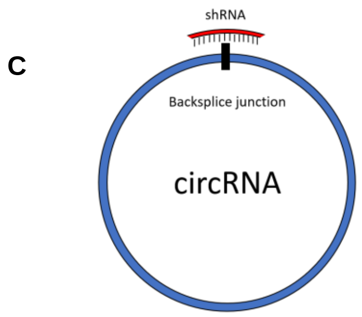


Figure 4-2 shRNA design for targeting circRNA backsplice junction

(A) Design of shRNA oligo including restriction enzyme sites (AgeI and EcoRI) for cloning into pLKO.1 shRNA vector (B) The shRNA construct is expressed by Pol III under the U6 promotor to generate the hairpin shRNA molecule (*Addgene: Protocol - pLKO.1 – TRC Cloning Vector*, no date). (C) The shRNA

The pLKO.1 vector is a third generation lentiviral vector, and is co-transfected with psPAX2 and pCMV-VSV-G, into HEK293FT cells to generate lentivirus particles. The cell culture supernatant was collected and used to infect hESCs. A pLKO.1 empty vector (EV) and pLKO.1 vector expressing a non-targeting shRNA (scrambled) were generated to use as controls for experiments. Several scrambled shRNAs that were generated were found to alter H9 hESC cell morphology, while the EV control did not induce a phenotype as hESCs maintained their morphology. Eventually, a scrambled shRNA sequence was generated that did not illicit any adverse affects, the same as the EV control. For this reason, the inclusion of the scrambled shRNA for qRT-PCR analysis was only included in the quantification of one shRNA KD, *circACVR2A,* in H9 hESCs, as this eventually became a focus for further functional profiling. Two independent shRNAs were designed for most circRNA candidates to increase confidence that any observed phenotype is due to the circRNA KD. After lentiviral transduction and selection of puromycin resistant cells, qRT-PCR was performed to quantify circRNA KDs in the stable KD H9 hESCs.

4.2.3 Stable knockdown of circRNAs in H9 hESCs by targeting backsplice junction

Thirteen of the twenty seven circRNA candidates, identified from circRNA-seq across hESC differentiation, were KD in H9 hESCs with lentivirus harbouring the circRNA-specific shRNA cassette. Remarkably, five out of the thirteen individual circRNA KDs, from initially screening with only one designed shRNA, induced significant morphological changes to hESCs (Table 4-1). These affects fall into the the categories of embryoid body (EB) production. loss of compact colonies and altered cell morpholgy. For circRNAs where a phenotypic effect was observed, targeting the circRNA with the second individual shRNA was attempted to confirm the phenotype.

Table 4-1 Candidate circRNAs with cellular affect after knockdown

CircRNA targets and effects on hESCs after shRNA transduction and puromycin selection.

CircRNA target	Cellular effect	No. of shRNAs
circACVR2A	Embryoid body production	2 (2 confirmed KDs)
circPIF1	Embryoid body production	2 (2 confirmed KDs)
circCDYL	Loss of compact colonies	2 (2 confirmed KDs)
circFAT3	Loss of compact colonies	1 (1 confirmed KD)
circZNF609	Altered cell morphology	1 (1 confirmed KD)
circCDK13	No observable effect	1 (no confirmed KD)
circROR2	No observable effect	1 (no confirmed KD)
circSEMA3E	No observable effect	1 (no confirmed KD)
circMYBPC3	No observable effect	1 (no confirmed KD)
circRIMS1	No observable effect	1 (no confirmed KD)
circSETBP1	No observable effect	1 (no confirmed KD)
circSP100	No observable effect	1 (no confirmed KD)
circADAMTSL5	No observable effect	1 (no confirmed KD)

RNAi is a common method for the reduction of RNA transcripts in mammalian cells and the use of correct controls permits accurate interpretation of reduction to RNA levels. Alongside the use of shRNAs, an empty vector control mimics the transduction process that cells undergo and also imparts cells with the same plasmid and antibiotic selection that an shRNA producing plasmid contains. However, this plasmid does not activate the RNAi pathway. The use of a scrambled or non-targeting shRNA molecule will produce a RNA molecule that can associate with RISC, however does not target any human gene. To this end, an empty vector (EV) and non-targeting control (scrambled) is used and both were shown not to affect hESCs growing on vitronectin ECM, as colonies were dense and phase bright edges were observed (Figure 4-3).

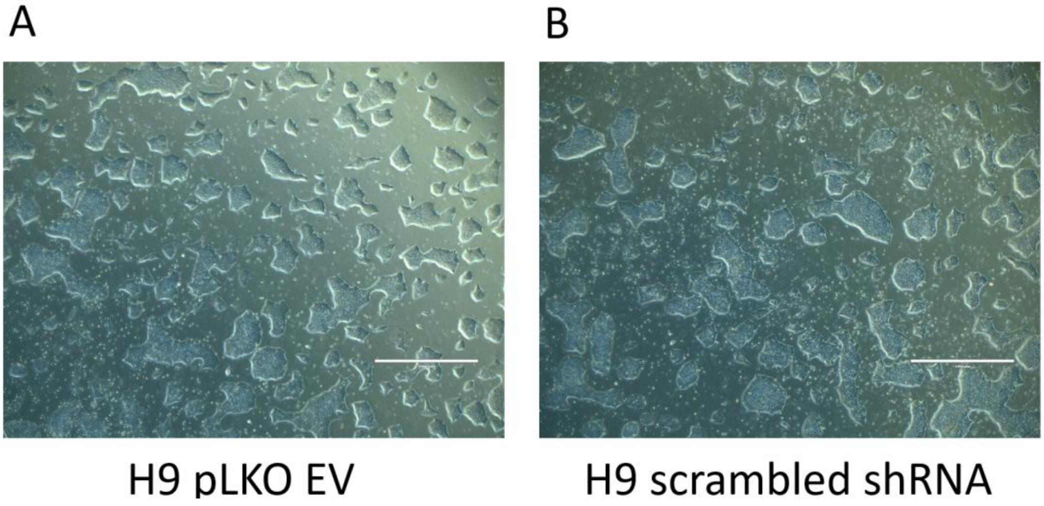


Figure 4-3 Empty Vector and scrambled control shRNA plasmids show no altered phenotype in H9 hESCs

Images of H9 hESCs transduced with empty vector and scrambled shRNA pLKO.1 generated lentivirus. (A) H9 hESC pLKO.1 empty vector control. (B) H9 hESC pLKO.1 scrambled control. Images were taken using the EVOS™ Imaging system. Scale bar 1000 µm.

4.2.4 Knockdown of *circFAT3*

Significant KD of *circFAT3* (95.4% KD) in H9 hESCs (Figure 4-4B), cultured on vitronectin, affected the density of colonies and defined phase bright edges were lost. Individual stem cells could be distinguished within colonies as well as cells growing separately from a colony (Figure 4-4A). Although a phenotype was observed for *circFAT3*, a second shRNA was not tested to confirm this phenotype.

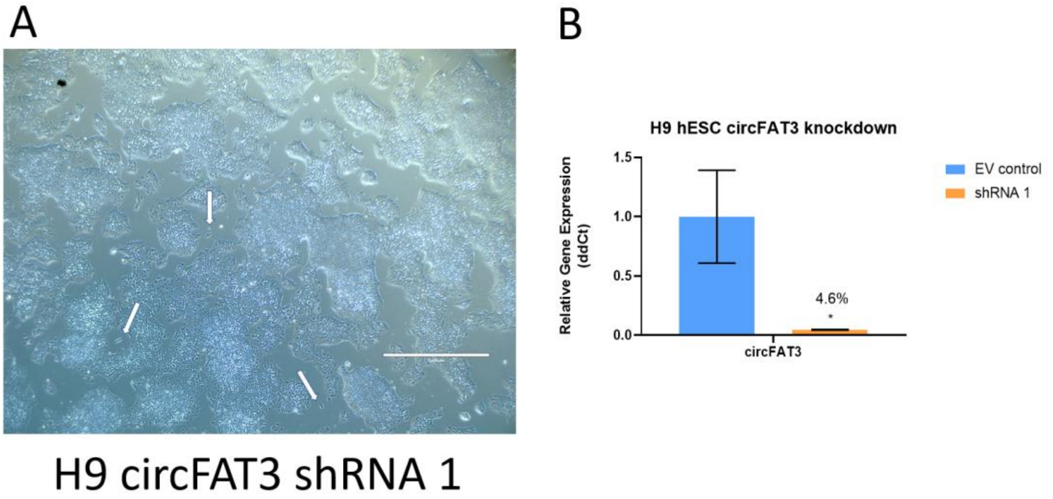


Figure 4-4 Reduction of *circFAT3* by shRNA induces the loss of compact colonies in H9 hESCs

Image of H9 hESCs transduced with an shRNA targeting *circFAT3*. (A) pLKO.1 *circFAT3* shRNA 1 H9 hESCs. White arrows mark cells that are growing separately from colonies. (B) qRT-PCR validation of *circFAT3* KD in H9 hESCs. Relative to *YWHAZ* expression. Statistical analysis was performed with an unpaired one-tailed t test. (* = $p<0.05$) (n=3; Mean+SEM) Images were taken using the EVOS™ Imaging system. Scale bar 1000 µm.

4.2.5 Knockdown of *circCDYL*

The reduction of *circCDYL* abundance was performed with 2 independent shRNAs (shRNA 1 = 53.9%, shRNA 2 = 83.4% *circACVR2A* KD) in H9 hESCs (Figure 4-5C). Similarly to *circFAT3* KD, the reduction of *circCDYL* affected the density of hESC colonies and defined phase bright edges were lost for some colonies (Figure 4-5A and B). Furthermore, individual stem cells growing apart from the colonies were observed. As there is a possibility that shRNAs can target the mRNA, *CDYL* mRNA was also quantified. PCR primers were designed to detect all alternatively spliced isoforms of *CDYL* as these primers would amplify from within the last exon and 3’ untranslated region (UTR), common to all mRNA variants. QRT-PCR analysis confirmed that the additive abundance of all *CDYL* mRNA transcripts was not affected by *circCDYL* KD (Figure 4-5D).

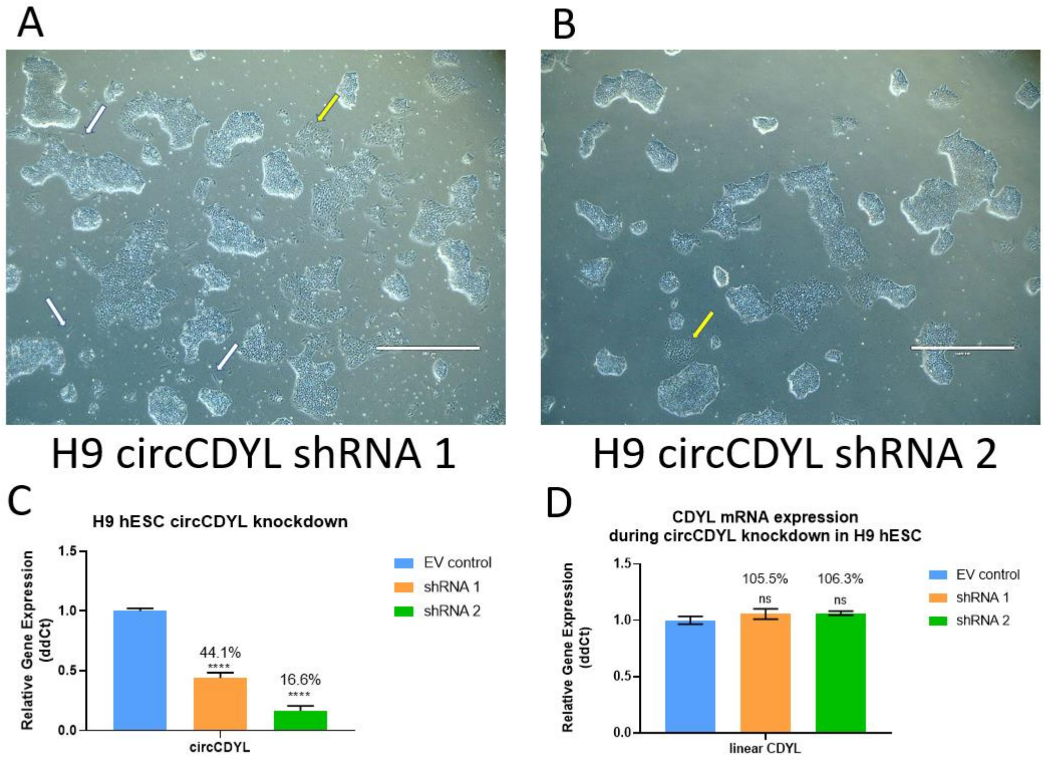


Figure 4-5 Reduction of *circCDYL* by shRNA induces the loss of compact colonies in H9 hESCs

Image of H9 hESCs transduced with shRNAs targeting *circCDYL*. (A) pLKO.1 *circCDYL* shRNA 1 H9 hESCs. (B) pLKO.1 *circCDYL* shRNA 2 H9 hESCs. White arrows mark cells that are growing separately from colonies. Yellow arrows mark colonies that have lost phase bright edges. (C) qRT-PCR validation of *circCDYL* KD in H9 hESCs. (D) qRT-PCR comprising all *CDYL* mRNA variants during *circCDYL* KD in H9 hESCs. Relative to *YWHAZ* expression. Statistical analysis was performed with an Ordinary one-way ANOVA with Fisher's LSD test. (**** = p<0.0001) (n=3; Mean+SEM) Images were taken using the EVOS™ Imaging system. Scale bar 1000 µm.

4.2.5.1 CircCDYL drives alternative splicing of CDYL

Chromodomain Y like (CDYL) is a transcriptional repressor that regulates histone lysine crotonylation and has multiple alternatively spliced mRNA variants (Liu *et al.*, 2017) (Figure 4-6A). Isoform 2a is the most abundant isoform and has a functional chromodomain, while isoform 1, the longest isoform, has a N-terminal extension that inactivates the chromodomain. Isoform 3 does not contain a chromodomain and has a downstream protein start site. Transcription of isoform 3 starts from exon 4, the same exon as isoform 2a, however an exon skipping splicing event excludes exon 5, which intriguingly, is the exon that *circCDYL* arises via backsplicing (Figure 4-6A). Evidence suggests that *CDYL* isoform 3 is able to displace *CDYL* isoform 2a from heterochromatin by forming a multimer with isoform 2a (Franz *et al.*, 2009).

Although no changes to *CDYL* mRNA was observed using primers that capture all spliced variants, this primer pair does not distinguish if reduction of *circCDYL* is able to affect alternative splicing of its parental gene. To identify if this occured, multiple primer pairs were designed to amplify the individual isoforms of *CDYL*, however only isoform 2a and 3 transcripts were detected. To give the best chance of seeing any splicing changes, *circCDYL* was also overexpressed in H9 hESCs using published protocols (see materials and methods).

KD of *circCDYL* in H9 hESCs with 1 of the 2 shRNAs was found to reduce *CDYL* isoform 3 by 20.3% (Figure 4-6B). Overexpression of *circCDYL*, over 2.5 fold above normal cellular levels (Figure 4-6C), was found to increase *CDYL* isoform 3 by 3.5 fold, while also reducing isoform 2a and all *CDYL* variants overall (Figure 4-6D). Therefore, any alteration to *circCDYL* abundance also changes the alternative splicing of *CDYL* transcripts.

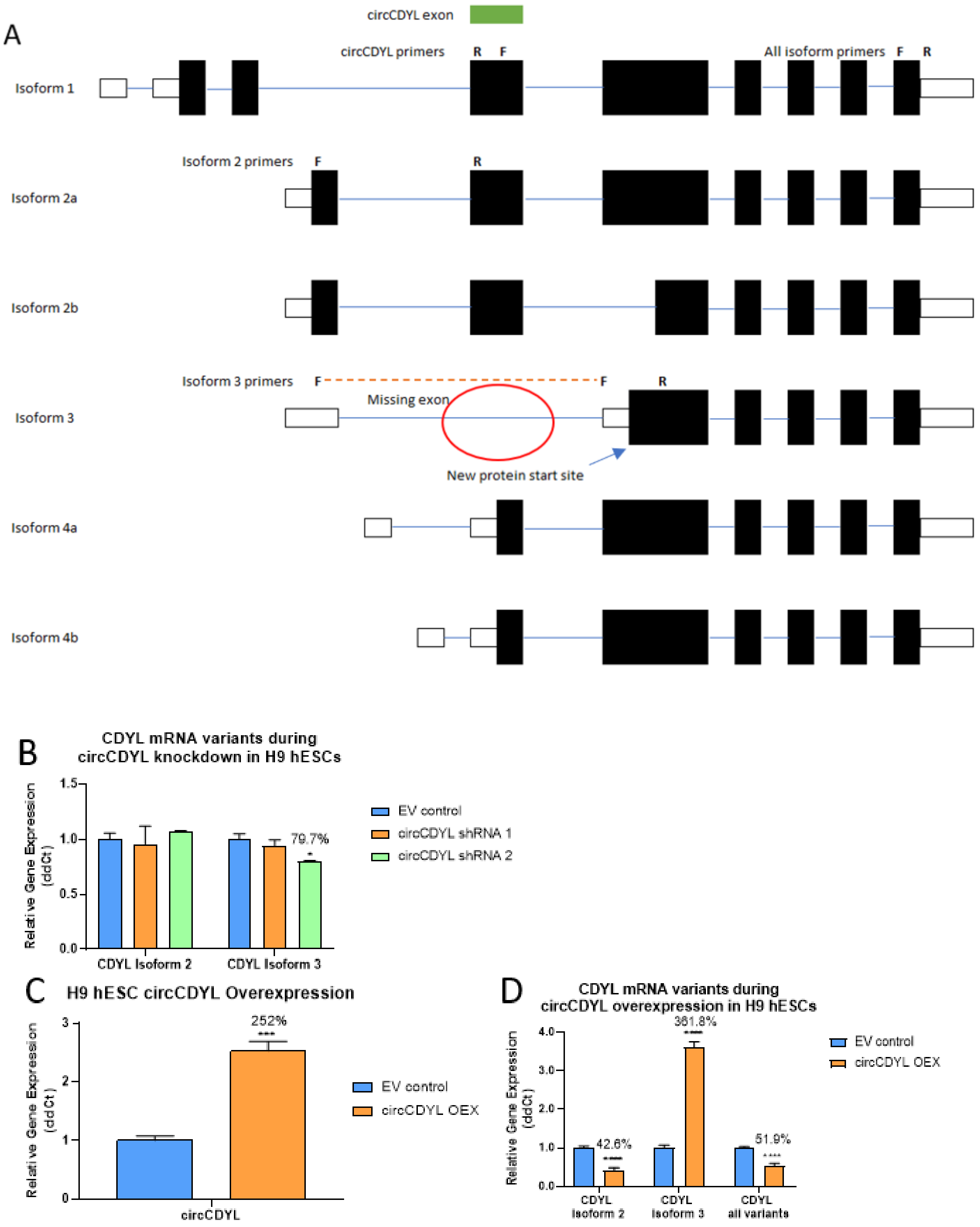


Figure 4-6 Changing the levels of *circCDYL* abundance alters the splicing *CDYL* mRNA isoforms in H9 hESCs.

(A) Representation of *CDYL* mRNA isoforms. *CircCDYL* is generated from the backsplicing of exon 5, the same exon that is spliced out from *CDYL* isoform 3. Divergent primers were designed to detect circCDYL. Isoform-specific primers were designed to detect the isoforms of interest, and primers was also designed to detect all CDYL isoforms. The black boxes represent the coding exons, the white boxes represent the non-coding untranslated regions of the isoforms, the blue lines represent introns, and the dotted orange line represents that the isoform 3 forward primer is one primer that spans the first intron of isoform 3. (B) qRT-PCR of *CDYL* isoform 2 and 3 during *circCDYL* KD in H9 hESCs. (C) qRT-PCR validation of *circCDYL* overexpression in H9 hESCs. Statistical analysis was performed with an unpaired one-tailed t test. Relative to *YWHAZ* expression. (D) qRT-PCR of *CDYL* isoform 2, 3 and primers that detect all *CDYL* mRNA variants, during *circCDYL* KD in H9 hESCs. (B and D) Relative to *YWHAZ* expression. Statistical analysis was performed with an Ordinary one-way ANOVA with Fisher's LSD test. (*** = p<0.001**** = p<0.0001) (n=3; Mean+SEM).

4.2.6 Knockdown of *circPIF1*

KD of *circPIF1* in H9 hESCs was effective, with 2 independent shRNAs targeting the backsplice junction, as the circRNA was no longer detectable by qRT-PCR (Figure 4-7C). Both KD hESCs of *circPIF1* induced adherent EB formation within 5 days post lentiviral transduction, as hESC colonies would become 3D aggregates of cells, however, remained adherent to the vitronectin-coated culture dish (Figure 4-7A and B). The reduction and overexpression of *circCDYL* was found to alter splicing of *CDYL*, therefore quantification of *PIF1* mRNA levels was also completed. Both *circPIF1* KD hESC lines reduced the level of *PIF1* mRNA (shRNA 1 = 58.1%, shRNA 2 = 61.4%) (Figure 4-7D). The *PIF1* gene does have multiple mRNA variants that are generated from alternative splicing, however, determining if changing the abundance of *circPIF1* is able to alter splicing of its parental gene was not completed. Therefore, KD of *circPIF1* may directly affect the transcription of *PIF1* mRNA.

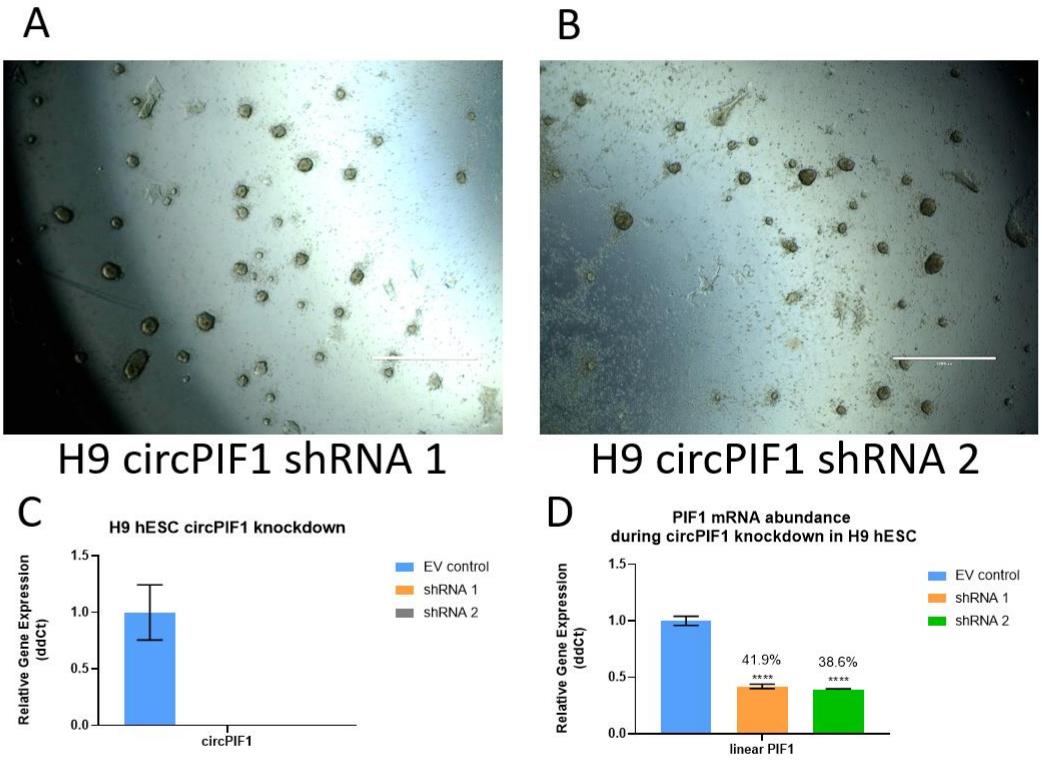


Figure 4-7 Reduction of *circPIF1* by shRNA induces adherent EB formation in H9 hESCs

Image of H9 hESCs transduced with shRNAs targeting *circPIF1* 5 days post-transduction. (A) pLKO.1 *circPIF1* shRNA 1 H9 hESCs. (B) pLKO.1 *circPIF1* shRNA 2 H9 hESCs. (C) qRT-PCR validation of *circPIF1* KD in H9 hESCs. (D) qRT-PCR comprising all *PIF1* mRNA variants during *circPIF1* KD in H9 hESCs. Relative to *YWHAZ* expression. Statistical analysis was performed with an Ordinary one-way ANOVA with Fisher's LSD test. (**** = p<0.001) (n=3; Mean+SEM) Images were taken using the EVOS™ Imaging system. Scale bar 1000 µm.

4.2.7 Knockdown of *circZNF609*

KD of *circZNF609* was completed in H9 hESCs with only one shRNA and a reduction of 75.7% was quantified by qRT-PCR (Figure 4-8B). These cells did not decrease adherence as uniformly as *circPIF1* KD cells as colonies were still partially adherent (Figure 4-8A). Parental *ZNF609* mRNA was quantified, and no statistical change was observed (Figure 4-8C). This circRNA was quantified before mRNA-seq was completed in chapter 3, therefore the house-keeping gene *B2M* was used instead of *YWHAZ*, as this was documented to be a good house-keeping gene for hESCs (Vossaert *et al.*, 2013). Due to this being the worst performing house-keeping gene that was analysed, qRT-PCR would need to be completed again with *YWHAZ* for normalisation in qRT-PCR.

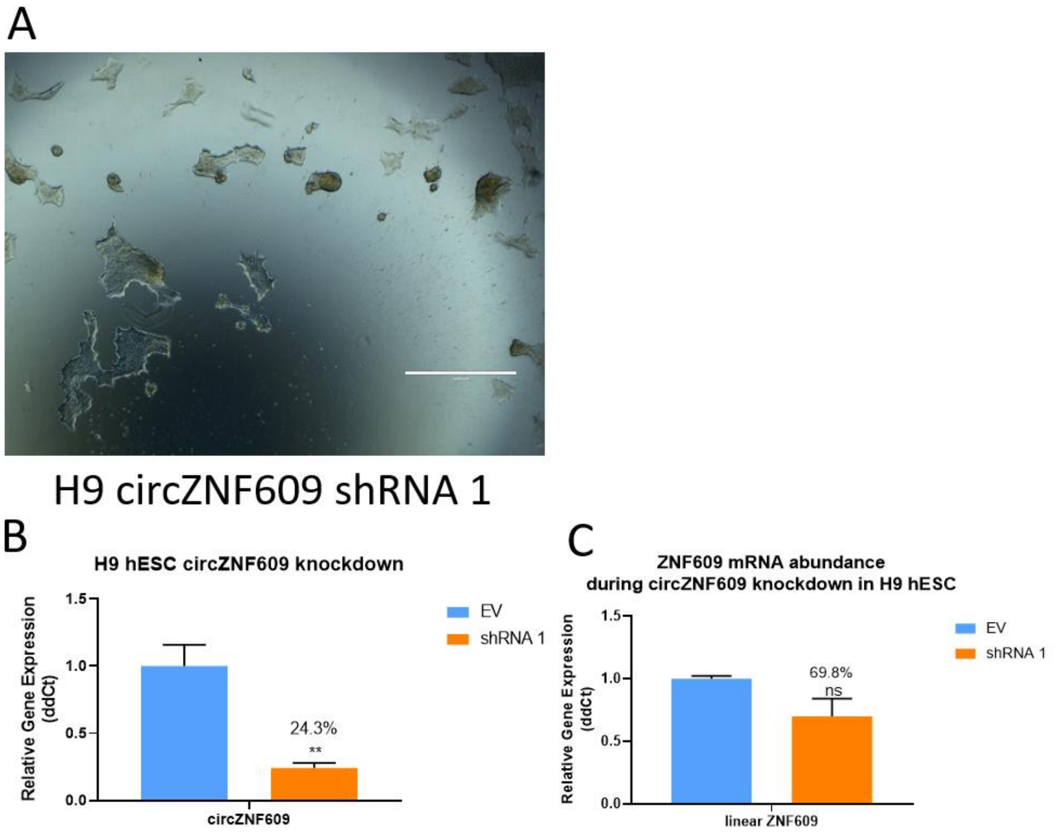


Figure 4-8 Reduction of *circZNF609* by shRNA induces altered cell morphology in H9 hESCs

Image of H9 hESCs transduced with an shRNA targeting *circZNF609*. (A) pLKO.1 *circZNF609* shRNA 1 H9 hESCs. (B) qRT-PCR validation of *circZNF609* KD in H9 hESCs. (C) qRT-PCR of *ZNF609* mRNA variants during *circZNF609* KD in H9 hESCs. Relative to *B2M* expression. Statistical analysis was performed with an unpaired one-tailed t test. (* = $p<0.05$) (n=3; Mean+SEM) Images were taken using the EVOS™ Imaging system. Scale bar 1000 µm.

4.2.8 Knockdown of *circACVR2A*

The initial set of qRT-PCR primers that were designed using SYBR Green for the detection of *circACVR2A* were inefficient despite being a highly expressed circRNAs in H9 hESCs. A dilution series found that these primers were at most only 80% efficient as standard curves were produced with annealing temperatures at 54°C, 56°C and 58°C with a PCR template (Appendix 4). Therefore, a PrimeTime probe assay and primers were designed using the PrimerQuest® Tool (Integrated DNA Technologies). The assay showed a 94% efficiency by qRT-PCR using a standard curve of five, serial 10-fold dilutions of purified PCR template with an average CV of 0.89% (Figure 4-9).

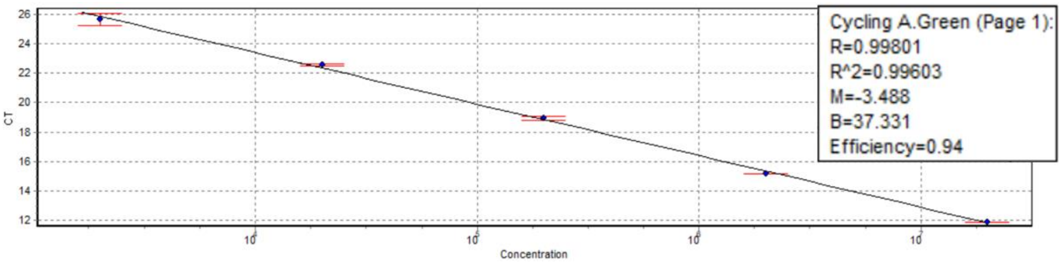


Figure 4-9 *circACVR2A* primetime qPCR standard curve

CircACVR2A PrimeTime assay (IDT) standard curve by 2-step qRT-PCR using 60°C combined annealing and elongation temperature with 2X PrimeTime master mix. Concentration is copies per qRT-PCR reaction with 3 technical reps. Error bars are SEM.

KD of *circACVR2A* was performed with 2 independent shRNAs that were effective at reducing circRNA abundance in H9 hESCs (sRNA 1 = 86.8%, shRNA 2 = 80.6%) (Figure 4-10C). *CircACVR2A* was detected at the same level in EV control and scrambled control further supporting specificity of the *circACVR2A*-targeting shRNAs. Both shRNA treatments in hESCs, similar to *circPIF1* KD, induce EB formation (Figure 4-10A and B), however EBs would eventually detach and become suspended in the culture dish. EB production occurred quicker for shRNA 2 than shRNA 1 treated hESCs, as shRNA 1 colonies were larger and took longer to form EBs, however colonies could be observed to be dark and rising from the culture dish. Profiling *ACVR2A* mRNA expression during *circACVR2A* KD shows that in shRNA 1 treatment, *ACVR2A* mRNA expression was reduced by 41.8%, however no statistical reduction was detected for shRNA 2 treated cells (Figure 4-10D).

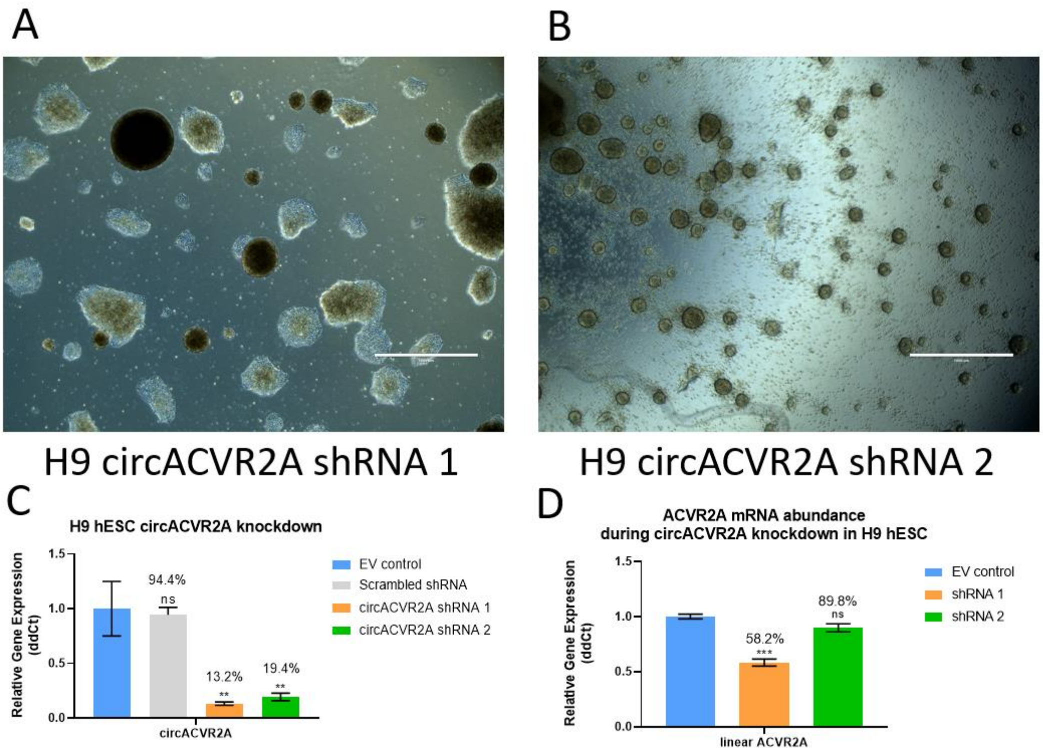


Figure 4-10 Reduction of *circACVR2A* by shRNA induces EB formation in H9 hESCs

Image of H9 hESCs transduced with shRNAs targeting *circACVR2A* 5 days post-transduction. (A) pLKO.1 *circACVR2A* shRNA 1 H9 hESCs. (B) pLKO.1 *circACVR2A* shRNA 2 H9 hESCs. (C) qRT-PCR validation of *circACVR2A* KD in H9 hESCs. (D) qRT-PCR comprising all *ACVR2A* mRNA variants during *circACVR2A* KD in H9 hESCs. Relative to *YWHAZ* expression. Statistical analysis was performed with an Ordinary one-way ANOVA with Fisher's LSD test. (**** = p<0.001) (n=3; Mean+SEM) Images were taken using the EVOS™ Imaging system. Scale bar 1000 µm.

To further confirm that the loss of *circACVR2A* was inducing EB formation a third shRNA targeting the backsplice junction was designed and incorporated into hESCs by lentiviral transduction. A significant reduction of *circACVR2A* was quantified by qRT-PCR (Figure 4-11B) and EB formation was observed (Figure 4-11A), supporting that a reduction of *circACVR2A* is driving EB formation. Detection of *ACVR2A* mRNA in shRNA 3 treated hESCs was not completed.

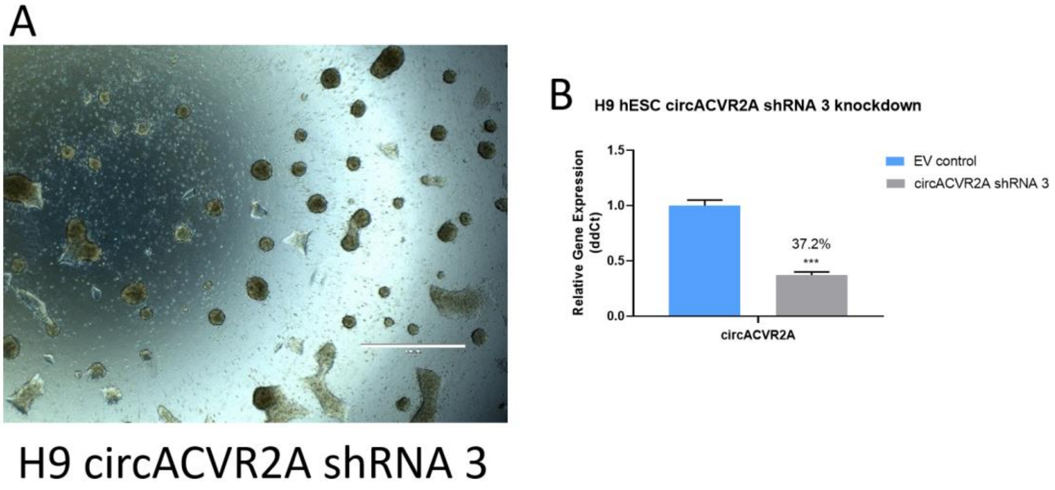


Figure 4-11 Reduction of *circACVR2A* by a third shRNA induces EB formation in H9 hESCs

Image of H9 hESCs transduced with a third shRNA targeting *circACVR2A*. (A) pLKO.1 *circACVR2A* shRNA 3 H9 hESCs. (B) qRT-PCR validation of *circACVR2A* KD in H9 hESCs. Relative to *YWHAZ* expression. Statistical analysis was performed with an unpaired one-tailed t test. (*** = $p<0.001$) (n=3; Mean+SEM)Images were taken using the EVOS™ Imaging system. Scale bar 1000 µm.

KD of *circACVR2A* (shRNA 1 86.8%; shRNA 2 80.6%) was first found to induce EB formation in H9 hESCs cultured on vitronectin (Figure 4-10A and B). However, doubling the lentiviral concentration, increasing the number of successfully transduced cells, allows for larger colonies to survive through puromycin selection without initially producing EBs and maintaining larger colonies, possibly due to a larger surface area with the ECM. Although a higher proportion of colonies remained adherent using increased lentiviral concentration, they did not maintain their homogenous undifferentiated morphology as neural rosette structures (Figure 4-12) were observed after the completion of puromycin selection. This phenomenon was observed with H9 hESCs grown on either vitronectin (Figure 4-12) or geltrex matrix (Figure 4-14), a supportive laminin-based ECM. Additionally, halving the amount of virus used during lentiviral transduction for *circACVR2A* shRNA 1, increased the proportion of EBs after puromycin selection similarly to shRNA 2 and shRNA 3 treatments (Figure 4-13). Therefore, the phenotypic differences observed between shRNA treatments is likely due to the efficiency of lentiviral transduction with larger successfully transduced colonies having more surface area with the ECM.

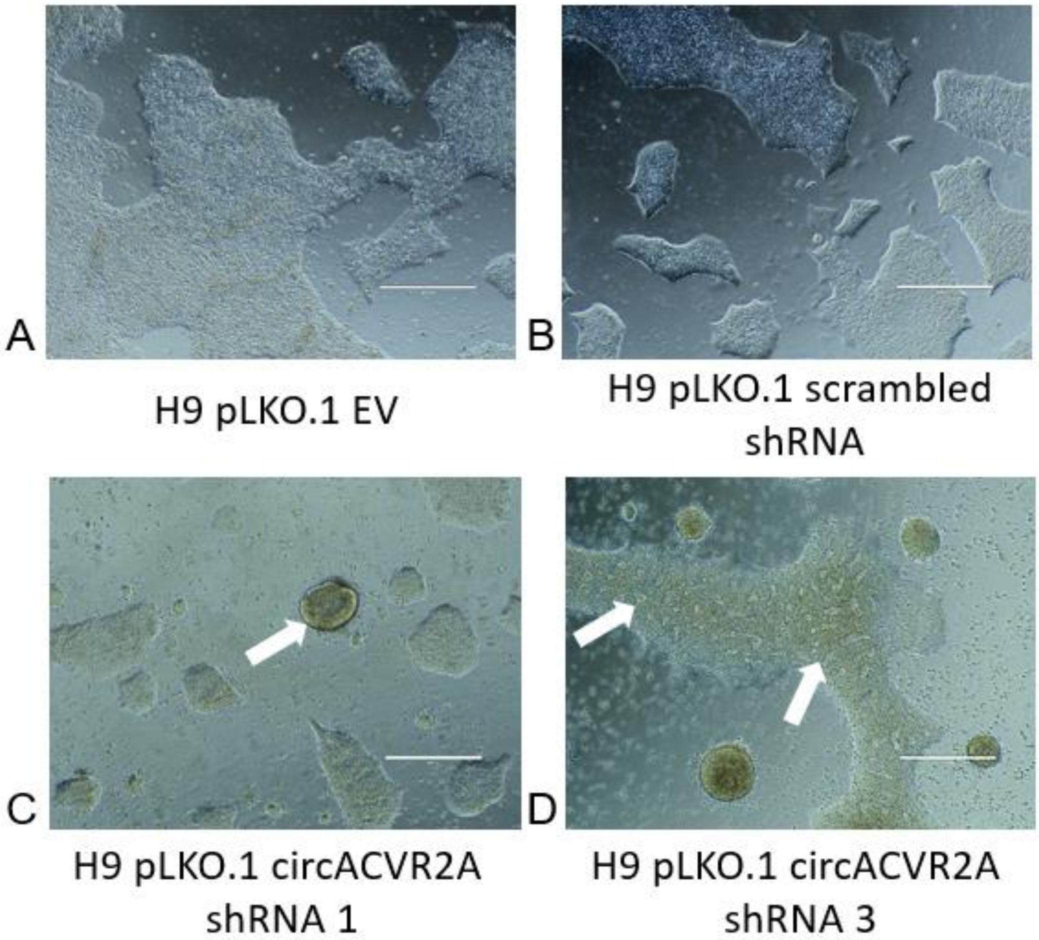

A
H9 pLKO.1 EV
B
H9 pLKO.1 scrambled shRNA
C
H9 pLKO.1 circACVR2A shRNA 1
D
H9 pLKO.1 circACVR2A shRNA 3

Figure 4-12 Knockdown of *circACVR2A* in hESCs cultured on vitronectin

Images of neural rosette formation from H9 cells after KD of *circACVR2A* on vitronectin using 2X virus concentration. (A) H9 Empty vector. (B) H9 scrambled shRNA. (C) H9 *circACVR2A* shRNA 1. (D) H9 *circACVR2A* shRNA 3. Neural rosettes marked by white arrows. Images were taken using the EVOS™ Imaging system. Scale bar 400 µm.

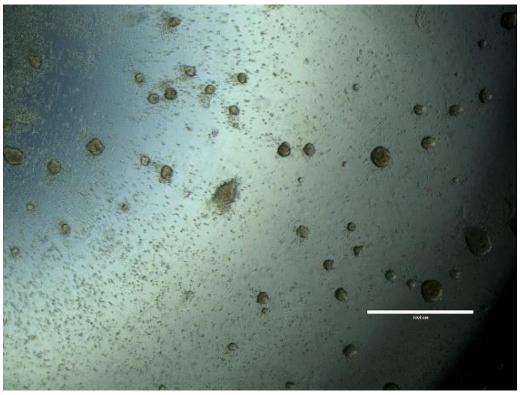


Figure 4-13 Increased hESC EB proportion using half *circACVR2A* shRNA 1 lentivirus

Image of hESCs cultured on vitronectin and transduced with *circACVR2A* shRNA 1 using half lentivirus concentration. Images were taken using the EVOS™ Imaging system. Scale bar 1000 µm.

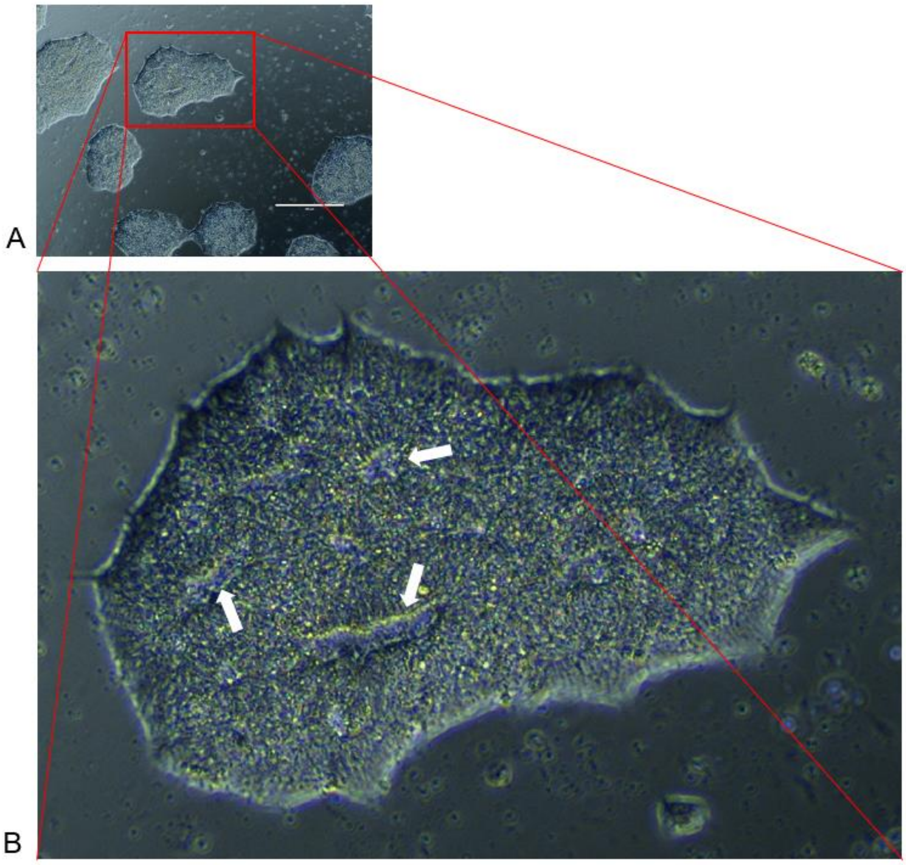


Figure 4-14 Neural rosette formation in hESC colony from *circACVR2A* shRNA 2 Knockdown on geltrex

Neural rosette formation of hESC colony cultured on geltrex with KD of *circACVR2A* with shRNA 2 (7 days post-transduction). Neural rosettes marked by white arrow. Image was taken using the EVOS™ Imaging system. Scale bar 400 µm.

Neural rosette formation occurred before (Figure 4-12C and D) or during EB formation (Figure 4-15), however all *circACVR2A* KD colonies produced EBs that eventually detach from vitronectin coated dishes. In fact, some EBs already detach during media changes when adding fresh puromycin-containing media during the 72-hour antibiotic selection. To determine if any further changes occurred, these EBs were maintained in suspension culture for up to 10 days. Three independent shRNAs for *circACVR2A* were used and each produced EBs. While EBs from shRNA 2 and 3 become spherical and dark during observation (Figure 4-16F and G), shRNA 1 EBs became mostly dark, with EBs containing bright edges and dark centres (Figure 4-16E). Bright and dark EB morphologies (complex EBs) have been documented previously (Kim *et al.*, 2011), with EBs generated by canonical methods (culturing hESCs in low adherence), producing several different EB morphologies (Cystic, light cavity, dark cavity and complex cavity) (Figure 4-16A). Despite the slight differences between the three shRNA treatments, all shRNA treatments produced colonies or EBs with neural rosette structures (Figure 4-12 and Figure 4-14). These structures were only noticeable in early stage EBs, as later stage EBs became too dark for light microscopy.

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

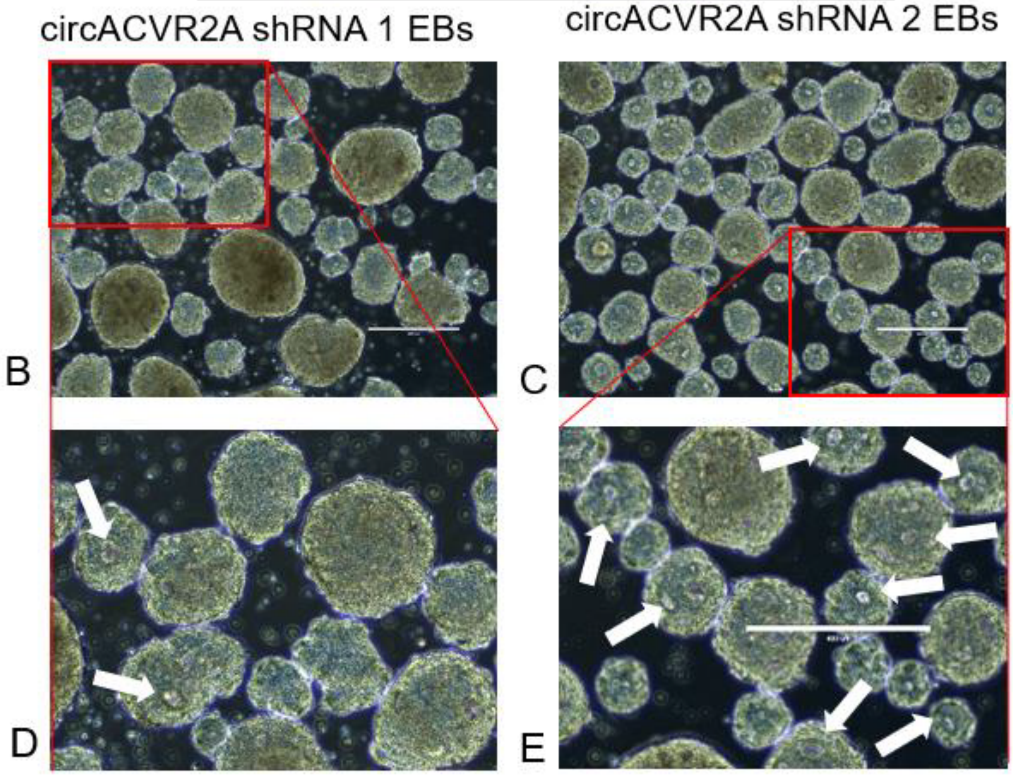
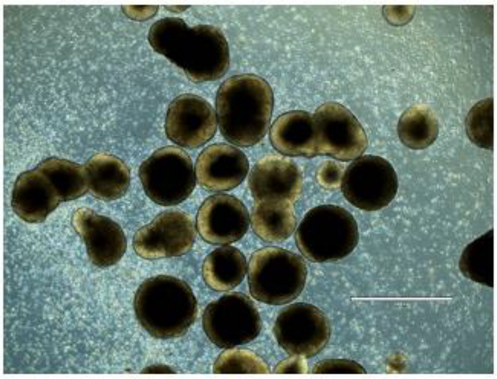


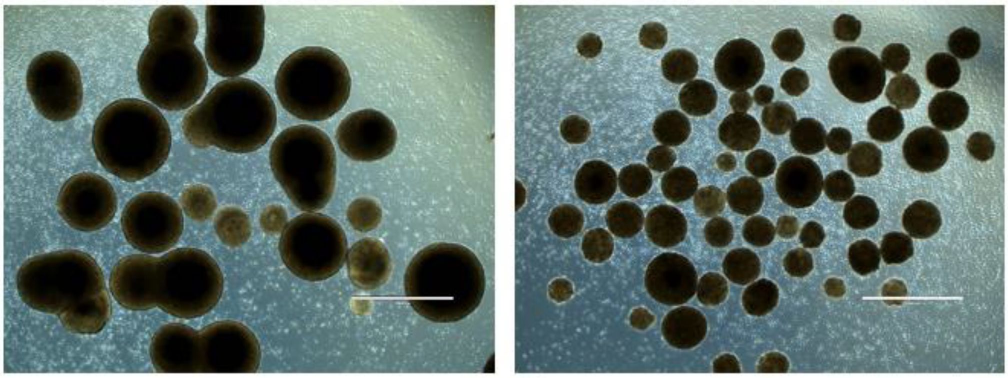
Figure 4-15 Neural rosette structures observed in *circACVR2A* knockdown induced EBs

(A) Bright field image of neural rosette formation from dual TGFβ inhibition with dorsomorphin and SB431542 (Image: (Shang, 2018)). (B and C) Neural rosette formation in Embryoid Bodies 5 days post-lentiviral transduction of *circACVR2A* shRNA KD. Images were taken using the EVOS™ Imaging system. Scale bar 400 µm. Neural rosettes marked by white arrow.

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.



E circACVR2A shRNA 1 EBs



F circACVR2A shRNA 2 EBs

G circACVR2A shRNA 3 EBs

Figure 4-16 Morphological classification of EBs following *circACVR2A* knockdown

(A-C) 10-day EB morphology *in vitro* including cystic, bright, and dark (Image: (Kim *et al.*, 2011)) and (D) percentage of each at day 10. (E-G) EBs generated from KD of *circACVR2A* with 3 independent shRNAs. Images of EBs are taken 13 days post-lentiviral transduction as EBs generally arise about 3 days after transduction during puromycin treatment. Images were taken using the EVOS™ Imaging system. Scale bar is 400 µm for A, B and C. Scale bar is 1000 µm wide for E, F and G.

4.2.8.1 Knockdown of ACVR2A does not induce EB formation

To ensure that EB formation is not a result from a reduction of *ACVR2A* mRNA, as KD of *circACVR2A* with one shRNA reduced *ACVR2A* expression, 2 independent shRNAs were designed to target *ACVR2A* mRNA. These shRNAs were designed to target *ACVR2A* mRNA downstream of the exons that are spliced to produce *circACVR2A* (Figure 4-17).

Efficient KD of linear RNA with both shRNAs (shRNA 1=59.1% and shRNA 2=93.4% KD) (Figure 4-18C) resulted in no phenotypic changes in H9 hESC cells (Figure 4-18A and B). Interestingly, *circACVR2A* was observed to increase in response to a reduction of *ACVR2A* mRNA by shRNAs proportionally to the level of *ACVR2A* knockdown (Figure 4-18D).

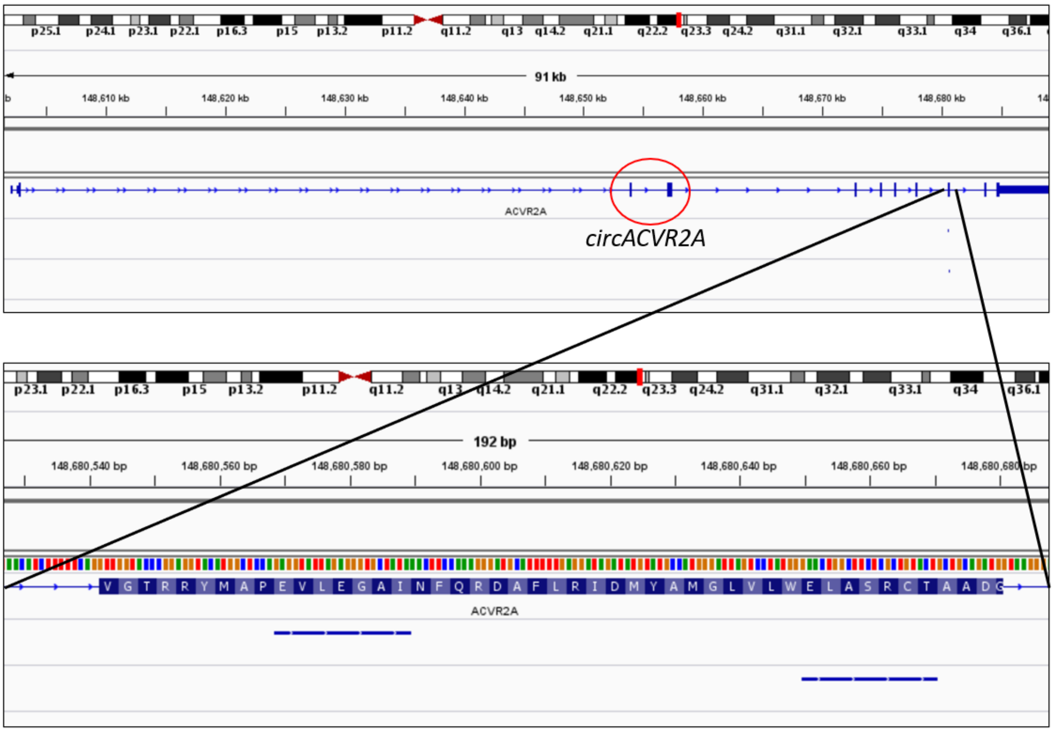

p25.1
p24.1
p23.1
p22.1
p16.3
p15
p13.2
p11.2
q11.2
q13
q14.2
q21.1
q22.2
q23.3
q24.2
q31.1
q32.1
q33.1
q34
q36.1
91 kb
148,610 kb
148,620 kb
148,630 kb
148,640 kb
148,650 kb
148,660 kb
148,670 kb
148,680 kb
ACVR2A
circACVR2A
192 bp
148,680,540 bp
148,680,560 bp
148,680,580 bp
148,680,600 bp
148,680,620 bp
148,680,640 bp
148,680,660 bp
148,680,680 bp
V G T R R Y M A P E V L E G A I N F Q R D A F L R I D M Y A M G L V L W E L A S R C T A A D G
ACVR2A

Figure 4-17 *ACVR2A* mRNA shRNAs were designed downstream of *circACVR2A*

Integrative Genomics Viewer display of *ACVR2A* genomic locus and location where shRNAs bind to *ACVR2A* mRNA for RISC mediated degradation.

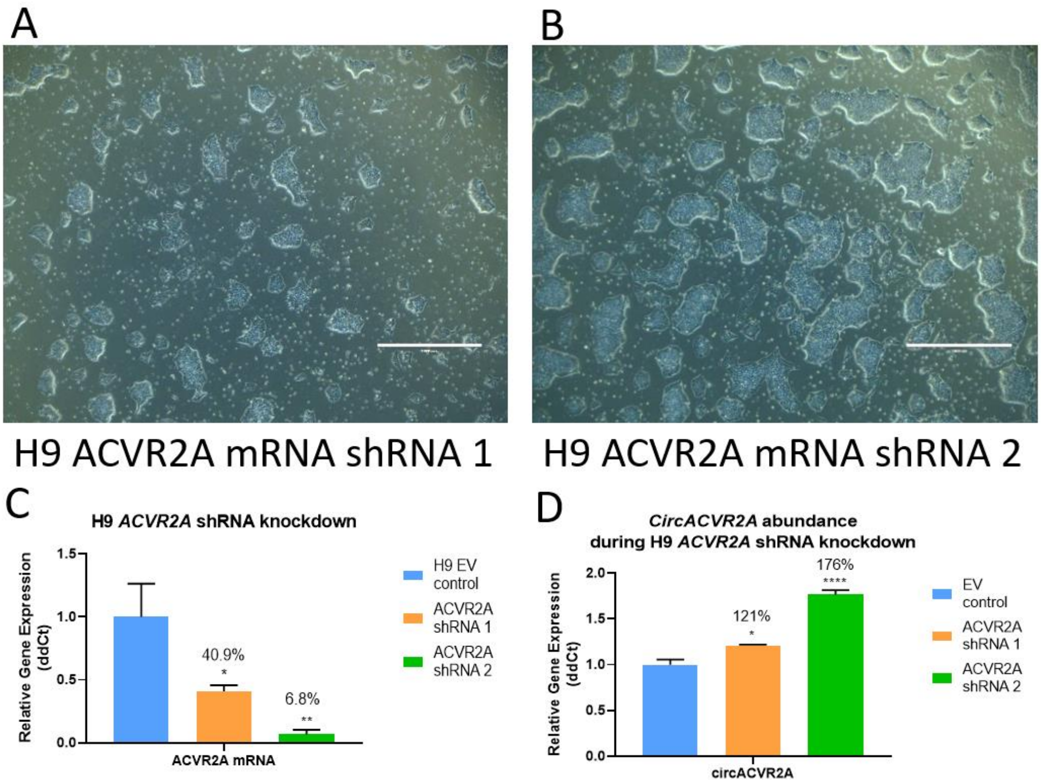


Figure 4-18 Knockdown of *ACVR2A* mRNA by shRNA does not alter hESC morphology

KD to *ACVR2A* mRNA did not produce EBs. (A) Images of hESCs with KD by *ACVR2A* shRNA 1. (B) Images of hESCs with KD by *ACVR2A* shRNA 2. Scale bar 400 µm. (C) Efficiency of *ACVR2A* KDs quantified by qPCR. (D) *circACVR2A* expression in H9 *ACVR2A* shRNA KD. Relative to *YWHAZ*. Statistical was performed with an Ordinary one-way ANOVA with Fisher's LSD test. * = $p<0.05$, **= $p<0.01$. (n=3; Mean+SEM) Images were taken using the EVOS™ Imaging system.

4.2.9 Knockdown of *circACVR2A* with shRNAs targeting non-backsplice sequences

As KD of *ACVR2A* did not result in an EB phenotype, multiple shRNAs targeting different locations of *circACVR2A* were designed (Figure 4-19) to determine if not targeting the backsplice junction still results in EB formation. This would also target *ACVR2A*, however, KD of *ACVR2A* does not induce EB formation. Nonetheless, lentiviral transductions were performed, and surprisingly, no EB formation was observed with KD of *circACVR2A* at locations other than the backsplice junction (Figure 4-20). Quantification of *circACVR2A* KDs revealed that these shRNAs were not as effective at reducing *circACVR2A* levels compared to the backsplice-targeting shRNAs with shRNA 7 reducing circRNA levels the most with a 61.9% reduction. Quantification of *ACVR2A* mRNA levels was not performed.

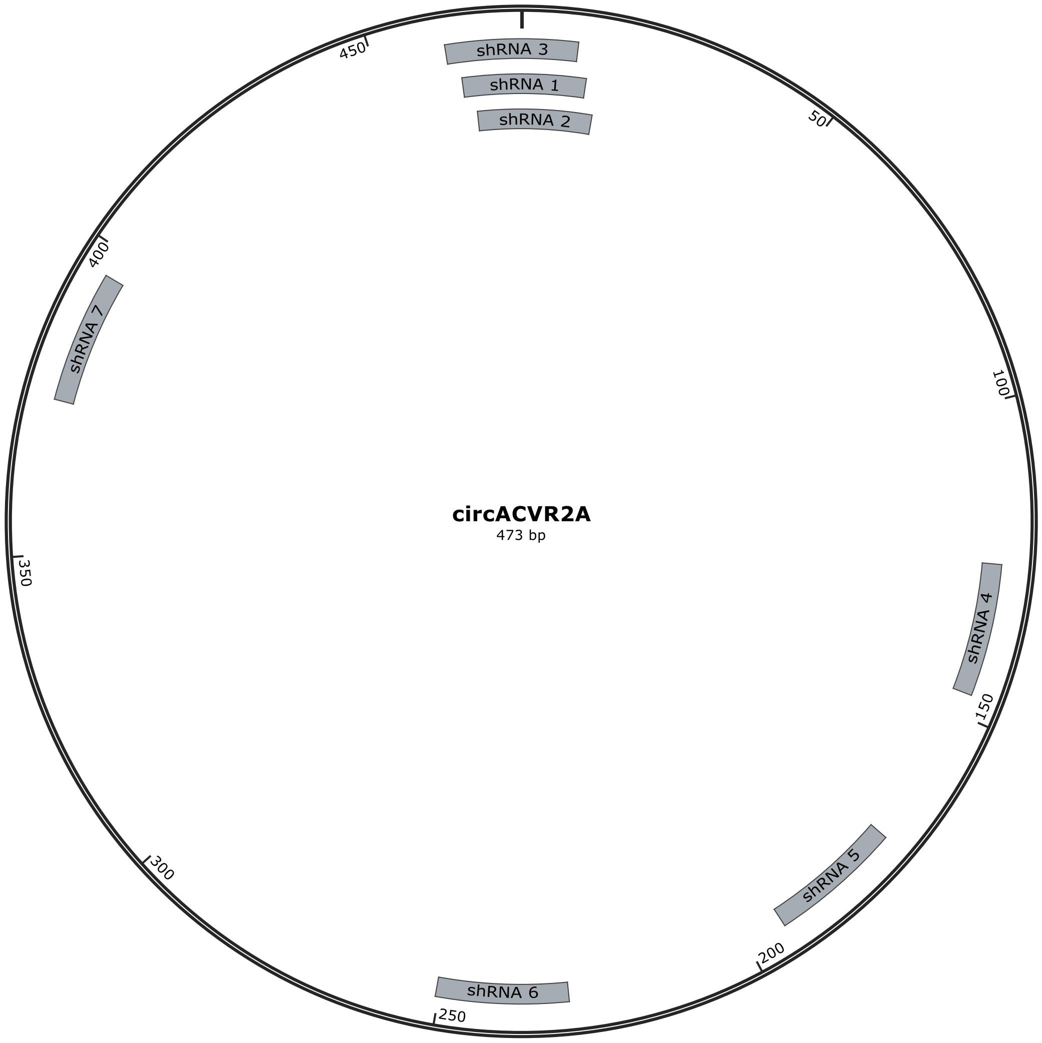

shRNA 3
shRNA 1
shRNA 2
50
100
shRNA 4
150
shRNA 5
200
shRNA 6
250
300
350
shRNA 7
400
450
circACVR2A
473 bp

Figure 4-19 Image of *circACVR2A* and shRNA locations

Illustration of shRNA locations around *circACVR2A*. shRNAs 1-3 are located across the backsplice, shRNA 4 is located adjacent to the backsplice junction and shRNAs 5-9 are located around the circRNA.

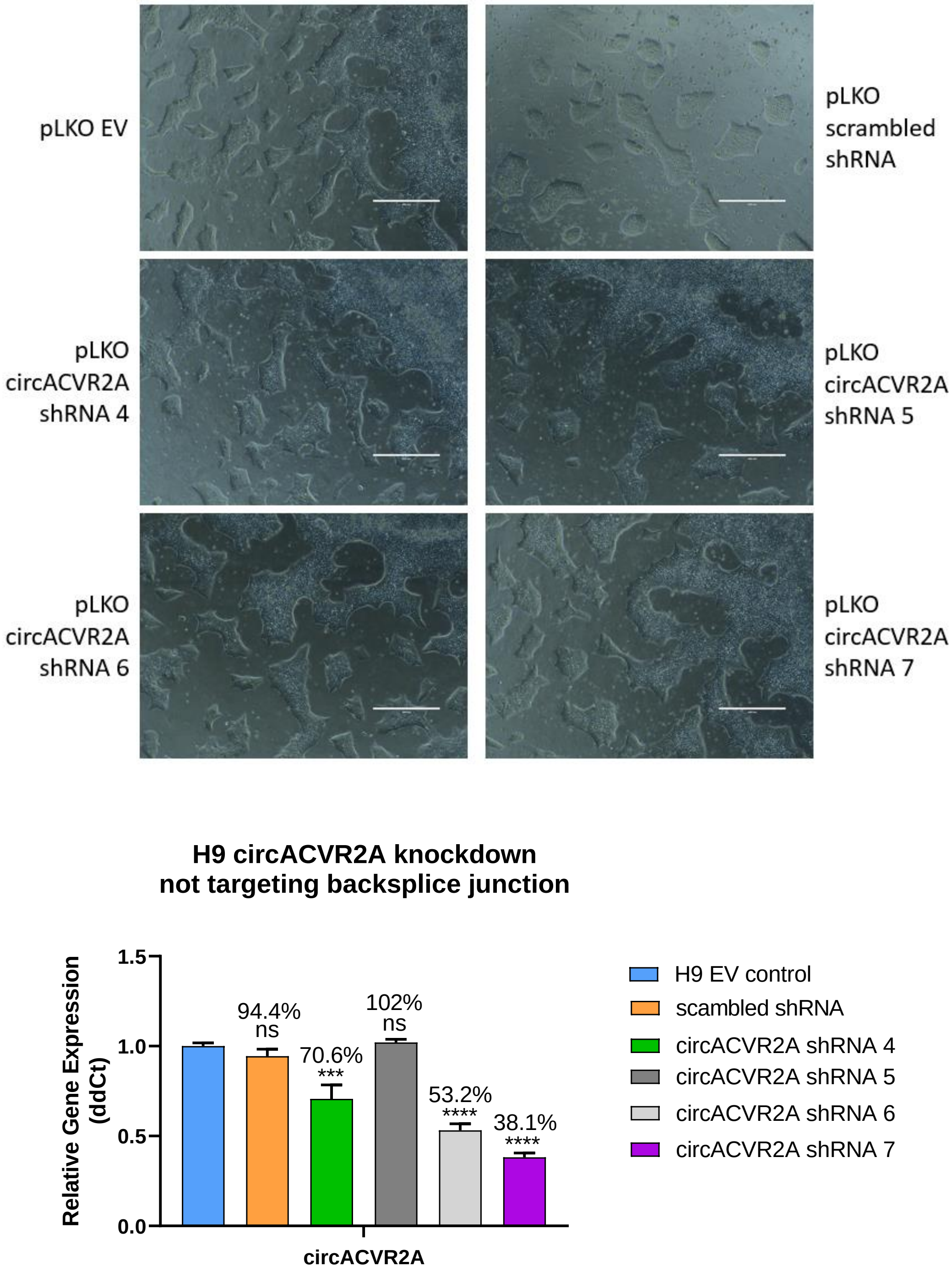


Figure 4-20 Knockdown of *circACVR2A* with non-backsplice targeting shRNAs

Images of *circACVR2A* KD by shRNAs not targeting the backsplice junction. (A) pLKO Empty Vector. (B) pLKO scrambled shRNA. (C) pLKO *circACVR2A* shRNA 4. (D) pLKO *circACVR2A* shRNA 5. (E) pLKO *circACVR2A* shRNA 6. (F) pLKO *circACVR2A* shRNA 7. (G) pLKO *circACVR2A* shRNA 8. (H) pLKO *circACVR2A* shRNA 9. (I) qPCR of *circACVR2A* KD of shRNA 4 and 5. (J) qPCR of *circACVR2A* KD of shRNA 6-9. Relative to *YWHAZ* expression. Statistical analysis was performed with an ordinary one-way ANOVA (* = $p<0.05$, **= $p<0.01$, ***= $p<0.001$) (n=3; Mean+SEM) Scale bar is 400µm wide.

4.3 Discussion

Genetic manipulation of circRNA levels in hESCs has identified an important role for specific circRNAs in maintaining stem cell morphology as well as retaining an undifferentiated state. CRISPR Cas13b is a novel alternative to RNAi in mammalian systems that offers superior targeting and comparable KD efficiencies. Although this is the case for many cell lines (Tng *et al.*, 2020; Fareh *et al.*, 2021; Tang *et al.*, 2021; Zhang *et al.*, 2021), Cas13b did not efficiently reduce *circACVR2A* levels in hESCs, as supported by collaborators that have been unable to produce efficient KD efficiencies with Cas13b in hESCs. This is most likely due to the common phenomenon of transgene silencing that occurs, notably in hESCs, leading to the suppression of transgene expression. The loss of the expression of Cas13b after lentiviral transduction was not investigated, as confirming this known phenomenon would not assist in the project aims of knocking down circRNA transcripts. Furthermore, personal communications with Ling Ling Chen from Associate Professor Joseph Rosenbluh’s lab from Monash University confirmed that they were also unable to generate hESCs with sufficient KD using Cas13b. However, shRNA KD of five independent circRNAs was effective and has induced phenotypes that are strongly suggestive of differentiation and/or EMT.

A variety of methods to differentiate stem cells exist and choosing which method requires careful consideration of many factors. One common method to differentiate ESCs involves an intermediate EB formation stage, which requires removing the cells from normal culturing conditions on ECM and seeding them as single cells in low adherence dishes, with some protocols also altering the presence of growth factors in the media (He *et al.*, 2017). After 1 day in low adherence, cells form EBs and continued passive culture or culturing these aggregates with specific growth factor combinations can produce cells from the three germ lineages (Ectoderm, Mesoderm and Endoderm) (Khoo *et al.*, 2005). Another method is to remove growth factors including fibroblast growth factor 2 (FGF2), which is present in normal culture media, to maintain cells in an undifferentiated state, as well as the addition of other factors which allows for adherent differentiation into specific lineages (Yao *et al.*, 2006, p.). EMT activation has been observed in both methods for differentiating embryonic stem cells (Eastham *et al.*, 2007; Boyd *et al.*, 2009; Chan *et al.*, 2012; Feng *et al.*, 2012; Liu *et al.*, 2012, p. 0; Bjørlykke *et al.*, 2019).

Remarkably, KD of *circZNF609*, *circACVR2A* and *circPIF1* in hESCs, cultured on vitronectin-coated dishes, have produced cells with altered adherence. Embryoid body-like aggregates were observed for *circACVR2A* and *circPIF1* without first seeding cells in low adherence. Furthermore, stem cells are cultured on ECM and in media that contain high levels of certain growth factors (e.g., FGF2) and other molecules which all function to keep cells in an undifferentiated state. Despite these culturing conditions, the KD of *circPIF1* and *circACVR2A* have induced the formation of an intermediate EB stage. Of note, reduction of *circACVR2A* in hESCs cultured on geltrex, a more supportive ECM, EB formation was not observed. This was not completed for *circPIF1* and *circZNF609*, however, this does indicate that, for hESCs specifically, the culture conditions can impact on the observed phenotypes of hESCs.

Before mRNA sequencing was performed, qRT-PCR was normalised using *B2M* transcript levels as it was reported to be a stable reference gene in differentiating hESCs (Vossaert *et al.*, 2013). However, this study seemed to be an exemption, as mRNA-seq in chapter 3, and others, have found it to be unsuitable for normalisation in ESCs and PSCs (Holmgren *et al.*, 2015; Panina, Germond and Watanabe, 2020). Therefore, qRT-PCR quantification of *circZNF609* and *ZNF609* mRNA during *circZNF609* H9 hESC KD was normalised using *B2M* and would need to be performed again using *YWHAZ* for a more accurate quantification.

Focusing on *circACVR2A*, neural rosette structures were observed with phase contrast microscopy 5 days post lentiviral transduction using 3 independent shRNAs, in adherent conditions and with using 2 different ECMs (vitronectin and geltrex). These structures are indicative of neuroectoderm and neural precursors and has also been observed in suspended EBs (Ma *et al.*, 2008) or attached EBs cultured in varying neural differentiation protocols. (Zhang *et al.*, 2001; Elkabetz *et al.*, 2008; Mariani *et al.*, 2012, 2015, p. 1; Shang, 2018). Additionally, neural rosettes are observed in adherent conditions (Rust, Sadasivam and Dunn, 2006; Yao *et al.*, 2006; Ma *et al.*, 2008; Chambers *et al.*, 2009; Shi *et al.*, 2012) with most methods involving the inhibition of tumour growth factor beta (TGFβ) signalling to induce neural differentiation.

Differences were observed between the shRNA treatments for *circACVR2A* KD, whereby colonies from shRNA 1 would take longer to form EBs than shRNA 2 or shRNA 3. Furthermore, EBs from shRNA 1 are not as uniformly round or as dark as shRNA 2 or 3, even though the shRNA 1 KD was just as effective as shRNA 2. Therefore, a third independent shRNA was designed to overlap the backsplice junction which also generated dark EBs. Light and dark cavity EBs have the capacity to differentiate into all cell types, with cystic EBs composing of more endodermal populations. Interestingly, the dark EB phenotype is also observed when EBs are grown in neural differentiation medium for 10 days (Ma *et al.*, 2008). The morphological differences between the different shRNA treatments (not uniformly generated at the same time or not all dark EBs) may be due to the titre of lentivirus used for transduction as modulating the virus concentration was found to to alter how quickly EBs formed after antibiotic selection. A reduction of lentivirus concentration was found to decrease the number of cells that were successfully transduced. Therefore, smaller colonies remained after selection, which in turn increased the proportion of EBs produced from shRNA 1. This is most likely due to smaller colonies having less surface area that could remain bound to the vitronectin-coated dishes, and therefore would become suspended quicker. Furthermore the design of the shRNAs could affect how quickly the shRNA was transcribed as Pol III promotors have a strong preference for starting transcription with a purine (Ros and Gu, 2016). Sub-optimal nucleotide selection may also produce shRNAs with non-ideal secondary structures. Other factors that could affect shRNA efficiency include a loop structure within the shRNA, thermodynamic properties of the hairpin, secondary structure, and the surrounding sequences (Taxman *et al.*, 2006; Moore *et al.*, 2010; O'Keefe, 2020).

EBs from *circACVR2A* shRNA KD were dark and presented with neural rosette-like structures, suggesting that the cells may be differentiating into neural precursors as neural rosettes structures are also observed in suspended EBs that are cultured in neural differentiation medium (Ma *et al.*, 2008). However, neural rosettes are more frequently observed after EB adherence to a laminin-based ECM (Ma *et al.*, 2008). To determine if the type of attachment matrix could impact the phenotype of *circACVR2A* KD, this transduction was also performed on cells that were grown on a laminin-based matrix (geltrex), known to be a more supportive ECM and the preferred matrix for neuronal generation (Ma *et al.*, 2008). During this transduction, hESCs maintained their contact with the ECM and did not form

EBs, however, the colonies that grew formed neural rosettes within adherent colonies. Therefore, EB formation may not be the only initial consequence of *circACVR2A* KD, and the phenotype observed may also be dependent on the use of different ECMs. EB formation may occur only on vitronectin due to the differentiating cells that arise requiring other ECM components found in geltrex to promote adhesion and maintain colonies. In fact, many studies have investigated the effect of different ECMs on neural progenitor formation and conclude that proteoglycans and laminins to be vital components for neural development (Long and Huttner, 2019).

CircCDYL and *circFAT3* KDs produced a different phenotype to KDs of *circPIF1*, *circZNF609* and *circACVR2A*, yet this significant morphological change mirrors the phenotypes of other stem cells undergoing differentiation (Yao *et al.*, 2006; Eastham *et al.*, 2007; Boyd *et al.*, 2009; Chen *et al.*, 2012; Feng *et al.*, 2012; Liu *et al.*, 2012; Bjørlykke *et al.*, 2019). These cells remain adherent to the culture dish/ECM however colonies lose phase bright edges and individual cell boundaries are able to be distinguished within colonies and in between colonies. Further investigation would need to be completed to determine if this differentiation is lineage-specific or if these cells maintain undifferentiated stem cell properties while showing morphological changes. Although these individual morphological changes are different (Loss of adherence, EB formation and loss of dense colonies), all phenotypes have been linked to EMT and differentiation (Yao *et al.*, 2006; Eastham *et al.*, 2007; Boyd *et al.*, 2009; Chan *et al.*, 2012; Feng *et al.*, 2012; Liu *et al.*, 2012, p. 0; Bjørlykke *et al.*, 2019; P. Chen *et al.*, 2020). The reduction of all five circRNAs may be able to activate EMT and/or differentiation, however, activation of different EMT/differentiation TFs are possibly influencing the observed phenotypic responses.

4.3.1 CircRNAs regulating parental gene transcripts

Initial investigations into the functional consequences of circRNA KDs has revealed that altering the abundance of *circCDYL* affects the alternative splicing of *CDYL* mRNA. CircRNA regulation of parental gene alternative splicing has only been reported for 2 circRNAs and is not a comprehensively studied phenomenon. CircRNA *SEP3* from the SEPALLATA3 gene in *Arabidopsis Thaliana* regulates splicing of its parental gene by binding to its genomic locus forming an R-loop with the DNA, causing a stalling of transcription, and leading to the recruitment of splicing factors and upregulation of an exon skipped variant (Conn *et al.*, 2017). Working via a different mechanism, *circMBL*, as discussed in chapter 1, is able to compete with linear splicing and regulate MBL levels due to multiple MBL binding motifs (Ashwal-Fluss *et al.*, 2014). It is currently not known if *circCDYL* is able to bind to DNA, RNA or protein, however it is possible that any of these interactions can alter the splicing of it's parental gene. *CDYL* alternative splicing generates an mRNA transcript that skips the exon that comprises the sequence of *circCDYL*, and so the mechanism with which *circCDYL* can regulate splicing is most likely due to the circRNA being complementary to the genomic locus, similar to circular RNA *SEP3*. Further investigation into how *circCDYL* can regulate splicing of its parental gene would need to be completed, to uncover if this is regulating splicing by currently published mechanisms or a novel process.

Reduction of *circPIF1* has also been found to reduce *PIF1* mRNA in hESCs. This may be due to *circPIF1* regulating the transcription of *PIF1* rather than altering alternative splicing. This observation of *circPIF1* formed an honours project during 2020, completed by Jordyn Binelli, and it was further discovered that KD of *PIF1* mRNA also induced EB formation. Therefore, EB formation may be directly due to a reduction of *PIF1* during *circPIF1* KD. It is currently not known if *circPIF1* is directly affecting *PIF1* expression or if the *circPIF1* shRNAs can non-specifically bind the parental mRNA leading to its degradation, and therefore induce EB formation. Additional experiments would need to validate the specificity of *circPIF1* shRNAs to confirm that the specific KD of *circPIF1* is responsible for the phenotypes observed. Furthermore, a relationship between *circPIF1* KD and *PIF1* mRNA would need to be further investigated to determine the method for circRNA regulation on parental mRNA levels.

The gene *ACVR2A* produces a cell surface receptor that, when bound by specific ligands, activates TGFβ signalling within cells and is a strong activator of differentiation as well as EMT. Inhibition of this pathway is a common method for the exclusive production of neural cells, however this process can take up to 3-4 weeks (Dhara and Stice, 2008; Francis and Wei, 2010). As *circACVR2A* KD produces EBs with a possible neural focus, culturing the cells longer may allow for further investigation into the outcomes of *circACVR2A* KD. Successful KD of *ACVR2A*, did not recapitulated the phenotype, indicating that the phenotype is *circACVR2A* specific. Interestingly, *circACVR2A* expression was increased during *ACVR2A* shRNA KD in hESCs using two independent shRNAs. This may be explained by an increase in transcription of the endogenous *ACVR2A* mRNA in the nucleus to replace degraded cytoplasmic mRNA. Due to an increase in transcription of *ACVR2A*, *circACVR2A* may increase above normal levels, as more pre-mRNA is present, and backsplicing, along with linear splicing occurs before the RNA is exported from the nucleus. This hypothesis was not investigated, however, if KD of mRNAs is able to increase circRNA production from the same gene, then many current published mRNA KD experiments could be driving gene-specific increases to circRNAs, leading to unintended cellular consequences.

Targeting non-backsplice sequences of circRNAs by RNAi is not commonly performed as this would also target the parental mRNA. In the case of *ACVR2A*, KD by 2 independent shRNAs did not induce a phenotype and so attempts to reduce *circACVR2A* in other locations were undertaken. Four independent shRNAs targeting *circACVR2A* in locations other than the backsplice region, did not induce EB formation. However, these shRNAs were not as efficient at reducing *circACVR2A* levels compared to the backsplice junction targeting shRNAs. This may be due non-ideal secondary structure of the shRNAs, secondary structure of *circACVR2A* impeding shRNA binding, or possibly due to less available shRNAs targeting the circRNA, as the expressed shRNAs were also targeting the mRNA transcripts. Additionally, reduction of *ACVR2A* has been shown to increase *circACVR2A* levels in H9s hESCs and so splicing of *circACVR2A* may increase when targeting non-backsplice sequences. If the EB phenotype observed is due to a loss of *circACVR2A* specifically in the nucleus, then an increase of nuclear *circACVR2A* as a result of *ACVR2A* knockdown may not induce the EB phenotype as nuclear levels may have increased even though total cellular *circACVR2A* has decreased. Furthermore, it might be possible that normal cellular levels of *ACVR2A* may be required to induce EB formation during KD of *circACVR2A*.

4.4 Chapter 4 Conclusion

KD of circRNAs has induced phenotypes indicating involvement with EMT or differentiation as epithelial hESC colony morphology is lost. Furthermore, *circCDYL* may play a role in regulating it’s parental gene alternative splicing and transcription.

ACVR2A is a type 2 receptor for TGFβ signalling is involved with the TGFβ signalling pathway, a commonly inhibited pathway for differentiation into neural lineages in hESCs and iPSCs. Interestingly KD of *circACVR2A* can induce neuroectodermal differentiation, possibly by inhibiting TGFβ signalling downstream of ACVR2A, it’s parental gene protein. As *circACVR2A* KD produces EBs with already neural characteristics, this was chosen as the primary focus for the remainder of my thesis. Further profiling, by increasing the culture time for maturing EBs, may illuminate further on the consequences of *circACVR2A* KD in hESCs.

Chapter 5 Differentiation of Human Embryonic Stem Cells Induced by Knockdown of *CircACVR2A*

5.1 Introduction

CircRNAs have been shown to affect EMT and differentiation via multiple pathways, including the regulation of TGFβ signalling (Yu and Kuo, 2019). In hESCs, KD of *circACVR2A* induces spontaneous EB production from hESCs even under adherent conditions on vitronectin ECM. EB morphology from independent shRNA treatments indicate the presence of neural precursors as neural rosettes are observed 5 days post lentiviral KD of *circACVR2A*. Furthermore, KD of *circACVR2A* in hESCs, cultured on a laminin-based matrix, also generates neural rosettes, however the colonies remain adherent. Neural rosettes have been observed by others to arise from hESCs by dual inhibition of TGFβ by Noggin and SB431542, two inhibitors of SMAD signalling (Chambers *et al.*, 2009; Wattanapanitch *et al.*, 2014). Continued profiling and attachment of EBs to culture dishes may further indicate differentiation into neuronal cells.

An investigation by Dong *et al.* (2019) found *circACVR2A* to be lower in bladder cancer and correlates with aggressive clinicopathological characteristics. Both *in vitro* and *in vivo* data suggest that *circACVR2A* overexpression suppressed the proliferation, migration and invasion of bladder cancer cells by sponging miRNA *hsa-mir-626*, regulating *EYA4* expression. KD of *circACVR2A* with siRNAs was able to reverse this effect by increasing the above cellular characteristics. Wang *et al.* (2020) found that *circACVR2A* expression is also lower in Non-Small-Cell Lung Cancer (NSCLC) tissues, and expression of *circACVR2A* could inhibit *in vivo* tumour growth without altering *in vitro* proliferation. Mechanistically, they found that *circACVR2A* could sponge *hsa-mir-483-3p*, thereby promoting the expression of chemerin, a chemoattractant, which recruits subsets of immune cells, inhibiting *in vivo* tumorigenesis. Investigation into the levels of downstream targets of miRNAs with putative binding sites may give evidence for *circACVR2A* acting as a miRNA sponge, however generally low expression of circRNAs, and low number of miRNA binding sites, do not present with favourable conditions to effectively sequester miRNAs from other targets (Jarlstad Olesen and S. Kristensen, 2021).

5.2 Results

5.2.1 *CircACVR2A* EBs have altered adherence properties and neural differentiation

Embryoid bodies are 3D aggregates of differentiating stem cells and have the capacity to differentiate into all three cell lineages. KD of *circACVR2A* induced EB formation from H9 hESCs cultured on vitronectin-coated dishes. Additionally, EB formation was also observed with *circACVR2A* KD by shRNAs in male H1 hESC cells, however qRT-PCR validation of *circACVR2A* KD was not completed for this cell line (Appendix 5).

5.2.1.1 CircACVR2A knockdown embryoid bodies are non-adherent

Initial attempts to allow *circACVR2A* shRNA KD induced-EBs, to attach to glass slides revealed significant differences between these, and EBs generated through the standard low adherence culture method with an EV (pLKO.1) control cell line. The resultant EV EBs have the capacity to attach to glass slides if undisturbed, without the need for ECM. After 12 days of culturing *circACVR2A* shRNA 2 EBs and control EV EBs in individual glass slides, with media changes every 2 days, 12 of the 24 EV EBs adhered to the slide. In contrast, no *circACVR2A* shRNA 2 KD induced-EBs were adherent and remained suspended in the individual glass slide (Table 5-1).

Table 5-1 Adherence of H9 *circACVR2A* shRNA 2 EBs and EV EBs

Number of control EBs and H9 *circACVR2A* shRNA 2 EBs that were able to adhere to non-coated glass slides after 12 days of culture.

Cell line	Number of adherent EBs	Total number of EBs	Percentage
Empty vector control EBs	12	24	50%
CircACVR2A KD EBs	0	24	0%

5.2.1.2 CircACVR2A knockdown embryoid body differentiation on different ECMs

Although *circACVR2A* shRNA 2 EBs did not attach to uncoated slides, these EBs are able to attach to vitronectin-coated dishes, as well as dishes coated with geltrex, a more supportive extracellular matrix (Figure 5-1). Geltrex consists of laminin, collagen IV, entactin, and heparin sulfate proteoglycans. These ECM proteins allow for a greater variety of integrin binding that promote adhesion to the surface of the dish. Attachment of *circACVR2A* shRNA 2 EBs and control EBs to a culture dish surface, permits the observation of cells that grow away from the EBs so that the cellular composition of the EBs can be studied.

To investigate cell differentiation from *circACVR2A* shRNA 2 EBs, individual EBs were cultured in separate wells so the EBs would not be affected by differentiation signals from neighbouring EBs. Control EBs adhered to any surface (non-coated slides, Vitronectin and Geltrex) and deposited differentiated cells with a variety of morphologies observed on each surface type (Figure 5-1A, Figure 5-1C and Figure 5-1E). These include cells that show long axonic projections (Figure 5-1A), cells that expand as single cells (Figure 5-1B) and cells that expand with close cell-cell connections (Figure 5-1E). As mentioned previously, *circACVR2A* shRNA 2 EBs do not adhere to glass slides (Figure 5-1B), however, these EBs do adhere to freshly coated vitronectin and geltrex dishes. *CircACVR2A* shRNA 2 EBs adhered to vitronectin dishes and some cells surrounding the attached EB were observed (Figure 5-1D). An increased magnification revealed that axonal cells (white arrows) were observed within this population of cells surrounding the EB, however they were unable to expand further away from the EB. Of note, *circACVR2A* shRNA 2 EBs that adhered to geltrex, deposited a ring of cells, like on vitronectin ECM, however, neuronal-like cells also emerged within 24 hours of EB adherence and spread away from the attached EB. Continued culture of these cells did not yield any additional cell types, with these cells continuing to grow and migrate away from the EB (Figure 5-5F).

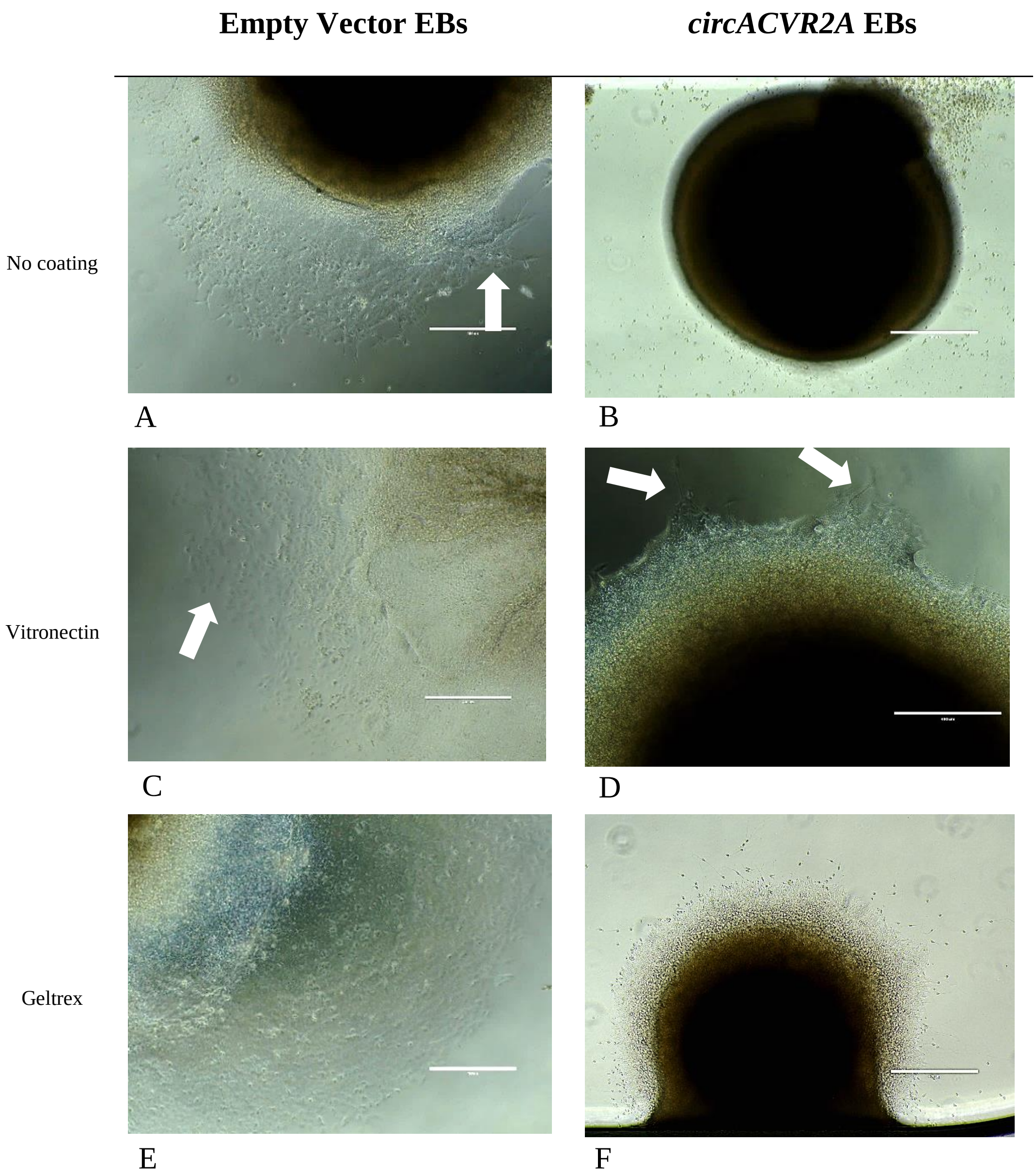
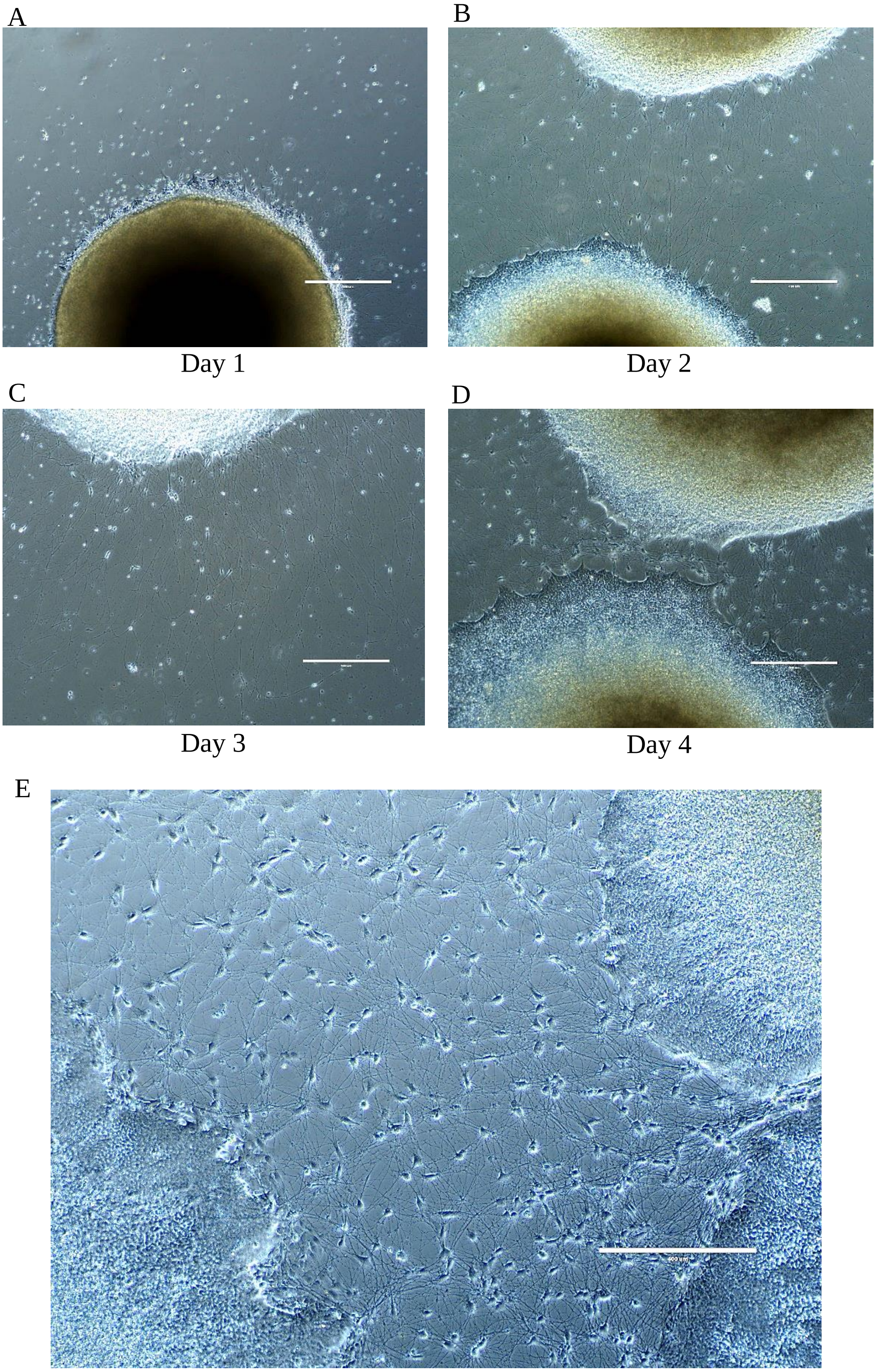


Figure 5-1 Embryoid Body attachment and outgrowth phenotypes on different binding matrices

Comparison of EB attachment and outgrowth of *circACVR2A* shRNA EBs to empty vector EBs on different matrices (non-coated, vitronectin and geltrex coated slides). (A) EV EBs adherent on no coating dishes show different cell morphologies. White arrow is indicating axonal projections. (B) *CircACVR2A* shRNA 2 EBs not adherent on non-coated dishes and stay suspended (C) EV EBs adherent on vitronectin and deposit different cell types. White arrow is indicating single cells. (D) *CircACVR2A* shRNA 2 EBs adherent to vitronectin and cells are observed to surround EB but expand very slowly. Scale bar is 400µm (E) EV EBs adherent on geltrex and deposit many cells types. (F) *CircACVR2A* shRNA 2 *EBs* adherent to geltrex and deposit cells with neuronal traits. Images were taken using the EVOS™ Imaging system. All scale bars are 1000µm except when specified.

5.2.1.3 Neural-like cell production by circACVR2A shRNA EBs

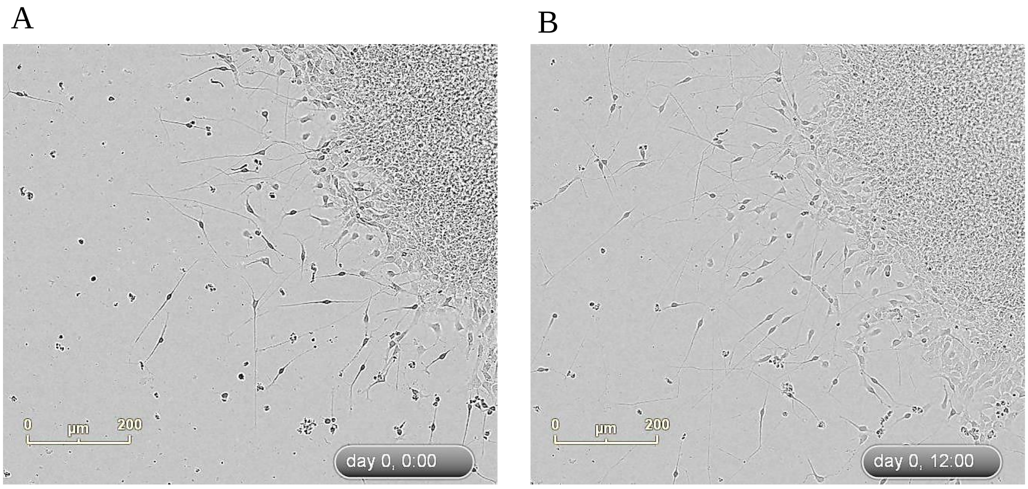
Both independent *circACVR2A* shRNA treatments of hESCs induced EB formation during culture of hESCs on vitronectin (Figure 5-1**,** Figure 5-2). To further image the cells that are deposited by the *shACVR2A* shRNA induced EBs, EBs were allowed to attach to geltrex coated dishes in the centre of the dish to allow for improved image capture under phase-contrast microscopy (Figure 5-2). On geltrex coated dishes, all *circACVR2A* shRNA 2 EBs would attach and within 1 day, small projections from cells could be observed on the dish (Figure 5-2A). These projections, and their cell bodies, move out from the adherent EB and many cell bodies are observed after 7 days on the geltrex coated dish (Figure 5-2E). This is reciprocated for both *circACVR2A* shRNA 2 (Figure 5-2) and shRNA 1 (Figure 5-3). Supporting the appearance of neural-like cells, these images look comparable to culturing methods performed by Francis and Wei, (2010) involving single cell plating of hESCs to from floating aggregates called neurospheres that are cultured with bFGF and the BMP inhibitor Noggin for 42 days (Figure 5-4).



Day 1
Day 2
Day 3
Day 4
Day 7

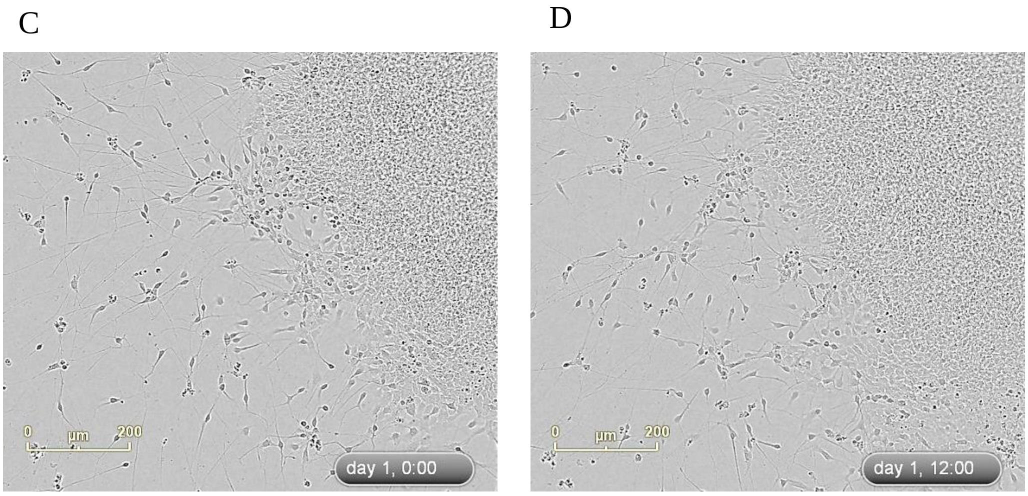
Figure 5-2 Neural-like outgrowth of *circACVR2A* shRNA 2 EBs

CircACVR2A shRNA 2 EBs attach to geltrex matrix and differentiate to produce neural-like cells. Images were taken on (A) day 1 (B) day 2 (C) day 3 (D) day 4 (E) Day 7. Images were taken using the EVOS™ Imaging system.


A
B
0
µm
200
day 0, 0:00
0
µm
200
day 0, 12:00

24 hours after adherence

36 hours after adherence


C
D
0
µm
200
day 1, 0:00
0
µm
200
day 1, 12:00

48 hours after adherence

60 hours after adherence

Figure 5-3 Neural-like outgrowth of *circACVR2A* shRNA 2 EBs

CircACVR2A shRNA 1 EBs attach to geltrex matric and differentiate to produce neural-like cells. Images were automatically taken after adherence at (A) 24 hours (B) 36 hours (C) 48 hours and (D) 72 hours. Images were taken using the IncuCyte® FLR system.

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 5-4 Differentiation of neural cells from Neurosphere

Neural differentiation of UCO6 hESCs by Francis and Wei, 2010. (f) Twenty-eight day floating neurosphere. (g) Cells exhibiting increasing neural morphology after (g) 1 day and (i) 14 days. (j) cells exhibit neuritic extensions and dendritic spine formation (white arrows) (Image: (Francis and Wei, 2010)).

5.2.1.4 QRT-PCR analysis of EBs and outgrown neural-like cells

QRT-PCR of *circACVR2A* shRNA induced EBs for pluripotency and neural transcripts was kindly performed by Jenne Tran from Associate Professor Cedric Bardy’s laboratory at South Australian Health and Medical Research Institute (SAHMRI). Early *circACVR2A* shRNA EBs, 5 days post transduction, as well as late *circACVR2A* shRNA EBs (attached colonies and neuronal cells separately), RNA samples were given to SAHMRI for qRT-PCR analysis (Table 5-2).

Table 5-2 Samples for qRT-PCR analysis by SAHMRI

List of RNA samples given to SAHMRI for qRT-PCR analysis of pluripotency and neural transcripts.

Samples for qRT-PCR by SAHMRI
H9 EV (undifferentiated)
H9 *circACVR2A* shRNA 1 early EBs (5 days post transduction)
H9 *circACVR2A* shRNA 1 attached EBs (17 days post transduction)
H9 *circACVR2A* shRNA 1 neurons (17 days post transduction)
H9 *circACVR2A* shRNA 2 early EBs (5 days post transduction)
H9 *circACVR2A* shRNA 2 attached EBs (17 days post transduction)

These qRT-PCR results shows that the H9 EV cells used in this study had higher pluripotency transcripts (*OCT4* and *NANOG*) than the hESCs used for comparison for the qRT-PCR (Figure 5-5). Therefore, changes to transcripts by qRT-PCR should be analysed relative to H9 EVs and not the H9 ESC guide control included in the graph. *OCT4* and *NANOG* pluripotency markers were reduced in all EB samples, as well as attached EBs and outgrown neural-like cells, indicating a loss of stemness during differentiation. *SOX2* embryonic stem cell marker was slightly reduced in most EBs, outgrown colonies and individual neuronal-like cells, however was not lost, similarly to the neural cells included in the analysis. Neuronal markers *MAP2* and *SLC17A6* (glutamatergic neuron marker) were detected in the *circACVR2A* shRNA 1 EBs that had attached to geltrex ECM and also in the outgrown neural-like cells. Additionally, *MAP2* was also detected in *circACVR2A* shRNA 2 attached EBs, however, this was to a lesser extent. The neural stem cell marker *NES* and mature neuron marker *RBFOX3* showed minimal change in all treatments, while *SYN1* (neuronal development marker) was detected in all *circACVR2A* EBs and outgrown cells, and with expression levels comparable to Neural progenitors. Lastly, *GFAP*, an astrocyte specific marker was not detectable, and was only detectable in the day 30 and day 60 neuron samples supplied by SAHMRI.

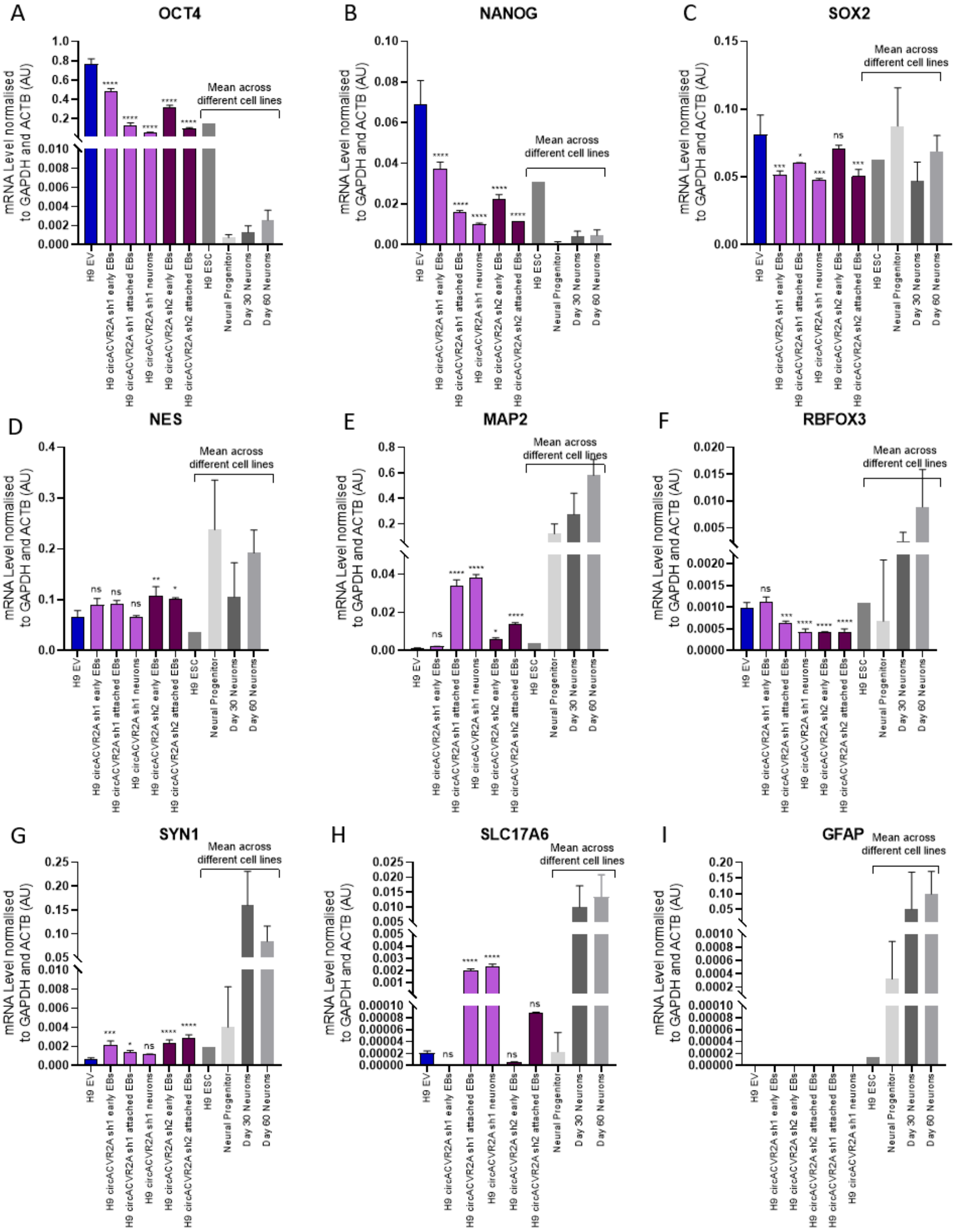


Figure 5-5 Pluripotency and neuronal specific QRT-PCR analysis of *circACVR2A* shRNA induced EBs

QRT-PCR analysis of (A-C) pluripotency and (D-I) markers of different cells neural cells of *circACVR2A* shRNA induced EBs. Relative to *GAPDH* and *ACTB*. qRT-PCR was performed using technical triplicates. Data is displayed as mean with standard deviation. Statistical significance was determined by one-way ANOVA with multiple comparisons using Dunnett's test. (* = p<0.5, ** = p<0.01, *** = p<0.001, **** = p<0.0001, ns = not significant) Performed by Jenne Tran at SAHMRI.

5.2.1.5 Embryoid body immunofluorescent analysis

To further confirm the differentiation of the *circACVR2A* shRNA EBs, at the protein level, immunofluorescence for EV and *circACVR2A* shRNA 2 EBs was performed. EV EBs and *circACVR2A* shRNA 2 EBs were fixed with 4% paraformaldehyde and paraffin embedded. 5 µm sections of each sample were placed onto glass slides for immunofluorescent analysis. As EBs generated from the KD of *circACVR2A* did not appear until approximately 3 days post transduction, *circACVR2A* shRNA 2 EBs at 5 days post transduction ("early" *circACVR2A* shRNA 2 EBs) were found to be a similar size to 2 day post low adherence seeding H9 EV EBs ("early" H9 EV EBs). Therefore, despite the differences in the generation of the EV control and *circACVR2A* KD EBs, similar sized EBs were compared during immunofluorescence (Figure 5-6). EBs were matured for a further 11 days and were subsequently fixed and sectioned for a late stage EB ("Late" H9 EV EBs and "Late" *circACVR2A* shRNA 2 EBs).

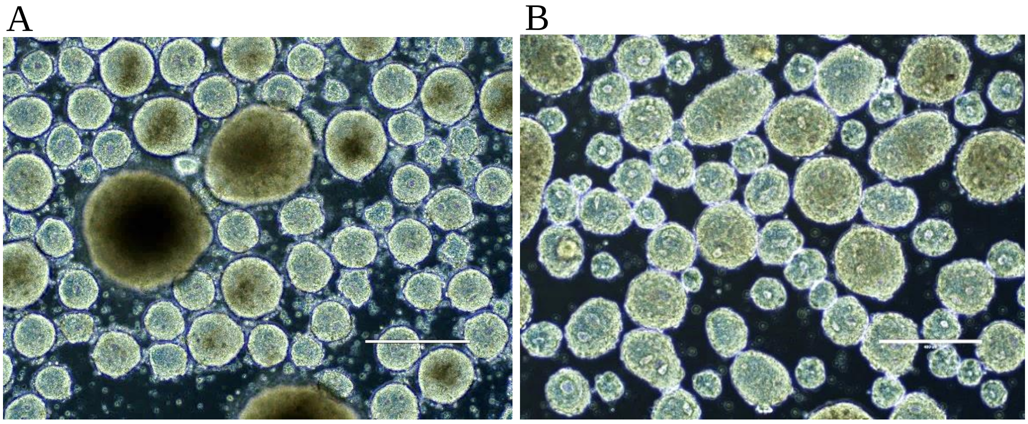


Figure 5-6 EV EBs and *circACVR2A* shRNA 2 EBs are similar sizes for Immunofluorescence comparisons

(A) H9 EV EBs at 2 days post low adherence seeding compared to (B) 5 days post transduction H9 *circACVR2A* shRNA 2 EBs.

Table 5-3 Maturation of control and *circACVR2A* knockdown EBs used for Immunofluorescence

List of H9 EV EBs and H9 *circACVR2A* shRNA 2 EBs that were embedded in paraffin and sectioned for immunofluorescent staining.

Samples for Immunofluorescence
H9 EV early EBs (2 days post low adherence)
H9 EV late EBs (13 days post low adherence)
H9 *circACVR2A* shRNA 2 early EBs (5 days post transduction)
H9 *circACVR2A* shRNA 2 late EBs (16 days post transduction)

Initially sections were stained for NANOG and POU5F1 (OCT3/4) to confirm if *circACVR2A* shRNA 2 EBs lost pluripotency factors during maturation and also to discover if these EBs show altered immunostaining for multiple markers compared to canonically produced H9 EV EBs. Master pluripotency regulator NANOG expression in early *circACVR2A* shRNA 2 EBs (Figure 5-7), was 62.3% lower than H9 EV control EBs. POU5F1 was still expressed in the nucleus of cells within early EV EBs and early *circACVR2A* shRNA 2 EBs (Figure 5-8). POU5F1 was also expressed in late *circACVR2A* shRNA 2 EBs, however H9 EV control EBs were 61.4% lower in POU5F1 expression compared to shRNA 2 *circACVR2A* EBs.

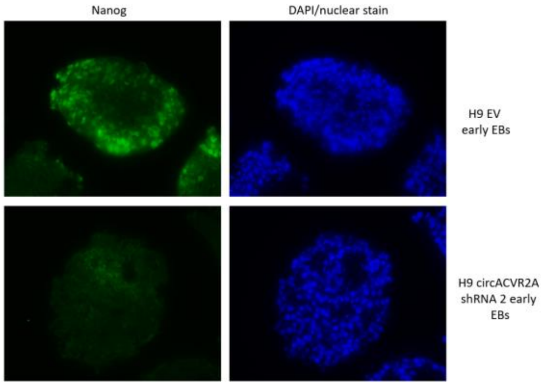


Figure 5-7 NANOG pluripotency factor is decreased in early *circACVR2A* shRNA 2 EBs

Immunofluorescence staining for NANOG in early H9 EV EBs and H9 *circACVR2A* shRNA 2 EBs. Images were taken on the Olympus AX70.

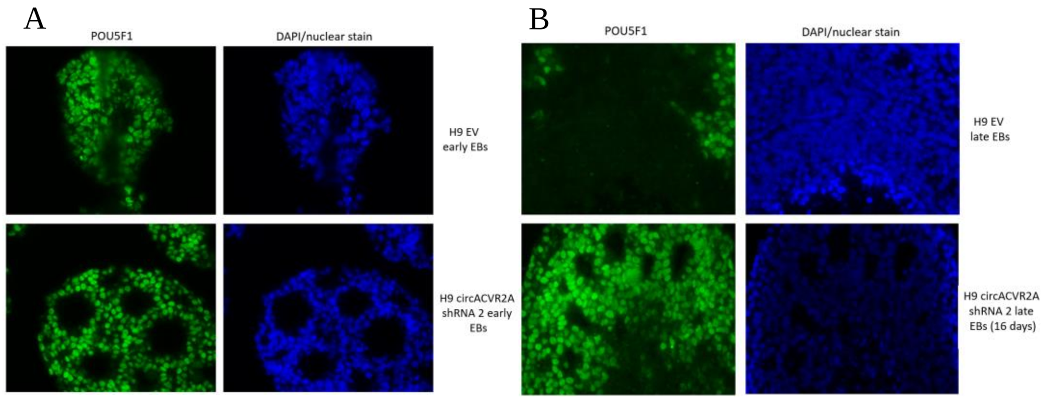

A
POU5F1
DAPI/nuclear stain
H9 EV
early EBs
H9 circACVR2A
shRNA 2 early
EBs
B
POU5F1
DAPI/nuclear stain
H9 EV
late EBs
H9 circACVR2A
shRNA 2 late
EBs (16 days)

Figure 5-8 POU5F1 pluripotency factor is retained in late *circACVR2A* shRNA 2 EBs

Immunofluorescence of (A) early and (B) late (16-days post transduction) H9 EV EBs and H9 *circACVR2A* shRNA 2 EBs for POU5F1 (OCT3/4). Images were taken on the Olympus AX70.

Analysis of immunofluorescent staining found a 52% higher expression of FOXA2 midbrain floorplate marker protein in early *circACVR2A* shRNA 2 EBs that is not present in control EV EBs. FOXA2 expression in late shRNA 2 circACVR2A EBs decreases to be only 29% higher than in late EV*circACVR2A* EBs (Figure 5-9).

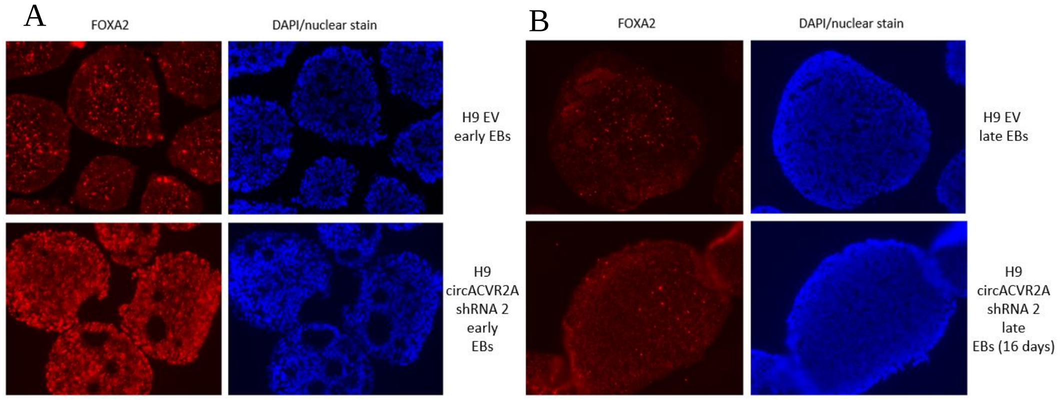

A
FOXA2
DAPI/nuclear stain
H9 EV
early EBs
H9
circACVR2A
shRNA 2
early
EBs
B
FOXA2
DAPI/nuclear stain
H9 EV
late EBs
H9
circACVR2A
shRNA 2
late
EBs (16 days)

Figure 5-9 FOXA2 midbrain floorplate marker is expressed in early *circACVR2A* shRNA 2 EBs

Immunofluorescence of FOXA2 of (A) early and (B) late (16-days post transduction) EV control EBs and *circACVR2A* shRNA 2 EBs. Images were taken on the Olympus AX70.

5.2.2 RNA-seq profiling indicates *circACVR2A* knockdown embryoid bodies have undergone neural differentiation

5.2.2.1 Total-RNA sequencing of EBs

Interestingly, KD of *circACVR2A* has induced EB formation and differentiation into neuron-like cells with qRT-PCR and immunofluorescence indicating the presence of neural cell types within 5 days post lentiviral transduction. Whether EMT is occurring during EB formation and EB maturation, is still unknown. To identify changes to EMT transcripts and to determine the transcriptional consequences of *circACVR2A* KD in H9 hESCs, total RNA-seq was performed on early and late *circACVR2A* shRNA EBs and also 1 replicate of H9 *circFAT3* and *circCDYL* KD cell lines. *CircACVR2A* shRNA 1 treatment of H9 hESCs did not produce as many EBs as shRNA 2, therefore some colonies were still attached to the bottom of the culture dish when the cells were harvested for total-RNA seq. Total RNA seq was performed on cells either as single samples or in duplicates (Table 5-4). To understand how normal embryoid bodies develop and to identify how the H9 *circACVR2A* KD EBs may differ, early EV EBs were generated by canonical methods, via plating cells in low adherence, and including these for sequencing. Total RNA sequencing was chosen over mRNA-seq to capture changes to circRNA levels during shRNA KDs of hESCs, however, few circRNAs were identified and did not give quantifiable data with only single digit backsplice reads identified.

Table 5-4 List of samples for total-RNA sequencing

Sample names	Number of replicates
H9 hESC EV control	2

H9 EV early EB (2 days post low adherence seeding)	1
H9 early *circACVR2A* shRNA1 (5 days post transduction)	2
H9 early *circACVR2A* shRNA2 (5 days post transduction)	2
H9 late *circACVR2A* shRNA1 (18 days post transduction)	1
H9 late *circACVR2A* shRNA2 (19 days post transduction)	1
H9 *circCDYL* shRNA 1	1
H9 *circFAT3* shRNA 1	1

5.2.2.2 RNA-seq Analysis of H9 circACVR2A knockdown EBs

Early H9 EV EBs were included in the analysis to determine if *circACVR2A* shRNA EBs were undergoing transcriptional changes that are observed during canonical EB formation. The number of statistically differentially expressed transcripts in each sample was determined. It was found that *circACVR2A* shRNA 2 treatment had more differentially expressed transcripts than shRNA 1, most likely due to some cells remaining as colonies in the shRNA 1 treatment cells (Table 5-5).

Table 5-5 Number of statistically differentiated genes from H9 EV control (p<0.05)

Samples	Increased differentially expressed transcripts	Decreased differentially expressed transcripts	Total differentially expressed transcripts
H9 early EV EBs	147	1669	1816
H9 early *circACVR2A* sh1 EBs	344	721	1065
H9 early *circACVR2A* sh2 EBs	910	1344	2254
H9 late *circACVR2A* sh1 EBs	529	151	680
H9 late *circACVR2A* sh2 EBs	1717	1155	2872

Early shRNA 1 and shRNA 2 *circACVR2A* EBs show no change to pluripotency markers except for a decrease to NANOG expression in shRNA 2 EBs (Figure 5-10) also observed at the protein level by immunofluorescence. Other pluripotency markers decrease in expression over time in late *circACVR2A* shRNA 2 EBs. *POU5F1* (*OCT3/4*), a master regulator of stem cell pluripotency, does not change in expression for either shRNA 1 or shRNA 2, however this is also observed by immunofluorescence as the protein is retained in late *circACVR2A* shRNA 2 EBs.

Differential expression analysis shows that there is an increase in mesenchymal markers (Figure 5-10), including Vimentin (*VIM*) and Zinc Finger E-Box Binding Homeobox 2 (*ZEB2* also known as Smad-interacting protein 1 or *SIP1*) in both early and late stage EBs. This is reciprocated with a decrease in epithelial markers (Figure 5-10) including E-Cadherin (*CDH1*) and Epithelial Splicing Regulatory Protein 1 (*ESRP1*). These transcriptional changes are seen to a greater extent in *circACVR2A* shRNA 2 treatment which produced embryoid body-like aggregates earlier than treatment with shRNA 1. These transcriptional changes are further enhanced in the EB-like aggregates that were allowed to culture up to 18-19 days post transduction. Therefore, both shRNA treatments show an increase to mesenchymal markers and shRNA 2 EBs displaying a decrease to epithelial markers.

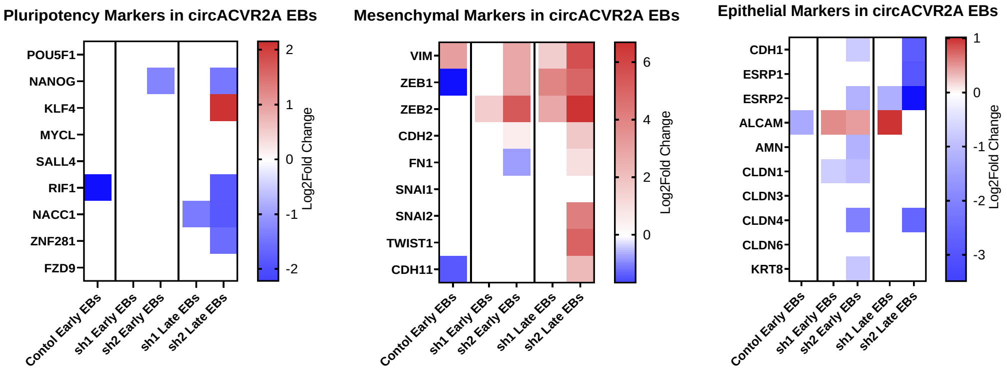


Figure 5-10 Changes to pluripotency and EMT transcripts through RNA-seq with *circACVR2A* shRNA knockdown in hESCs

Heat maps of Log2Fold changes of Pluripotency, Mesenchymal and Epithelial markers in early and late EBs from KD of *circACVR2A* compared to Empty vector control hESCs. Coloured squares are statistically significant Log2Fold changes after normalisation.

To further support findings of neural cell differentiation, analysis of lineage markers for hESCs, and the 5 other lineages investigated, were included to determine if any other lineages are present in the differentiating shRNA1 and shRNA 2 circACVR2A EBs. Of note, DE or Cardio markers were not upregulated, however, Liver, FP and Dopa were upregulated compared to EV H9 hESCs (Figure 5-11). Specifically, in late shRNA2 EBs, only neuronal and neuronal precursor (FP) markers were upregulated with an absence of any other lineages upregulated. Therefore, pluripotency markers are downregulated, EMT markers support an increase in expression to mesenchymal, with a reciprocal decrease to epithelial genes, and specific upregulation of neuronal and neuronal precursor markers.

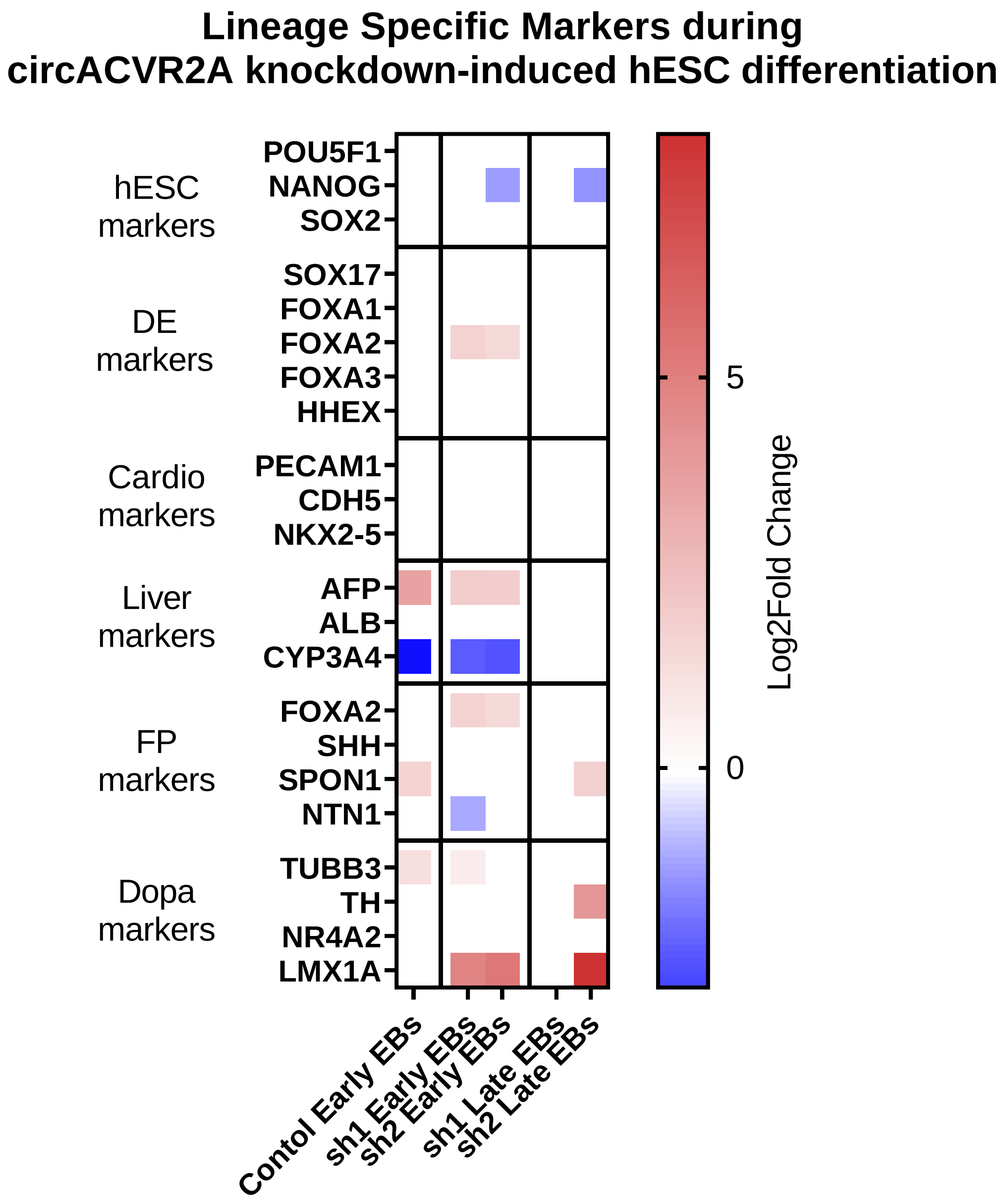


Figure 5-11 Changes to lineage specific transcripts through RNA-seq with *circACVR2A* shRNA knockdown in hESCs.

Heat maps of Log2Fold changes of lineage specific markers in early and late EBs from KD of *circACVR2A* compared to Empty vector control hESCs. Coloured squares are statistically significant Log2Fold changes after normalisation.

5.2.2.3 Gene Ontology Enrichment analysis

Towards ascribing biological significance to these transcriptional changes, a gene ontology (GO) enrichment analysis was undertaken. Utilising the RNA-seq data, a list of commonly differentially expressed mRNA transcripts (observed to change in both shRNA treatments compared to the H9 EV control) was generated. As stated in chapter 2, non-adjusted p-values ($p<0.05$) were used to determine whether expression changes were significant compared to the empty vector control cells. This is especially important when only using 2 replicates for each treatment as statistical power is lower than when using 3 replicates. Using the list of differentially expressed genes, GO enrichment analysis was performed (http://geneontology.org/). This identifies biological processes that are enriched in both shRNA treated cells that form EBs based on a list of gene names.

GO analysis of statistically increased differentially expressed genes, observed in both shRNA treatments for *circACVR2A* KD, has shown an enrichment for genes involved with the regulation of neuron development/differentiation (Figure 5-12). This includes GO terms for brain development, neuron projection morphogenesis and cell morphogenesis involved in neuron differentiation. Interestingly, genes involved in response to transforming growth factor beta (TGFβ) stimulus are enriched, the pathway with which *circACVR2A's* parental gene, *ACVR2A*, is involved. The largest enrichment is for negative regulation of cardiocyte differentiation and suggests an inhibition to cardiocyte differentiation, possibly as cells are undergoing neuronal differentiation. GO analysis of statistically decreased differentially expressed transcripts shows a decrease to cardiac processes including cardiac muscle contraction, circulation and angiogenesis (Figure 5-12), further supporting an inhibition of cardiac processes during neural differentiation.

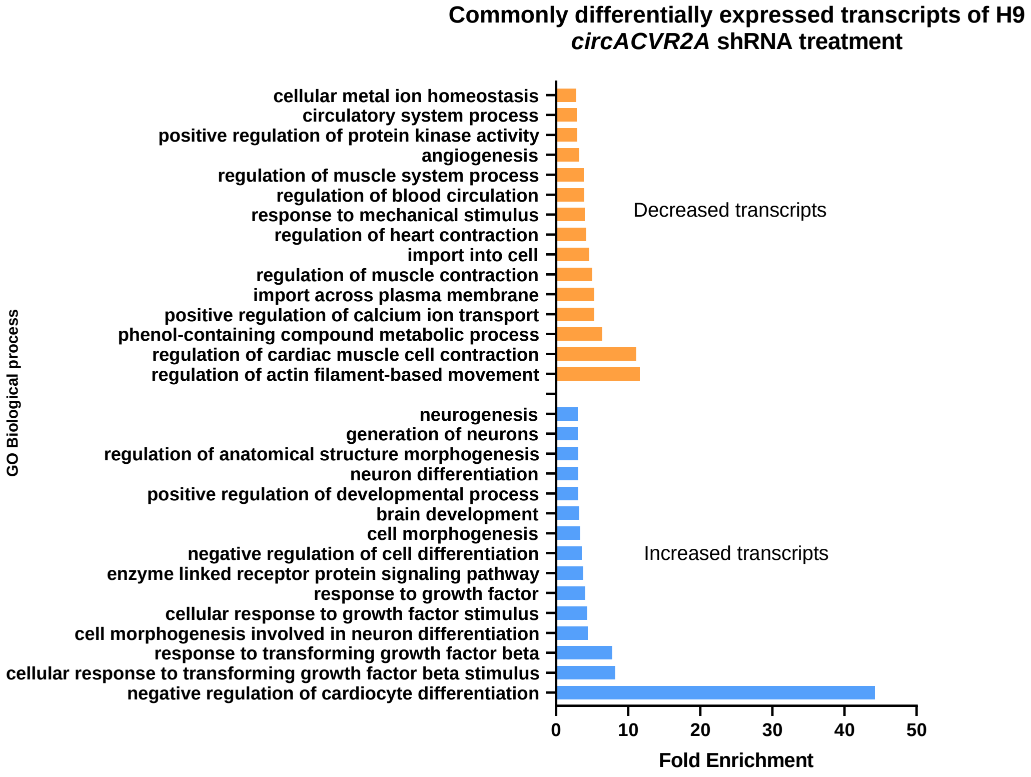
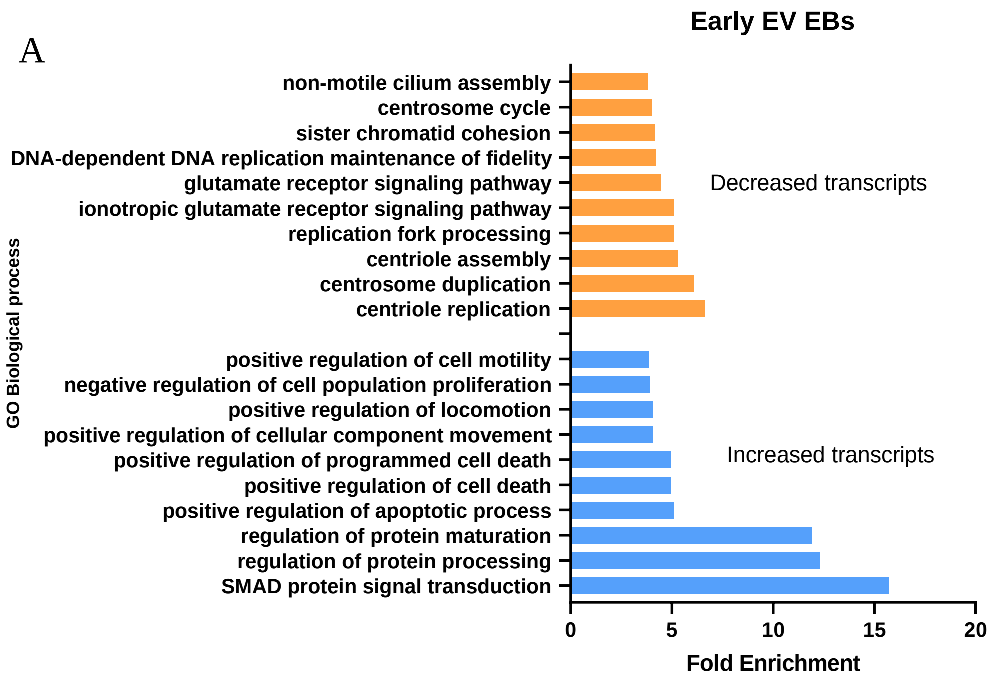


Figure 5-12 GO analysis of differentially expressed RNA transcripts common to both shRNA treatments compared to H9 EV hESCs

GO analysis of 167 significantly increased differentially expressed RNA transcripts ($p<0.05$) and 411 commonly significantly decreased RNA transcripts common to both shRNA treatments of *circACVR2A* KD compared to H9 EV hESCs. Displaying the top 15 GO terms only.

In contrast, empty vector EBs do not show high enrichment for any particular lineages from H9 EV hESCs, instead changes to apoptotic processes, cellular motility and proliferation are enriched (Figure 5-13A). Downregulated transcripts also show no suggestion of any change to lineage specific genes during this stage of EB maturation, while GO terms for critical cellular processes including DNA and chromosome maintenance are enriched (Figure 5-13A).

As *circACVR2A* shRNA 2 treatment is more potent at generating EBs, as nearly all cells form EBs 3 days after transduction, GO analysis was performed for shRNA 2 for both early and late EBs individually. GO analysis of early *circACVR2A* shRNA 2 EBs strongly suggests neurogenesis and brain development are enriched processes compared to undifferentiated hESCs, while GO terms associated with the development of other lineages (Megakaryocyte, bone, smooth muscle) are enriched in significantly decreased transcripts (Figure 5-13B). Late stage EBs follow this trend, with extensive neuron-specific GO terms enriched in increased differentially expressed RNA transcripts. Interestingly, some GO terms associated with Kidney formation (Nephron tube identity, Renal system specification) are enriched in late *circACVR2A* shRNA 2 EBs (Figure 5-13C). It is known that there are some neural genes (*ELAVL3*, *ELAVL4*, *NCAM1*, etc.) that are expressed during kidney organoid culture and some evidence to suggest neural precursors being present during human kidney development (Combes *et al.*, 2019). In concordance with the GO analysis of both shRNA treatments, cardiac muscle processes are still heavily enriched in the downregulated RNA transcripts (Figure 5-13C).



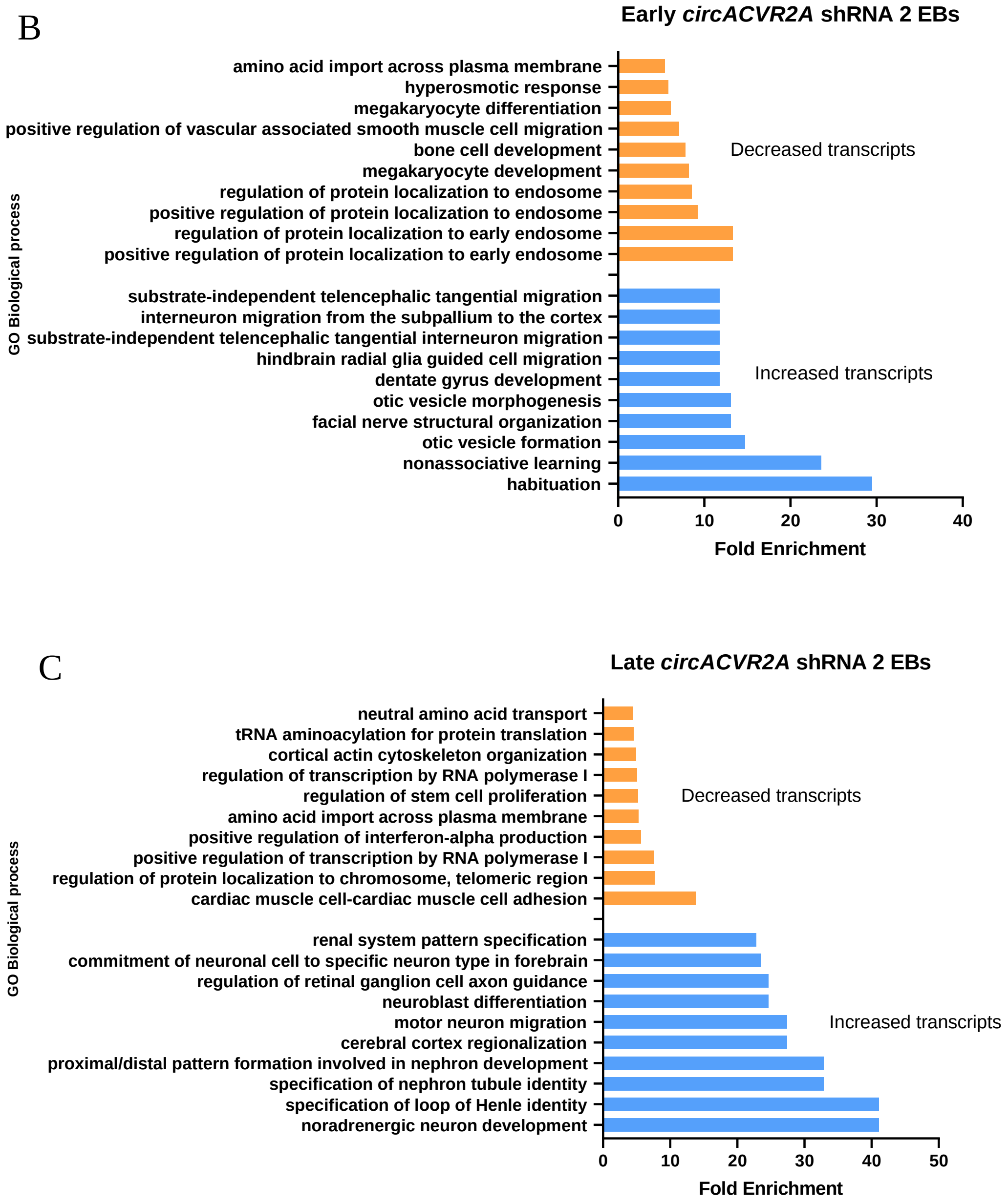


Figure 5-13 GO analysis of differentially expressed RNA transcripts of EV control EBs, and early and late *circACVR2A* shRNA 2 EBs

GO analysis of significantly differentially expressed RNA transcripts ($p<0.05$) of (A) EV control EBs, *circACVR2A* shRNA 2 (B) early and (B) late EBs from H9 EV hESCs. Only showing the top 10 GO terms.

5.2.2.4 CellNet analysis of H9 circACVR2A shRNA EBs

CellNet (Cahan *et al.*, 2014) is a bioinformatics tool that enables the analysis of RNA-seq or microarray data to give insight into Gene Regulatory Networks (GRNs) expressed by the cells or tissues of interest. CellNet is predicated on the discovery of GRNs that act as major molecular determinants of cell-type identity (Davidson and Erwin, 2006). CellNet measures the establishment of cell and tissue-specific GRNs to classify samples by their similarity to target cells and tissues, however this is restrictive and only certain cell and tissue-specific GRNs are included in the analysis.

CellNet is able to complete a stringent analysis of RNA-seq data comparing to a set of training data that, although is tissue specific, comprises of heterogenous populations of cells. This primary analysis scores samples which indicate the extent to which a query sample is indistinguishable in their expression profile from each of the reference C/T types. To test that the CellNet analysis can distinguish different cell types, the mRNA-seq data that comprises the 5 differentiated cell types and hESCs was included in the analysis (Figure 5-14). This analysis has shown that the classification is able to distinguish some cell types however some cell lines show multiple cell type classifications. Dopaminergic Neurons is the only differentiated cell type that selectively classifies with the correct cell type. Cardio cells show some classification of heart cells while also still showing some ESC classification. Liver cells show classification for liver GRNs however also shows some intestine_colon classification. Floorplate Precursors (FP) do not match with any cell type as this cell type is not within the training data for CellNet, however they have less ESC classification. Endoderm cells also do not classify as anything other than ESCs as this is not a cell type that is within the training data however less ESC classification is observed. Overall, CellNet is able to predict the cell type that is being analysed as long as the training data includes the cell type being analysed. There is some cross over of GRNs with other cell types (Hepatocytes and

Cardiomyocytes), however this may just be due to overlapping GRNs, methods for how the cells were differentiated, or due to the heterogeneity of the training data.

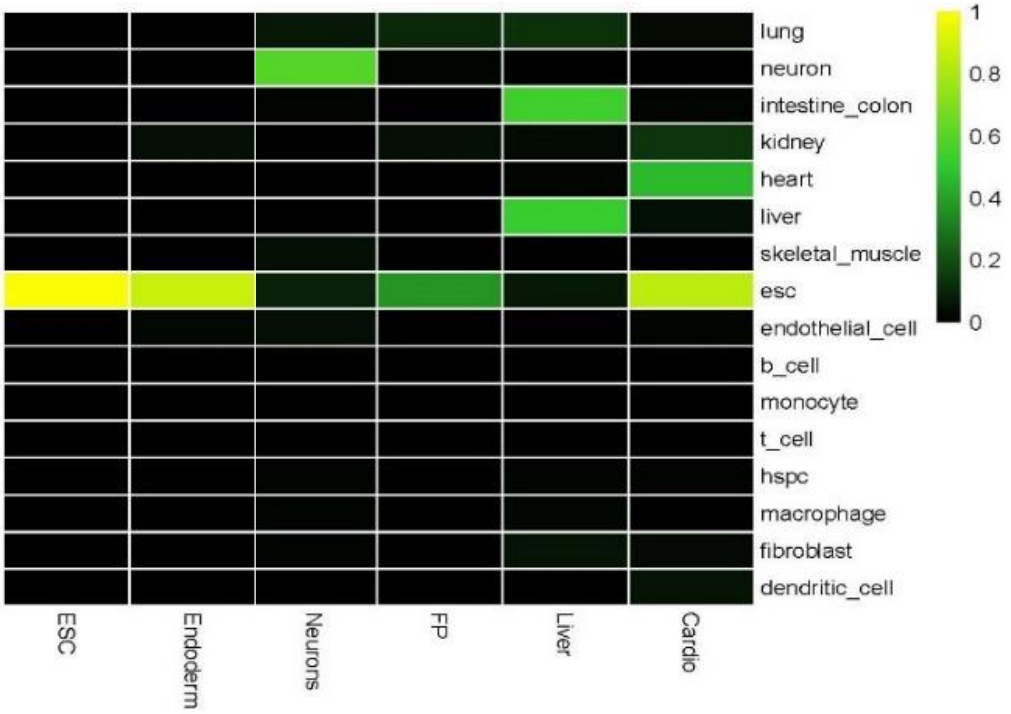


Figure 5-14 CellNet analysis of differentiate cells used in mRNA-seq
CellNet analysis of hESCs and differentiated cell types from mRNA-seq data.

CellNet analysis of the total RNA-seq data for the *circACVR2A* shRNA EBs shows no increase in other classifications of other cell types, however there are observable decreases to the ESC status of these EBs compared to EV hESCs as the samples observed within the heatmap (Figure 5-15). As stated, this analysis is determining if the query samples are indistinguishable to the training data and therefore these samples still classify as ESCs with partial reduction in ESC status.

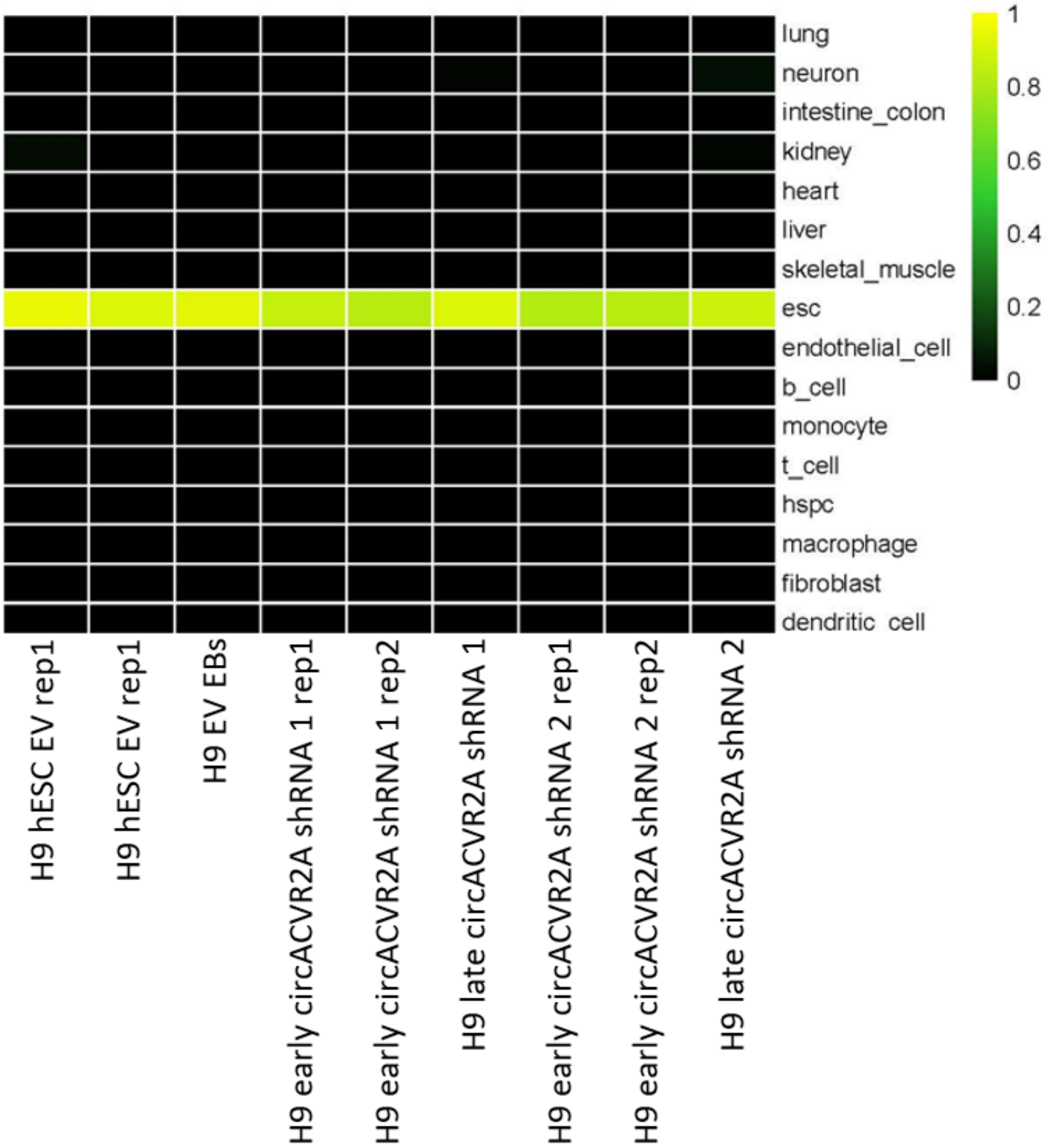


Figure 5-15 CellNet analysis of *circACVR2A* shRNA knockdown EBs

CellNet analysis of hESCs and *circACVR2A* shRNA EBs. Analysis also includes RNA-seq data for shRNA1 of *circFAT3* and *circCDYL*.

Further analysis can be performed by CellNet as the program is able to measure the extent to which a cell and tissue (C/T)-specific GRN is established in a query sample relative to the corresponding C/T in the training data. Late c*ircACVR2A* shRNA 2 EBs shows increased correlation with neuronal GRNs (Figure 5-16) and supports that the cells differentiating in the EB include neuronal cell types. This sample also showed an increased GRN for Kidney cells. As stated earlier, Kidney and neural cells share some overlapping transcriptional changes (Combes *et al.*, 2019). Furthermore, a quantifiable increase to kidney GRNs is observed in mouse ESC-derived neural precursors using CellNet analysis (Cahan *et al.*, 2014). Together, neural morphology, qRT-PCR analysis of neural genes, IF, GO analysis and CellNet analysis strongly indicates that *circACVR2A* shRNA EBs are producing neural precursors.

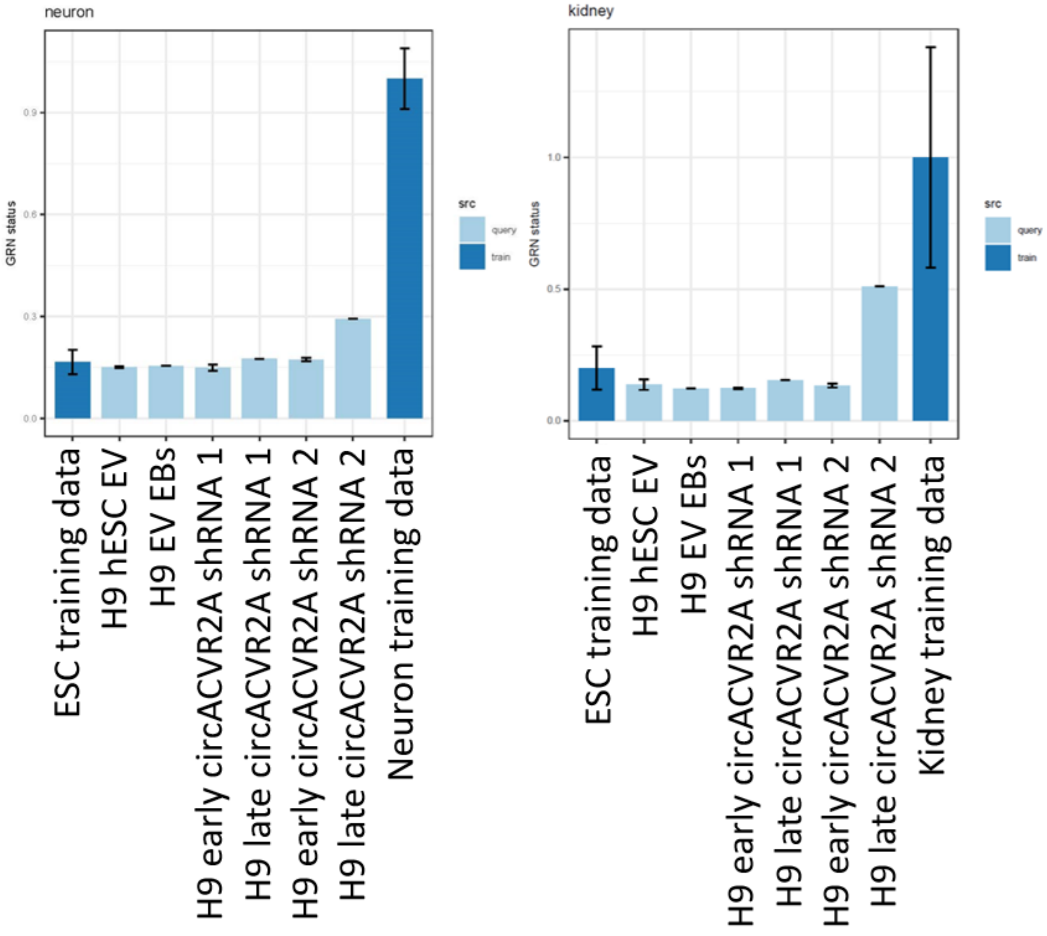


Figure 5-16 CellNet analysis of GRN status of *circACVR2A* shRNA knockdown EBs

CellNet analysis of displaying the extent of (left) Neuronal and (right) Kidney GRNs present in the *circACVR2A* shRNA EBs. Analysis also includes RNA-seq data for shRNA1 of *circFAT3* and *circCDYL*.

5.2.2.5 Neuronal markers by RNA-seq

For a direct transcriptional analysis of neuronal differentiation, an array of neuronal genes was obtained from the Abcam website (Abcam, no date), a supplier of biological reagents and antibodies. Analysis of the RNA-seq data for these neural genes has revealed an array of statistically increased neural transcripts in late *circACVR2A* shRNA 2 EBs (Figure 5-17). Of the 68 neural genes included in the heat map, 21 transcripts were increased while on only 6 were decreased. Of note, both intermediate progenitor transcripts were increased, with this population of cells known to produce multiple neuron types (Martínez-Cerdeño, Noctor and Kriegstein, 2006). Markers for multiple neuronal cells are also seen to increase including immature neurons, mature neurons, glutamatergic, GABAergic and dopaminergic neurons. No astrocyte transcripts were observed to increase suggesting that the differentiating cells are favouring the production of neurons and not astrocytes or oligodendrocytes. Probing for the differentiation expression of neural transcripts further supports neuronal differentiation occurring after the reduction of *circACVR2A* in hESCs.

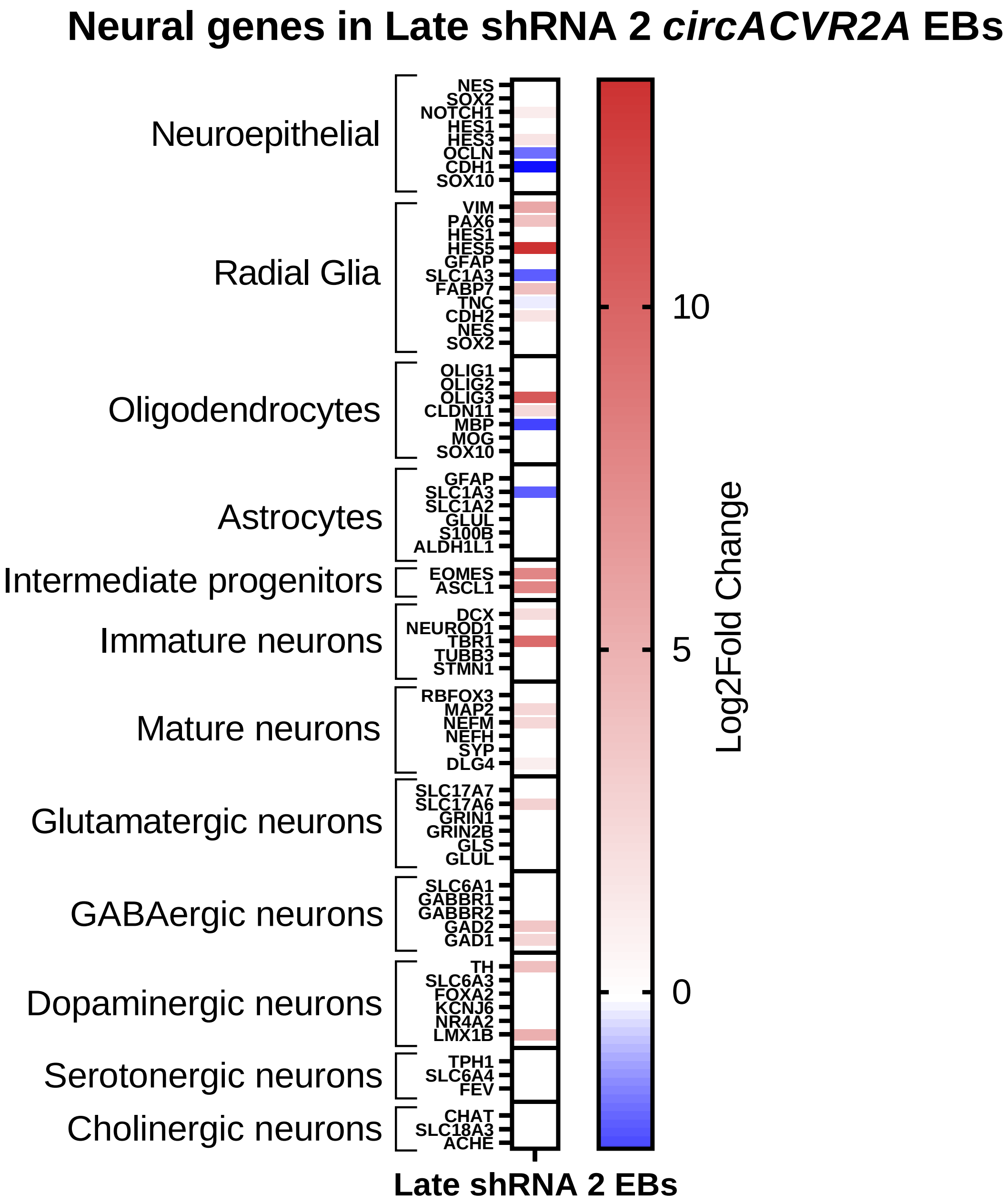


Figure 5-17 Changes to neural RNA transcripts in late *circACVR2A* shRNA EBs

Statistical increases ($p<0.05$) to many neural genes in late *circACVR2A* shRNA2 EBs in RNA-seq data set. Data is only given if the change from H9 hESCs is statistically significant ($p<0.05$).

5.2.2.6 Adhesion markers by RNA-seq

The first phenotypically significant phenotype observed after the lentiviral transduction of hESCs with shRNAs targeting *circACVR2A*, cultured on Vitronectin and not on Geltrex coated dishes, is the production of EBs that lose adherence and detach from the coated tissue culture dish. RNA-seq analysis of adherence molecules that support attachment to matrices suggest that in the early stages of EB formation, several adherence gene transcripts are lost while others are gained. ITGA2, ITGB5 and PCDH12 are decreased in both shRNA treatments in early *circACVR2A* KD EBs while ITGA8 and ITGB8 are increased (Figure 5-18). Late *circACVR2A* shRNA EBs have a different adhesion profile to early EBs with increases to several Protocadherins and cell adhesion molecules in both shRNA treatments. Of note, ITGA8, an integrin known to be highly expressed in neural lineages, is upregulated in both shRNA treatments in late *circACVR2A* shRNA EBs and is the only adhesion gene transcript that is consistently and uniformly differentially expressed in early and late circACVR2A shRNA EBs. Additionally, ITGB5, a required integrin for cells binding to Vitronectin, has reduced transcript levels in early *circACVR2A* shRNA EBs, offering an explanation for the conflicting phenotypes observed when performing the transduction on Vitronectin and Geltrex.

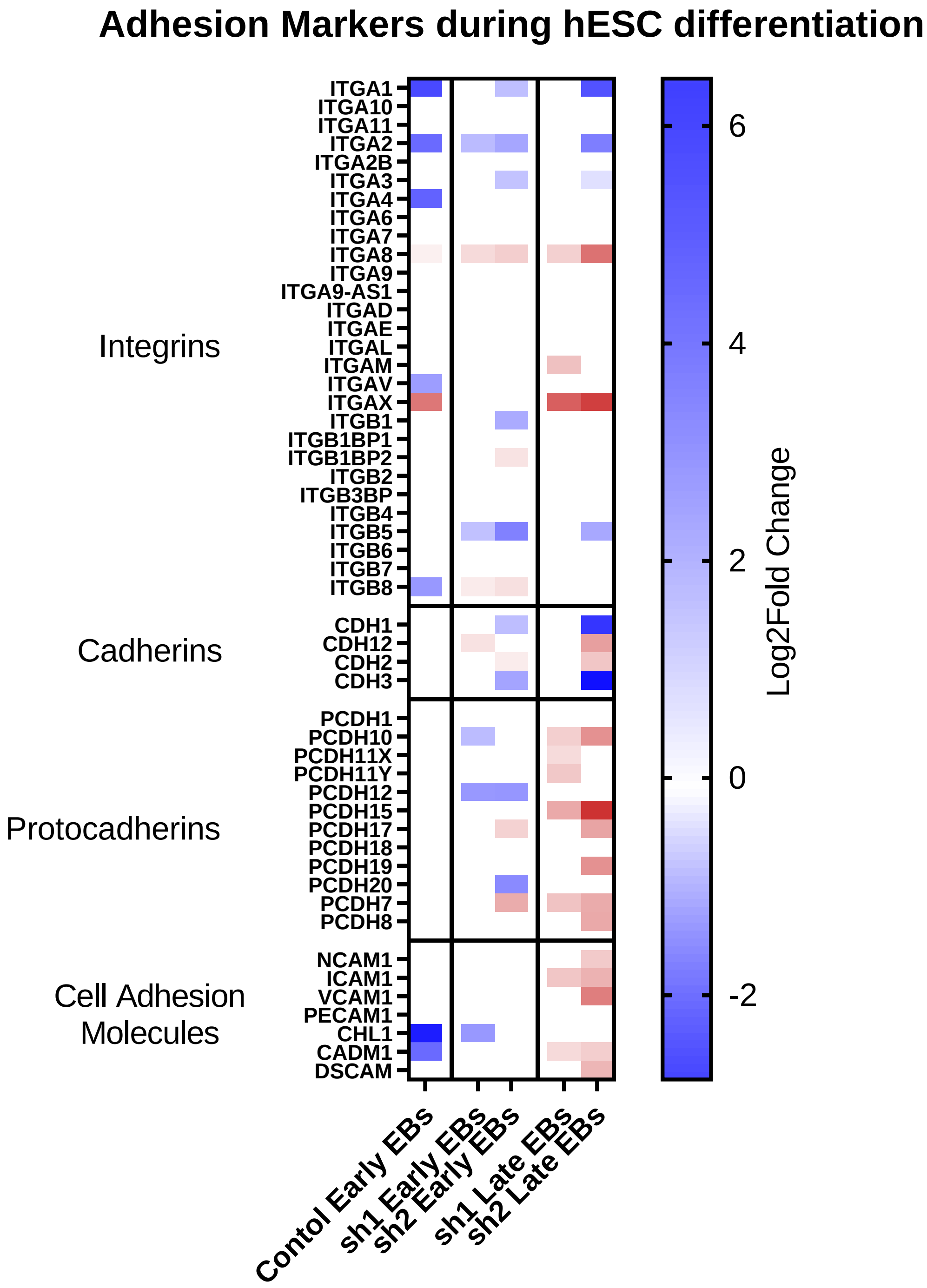


Figure 5-18 Changes to Adhesion RNA transcripts in *circACVR2A* shRNA EBs.

Differential expression of Adhesion molecule transcripts in shRNA 1 and shRNA 2 *circACVR2A* shRNA EBs from EV H9 hESCs in total RNA-seq data set. Data is only given if the change from H9 hESCs is statistically significant (p<0.05).

5.2.2.7 Changes to hESC maintenance and differentiation signalling pathways

As activation or inhibition of multiple signalling pathways can promote stemness or induce differentiation of stem cells, the transcriptional profile of WNT, FGF and TGFβ signalling pathways were investigated. Heatmaps containing only significant transcriptional changes reveals that more TGFβ signalling pathway transcripts (Figure 5-19) are significantly altered in both shRNA treatments than FGF (Figure 5-20) and WNT (Figure 5-21) signalling pathway transcripts. Although less FGF and WNT signalling pathway genes exhibit significant transcriptional changes, the changes observed are suggestive of neuronal differentiation, with FGF10 being increased and WNT8B decreased for both shRNA treatments in early *circACVR2A* shRNA EBs.

Inhibition of Activin/Nodal/BMP (TGFβ signalling family) signalling is a common method to differentiate ESCs and iPSCs into neural lineages, and therefore, transcriptional profiling of TGFβ signalling genes may suggest if an overall inhibition is occurring in H9 circACVR2A shRNA EBs. A heatmap of important TGFβ transcripts was generated (Figure 5-19), and displays increased and decreased levels of genes, without a consistent uni-directional change. An increase of RNA transcripts for factors involved with activating or stimulating TGFβ signalling are observed in early *circACVR2A* shRNA 2 EBs, including an increase to *BMP* transcripts, decrease in a Nodal inhibitors (*LEFTY1* and *LEFTY2*), decrease in inhibitory Smads (*SMAD6* and *SMAD7*), and a decrease to an adhesive glycoprotein that is sensitive to TGFβ treatment (*THBS1*). Conversely, there are some changes that could indicate a negative regulation of TGFβ signalling including a decrease to a Type 1 activin receptor (*ACVRL1*), increase to an Activin inhibitor (*FST*), decrease in TGFβ receptor ligand (*NODAL*), decrease to R-SMAD (*SMAD5*), and a decrease to a type 2 TGFβ receptor (*TGFBR2*). TGFβ signalling factors that have changed in both shRNA treatments are *ACVRL1*, *BMP2*, *BMP7*, *CTGF*, *FST*, *SMAD6*, and *VWC2*.

TGFβ signalling comprises largely of ligands TGFβ, Activin, Nodal, and BMP, with each ligand mediating different downstream transcriptional effects. Inhibition of Activin/Nodal signalling as well as BMP signalling have both independently been shown to promote neuroectoderm specification during differentiation as well as combined inhibition. Looking only at transcriptional changes that could affect Activin/Nodal signalling in both shRNA treatments identifies FST as a possible mediator that is driving an inhibition of signalling. In comparison, BMP signalling still has multiple factors that are transcriptionally up- or down-regulated, however it can be noted that a BMP ligand (BMP7) and cell surface receptor (ACVRL1/ALK1) are downregulated.

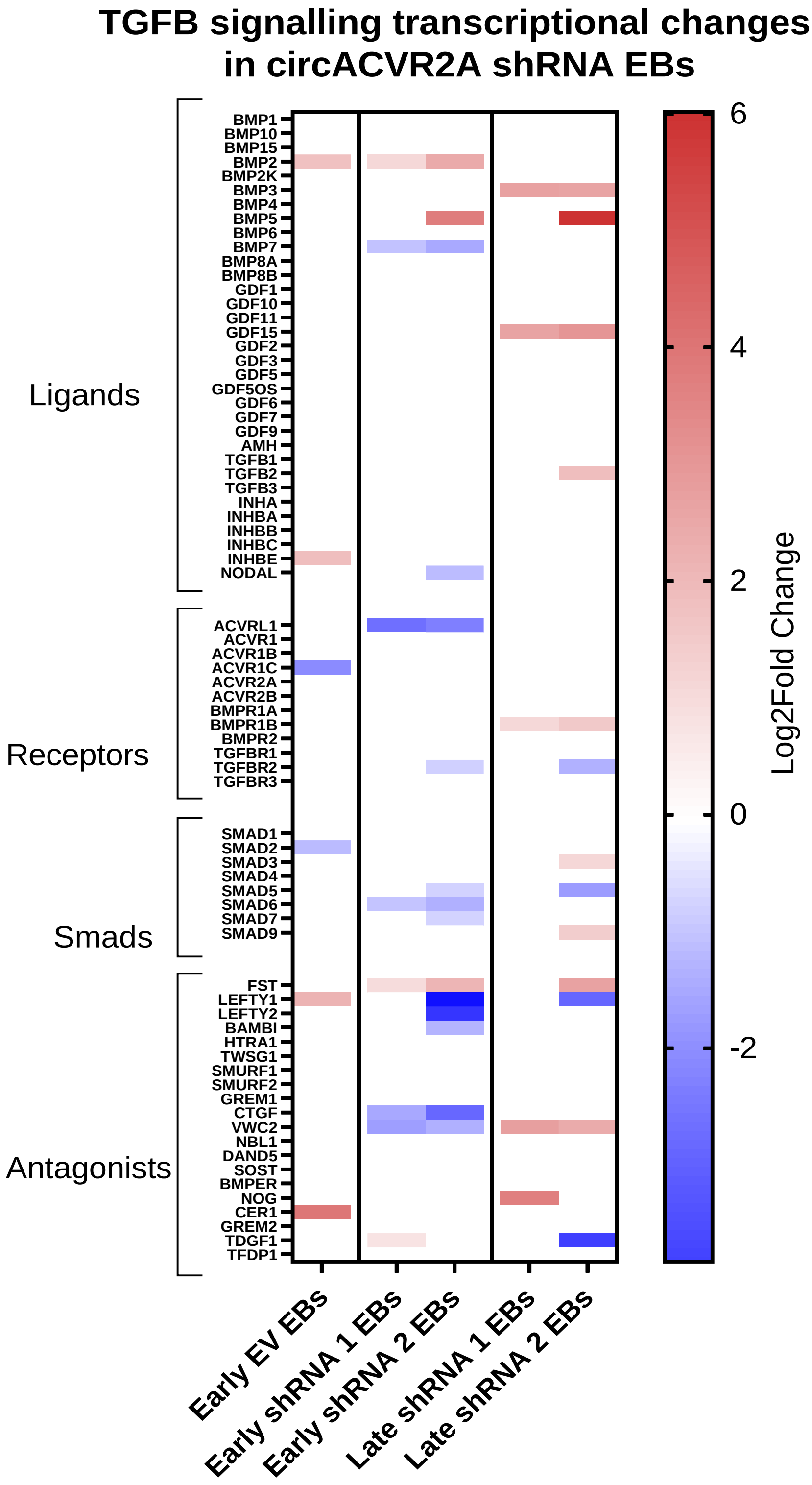


Figure 5-19 Changes to TGFβ signalling pathway transcripts in *circACVR2A* shRNA EBs by RNA-seq

Differential expression of TGFβ signalling transcripts in late *circACVR2A* shRNA EBs in total RNA-seq data set. Data is only given if the change from H9 hESCs is statistically significant ($p<0.05$).

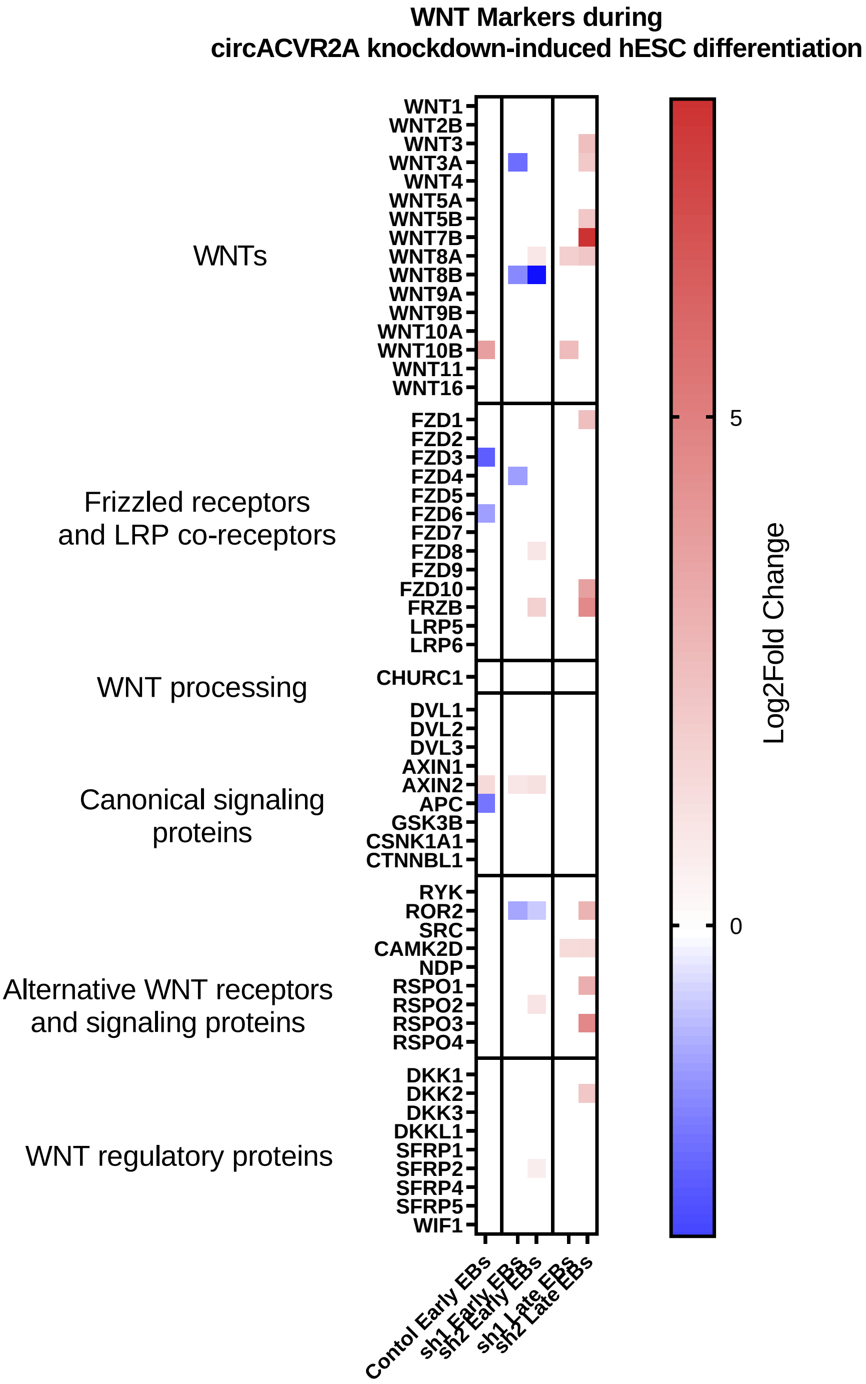


Figure 5-20 Changes to WNT signalling pathway transcripts in *circACVR2A* shRNA EBs by RNA-seq

Differential expression of WNT signalling transcripts in late *circACVR2A* shRNA EBs in total RNA-seq data set. Data is only given if the change from H9 hESCs is statistically significant (p<0.05).

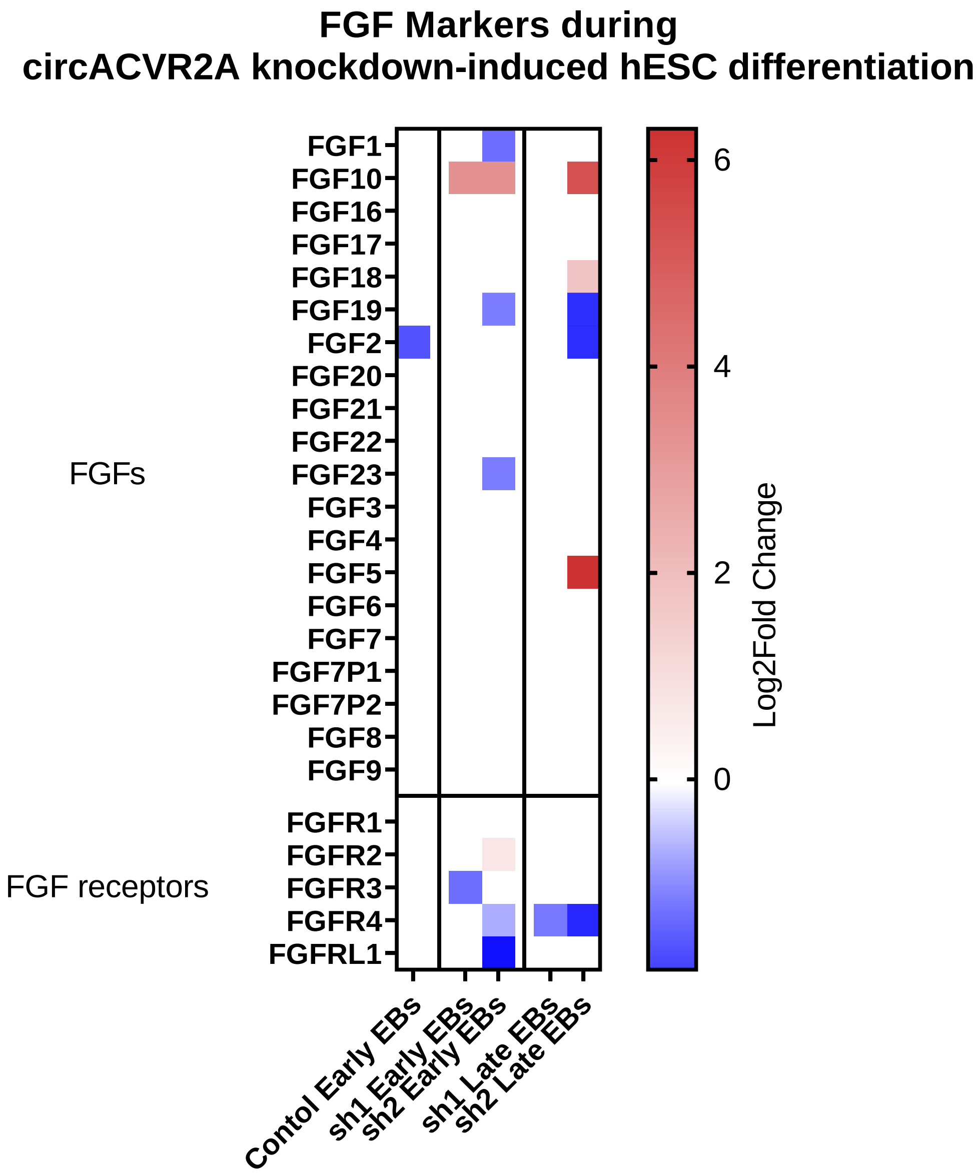


Figure 5-21 Changes to FGF signalling pathway transcripts in *circACVR2A* shRNA EBs by RNA-seq

Differential expression of FGF signalling transcripts in late *circACVR2A* shRNA EBs in total RNA-seq data set. Data is only given if the change from H9 hESCs is statistically significant ($p<0.05$).

5.2.2.8 CircACVR2A does not appear to act as a microRNA sponge in hESCs

To test the theory of a microRNA sponge mechanism for *circACVR2A*, increases in downstream mRNA targets of putative sponged miRNA can be determined with the use of RNA-seq data. Dong *et al.* (2019) utilised CircInteractome and miRanda to identify putative binding sites for *circACVR2A* (*hsa-mir-548p*, *hsa-mir-571*, *hsa-mir-626*, *hsa-mir-659-3p*, *hsa-mir-1200*, *hsa-mir-1243*, *hsa-mir-1265*, *hsa-mir-1279*) (Figure 5-22). To identify an interaction between these miRNAs and *circACVR2A*, Dong *et al.* designed a biotinylated probe for *circACVR2A* allowing for the isolation of overexpressed *circACVR2A* along with any bound RNA species. QRT-PCR identified specific miRNA species (*hsa-mir-626*, *hsa-mir-1265*, *hsa-mir-1279*) as binding to *circACVR2A*.

Although Dong *et al.* were able to show that these miRNAs binding sites are able to bind miRNA mimics in a luciferase assay and show that downstream miRNA targets genes are affected including *EYA4*, no miRNA target genes, identified by miRDB database (Chen and Wang, 2020), were found to be affected in both *circACVR2A* shRNA treatments in hESCs (Figure 5-23).

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 5-22 Putative miRNA binding sites on *circACVR2A*

Visualisation of putative binding miRNA binding sites of *circACVR2A* and the abundance of miRNAs quantified from pulldown of *circACVR2A* in bladder cancer cell lines. Data presented as the mean + SEM (Image adapted from: (Dong *et al.*, 2019)).

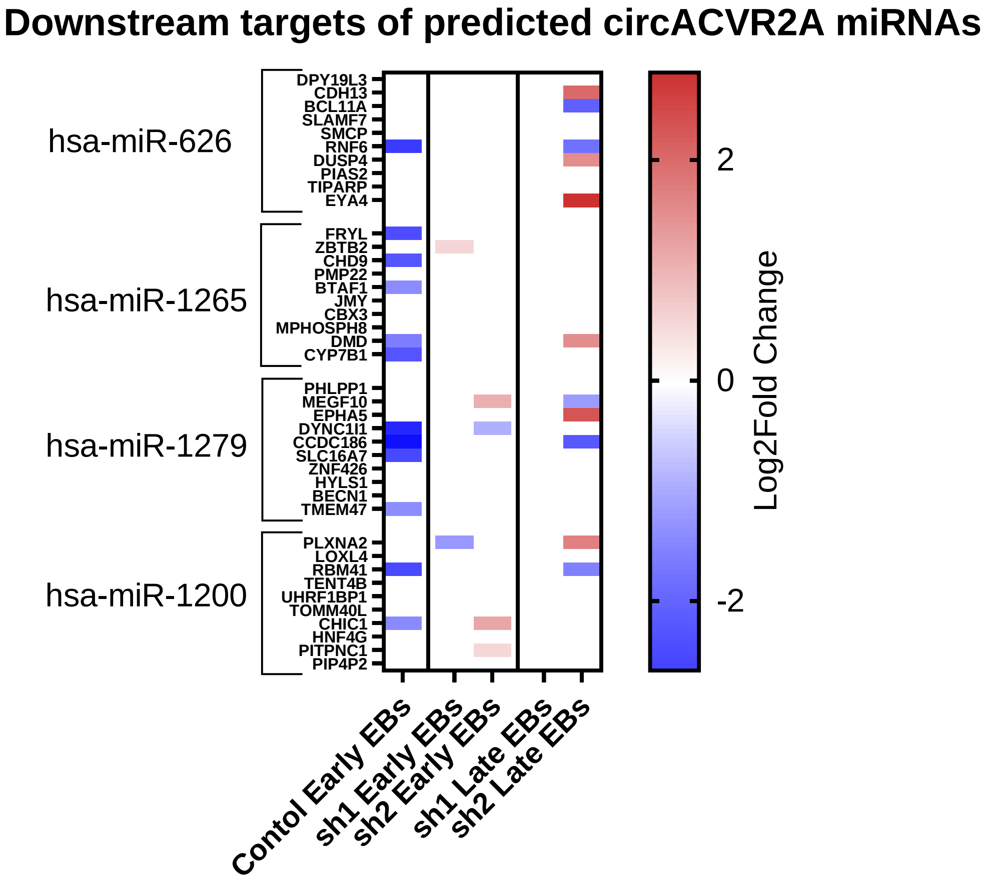


Figure 5-23 No changes in both *circACVR2A* shRNA treatments for miRNA target genes

RNA-expression of downstream putative miRNAs target genes in *circACVR2A* shRNA KD hESCs. Data is included only if statistical significance is reached (p<0.05).

5.2.3 RNA-seq analysis for *circFAT3* and *circCDYL*

Inclusion of samples for shRNA KD of *circFAT3* and *circCDYL* in hESCs was also included for RNA-seq and differential expression analysis. Differential expression analysis has identified hundreds of RNA transcripts that are statistically differentially expressed ($p<0.05$) from H9 EV hESCs (Table 5-6). As only 1 replicate was included for RNA-seq, less stringent statistical testing (non-FDR adjusted p-values) was used to identify statistically expressed RNA transcripts. Probing for EMT markers has revealed an increase to mesenchymal transcripts in both *circFAT3* and *circCDYL* KD hESCs. Minimal changes to pluripotency are observed for *circCDYL* shRNA 1 KD, however, master pluripotency factor *NANOG* is decreased in *circFAT3* shRNA 1 hESCs (Figure 5-24). Epithelial markers show minimal changes for *circFAT3* shRNA 1 KD hESCs, while for *circCDYL* hESCs, no decrease is observed, with epithelial cadherin increasing in abundance.

Table 5-6 Statistically differentially expressed RNA transcripts for *circFAT3* and *circCDYL* knockdown in H9 hESCs

	Statistically Significantly Increased transcripts ($p<0.05$)	**Statistically Significantly Decreased transcripts ($p<0.05$)**
circFAT3	765	558
circCDYL	516	150

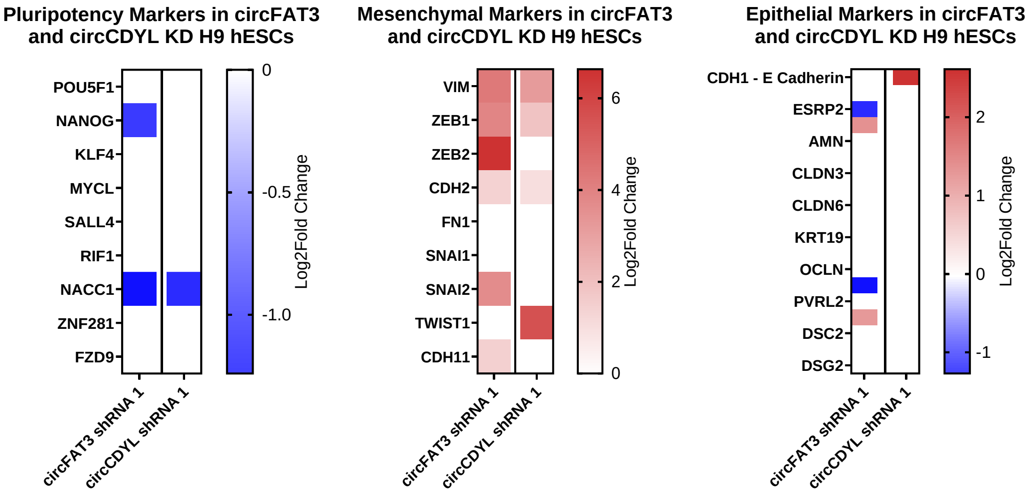


Figure 5-24 Pluripotency and EMT markers altered in *circFAT3* and *circCDYL* shRNA knockdown H9 hESCs

RNA-expression of pluripotency and EMT markers in *circFAT3* shRNA 1 and *circCDYL* shRNA 1 KD H9 hESCs. Data is included only if statistical significance is reached (p<0.05).

GO analysis of significantly increased transcripts (p<0.05) reveals that *circFAT3* KD hESCs have increased enrichment for neural processes, similar to *circACVR2A* KD hESCs as well as cell migration (Figure 5-25). Conversely, *circCDYL* KD hESCs have increased enrichment for hormone related pathways of the thyroid, corpus luteum and adrenal glands. Further profiling of the ability of these cells to undergo differentiation into the different lineages (endoderm, mesoderm and ectoderm) may provide evidence for altered differentiation capacities after circRNA KD.

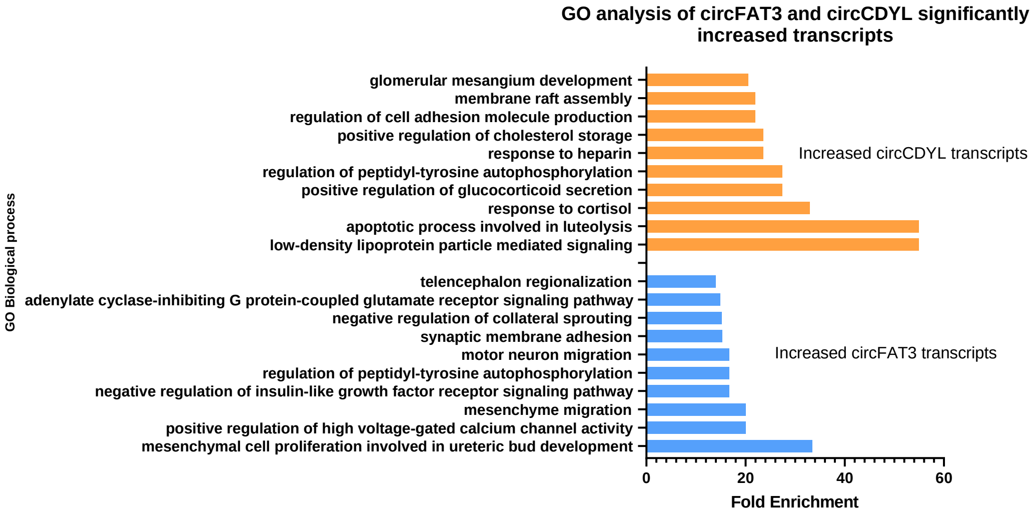


Figure 5-25 GO analysis of differentially expressed RNA transcripts from *circFAT3* and *circCDYL* knockdown hESCs compared to H9 EV hESCs

GO analysis of significantly differentially expressed RNA transcripts ($p<0.05$) of *circFAT3* and *circCDYL* KD H9 hESCs compared to H9 EV hESCs. Only showing the top 10 GO terms.

5.3 Discussion

Transcriptional profiling of *circACVR2A* shRNA EBs and further culturing and attachment of the EBs has uncovered neural cell differentiation after KD of *circACVR2A* with multiple shRNAs. This phenotype was not observed by efficient KD of the parental *ACVR2A* mRNA from chapter 4.

CircACVR2A shRNA 2 EBs were compared to EV EBs for adherence properties. As shRNA 1 EBs formed slower and over a number of days, these EBs were not selected for this profiling as the EBs. It was found that decreasing the lentiviral concentration increased shRNA 1 EB production, however this was determined after profiling the adherence of the EBs and also after the total-RNA sequencing.

CircACVR2A shRNA 2 EBs proved to be far less adherent than canonical EV EBs in culture as no *circACVR2A* shRNA 2 EBs were able to adhere back to non-coated TC-treated culture dishes after 12 days of culturing. In comparison, EV EBs had some binding capacity with 50% of EBs binding back to non-coated dishes. All EV control and *circACVR2A* shRNA 2 EBs were able to bind back to culture dishes coated with either vitronectin or geltrex, however differences were observed with the number of cells that were able to spread out from the re-attached EB. Vitronectin is a single protein matrix, however geltrex contains many factors including laminin, collagen IV, entactin, and heparin sulfate proteoglycans. Both ECMs support culturing and expansion of undifferentiated stem cells, however these may not best represent *in vivo* environments for differentiated cells, with laminin significantly enhancing neural progenitor generation and differentiation of neurons, compared to fibronectin and type I collagen substrates (Hynds and Snow, 2001; Ma *et al.*, 2008). As geltrex is more supportive for neural cells, this may explain the outgrowth of neural like cells on geltrex, yet slow expansion on vitronectin and the inability of these cells to move past the edge of the cells surrounding the EB on vitronectin. Both independent shRNA induced EBs produced neural-like cells after the EBs were attached to geltrex matrix and allowed to culture for a number of days. These neural cells look similar to cells that expand out from attached neurospheres (Francis and Wei, 2010). No other cell populations were observed to disperse from the flattened out EB.

QRT-PCR analysis performed by Jenne Tran at SAHMRI South Australia identified several neural transcripts increased in the expanded neural-like cell population and also the EBs that attached to the ECM, as a separate sample. This further indicates that the EBs and the cells that move away from the attached EB are neural cell types.

5.3.1 Immunofluorescence

NANOG protein was found to be decreased in *circACVR2A* shRNA 2 EBs compared to EV EBs, corroborated in the total-RNA seq data with a reduced *NANOG* mRNA expression. This is noteworthy, as downregulation of *NANOG* has been shown to promote specification into neuroectoderm and is controlled by TGFβ signalling (Vallier *et al.*, 2009), although loss of NANOG is not specific to neuroectoderm differentiation. POU5F1 protein was still present in early *circACVR2A* shRNA 2 EBs and in control EV EBs, as POU5F1 protein expression usually decreases in EBs over the course of 4-7 days (Sajini *et al.*, 2012). Interestingly, POU5F1 protein expression was retained in late *circACVR2A* shRNA 2 EBs, while in EV EBs, POU5F1 expression was diminished. Kim *et al.* described different types of EBs as dark, cystic or light EBs and it was found that dark EBs did have a higher level of POU5F1 expression and could explain for the increased expression compared to EV control EBs (Kim *et al.*, 2011). Alternatively, ES cells that haven't fully committed to differentiation can still show some POU5F1 expression with a transient decrease in NANOG expression (Chambers *et al.*, 2007). In alignment with the qRT-PCR analysis, detection of FOXA2 by immunofluorescence also suggests a differentiation into a neural lineage. FOXA2 is a midbrain floorplate marker (Metzakopian *et al.*, 2012; Pristerà *et al.*, 2015) but can also indicate differentiation and cell types of other lineages (Domanskyi *et al.*, 2014).

It is known that cells on the edge of growing hESC colonies express more POU5F1 and NANOG protein than cells in dense colony centres (Hamidi *et al.*, 2019), however overall pluripotency is not affected. This is also the case with EMT markers as E-cadherin, an epithelial cadherin protein, is expressed in the centre of colonies, while N-cadherin, an indicator of mesenchymal cell morphology, is expressed exclusively in the borders of hESC colonies (Ullmann *et al.*, 2007). Vimentin, a type 3 intermediate filament normally expressed in mesenchymal cells, is also found to be expressed in the periphery of hESC colonies. This may be a confounding factor for all methods of protein and RNA detection of hESCs and iPSCs as harvesting different sized colonies may have different ratios of pluripotency and EMT markers. Careful consideration on the size of colonies at the time of harvesting for protein and RNA collection, and possibly immunostaining, should be given.

5.3.2 RNA-seq analysis

Analysis of total RNA-seq supports the induction of EMT within *circACVR2A* KD EBs with an increase of mesenchymal markers and decrease to epithelial markers, most prominently in *circACVR2A* shRNA 2 EBs. ShRNA 1 hESCs did not induce EB formation to the same extent as shRNA 2, most likely due to differences in lentivirus concentration used on the cells for RNA-seq. Therefore, some shRNA 1 treated colonies were still attached when the cells were harvested for RNA-sequencing. This is a likely explanation for less differentially expressed transcripts for shRNA 1 compared to shRNA 2 treated cells. RNA-seq analysis of adhesion molecules also found that ITGB5 integrin, a cell surface receptor vital for vitronectin binding, was decreased in early *circACVR2A* shRNA EBs. ITGB5 was more downregulated in shRNA 2 than shRNA 1 hESCs which could explain for the more efficient EB production in shRNA2 treated cells. Further analysis of the adhesion markers reveals a loss of ITGA2, an integrin known to be highly expressed in hESCs, and would likely decrease during differentiation; ITGA8, which has an increased expression in neural lineages, and ITGB8 which is upregulated in neural crest and neuroepithelial precursor cells.

Interestingly, the mesenchymal transcripts increase further from early to late stage EBs and a greater number of genes transcripts increase. It is not yet known if KD of *circACVR2A* is inducing either EMT, differentiation, or both (EMT and differentiation). Sponging of miRNAs by *circACVR2A* was investigated by Dong *et al.* and Wang *et al.* and although they found downstream targets of miRNAs to be affected, with altering circRNA levels, these same transcripts were unaffected in *circACVR2A* KD EBs, suggesting a different interaction is inducing this phenotype in hESCs.

Early EBs for both shRNA treatments show no change to pluripotency markers except for a decrease to *NANOG* expression in H9 *circACVR2A* shRNA 2, also identified at the protein level by immunofluorescence. Other pluripotency markers decrease in expression over time as more pluripotency markers are decreased in late *circACVR2A* shRNA 2 EBs. Surprisingly, POU5F1 (OCT3/4), a master regulator of stem cell pluripotency, does not change in protein expression. This may be due to the *circACVR2A* shRNA EBs used for RNA seq analysis were 18-19 days post transduction, while for immunofluorescence, the EBs were 16 days post transduction. EBs generated from murine stem cells have shown to recover POU5F1 and NANOG after extended growth (Attia *et al.*, 2014) and this could explain for differences observed by RNA-seq and immunofluorescence. Furthermore, only 40% of the variance in protein expression can be explained by changes to the mRNA transcript (Abreu *et al.*, 2009), as a combination of posttranscriptional and post-translational regulation accounts for the remaining ~60% (Maier, Güell and Serrano, 2009; Vogel and Marcotte, 2012). Additionally, POU5F1 has also been observed to be more highly expressed in dark EBs than other types (cystic or light EBs) and this may explain why *POU5F1* is not seen to be altered at the transcriptional level in *circACVR2A* shRNA EBs (Kim *et al.*, 2011). Alternatively, ES cells that haven't fully committed to differentiation can still show some *POU5F1* expression with a transient decrease in *NANOG* expression (Chambers *et al.*, 2007). This is observed in both total RNA-seq data and immunofluorescence comparing EV control and late *circACVR2A* shRNA EBs. Overall, this suggests that pluripotency is being lost, however the OCT3/4 marker may be retained due to the dark *circACVR2A* shRNA EB morphology, recovery of OCT3/4, or changes in how *circACVR2A* shRNA EBs commit to differentiation compared to H9 EV EBs.

Analysis of the RNA-seq data has confirmed increases in mesenchymal markers and decreases to epithelial markers in H9 *circACVR2A* shRNA EBs. Additionally, investigation of lineage markers reveals no upregulation of other cell type transcripts and increases to neuron and neural precursor transcripts. CellNet analysis and GO enrichment analysis further supports neural precursor differentiation occurring within *circACVR2A* shRNA EBs leading to the deposition of neural-like cells only on a laminin-based ECM. Control EV EBs cultured in the same media and for the same time has produced an array of cell morphologies while *circACVR2A* shRNA EBs only produce neural cells suggesting that *circACVR2A* KD is inducing only neuroectoderm differentiation, driven by a reduction of *circACVR2A*. Increases in neuron markers and other neuroectoderm lineage markers identified by qRT-PCR and RNA-seq in *circACVR2A* shRNA EBs further identifies neural differentiation.

Chemical inhibitors of TGFβ signalling are commonly used to differentiate stem cells into neuroectoderm lineages in 2D culture (Kelava and Lancaster, 2016). As *circACVR2A* arises from a gene that encodes for a TGFβ cell surface receptor, and is co-expressed with its parental gene, *circACVR2A* may contribute to the downstream signalling of any of the TFGβ ligands (Activin, Nodal, BMPs, GDPs, TGFβ). KD of *circACVR2A* has induced the differentiation of hESCs into neural lineages and therefore may be required for signalling downstream of ACVR2A, whereby reduction of *circACVR2A* has inhibited TGFβ signalling. Loss of TGFβ signalling may be the major contributing factor that is driving neuroectoderm differentiation which may also be affecting an EMT response. RNA-seq analysis has confirmed changes in many TGFβ pathway transcripts, however no overall inhibition or activation is able to be concluded. Looking at only statistically significant transcriptional changes occurring in both shRNA treatments, FST, an activin inhibitor, is upregulated. Within the BMP signalling branch of the TGFβ signalling pathway, several transcripts are up and down regulated. Additionally, analysis of the WNT and FGF signalling pathways give support for induction of EMT and differentiation into a neural lineage. *WNT8b* expression is known to decrease in mesenchymal stem cells (Okoye, Malbon and Wang, 2008) and *WNT8b* transcripts were found to decrease in early *circACVR2A* shRNA EBs. Furthermore, *FGF10*, was upregulated in early *circACVR2A* shRNA EBs and is known to be expressed in neurons, neuroepithelial cells and radial glia cells(Sahara and O’Leary, 2009; Li *et al.*, 2016). Although transcriptional changes can give insight into possible changes in protein levels, RNA-seq does not give information on protein abundance, folding, post translational modifications and chemical modifications including phosphorylation status of downstream effectors in a pathway.

Analysis of *circCDYL* and *circFAT3* RNA-seq data has identified increased mesenchymal transcripts with little change to classical epithelial markers. Comparatively, epithelial transcripts were decreased in EBs generated by *circACVR2A* KD. Therefore, preservation of epithelial transcripts in *circCDYL* and *circFAT3* KD hESCs, may be the key difference between the two different phenotypic changes that occur (EB phenotype and loss of dense colonies) when hESCs are culture on vitronectin. Preservation of epithelial transcripts and proteins may maintain attachment of colonies, demonstrating the plastic nature of EMT and unique context dependent EMT regulatory factors. GO analysis of mRNAs that were statistically increased has identified pathways with are enriched within *circCDYL* and *circFAT3* KD hESCs. The GO terms identified indicate increases to transcripts involved with multiple cell types, unrelated to stem cell processes. Differentiation studies of these KD cell lines may illuminate if the differentiation propensities are altered due to knockdown of these circRNAs. Furthermore, these cells may already be undergoing differentiation and so culturing these cells for a longer period of time may give further indication into the cell types that these cells are becoming.

5.4 Chapter 5 Conclusion

CircACVR2A KD induces hESC differentiation exclusively into neural lineages. Dual inhibition of TGFβ signalling is the most common method for *in vitro* production of neural cells from hESCs and iPSCs. It is not yet known if *circACVR2A* is inducing EMT, differentiation, or both. Therefore, to tease apart these two processes that are intertwined in stem cells, investigation with an independent cell line that does not undergo differentiation, but can undergo TGFβ-induced EMT, will allow for the removal of differentiation as a confounding factor when analysing cell properties after KD of *circACVR2A*.

Chapter 6 Profiling *circACVR2A*-Protein Interactions and Functional Consequences in A549 Cells

6.1 Introduction

Adenocarcinomic human alveolar basal epithelial cells (A549s) are a useful cell line as they undergo EMT when treated with TGFβ1 or by other methods (Kasai *et al.*, 2005; Reka *et al.*, 2011; Zahedi *et al.*, 2018). Addition of TGFβ1 quickly induces the phosphorylation of ERK1/2 and SMAD2 and cells become more elongated and decrease in cell-cell contacts (Kasai *et al.*, 2005). Furthermore, alongside EMT traits, TGFβ1 treatment of A549 cells induces stemness characteristics with demethylation at EMT and stem cell marker promotors (Kim *et al.*, 2020). Without the consequential effects of cell differentiation, profiling *circACVR2A* KD in A549s may give insight into the cellular changes occurring that may explain for neural differentiation in hESCs or if *circACVR2A* can modulate EMT.

A miRNA sponge mechanism for *circACVR2A* was not established in hESC, by profiling downstream miRNA targets using RNA-seq. Therefore, profiling interactions between *circACVR2A* and proteins may illuminate on the functional consequences of *circACVR2A* KD. Dong *et al.* (2019) established that *circACVR2A* is mainly localised in the cytoplasm within bladder cancer cells by qRT-PCR and Fluorescence In Situ Hybridisation (FISH), however ~20% resides in the nucleus (Figure 6-1). This is in concordance with studies that observe exonic circRNAs to be predominantly localised in the cytoplasm (Salzman *et al.*, 2012; Jeck *et al.*, 2013; J. Zhang *et al.*, 2019), however this may be cell-type specific. Therefore, a *circACVR2A*-protein interaction can possibly occur in either the nucleus and/or cytoplasm.

This image has been removed due to copyright restriction. The image is available online. See figure legend for citation.

Figure 6-1 Analysis of *circACVR2A* localisation in bladder cancer

Analysis of *circACVR2A* localisation in the nuclear and cytoplasmic fractions of T24 and UM-UC-3 bladder cancer cells by (A) qRT-PCR and (B) FISH. Data presented as the mean + SEM (Image adapted from:(Dong *et al.*, 2019)).

6.2 Results

6.2.1 *CircACVR2A* is not regulated during hTGFβ1-induced EMT in A549 cells.

The human lung adenocarcinoma cell line, A549, is able to undergo rapid TGFβ1-induced EMT (0.5-5ng/ml TGF-β1), with detectable phosphorylation of TGFβ TF proteins SMAD2 and SMAD3 within 10-15 minutes (Kasai *et al.*, 2005; Ji *et al.*, 2016), and with converting to spindle-shaped mesenchymal cells within 24-48 hours (Kasai *et al.*, 2005; Kim *et al.*, 2007; Liu *et al.*, 2013; Tirino *et al.*, 2013; Ji *et al.*, 2016). This is unlike other mammalian cells, that can take up to 21 days to induce EMT with TGFβ1 (Conn *et al.*, 2015). To understand the expression changes of TGFβ-induced EMT of A549s, cells were treated with human TGFβ1 (hTGFβ1) for 24 hours and RNA was extracted. Canonical EMT TFs and cadherin RNA transcripts were quantified by qRT-PCR (Figure 6-2), demonstrating the changes that occur with increases to EMT TFs transcripts (ZEB1, SNAI1 and SNAI2), and an increase in a mesenchymal cadherin transcript (CDH2) and a decrease to an epithelial cadherin transcript (CDH1).

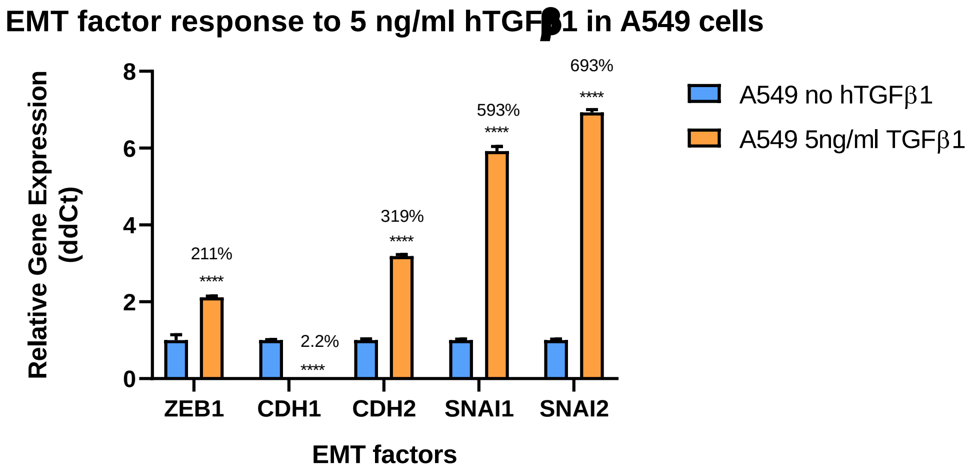


Figure 6-2 EMT factor response to hTGF β1 induced EMT in A549 cells

Canonical EMT factors were quantified by qRT-PCR for relative gene expression changes in A549 cells after 24 hours of 5 ng/ml of hTGFβ1. Statistical significance was determined by one-way ANOVA with Dunnett's test correcting for multiple comparisons. Relative to *GAPDH* and *ACTB*. (**** = $p<0.0001$) (n=3).

Although *circACVR2A* levels changed during differentiation in hESCs, addition of previously published TGFβ1 concentrations to A549s did not significantly alter the level of *circACVR2A* after 48 hours, a sufficient time for phenotypic and transcriptional changes to occur (Figure 6-3A). A549s were stably transduced with 3 independent shRNAs targeting *circACVR2A*. Two of the three shRNA treatments were effective in reducing *circACVR2A* levels in A549 cells (Figure 6-3B). Control A549s and A549s with decreased *circACVR2A* expression were incubated with 5 ng/ml hTGFβ1 for 48 hours to observe if cells were still able to undergo an EMT event. Control and *circACVR2A* KD cell lines were able to undergo EMT as cells became spindle shaped and elongated (Figure 6-4 right column). Therefore, a reduction of *circACVR2A* does not inhibit the ability of cells to undergo an EMT, moreover, the induction of EMT does not alter *circACVR2A* levels in A549 cells.

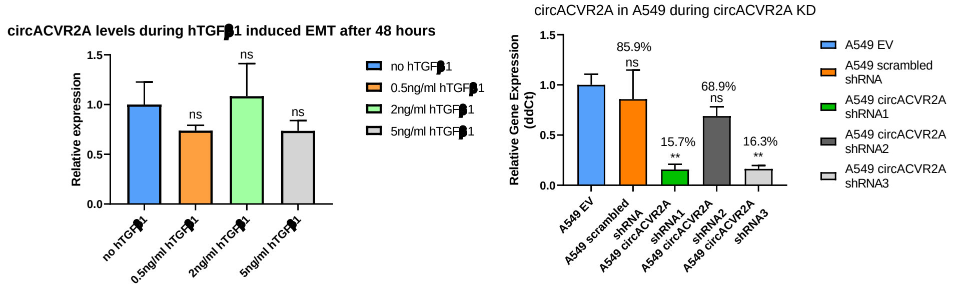


Figure 6-3 A549 *circACVR2A* shRNA knockdown efficiency and sensitivity to TGFβ1

(A) *CircACVR2A* displays no significant regulation during hTGFβ1 treatment of A549 cells. A594 cells were treated with varying levels of hTGFβ1 for 48 hours and RNA was collected. (B) qPCR quantification of *circACVR2A* KD in stable A549 cells. Relative to *GAPDH* and *ACTB*. Statistical significance was determined by one-way ANOVA (** = p<0.01) (n=3; Mean+SEM).

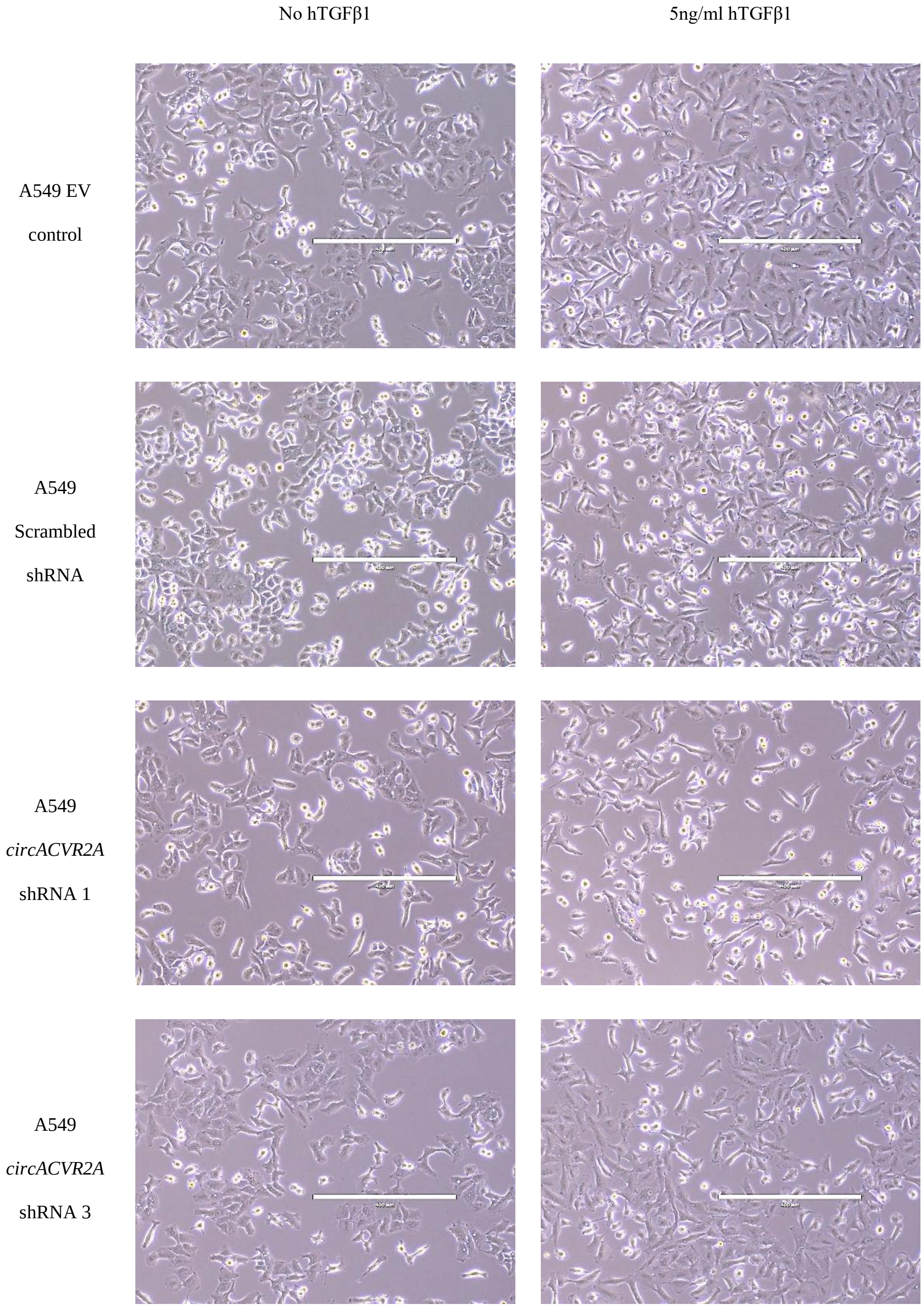

No hTGFβ1
5ng/ml hTGFβ1
A549 EV control
A549 Scrambled shRNA
A549 *circACVR2A* shRNA 1
A549 *circACVR2A* shRNA 3

Figure 6-4 A549 *circACVR2A* shRNA knockdown cells responsive to TGFβ1

Control A549 and *circACVR2A* KD cells were incubated with media containing 5ng/ml hTGFβ1 for 48 hours and images were taken. Images were taken using the EVOS™ Imaging system. Scale bar is 400µm.

6.2.2 A549 cells with decreased *circACVR2A* 549 cells show changes to EMT markers in response to hTGFβ1.

Quantifying EMT markers by qRT-PCR after hTGFβ1-induced EMT was performed to see if any individual EMT factors were affected, even though a morphological EMT was observed (Figure 6-5). *ZEB1*, *CDH1* and *CDH2* do not show any changes in both shRNA treatments. *SNAI1* and *SNAI2* markers in *circACVR2A* shRNA A549s are decreased compared to the EV A549s, however the scrambled shRNA also shows some changes to *SNAI1* and *SNAI2* (Figure 6-5). Both SNAI proteins are involved with G1 transition with SNAI1 regulating components of early to late G1 and the G1/S checkpoint (Vega, 2004), while SNAI2 is regulated (phosphorylated) during G1/S transition by cyclinE-CDK2 (Wang *et al.*, 2015).

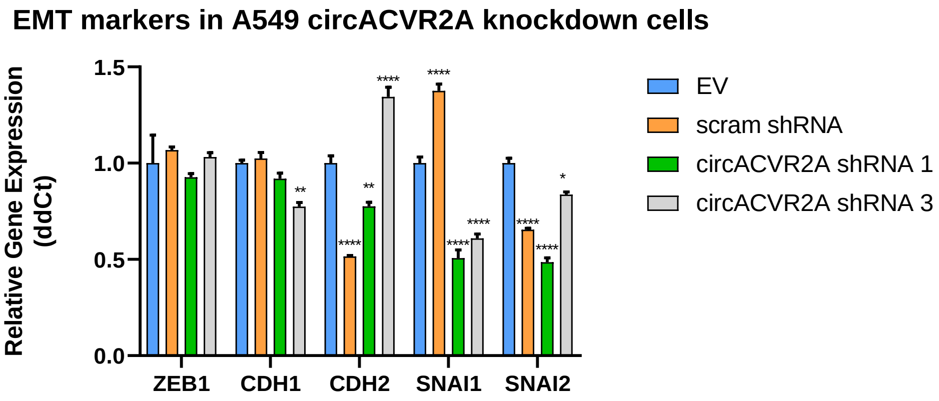


Figure 6-5 EMT marker changes with *circACVR2A* shRNA knockdown in A549s

Control A549 and *circACVR2A* KD cells were incubated with media containing 5 ng/ml hTGFβ1 for 48 hours and RNA was harvested. EMT markers were quantified between control and *circACVR2A* KD A549 cells lines. Statistical significance determined by two-way ANOVA with Dunnett's test correcting for multiple comparisons. Relative to *GAPDH* and *ACTB*. (** = p<0.01, *** = p<0.001, **** = p<0.0001) (n=3; Mean+SEM).

Control and *circACVR2A* KD A549s underwent a morphological hTGFβ1-induced EMT, however, these cells may have increased or decreased sensitivity to hTGFβ1. To test this, EMT TFs and cadherin transcripts were quantified before and after hTGFβ1-induced EMT to discover if these transcripts, during *circACVR2A* KD, increase or decrease to the same extent as in control A549s. After hTGFβ1 treatment, *ZEB1* and *CDH2* transcripts increased to the same extent as control EV A549s with no statistical difference, while *CDH1* transcripts were decreased to the same extent (Figure 6-6). *SNAI1* transcripts in EV A549s increased 8.2 fold, however during *circACVR2A* KD, *SNAI1* only increased 4.4 fold in shRNA 1 transduced A549s and 5.4 fold in shRNA 2 transduced A549s. Therefore, *circACVR2A* KD A549 cells have a lower abundance of *SNAI1*, and *SNAI1* is not increased to the same extent as control A549s during hTGFβ1 treatment. Although this change is significant in A549 cells, with the levels and sensitivity of *SNAI1* decreased, *SNAI1* and *SNAI2* transcript levels are not affected in H9 hESCs (Figure 5-10) and therefore, is most likely not the change that is occurring that induces EB formation.

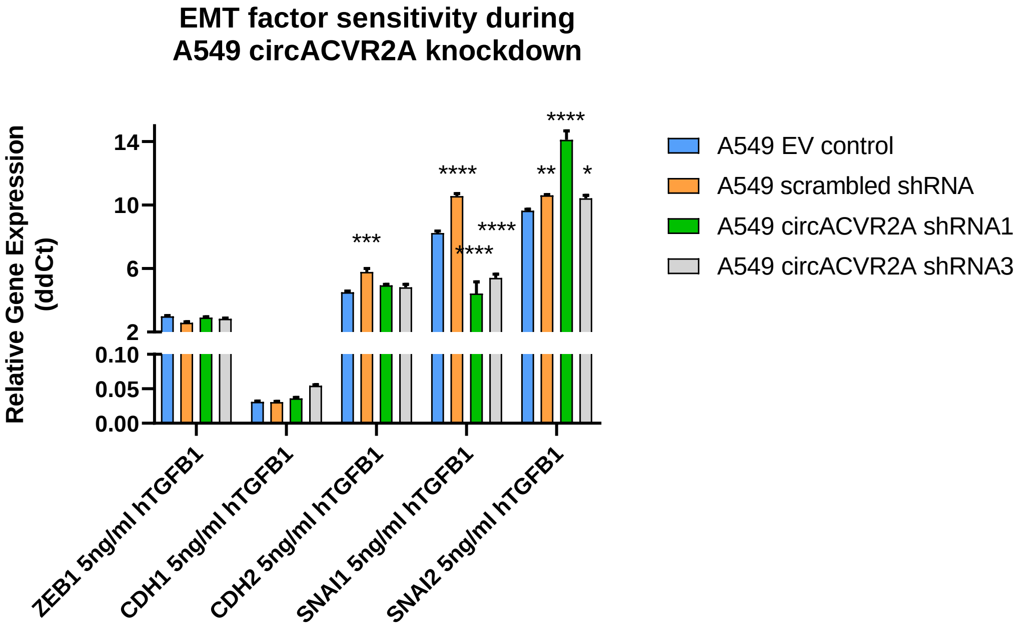


Figure 6-6 Sensitivity of EMT markers during TGFB induced EMT of *circACVR2A* knockdown A549s

The log2Fold change value in the graph is the movement of RNA transcript abundance after 48 hours of hTGFB1 treatment measured by qRT-PCR Relative to *GAPDH* and *ACTB*. (* = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$, **** = $p<0.0001$) (n=3; Mean+SEM).

6.2.3 Knockdown of *circACVR2A* does not affect SMAD levels

As SMAD2/3 are the effector molecules of TGFβ signalling, quantification was performed to ascertain if they are affected with *circACVR2A* KD as reduction of protein levels could explain an inhibition of TGFβ signalling and therefore, neuronal differentiation of hESCs. However, SMAD2/3 protein levels were not altered with KD of *circACVR2A* in H9 hESCs (Figure 6-7).

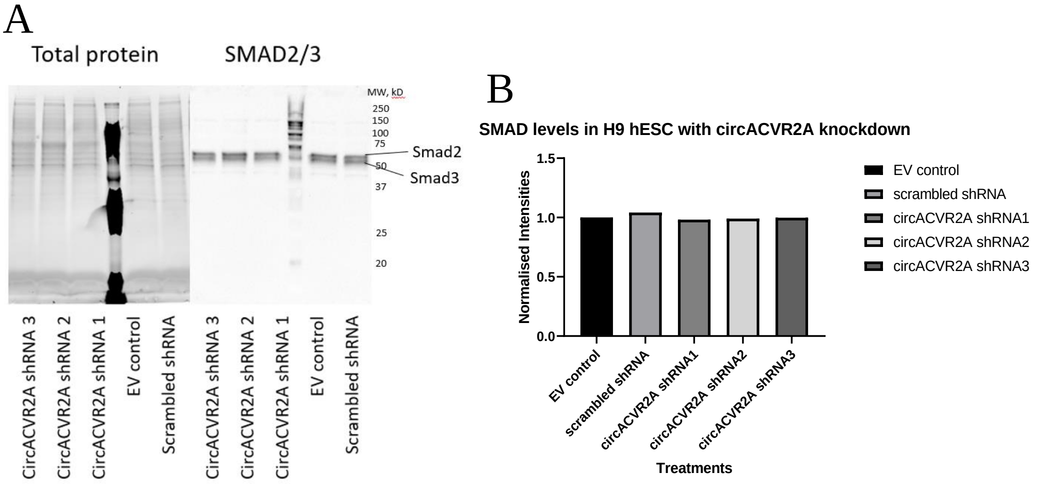


Figure 6-7 No changes to SMAD2/3 protein levels with *circACVR2A* shRNA knockdown in A549s

Quantification of SMAD2/3 in with *circACVR2A* KD. (A) Total protein normalisation and western blot of SMAD2/3. 1/1000 dilution of primary antibody. (B) Normalised intensities of SMAD2/3 levels in control and *circACVR2A* KD cell lines.

6.2.4 Knockdown of *circACVR2A* negatively affects proliferation of A549 cells.

While culturing the modified A549 cells lines, significant altered proliferation rates were apparent as the *circACVR2A* KD A549s required longer culturing times between routine passaging. To determine the effect on cell growth during KD of *circACVR2A*, with two independent shRNAs, a proliferation assay was performed. Reduction of *circACVR2A* was found to decrease the proliferation of A549 cells (Figure 6-8) to a statistically significant level over 48 hours. The doubling time of A549 *circACVR2A* shRNA 1 cells was 22% slower than A549 EV control cells, while A549 *circACVR2A* shRNA 2 cells were 16.8% slower over 24 hours (Table 6-1).

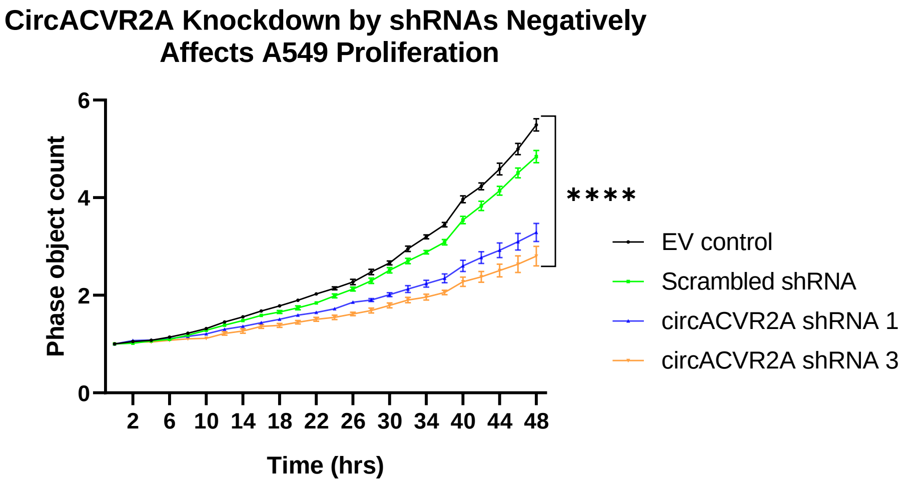

CircACVR2A Knockdown by shRNAs Negatively Affects A549 Proliferation

Figure 6-8 Proliferation assay of *circACVR2A* shRNA knockdown in A549s

Proliferation assay of control and *circACVR2A* KD A549 cells. Cells were seeded at the same density and imaged using an Incucyte S3. Statistical significance determined by two-way ANOVA. **** = p<0.0001 (n=4; Mean+SEM per well).

Table 6-1 Doubling time of control and *circACVR2A* shRNA knockdown A549s

Cell lines	Doubling time (hrs)	Relative to A549 EV
A549 EV	20.83	1
A549 Scrambled shRNA	22.18	1.065 (6.5% slower than EV)
A549 *circACVR2A* shRNA 1	25.41	1.220 (22% slower than EV)
A549 *circACVR2A* shRNA 3	24.33	1.168 (16.8% slower than EV)

6.2.5 Knockdown of *circACVR2A* negatively affects migration of A549 cells and nuclear size

To determine if migration is also affected, scratch wound (Figure 6-9A, Figure 6-9B) and transwell migration (Figure 6-10A) assays were performed. KD of *circACVR2A* decreased the migratory ability of A549 cells as determined with a scratch wound assay (Figure 6-9A and Figure 6-9B). Statistical significance by two-way ANOVA was reached by the first time point, 2 hours post scratch wound. Relative to A549 EV control cells, a 46.3% reduction of migration was observed in A549 *circACVR2A* shRNA 1, and 56.2% reduction in A549 *circACVR2A* shRNA 2 after just 2 hours. As the proliferation rates of *circACVR2A* shRNA 1 and shRNA 2 are only 22% and 16.8% respectively, the decrease of cell proliferation compared to A549 EV control cells, does not explain for the decrease in migration for both shRNA treatments (shRNA 1=46.3% and shRNA 2=56.2%).

Measurement of migration was also attempted by using a transwell system however this was labour intensive compared to the scratch wound assay. The transwell migration assay also showed decreased cell migration, however significant change was only observed for *circACVR2A* shRNA 1 A549s. A549 *circACVR2A* shRNA 3 cells did not meet statistical significance as the number of samples processed was low, however a low p-value of 0.075 indicates a trend for a decrease in migration (Figure 6-10). Although statistical significance was obtained by only one shRNA treatment in the transwell migration assay, nuclear area could also be measured by ImageJ of the DAPI stained nuclei, where the average nuclear size was reduced in *circACVR2A* KD A549s, with significance approaching $p<0.05$ (Figure 6-10C). Nuclear size has been found to correlate with stages in the cell cycle and as cells progress from G1 into G2, nuclear size increases (Fidorra *et al.*, 1981; Chu, Haley and Zidovska, 2017).

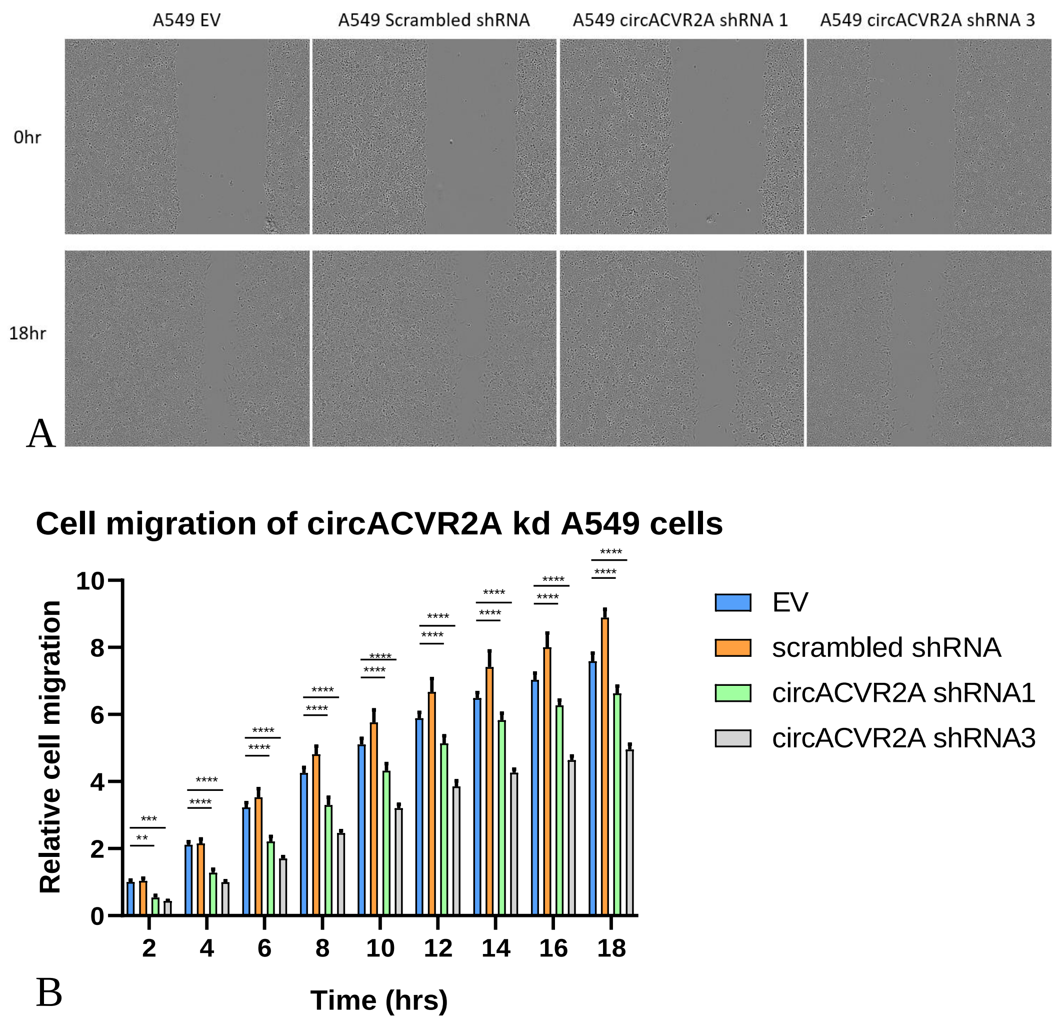


Figure 6-9 Migration assay of *circACVR2A* shRNA knockdown A549s

KD of *circACVR2A* has decreased migration of A549 cells in scratch wound assay. (A) Representative images of scratch wound at 0 hours and 18 hours. (B) Statistical analysis of migration compared to EV control. Two-way ANOVA with Dunnett test to correct for multiple comparisons. Relative to *GAPDH* and *ACTB*. ** = p<0.01, *** = p<0.001, **** = p<0.0001 (n=4; Mean+SEM per well).

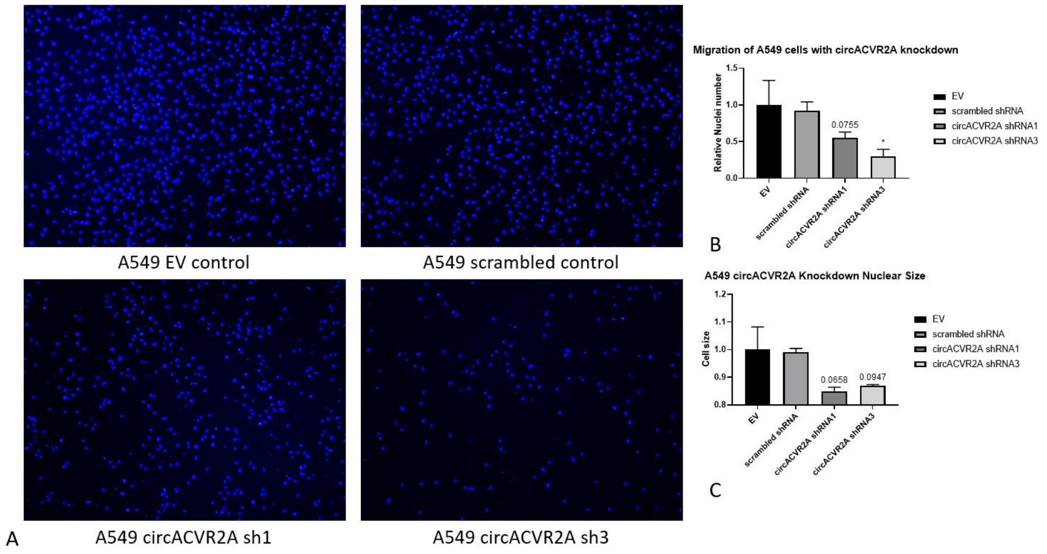


Figure 6-10 Migration assay of *circACVR2A* shRNA knockdown A549s

Migration of *circACVR2A* KD cells in transwell system after 16 hours. (A) Images of stained nuclei of migrated cells in transwell system. (B) Quantification of cell nuclei stained with DAPI. (C) Nuclear size of migrated cells. Quantified using ImageJ. Statistical analysis was performed with an Ordinary one-way ANOVA with Fisher's LSD test. Relative to *GAPDH* and *ACTB*. * = $p<0.05$ (n=2; Mean+SEM).

6.2.6 *CircACVR2A* pull-down and Mass Spectrometry

Thus far, no functional protein interactions for *circACVR2A* have been investigated that would indicate how a reduction of this circRNA is able to induce neural differentiation in hESCs and decrease the proliferation of A549s. In fact, a circRNA-protein interaction is not commonly investigated, even though select circRNAs have been proven to bind proteins. CircRNA-miRNA interactions remain the most commonly investigated interaction. To determine functional protein interactions with *circACVR2A*, pulldown techniques were employed to capture and identity bound proteins. Pulldown of overexpressed *circACVR2A* from HEK293T cells identified a number of potential interacting proteins via Mass Spectrometry (MS). HEK293T cells were utilised as these cells are vastly easier to culture to the required cell number to capture enough protein for the MS analysis and have greater transfection efficiency for overexpression of circRNAs.

This involved replicate pulldowns using a biotinylated *circACVR2A* sense DNA probe, that would bind to *circACVR2A* via complementary base pairing. Additionally replicate pulldowns using a non-*circACVR2A* targeting antisense probe was employed, to determine proteins that may have high binding specificity to biotinylated probes and would be recognised as background protein binding.

To filter the list of possible *circACVR2A* interacting proteins, identified by MS, stringent criteria were employed. Only proteins that were detected in all 3 replicate pulldowns and not detected in any of the 3 negative antisense pulldowns were selected as a possible interacting protein of *circACVR2A*. Using this criteria, four proteins were identified as possible *circACVR2A* binding proteins (Table 6-2**)**. Identified as one possible binding partner, interaction between *circACVR2A* and Cyclin H (CCNH), and any perturbation of this interaction, may interrupt cell proliferation and normal cell cycle processes. This includes CCNH formation within the CAK complex, activation of multiple cyclin/CDK complexes throughout the cell cycle (Figure 6-11), or the phosphorylation of RNA pol II by the CAK complex, as discussed in Chapter 1. This interaction may then impact on proliferation of cells if the cell cycle is affected by reduction of *circACVR2A*. Of note, both other members of the CAK complex (CDK7 and MAT1) were identified in 2 out of the 3 *circACVR2A* pulldowns, giving support for the interaction between CCNH and *circACVR2A*.

Table 6-2 Proteins identified from biotinylated probe pulldown and mass spec

Proteins identified from *circACVR2A* pulldown and MS. Detected in all 3 *circACVR2A* sense pulldowns and in no antisense pulldowns.

Gene Name	Gene alias	Protein Summary
Cyclin H	*CCNH*	Functions in CAK complex and phosphorylates cyclin/CDK complexes. Also functions in RNA transcription by phosphorylating RNA polymerase II
Translocase of Outer Mitochondrial Membrane 5	*TOMM5*	Involved with Mitochondrial protein import and Metabolism of proteins
RAB11 Family Interacting Protein 5	*RAB11FIP5*	Involved in Rab GTPase binding and gamma-tubulin binding
Casein Kinase 1 Gamma 3	*CSNK1G3*	Serine/threonine protein kinase. Participates in Wnt signalling

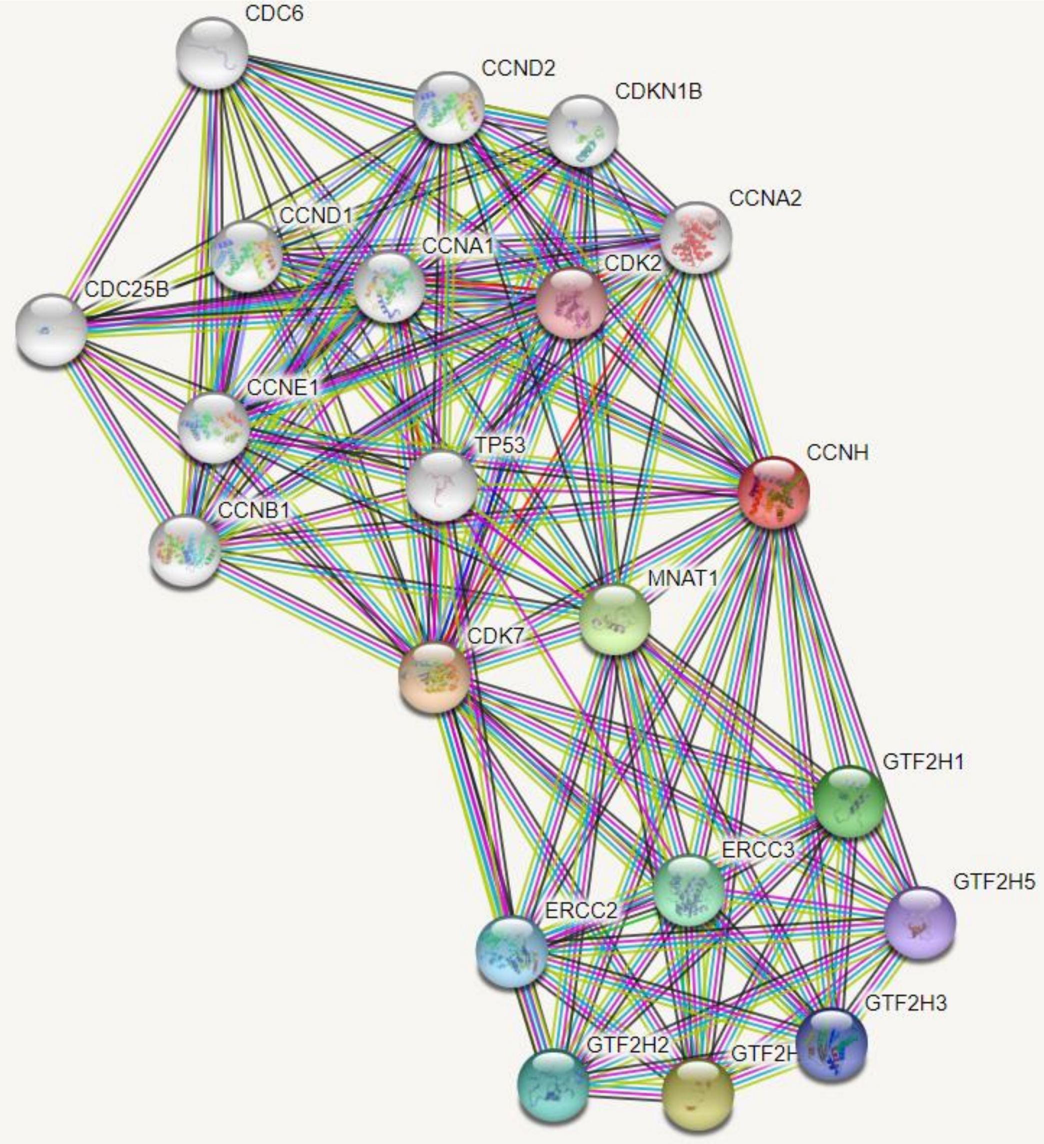

CDC6
CCND2
CDKN1B
CCNA2
CCND1
CCNA1
CDK2
CDC25B
CCNE1
TP53
CCNH
CCNB1
MNAT1
CDK7
GTF2H1
ERCC3
GTF2H5
ERCC2
GTF2H3
GTF2H2
GTF2H

Figure 6-11 String analysis of Cyclin H

String analysis of CCNH reveals extensive known and predicted protein-protein interactions with its complex proteins MNAT1 (MAT1) and CDK7 forming the CAK complex. The CAK complex is also shown to have functional relationships with other cyclin/CDK complexes (top group) and TFIIH basal TFs helicases and general transcription and DNA repair factor core complex (bottom group).

6.2.7 RNA immunoprecipitation of Cyclin H

Although pulldown of *circACVR2A* has uncovered candidate binding partners, MS is extremely sensitive and can identify amino acids from small traces of proteins. To validate the pulldown, and interaction between CCNH and *circACVR2A*, RNA immunoprecipitation (RIP) was performed to isolate CCNH and detect *circACVR2A* bound to the protein. IP using an anti-CCNH was not possible, hence CCNH was expressed with a C-terminal FLAG sequence and could be detected with both anti-CCNH and anti-FLAG antibodies (

Figure 6-12). CCNH FLAG was detected at a higher molecular weight than the endogenous CCNH, using the anti-CCNH antibody, due to the addition of a 20-peptide spacer and 8 peptides encoding for the FLAG tag.

Along with the expression of CCNH FLAG in HEK293T cells, *circACVR2A* was also expressed from pcDNA3.1 plasmid to increase the chance of detecting an interaction between CCNH and *circACVR2A* (Figure 6-13) with expression over 6000% times higher than the endogenous level. Transfection of cells with EV pcDNA3.1 was included, which allowed the determination of *circACVR2A* overexpression efficiency in the CCNH and *circACVR2A* overexpression lysate.

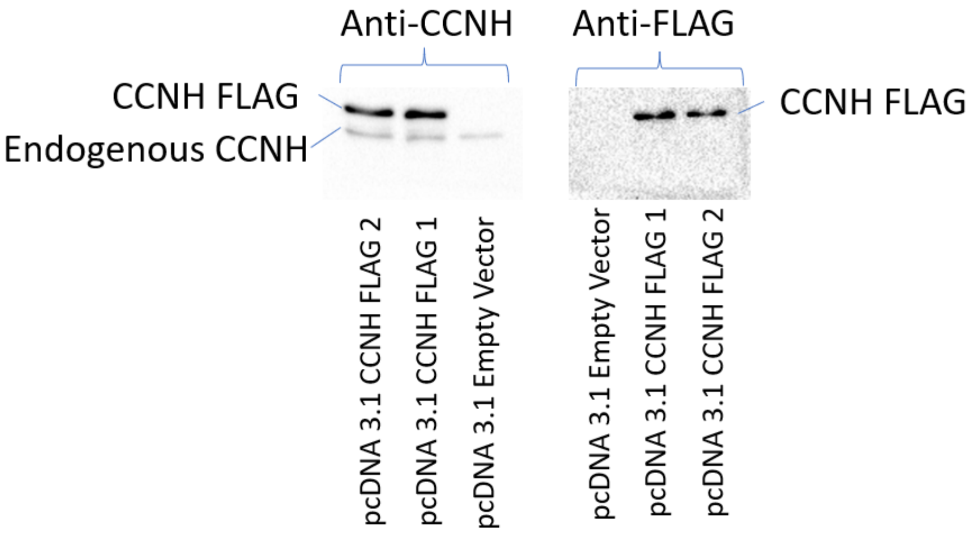

Anti-CCNH
Anti-FLAG
CCNH FLAG
Endogenous CCNH
CCNH FLAG
pcDNA 3.1 CCNH FLAG 2
pcDNA 3.1 CCNH FLAG 1
pcDNA 3.1 Empty Vector
pcDNA 3.1 Empty Vector
pcDNA 3.1 CCNH FLAG 1
pcDNA 3.1 CCNH FLAG 2

Figure 6-12 Western blot verifying pcDNA 3.1 CCNH expression with C-terminal FLAG tag

Western blot of transfected pcDNA 3.1 CCNH FLAG and pcDNA 3.1 empty vector control lysates using (A) anti-CCNH antibody and (B) anti-FLAG antibody.

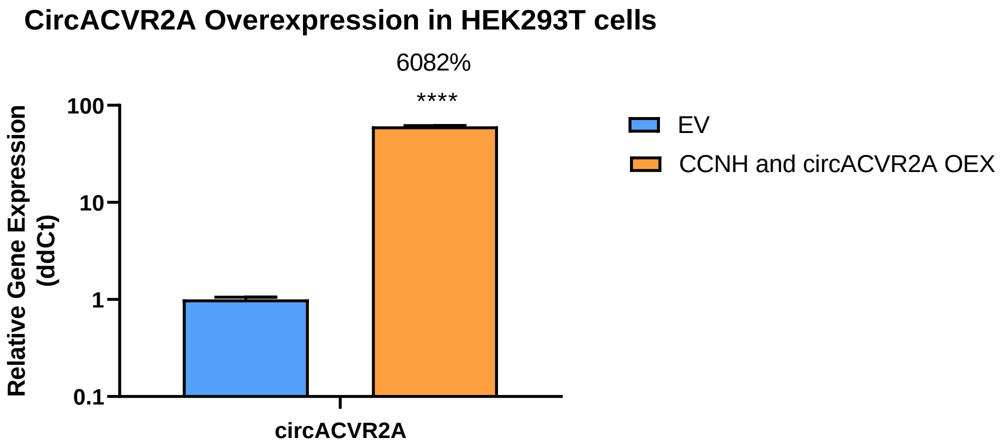


Figure 6-13 QRT-PCR of *circACVR2A* overexpression during CCNH FLAG RIP

QRT-PCR validation of *circACVR2A* overexpression in HEK293T cells in CCNH and *circACVR2A* overexpression cells used for RIP. Relative to *GAPDH* expression. Statistical analysis was performed with an Unpaired One-tailed t-test. (**** = p<0.0001) (n=3; Mean+SEM).

Two RIPs were performed using a cell lysate with overexpression of CCNH FLAG and *circACVR2A*. After removing some of the lysate for RNA and protein input levels, the remaining 1 ml lysate from 1 well of a 6 well plate was split in two. 500 µls was mixed with Dynabeads coated with anti-FLAG antibody (CCNH IP) and the other 500 µls mixed with mouse IgG coated Dynabeads (IgG IP) to use as a control, as non-specific IgG should not bind to CCNH FLAG. 75% of both IP lysates were used for CCNH FLAG detection by western blot, while the remaining 25% was used to extract RNA for *circACVR2A* quantification. CCNH IP was successful as CCNH was detected using the anti-FLAG antibody for IP, at the expected size of ~35kDa, and was not detected in the IgG control IP (Figure 6-14).

Using ImageJ, the intensities of the CCNH input and anti-FLAG IP CCNH protein bands were quantified, and, extrapolating the total amount of protein in the input lysate and CCNH IP, 48.58% of total cellular CCNH FLAG was captured in the CCNH IP. Furthermore, CCNH FLAG was depleted to a greater extent, as it was not detected, in the anti-FLAG unbound lysate after incubation of the lysate with anti-FLAG coated beads, as a faint band was still detected in the IgG IP. This further supports that the anti-FLAG antibody specificity binds to CCNH FLAG while the control mouse IgG antibody has less binding capacity for CCNH FLAG.

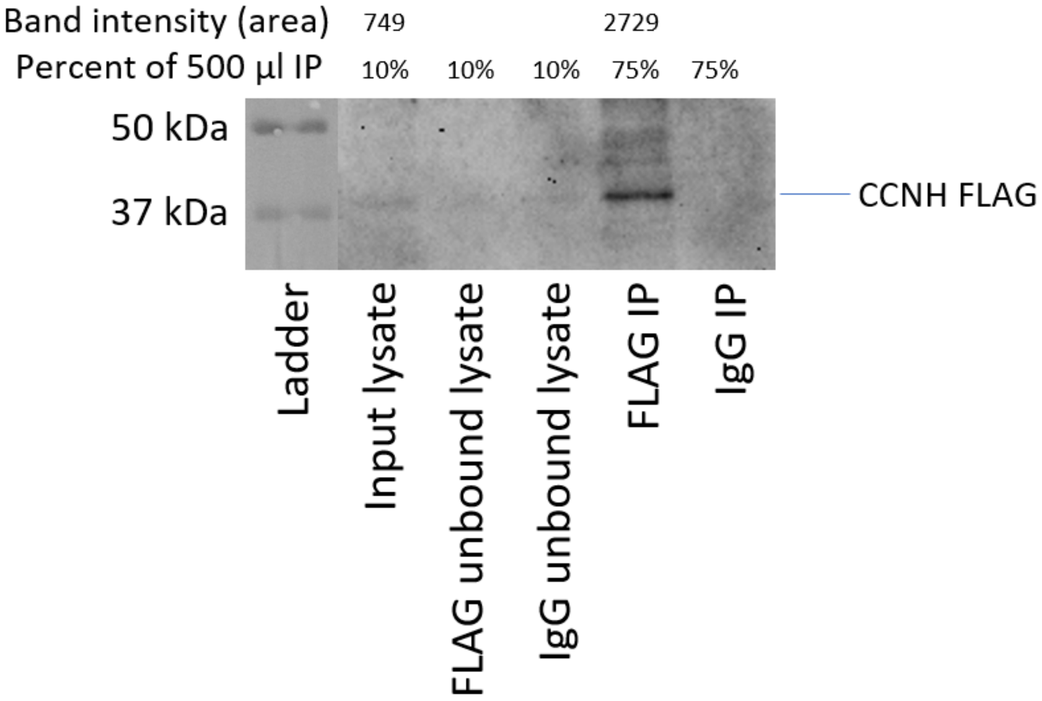

Band intensity (area)
749
2729
Percent of 500 µl IP
10%
10%
10%
75%
75%
50 kDa
37 kDa
CCNH FLAG
Ladder
Input lysate
FLAG unbound lysate
IgG unbound lysate
FLAG IP
IgG IP

Figure 6-14 Western blog of RNA immunoprecipitation of CCNH by FLAG antibody bound to Dynabeads

Lane 1. Precision Plus Protein Kaleidoscope Ladder Lane 2. 10% of input lysate. Lane 3. 10% of lysate after incubation with FLAG antibody coated Dynabeads. Lane 4. 10% of lysate after incubation with IgG control antibody coated Dynabeads. Lane 5. 75% of lysate bound to FLAG antibody coated Dynabeads. Lane 6. 75% of lysate bound to Mouse IgG control antibody coated Dynabeads.

To confirm that the pulldown of *circACVR2A* and identification of CCNH by MS was a genuine interaction, qRT-PCR for *circACVR2A* was performed. RNA was extracted from 25% of the IP after antibody-conjugated bead incubation with the lysate, and washing of the coated beads, for both CCNH and IgG IP. *CircACVR2A* was detected in the CCNH IP lysate by qRT-PCR and was not detected in the control IgG IP lysate (Figure 6-15). *U1* small nuclear RNA (snRNA), was used as a positive control as it has been previously shown to bind to CCNH (O'Gorman *et al.*, 2005). Confirming the success of the CCNH IP, *U1* was also detected in the CCNH IP by qRT-PCR. The presence of *GAPDH* in the CCNH IP was also tested by qRT-PCR to confirm the loss of a non-specific RNA during IP of CCNH (Figure 6-15). GAPDH was not detected in the FLAG or IgG IPs, supporting that *circACVR2A* and *U1* are specifically enriched during CCNH IP. Therefore, non-specific *GAPDH* RNA is not enriched in the CCNH IP, while *circACVR2A* and control *U1* transcripts are enriched, confirming an interaction between CCNH and *circACVR2A*.

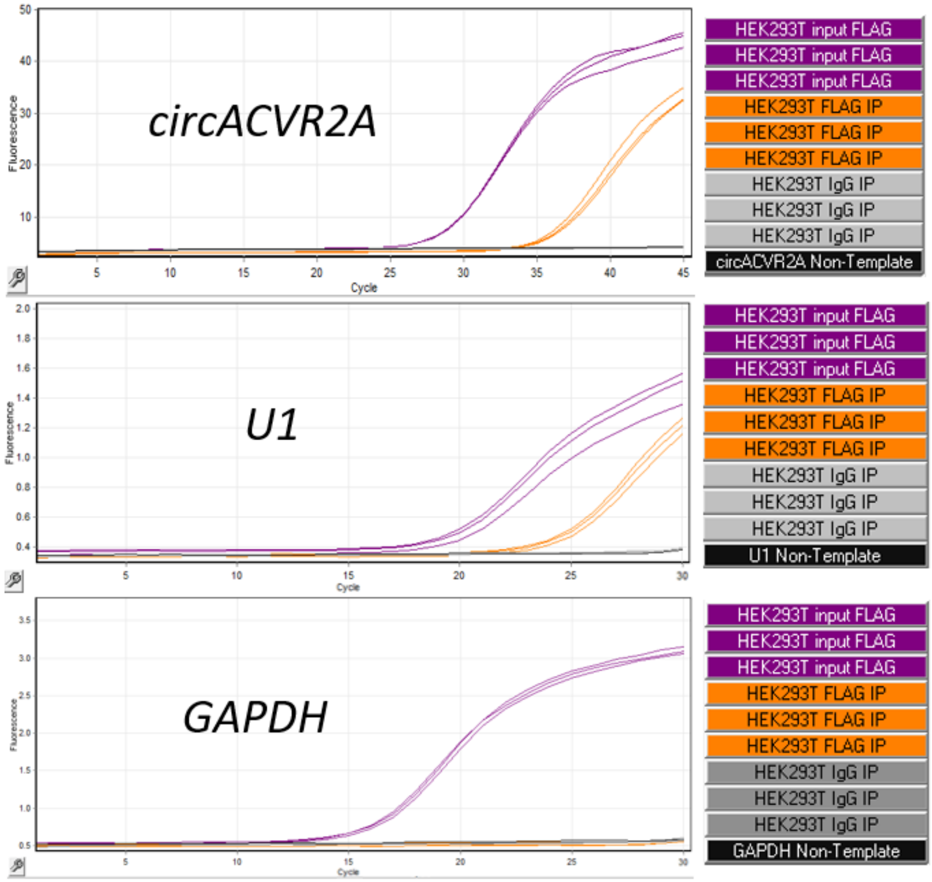


Figure 6-15 Raw fluorescence from qRT-PCR of CCNH FLAG RIP detecting *circACVR2A, U1* snRNA, and *GAPDH*.

The proportion of the total cellular *circACVR2A* and *U1* RNA transcript levels, as determined from the input lysate, found to be present in the CCNH IP, was determined by calculating the Delta Ct of the technical replicates in the CCNH adjusting for the amount of cDNA added into each PCR reaction (Appendix 6). Approximately 0.18% of total cellular *circACVR2A* was found to be bound to the 48.58% of IPed CCNH FLAG. In conjunction, 0.93% of total cellular *U1* was also bound to the 48.58% of IPed CCNH FLAG.

6.2.8 Cell cycle analysis of *CircACVR2A* knockdown A549s

Cell cycle analysis was performed to identify if the cell cycle is altered as a result of decreased *circACVR2A* in A549 cells, as *circACVR2A* was confirmed to bind to CCNH, and that a reduction of *circACVR2A* leads to a decrease in proliferation. Propidium Iodide staining of ethanol fixed A549s revealed an increased proportion of cells in the G1 phase of *circACVR2A* KD compared to EV and scrambled shRNA control A549s (Figure 6-16A). Both FlowJo analysed data (Figure 6-16B) and raw data from the CytExpert software after gating for the populations of cells were included (Figure 6-16C). FlowJo alters the raw data to match two different mathematical models. Using the recommended Watson Pragmatic model, FlowJo identified no cells in G2/M phase for *circACVR2A* shRNA 1 treated cells (Figure 6-16C) which would indicate that the cells are not dividing, however this cannot be true as cells are still proliferating. Analysis of the raw data using the CytExpert software, and not using mathematical models, where G2/M cells have double the amount of DNA than G1 cells, ~12% of cells are in G2/M. However, using either FlowJo or analysing the raw data from the CytExpert software, a higher proportion of cells in G1 is observed, indicating a block/inhibition in the cell cycle in G1.

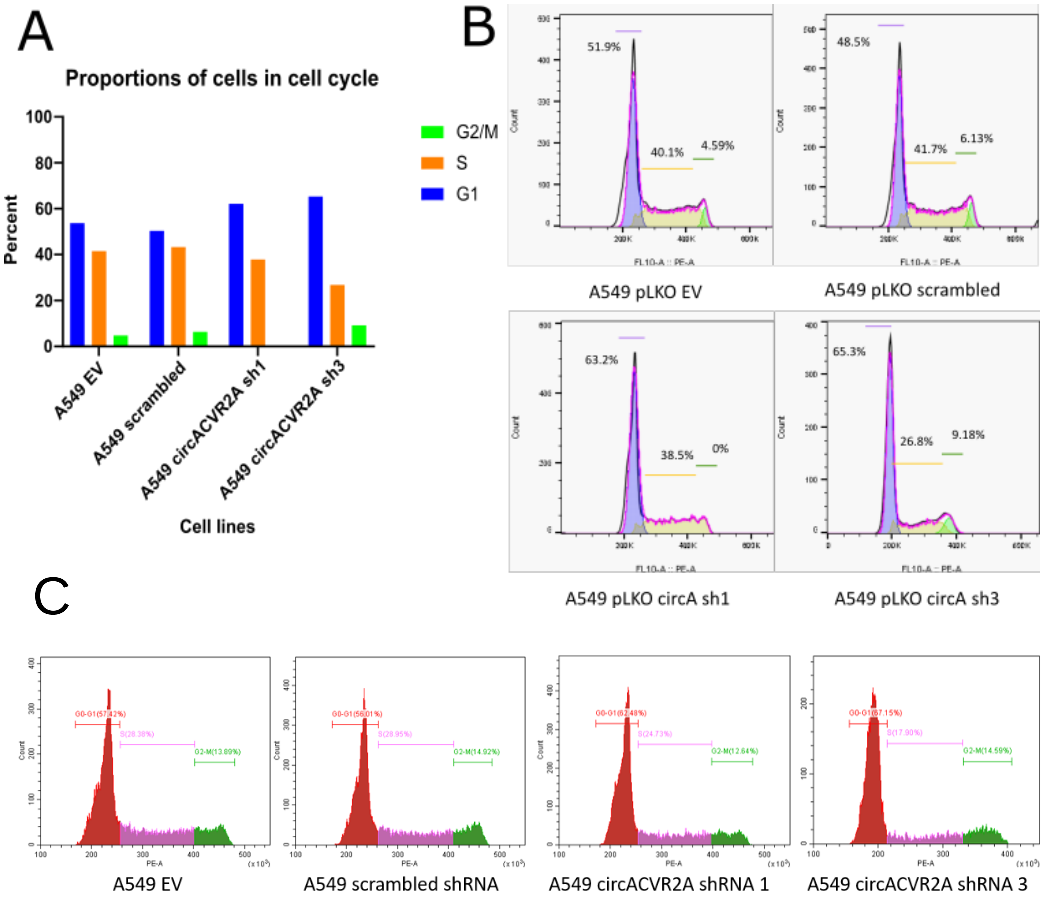


Figure 6-16 Cell cycle analysis of *circACVR2A* shRNA knockdown A549s

Cell cycle analysis of A549 control cells and *circACVR2A* KD cells with Propidium Iodide. (A) Graph representation of proportions of cell cycle phases using the FlowJo analysis. (B) Cell cycle histogram for each cell line. Modelling of cell cycle is determined using FlowJo and the Watson pragmatic model. (C) Raw data from CytExpert with manual gating of cell cycle phases where G2/M has double PI intensity compared to G1.

6.2.9 Regulation of *circACVR2A* during the cell cycle.

To identify if *circACVR2A* is regulated during the cell cycle, A549 cells were synchronised in G0/G1 by serum starvation. A549s were serum starved for 48 hours and released with fresh media containing 10% FBS to allow cells to progress through G1 phase and into S phase in synchronisation. Cells were collected for

RNA extraction and cell cycle analysis every 2 hours. Although the serum starvation was not completely effective at blocking cells in G0/G1 (only 75% of cells in G1), a population of cells did migrate together through G1, S and G2/M (Figure 6-17A and Figure 6-17B). The abundance of *circACVR2A* was quantified for each time point (Figure 6-17C) and a correlation of *circACVR2A* levels and the percentage of cells in G1 was observed (Figure 6-17E), however this was heavily reliant on the first time point (0hr) while *circACVR2A* expression was the highest. *CircACVR2A* is most expressed in G0/G1 during serum starvation and is found to decrease up until 16-18 hours after the release of serum starvation when the highest number of G2/M cells were recorded. After 20 hours the number of cells in G1 increases as the population of cells makes it through a cell division, although no immediate increase of *circACVR2A* is observed. As the serum starvation blocks the cells sometime during G0/G1, c*ircACVR2A* expression in early and late G1 cannot be delineated. In summary, *circACVR2A* is increased during the G0/G1 serum starvation block and a correlation is observed between *circACVR2A* with the percentage of cells in G1. However, the parental transcript is not correlated with the percentage of cells in G0/G1 (Figure 6-17F).

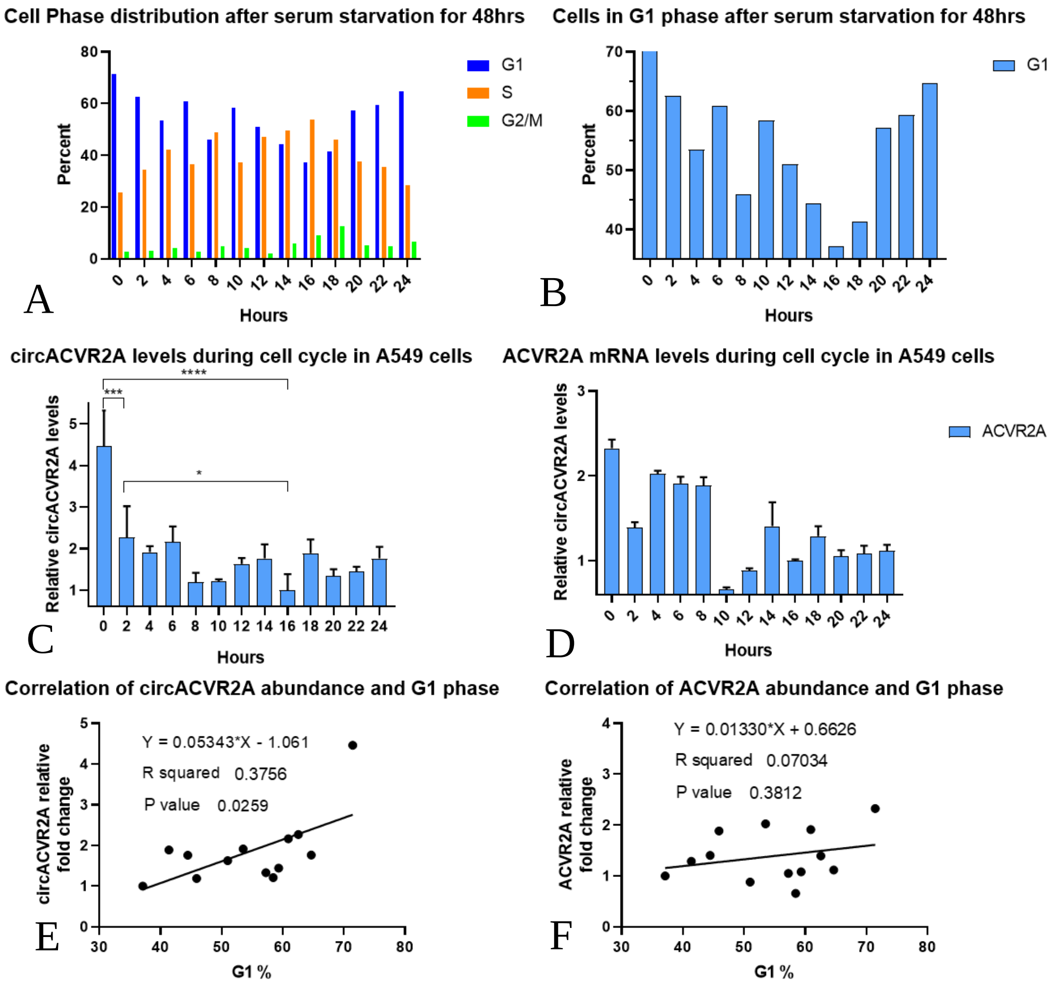


Figure 6-17 *circACVR2A* cell cycle profiling after serum starvation

Profiling of *circACVR2A* levels during cell cycle. A549 cells were serum starved for 48 hours, relaxed from the treatment with DMEM containing 10% FBS. RNA was harvested every 2 hours and cells were taken for cell cycle analysis. (A) Proportion of cells in each phase at each timepoint. (B) Number of cells in G1 phase of cell cycle. (C) Relative gene expression of *circACVR2A* levels in cell cycle. Relative to *GAPDH* and *ACTB*. Baseline expression is from timepoint 16hrs. Statistical analysis was performed with an Ordinary one-way ANOVA with Fishers LSD test. (n=3) (D) Relative gene expression of *ACVR2A* levels in cell cycle. Relative to *GAPDH* and *ACTB*. Baseline expression is from timepoint 16hrs. (n=3; Mean+SEM) (E) Two-tailed Pearson's Correlation of *circACVR2A* and percentage of cells in G1. (F) Two-tailed Pearson's Correlation of *ACVR2A* and percentage of cells in G1. * = p<0.05, *** = p<0.001, **** = p<0.0001.

To improve the resolution of *circACVR2A* expression during early G1 and late G1, Nocodazole cell cycle blockage was performed to synchronise cells. Nocodazole interferes with the polymerisation of microtubule assembly, impairing formation of the metaphase spindles and blocking G2/M transition before mitosis occurs (Blajeski *et al.*, 2002; Beswick, Ambrose and Wagner, 2006). Similarly to the serum starvation analysis, cells were harvested every 2 hours after release of nocodazole treatment for both cell cycle analysis and RNA extraction (Figure 6-18A). Cells in G2/M blockage by nocodazole were found to have the lowest level of *circACVR2A* which increased during progression of the cell cycle into G1 after mitosis had completed (Figure 6-18B and Figure 6-18C). *CircACVR2A* expression increased during G1 phase, and had no appreciable change for the quantified time points after 10 hours. Additionally, *circACVR2A* was observed to increase from before mitosis and up to 2-4 hours after mitosis (Figure 6-18C). *CircACVR2A* was also more abundant in late G1 (10hr) compared to early G1 (2hr) as well as G2/M cells. This supports the serum starvation data that *circACVR2A* has higher expression in G1 compared to cells in G2/M, but also that *circACVR2A* expression increases during G1 and is at its peak in late G1.

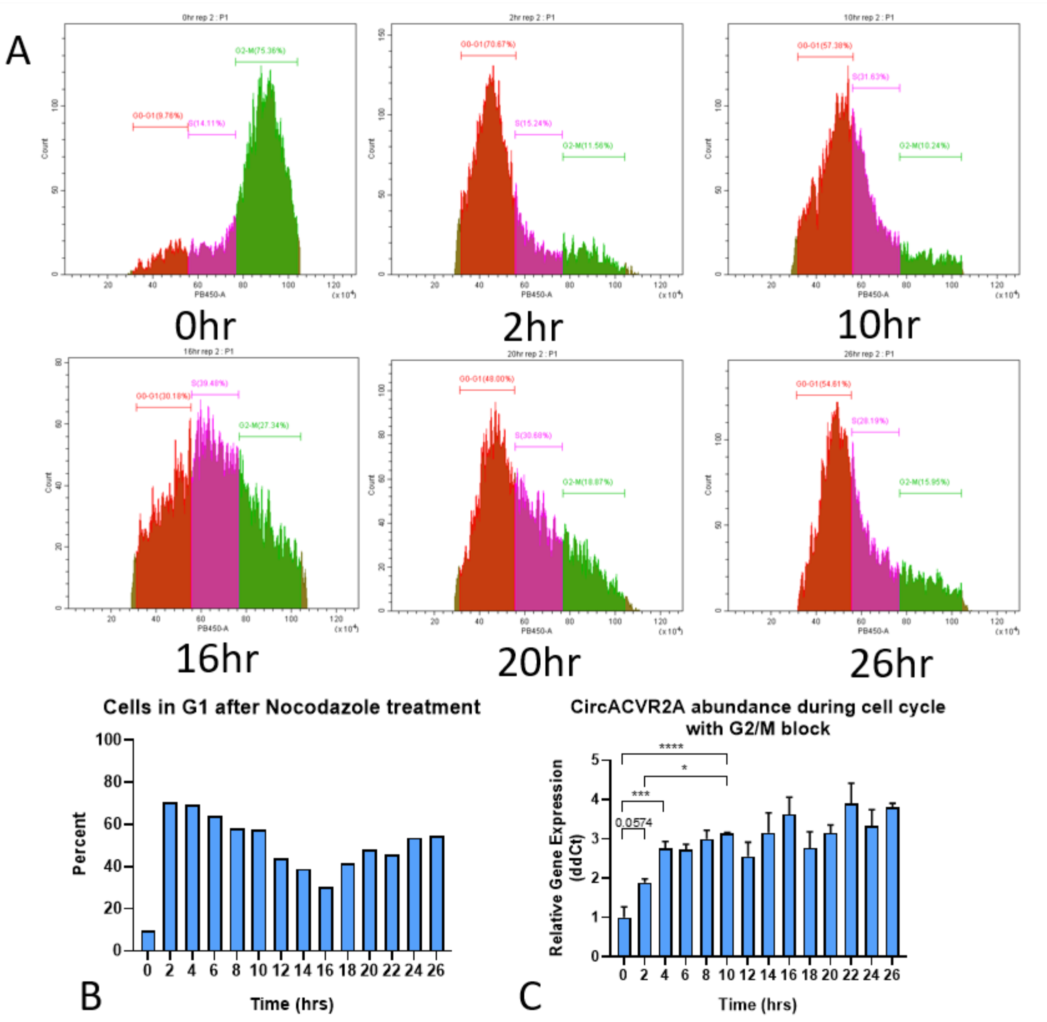


Figure 6-18 *circACVR2A* cell cycle profiling after Nocodazole treatment

Nocodazole treatment and QPCR analysis of *circACVR2A* during cell cycle. (A) Histogram plots of cell cycle proportions from some of the timepoints. Raw data from CytExpert. (B) Proportion of cells in G1. (C) Relative gene expression of *circACVR2A* levels in cell cycle. Relative to *GAPDH* and *ACTB*. Statistical analysis was performed with an Ordinary one-way ANOVA with Fisher's LSD test. * = $p<0.05$, *** = $p<0.001$, **** = $p<0.0001$ (n=3; Mean+SEM).

Directly comparing serum starved cells in G1 and nocodazole treated cells in G2/M (Figure 6-19B), *circACVR2A* as well as *ACVR2A* is more abundant in G1 cells than in G2/M cells (Figure 6-19A), confirming the validity of the two independent cell cycle analysis experiments above. Therefore, KD of *circACVR2A* induces a cell cycle blockage within the same cell cycle phase where an increase of *circACVR2A* expression is observed in unmodified A549s.

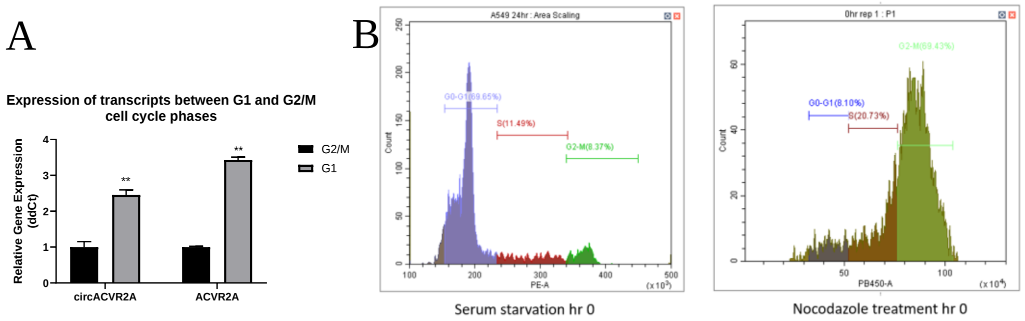


Figure 6-19 *circACVR2A* abundance in serum starved and nocodazole treated cells

CircACVR2A is more abundant in G1 cells than in G2/M cells. (A) Transcript expression in G1 and G2/M cells. Relative to *GAPDH* and *ACTB*. Statistical analysis was performed with a two-way ANOVA with Fisher's LSD test. ** = p<0.01 (n=3; Mean+SEM). (B) Cell cycle histograms of the two cell populations used in qPCR.

6.2.10 Decreased CDK1/2 activity during *circACVR2A* KD

During the cell cycle, cyclin/CDK complexes phosphorylate proteins in a organised and sequential manner to achieve cycling through G1, S, G2 and finally mitosis. CDK1/2 are highly functional during G1 and G1/S transition and perturbation of these CDKs and their bound cyclins can inhibit the progression into S phase. Of note, the CAK complex, comprising of CCNH, CDK7 and MAT1, is responsible for the phosphorylation and activation of CDK1/2 during the cell cycle in a cell cycle phase-dependent manner. Western blot analysis has shown a decrease to the amount of CDK1/2 phosphorylated substrates in A549 cells during *circACVR2A* KD (Figure 6-20A and Figure 6-20B). This further supports that the cell cycle is affected by KD of *circACVR2A* possibly through the interaction between *circACVR2A* and CCNH.

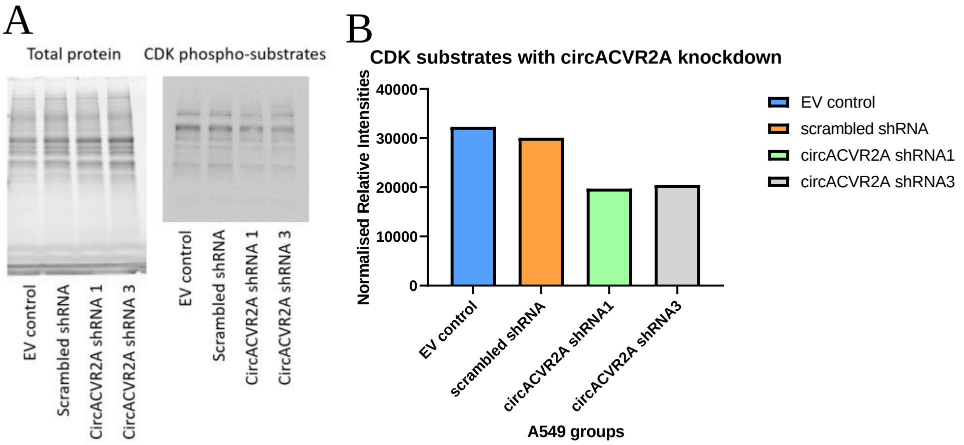


Figure 6-20 Western blot of phosphor-CDK substrates in *circACVR2A* shRNA knockdown A549s

Western blot of phospho-CDK substrates of control and *circACVR2A* KD A549 cells. (A) Total protein normalisation and blot for phosphor-CDK substrates. 1/1000 dilution. (B) Normalised relative intensities of total phospho-CDK substrates.

6.3 Discussion

As TGFβ signalling is a major driver of differentiation in stem cells and an activator of EMT, with an inhibition of TGFβ signalling inducing neural lineage differentiation, *circACVR2A* may have been involved with the regulation of TGFβ signalling. However, A549 cells with KD of *circACVR2A* show no differences in their ability to induce a phenotype during hTGFβ1-induced EMT. Furthermore, *circACVR2A* was not found to not be regulated during hTGFβ1-induced EMT in A549s.

KD of *circACVR2A* decreased the abundance and sensitivity of *SNAI1* RNA transcripts in A549 cells, however *SNAI1* RNA levels were not decreased during *circACVR2A* KD in H9 hESCs. Furthermore, EMT was induced with reduction of *circACVR2A* in H9 hESC, yet this was not reciprocated in A549 cells. Therefore, reduction of SNAI levels and sensitivity is unlikely driving hESC differentiation. Although *circACVR2A* expression did not change during TGFβ-induced EMT in

A549 cells, reduction of *circACVR2A* did negatively impact proliferation and migration.

CircACVR2A was shown to bind to CCNH, a component of the CAK complex, in HEK293T cells by a pulldown of *circACVR2A* and with identification of bound proteins by MS. In support of the *circACVR2A* pulldown and MS, RIP of FLAG tagged CCNH successfully detected *circACVR2A* and was not detected in the control IgG RIP. The control *U1* transcript, used to identify if the RIP was successful, was also found to be enriched in the CCNH IP. A higher percentage of the cellular pool of *U1* was found to be bound to CCNH, however *U1* is found to be more highly expressed that *circACVR2A* even after the circRNA overexpression. Furthermore, *U1* snRNA is localised in the nucleus of cells (Huang and Spector, 1992, p. 2), while *circACVR2A* is mainly localised to the cytoplasm (Dong *et al.*, 2019). CCNH has been shown to have varying degrees of nuclear and cytoplasmic localisation, depending on the cell line, organism or pathology (Devault *et al.*, 1995; Krempler *et al.*, 2005; Peng *et al.*, 2020), and a shift between nuclear and cytoplasmic CCNH has been observed in developing *Drosophila* embryos (Aguilar-Fuentes *et al.*, 2006). Of note, mouse ES cells have both cytoplasmic and nuclear CCNH, being present in both cellular fractions (Patel and Simon, 2010) with levels remaining steady over 8 days of differentiation. Therefore, the differences observed between the binding capacity of *U1* and *circACVR2A* may depend on the localisation of CCNH within cells. Additionally, being a regulator of the cell cycle, *circACVR2A* and *U1* may have different levels of binding to CCNH within different phases of the cell cycle as the CAK complex sequentially activates different CDK/cyclin complexes in a cell phase-dependent manner.

Cell cycle analysis of *circACVR2A* KD A549s revealed an increase to the G0/G1 phase cell population compared to cells in S, G2 and M. Knockdown of CCNH in an ovarian cancer cell line has also shown to induce a G1 block and a decrease to cell proliferation (Peng *et al.*, 2020) similarly to *circACVR2A* knockdown in A549 cells. Furthermore, reduction of CCNH *in vivo* decreased the growth of ovarian tumour growth in mice.

G1 is an important phase for hESCs as they can receive mitogenic signals for growth and differentiation. Stem cells are known to have a very short G1 phase as this allows them to be less receptable to mitogenic signals, maintaining an undifferentiated state. Perturbation of this phase can induce neuroectoderm differentiation of hESCs if TGFβ signalling is inhibited (Pauklin and Vallier, 2013) or if G1 is lengthened (Jang *et al.*, 2019). Thus, a block in G1 may be mediated by a dysregulation of CCNH with the reduction of *circACVR2A* levels. However, it is not known how a specific block in G0/G1 occurs if the CAK complex is involved with activating multiple CDKs (CDKs 1, 2, 4 and 6) at different stages in the cell cycle.

To further understand the regulation of *circACVR2A*, and how loss of *circACVR2A* can lead to a block in a specific phase of the cell cycle, the abundance of *circACVR2A* in the different phases of the cell cycle was determined using multiple cell synchronisation techniques. Both methods support that *circACVR2A* has higher expression in G1 than S and G2/M, with nocodazole synchronisation indicating that *circACVR2A* may be most expressed in late G1. Cell cycle synchronisation of A549s by serum starvation was moderately successful as a population of cells was observed to move through the cell cycle however complete

G1 blockage was not possible. Although serum starved cells are seen to be arrested in G1 phase using cell cycle analysis techniques, these cells are thought to be “out of the cell cycle”, and in a quiescent state called G0 (Cooper, 2003), however they classify as G1 cells by flow cytometry analysis as they have not undergone DNA synthesis in S phase. After relaxing the serum starvation by the addition of fresh culture media containing FBS, cells were observed to progress through the cell cycle. *CircACVR2A* was found to be expressed the highest in the serum starved cells and decreased in abundance as the cells progressed through to mitosis. Twenty hours post serum starvation, the number of cells in G1 increases again as a population of cells have undergone a cell division, although no immediate increase of *circACVR2A* was observed as would be expected if *circACVR2A* was upregulated in G1. This could potentially be due to the cells becoming less synchronised, or *circACVR2A* expression is increased only during serum starvation, while cells are in G0 and not actively progressing though G1.

To confirm if *circACVR2A* is increased in G1 and not during G0, and also to profile *circACVR2A* expression throughout G1, cells were blocked using nocodazole, impairing the ability of cells to enter mitosis. After withdrawal of nocodazole, cells will progress normally through G1 without entering G0 due to serum withdrawal. Nocodazole treatment of A549 cells was also moderately successful with ~75% of cells in G2/M, and 70% of cells in G1 after 2 hours. Validating that *circACVR2A* is expressed during G1 and not in G0, *circACVR2A* increased in abundance after mitosis and continued increasing through to late G1 and until S phase. By the end of G1, the expression of *circACVR2A* did not change as synchronisation was likely decreasing, similar to the serum starved cells. Even though cells became asynchronous, both serum starvation and nocodazole treatment show regulation of *circACVR2A* during the cell cycle with higher levels in G1 than in G2/M. Notably, *circACVR2A* was found to increase in abundance during G1, without cells entering G0 due to serum withdrawal. The nocodazole cell cycle analysis also gives evidence to suggest that *circACVR2A* is more highly expressed in late G1, an important phase with which neuroectoderm differentiation specifically occurs due to endogenous inhibition of TGFβ signalling before G1/S transition. Therefore, *circACVR2A* expression is the highest during late G1, the precise cell cycle phase that neuroectodermal signalling occurs in hESCs. Furthermore, *circACVR2A* associates with CCNH, a component of the CAK complex, involved with the phosphorylation of multiple cyclin/CDK complexes throughout the cell cycle. Additionally, KD of CCNH by shRNAs, also induces a G1 block, alike to *circACVR2A* KD.

Western blot analysis showed decreased levels of CDK1/2 phospho-substrates in *circACVR2A* KD A549s, suggesting that G1/S cell cycle phospho-substrates are not as activated (phosphorylated) or that a lower proportion of cells are within CDK1/2 activated phases. This may be due to a dysregulation of CCNH function with depletion of *circACVR2A*. A reduction to the level of activation of CDK1/2 may lengthen G1, allowing active CDK4/6 during late G1 to maintain the phosphorylation of SMAD2/3, blocking translocation to the nucleus, which predisposes hESCs to neural differentiation.

Although it is known that CAK is able to phosphorylate many CDKs, and that MAT1 can alter the specificity of CAK phosphorylation, the temporal regulation of CAK phosphorylation is not completely understood. Similarly to how the addition the MAT1 alters the specificity of the CAK complex, *circACVR2A* addition may act to alter the specificity of CAK, or may either stabilise, promote or dissociate the CAK complex in a cell cycle phase-dependent manner, as these are reported mechanisms for circRNAs (Zhang *et al.*, 2013; Ashwal-Fluss *et al.*, 2014; Z. Li *et al.*, 2015; Abdelmohsen *et al.*, 2017; F. Yang *et al.*, 2018). Of note, *CircFOXO3* is regulated during the cell cycle, being most abundant in G1, and binds to CDK2 and p21 to inhibit the interaction of CDK2 with its phase dependent cyclin to arrest cell cycle progression (William W Du *et al.*, 2017). Further analysis is required to determine the functional consequences to CCNH during *circACVR2A* KD.

6.4 Chapter 6 Conclusion

The reduction of *circACVR2A* in A549 cells has decreased cell proliferation, migration, and caused a cell cycle blockage in G1. MS has provided protein candidates, and successful identification lead to the discovery of *circACVR2A* binding to CCNH. The phenotypes observed by KD of *circACVR2A* in both hESCs and A549s may be mediated by the loss of interaction between *circACVR2A* and CCNH. Reduction of CCNH alone has been shown to also induce a G1 block, similar to *circACVR2A*. Early and late G1 are critical for hESC differentiation with cells in late G1 exclusively differentiating into neuroectoderm. *CircACVR2A* increases in expression during G1, peaks in late G1 and decreases during subsequent phases of the cell cycle. This cell cycle regulation may be a mechanism for phase-dependent regulation of CCNH, or the CAK complex, by *circACVR2A* specifically during G1. Therefore, reduction of *circACVR2A*, impacting CCNH function, a critical activator of the cell cycle, may be inducing the block in late G1 phase, inducing hESCs to undergo neuroectoderm differentiation. However, further investigation is required to confirm this hypothesis. This should include identifying if the CAK complex is stabilised or dissociated during *circACVR2A* KD or overexpression, or if altering *circACVR2A* levels changes the subcellular localisation of CCNH. Investigation into whether CDK4/6 activity may also give insightful information into if *circACVR2A* is able to affect specific cyclins and would give evidence to suggest that *circACVR2A* may be regulating the specificity of the CAK complex during the cell cycle. In addition, CDK7, as part of the CAK complex, phosphorylates the CTD domain of RNA pol II, promoting promotor escape. Profiling any effects that KD of *circACVR2A* may elicit on pol II function may also provide a mechanistic function for *circACVR2A* interacting with CCNH.

Chapter 7 Major Project Findings

7.1 Thesis summary

Profiling and characterisation of circRNAs during hESC differentiation was investigated by high throughput sequencing of circRNAs before and after hESC differentiation into multiple cell types. Enrichment of circRNAs, before RNA sequencing library preparation, has increased the variety of detected circRNAs and increases statistical power, required for differential expression analysis and characterisation between samples. CircRNA expression is cell-specific, and this expression profile alone is able to stratify cell populations similarly to mRNA expression profiles. A moderate correlation between circRNA abundance and mRNA abundance during stem cell differentiation into multiple lineages was discovered, however, for specific individual circRNAs, different regulatory mechanisms drive cell type-specific expression. Of note, circRNAs show positive or negative correlation with their parental mRNA transcripts, along with a variety of circRNAs that are independently regulated.

The reduction of circRNAs by targeting the backsplice junction with RNAi techniques illuminates the consequences of circRNA dysregulation that may be relevant in the fields of stem cell differentiation and development. Profiling *CDYL* mRNA variant abundances during *circCDYL* KD and overexpression in hESCs has uncovered an interesting and unpublished ability for *circCDYL* to regulate alternative splicing of its parental mRNA. In addition to this novel observation, hESCs with a reduction of *ACVR2A* mRNA was found to increased abundance of *circACVR2A*. An increase to circRNA abundance during mRNA KD is not a currently published phenomenon and requires further profiling to understand if this occurs for many genes that produce circRNAs. The consequences of not identifying circRNA dysregulation during the KD of any mRNA may have implications for many published experiments that use RNAi methodologies.

Profiling *circACVR2A* KD in hESCs has revealed a remarkable differentiation event which induces EB formation from adherent hESCs cultured on vitronectin ECM, and neuronal differentiation to produce neurons. Furthermore, KD of *circACVR2A* slows proliferation and migration of a human lung adenocarcinoma cell line and a G1 cell cycle blockage was observed. Profiling of *circACVR2A* expression during the cell cycle, identified *circACVR2A* as being the most expressed during late G1. Importantly, differentiating hESCs in late G1 or ES cells that have a longer than average G1 length, have altered differentiation propensity, as cells will exclusively differentiate into neuroectoderm.

In an effort to determine the functional interactors of *circACVR2A*, a circRNA pulldown and RIP were employed to verify protein interactions. This lead to a confirmed interaction between *circACVR2A* and CCNH. Although no consequences on CCNH function or complex formation during *circACVR2A* KD were performed in the context of this study, the phosphorylation of CDK1 and CDK2 substrates was markedly reduced in A549 cells along with an impact on cell cycle progression within the same phase that *circACVR2A* is most expressed.

The results of this study have provided several currently unpublished circRNA properties which have functional consequences for circRNA and mRNA regulation using current RNAi methodologies. Additionally, dysregulation of individual circRNAs that are regulated during hESC differentiation have shown to affect hESC pluripotency and can alter the morphology of hESCs. Characterisation of circRNAs during hESC differentiation has supplied an abundance of circRNAs that require further profiling to determine the functional significance of circRNAs in cell biology and epigenetic regulation.

Appendix 1

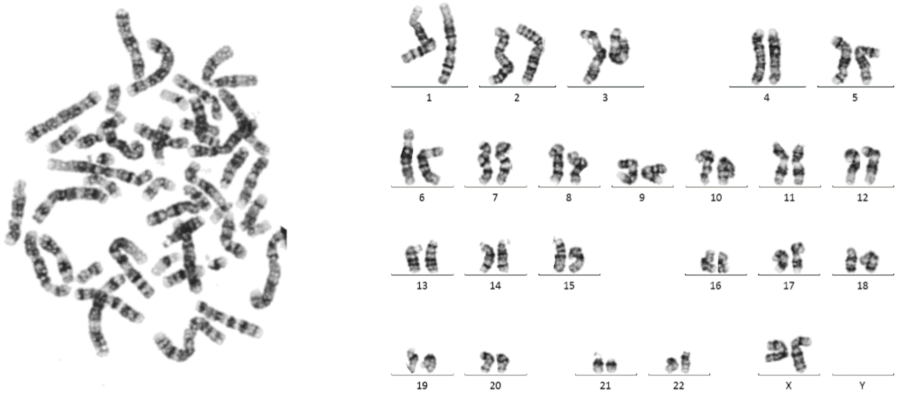

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y

Figure A- 1 Karyotype analysis of H9 hESCs

Karyotype analysis of H9 hESCs confirming a normal XX karyotype and 23 pairs of chromosomes with no abnormalities.

Appendix 2

Table A- 1 Primers, shRNA and gRNA sequences used in this study

circRNA Primers		
circACVR2A	Forward	AACATCCTGCTCTATTCCTTGG
	Reverse	CGCCGTTTATCTTTGTCACC
	PrimeTime Forward	CCAGTGGTCCTCGGACTT
	PrimeTime Reverse	CAAAGATAAACGGCGGCATTG
		TAGGTGTCTGGGTTGAAGACAGAACCAATC
circPIF1	PrimeTime Forward	CTCCTCAGTCGGCAGCA
	PrimeTime Reverse	CTAACAGGACACTGGGCATC
	PrimeTime Probe	CCATCCACAAGAGCCAAGTAAGGTACACAGATTT
circCDYL	Forward	GATCCTCGTGCCTAAAAGCC
	Reverse	TCCTTTCAACCTTTCCCGTT
circFAT3	Forward	ATCAGGGATGGCAGTGGTCT
	Reverse	GATGAGGCAACAAGCAGGAG
circZNF609	Forward	TCGCCTGCTAAAGAAAGTCAA
	Reverse	ATGTCCCATTCATCCCCACT
mRNA Primers		
ACVR2A	Forward	TCCAGAGGTATTAGAGGGTGCT
	Reverse	CGAGAAGCCAGTTCCCATAG
All CDYL	Forward	GAATCCAAAGCCCTCGTG
	Reverse	CGGTGATGTTCTCCTGCTC
CDYL iso 2	Forward	GGGAAAGGAGTGACTGCCA
	Reverse	AACAAGATGTGGGTGAAGCC
CDYL iso3	Forward	TACGAGGTACATCTCCGTTCA
	Reverse	TCCTGCTTCCTGACCACAAT
PIF1	Forward	GGAGAGGCAATGAAGACAGG
	Reverse	TTTGTTCTGGGCTGGGCTA
ZEB1	Forward	TTCAAACCCATAGTGGTTGCT
	Reverse	TGGGAGATACCAAACCAACTG
CDH1	Forward	TGCTGAGGATGATTGAGGTG
	Reverse	GGGATTCTGGGCTTTGAGTA
CDH2	Forward	CAACTTGCCAGAAAACTCCAGG
	Reverse	ATGAAACCGGGCTATCTGCTC
SNAI1	Forward	GCACATCCGAAGCCACAC
	Reverse	GGAGAAGGTCCGAGCACAC
SNAI2	Forward	TGGTCAAGAAGCATTTCAACG
	Reverse	GGAGGTGTCAGATGGAGGAG
U1	Forward	GGGCGAGGCTTATCCATT
	Reverse	CCCCACTACCACAAATTATGC

YWHAZ	Forward	CGAAGCTGAAGCAGGAGAAG
	Reverse	TTGTGGGACAGCATGGATG
GAPDH	Forward	CAGTGAGCTTCCCGTTCAG
	Reverse	ACCAGAAGACTGTGGATGG
ACTB	Forward	GGCCAACCGCGAGAAGAT
	Reverse	ATCACGATGCCAGTGGTACG
FAT3	Forward	GACTACGAGAGCGTGGGAGAG
	Reverse	TTGCTTGTTGGGACAGTATGAC
B2M	Forward	TGCTGTCTCCATGTTTGATGTA
	Reverse	TCTCTGCTCCCCACCTCTAA
Plasmid sequencing primers		
pLKO.1 plasmid	Forward	CAAGGCTGTTAGAGAGATAATTGGA
	Reverse	TCTCTGCTGTCCCTGTAATAAACC
shRNAs (Sense sequence)		
Scrambled shRNA		GGAATAAAGCGGAGAGAGTTA
circCDYL	shRNA 1	AACGGGAAAGGTTGAAAGGAT
	shRNA 2	CGGGAAAGGTTGAAAGGATTG
circACVR2A	shRNA 1	TCCAACTCAAGTGCTATACTT
	shRNA 2	CAACTCAAGTGCTATACTTGG
	shRNA 3	GTTCCAACTCAAGTGCTATAC
	shRNA 4	TTTGCTACCTGGAAGAATATT
	shRNA 5	AGGTTGTTGGCTGGATGATAT
	shRNA 6	AGACAGCCCTGAAGTATATTT
	shRNA 7	TTGGTGCCACTTATGTTAATT
ACVR2A	shRNA 1	GGACCTGGCTAATGGAGTGTT
	shRNA 2	GCTATGCTTAGTGCCAACAAT
circPIF1	shRNA 1	CAAGAGCCAAGTAAGGTACAC
	shRNA 2	AGAGCCAAGTAAGGTACACAG
circFAT3	shRNA 1	GACGAGAGTGGATGGAAGTAT
circSP100	shRNA 1	GATGAAGAAAGATGTTCACGG
circZFX	shRNA 1	ATGATTTCCTGAGCTGTGACT
circZNF608	shRNA 1	GGGGTCAACCTTCTTCTGTTT
circADAMTSL5	shRNA 1	TTCCATGGGGGCCCCACCTCT
circCDK13	shRNA 1	AATTAATAAGACGGTCTGGAA
circROR2	shRNA 1	ATAACTTTCAGTGAAGTGGAG
circSEMA3E	shRNA 1	TCATTTGGAGAGCTCTTGAAT
circMYBPC3	shRNA 1	TGAAAAAGAGGTCTATCTGTT
circRIMS1	shRNA 1	AGTCAAACAGCAAGCTTAGTA
circSETBP1	shRNA 1	ATGTCTCCAGGCTTACGAGAG
circLAMC3	shRNA 1	GGGTCCCCTTGTAGGAGGGAG
circZNF609	shRNA 1	GTCTGAAAAGCAATGATGTTG

Appendix 2

Cas13b gRNAs sequences		
circACVR2A	gRNA 1	TACCAAGTATAGCACTTGAGTTGGAACAAG
	gRNA 2	TGATCTACCAAGTATAGCACTTGAGTTGGA
	gRNA 3	AGTATAGCACTTGAGTTGGAACAAGTACAG

Appendix 3

LC-MS System Details

uHPLC System
Dionex UltiMate 3000 RSLCnano (Thermo Scientific)
Device: NCS-3500RS Nano ProFlow
S/N: 8157197
Dionex UltiMate 3000 RS Autosampler (Thermo Scientific)
Device: WPS-3000TPL RS
S/N: 8157386

Mobile Phases
Buffer A: Optima® LC/MS 0.1% FA in water (Fisher Chemical #LS118-4)
Buffer B: Optima® LC/MS Acetonitrile (Fisher Chemical #A955-4) with 0.1% Formic Acid
(Fluka Analytical P/N: 94318)

Sample Vials
1.1ml Microliter Short Thread Vial ND9, 32 x 11.6mm, clear glass (Thermo Scientific P/N THC11 09 0620)

Pre-column
µ-Precolumn, 300 µm ID x 5mm, C18 PepMap 100, 5 µm, 100 Å (Thermo Scientific P/N: 160454)

Analytical Column Specifications
Column Type: Reversed Phase
Column Source: In house manufactured
Capillary: 75 µm ID, 363 µm OD VSD Tubing (Trajan # 175352)
Length: 40cm
Packing Material: ReproSil-Pur 120 C18-AQ, 1.9 µm (Dr Maisch GmbH P/N: r119.aq.0001)
Bead size: 1.9 µm
Pore Size: 120 Å
Stationary Phase: C18

Spray Emitter
PicoTip™ Emitter 10± µm (New Objective P/N: FS360-20-10-N-5-105CT)

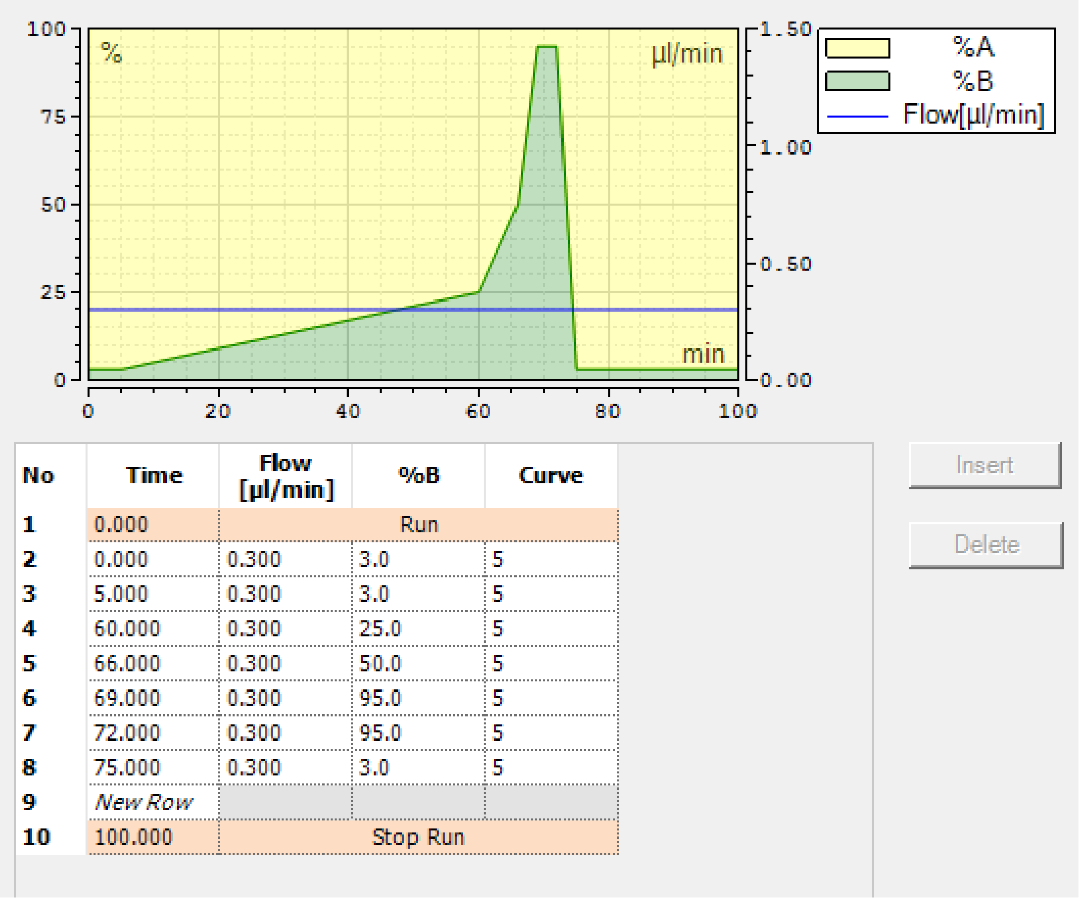
Mass Spectrometer
Model: Orbitrap Exploris 480 (Thermo Scientific)
S/N: MA 10083N
Year of manufacture: 2019

Ion Source
Nanospray Flex (Thermo Scientific)
S/N: ES010777

Ion Mobility Interface

FAIMS PRO (Thermo Scientific)
S/N: FAIMS 20210

Analytical Gradient Details



No	Time	Flow [µl/min]	%B	Curve
1	0.000	Run		
2	0.000	0.300	3.0	5
3	5.000	0.300	3.0	5
4	60.000	0.300	25.0	5
5	66.000	0.300	50.0	5
6	69.000	0.300	95.0	5
7	72.000	0.300	95.0	5
8	75.000	0.300	3.0	5
9	*New Row*			
10	100.000	Stop Run		

Mass Spectrometer Method Summary

Method Settings

Application Mode: **Peptide**
Method Duration (min): **100**

Global Parameters

Ion Source

Ion Source Type: **NSI**
Spray Voltage: **Static**
Positive Ion (V): **1700**
Negative Ion (V): **600**
Ion Transfer Tube Temp (°C): **300**
Use Ion Source Settings from Tune: **False**
FAIMS Mode: **Standard Resolution**
FAIMS Gas: **Static**
FAIMS Gas (L/min): **0**

MS Global Settings

Infusion Mode: **Liquid Chromatography**
Expected LC Peak Width (s): **60**
Advanced Peak Determination: **True**

Default Charge State: **2**
Internal Mass Calibration: **User-defined Lock Mass**
Current Lock Mass: **Current**

Experiment#1 [MS cv-50]

Start Time (min): **15**
End Time (min): **100**
Master Scan:
Full Scan
Orbitrap Resolution: **60000**
Scan Range (m/z): **350-1200**
FAIMS Voltages: **On**
FAIMS CV (V): **-50**
RF Lens (%): **40**
AGC Target: **Custom**
Normalized AGC Target (%): **300**
Maximum Injection Time Mode: **Auto**
Microscans: **1**
Data Type: **Profile**
Polarity: **Positive**
Source Fragmentation: **Disabled**
Scan Description: **-50**
Filters:
MIPS
Monoisotopic peak determination: **Peptide**
Relax restrictions when too few precursors are found: **True**
Intensity
Filter Type: **Intensity Threshold**
Intensity Threshold: **5.0e4**
Charge State
Include charge state(s): **2-8**
Include undetermined charge states: **False**
Dynamic Exclusion
Dynamic Exclusion Mode: **Custom**
Exclude after n times: **1**
Exclusion duration (s): **30**
Mass Tolerance: **ppm**
Low: **10**
High: **10**
Exclude isotopes: **True**
Perform dependent scan on single charge state per precursor only: **False**
Data Dependent
Data Dependent Mode: **Number of Scans**
Number of Dependent Scans: **38**
Scan Event Type 1:
Scan:
ddMS^2
Multiplex Ions: **False**
Isolation Window (m/z): **1.6**

Isolation Offset: **Off**
Collision Energy Mode: **Fixed**
Collision Energy Type: **Normalized**
HCD Collision Energy (%): **30**
Orbitrap Resolution: **15000**
TurboTMT: **Off**
Scan Range Mode: **Define First Mass**
First Mass (m/z): **120**
AGC Target: **Custom**
Normalized AGC Target (%): **75**
Maximum Injection Time Mode: **Auto**
Microscans: **1**
Data Type: **Centroid**
Scan Description:

Experiment#2 [MS cv-70]

Start Time (min): **15**
End Time (min): **100**
Master Scan:
Full Scan
Orbitrap Resolution: **60000**
Scan Range (m/z): **350-1200**
FAIMS Voltages: **On**
FAIMS CV (V): **-70**
RF Lens (%): **40**
AGC Target: **Custom**
Normalized AGC Target (%): **300**
Maximum Injection Time Mode: **Auto**
Microscans: **1**
Data Type: **Profile**
Polarity: **Positive**
Source Fragmentation: **Disabled**
Scan Description: **-70**
Filters:
MIPS
Monoisotopic peak determination: **Peptide**
Relax restrictions when too few precursors are found: **True**
Intensity
Filter Type: **Intensity Threshold**
Intensity Threshold: **5.0e3**
Charge State
Include charge state(s): **2-8**
Include undetermined charge states: **False**
Dynamic Exclusion
Dynamic Exclusion Mode: **Custom**
Exclude after n times: **1**
Exclusion duration (s): **20**
Mass Tolerance: **ppm**
Low: **10**
High: **10**
Exclude isotopes: **True**
Perform dependent scan on single charge state per precursor only: **False**

Data Dependent
Data Dependent Mode: **Number of Scans**
Number of Dependent Scans: **28**
Scan Event Type 1:

Scan:
ddMS²
Multiplex Ions: **False**
Isolation Window (m/z): **1.6**
Isolation Offset: **Off**
Collision Energy Mode: **Fixed**
Collision Energy Type: **Normalized**
HCD Collision Energy (%): **30**
Orbitrap Resolution: **15000**
TurboTMT: **Off**
Scan Range Mode: **Define First Mass**
First Mass (m/z): **120**
AGC Target: **Custom**
Normalized AGC Target (%): **75**
Maximum Injection Time Mode: **Auto**
Microscans: **1**
Data Type: **Centroid**
Scan Description:

Data Analysis Method Summary

Analysis Software

Thermo Proteome Discoverer 2.4.0.305

Processing Workflow

Workflow based on template: PWF_QE_Basic_SequestHT_Alex
Creation date: 3/18/2020 12:53:12 PM
Created with Discoverer version: 2.4.0.305

```
----------------------------------------------------------------
The workflow tree:
----------------------------------------------------------------

 |-(0) Spectrum Files RC
   |-(1) Spectrum Selector
     |-(2) Precursor Detector
       |-(3) Sequest HT
         |-(4) Percolator

           |-(5) Spectrum Confidence Filter
   |-(6) Minora Feature Detector

----------------------------------------------------------------
```

Appendix 3

```
Processing node 0: Spectrum Files RC
---------------------------------------------------------------
1. Search Settings:
- Protein Database:  Homo sapiens (SwissProt TaxID=9606_and_subtaxonomies) (v2017-10-25)
- Enzyme Name:  Trypsin (Full)
- Precursor Mass Tolerance:  20 ppm
- Fragment Mass Tolerance:  0.5 Da

2. Regression Settings:
- Regression Model:  Non-linear Regression
- Parameter Tuning:  Coarse

---------------------------------------------------------------
Processing node 1: Spectrum Selector
---------------------------------------------------------------
1. General Settings:
- Precursor Selection:  Use MS1 Precursor
- Use Isotope Pattern in Precursor Re-evaluation:  True
- Provide Profile Spectra:  Automatic

2. Spectrum Properties Filter:
- Lower RT Limit:  0
- Upper RT Limit:  0
- First Scan:  0
- Last Scan:  0
- Lowest Charge State:  0
- Highest Charge State:  0
- Min. Precursor Mass:  350 Da
- Max. Precursor Mass:  5000 Da
- Total Intensity Threshold:  0
- Minimum Peak Count:  1

3. Scan Event Filters:
- Mass Analyzer:  Is FTMS
- MS Order:  Is Not MS1
- Activation Type:  Is HCD
- Min. Collision Energy:  0
- Max. Collision Energy:  1000
- Scan Type:  Is Full

4. Peak Filters:
- S/N Threshold (FT-only):  1.5

5. Replacements for Unrecognized Properties:
- Unrecognized Charge Replacements:  Automatic
- Unrecognized Mass Analyzer Replacements:  ITMS
- Unrecognized MS Order Replacements:  MS2
```

- Unrecognized Activation Type Replacements: CID
- Unrecognized Polarity Replacements: +
- Unrecognized MS Resolution@200 Replacements: 60000
- Unrecognized MSn Resolution@200 Replacements: 30000

6. Precursor Pattern Extraction:
- Precursor Clipping Range Before: 2.5 Da
- Precursor Clipping Range After: 5.5 Da

Processing node 2: Precursor Detector
----
1. General:
- S/N Threshold: 1.5
- S/N Threshold for most abundant peak: 1.5

Processing node 3: Sequest HT
----
1. Input Data:
- Protein Database: Homo sapiens (SwissProt TaxID=9606_and_subtaxonomies) (v2017-10-25)
- Enzyme Name: Trypsin (Full)
- Max. Missed Cleavage Sites: 2
- Min. Peptide Length: 6
- Max. Peptide Length: 150
- Max. Number of Peptides Reported: 10

2. Tolerances:
- Precursor Mass Tolerance: 10 ppm
- Fragment Mass Tolerance: 0.02 Da
- Use Average Precursor Mass: False
- Use Average Fragment Mass: False

3. Spectrum Matching:
- Use Neutral Loss a Ions: True
- Use Neutral Loss b Ions: True
- Use Neutral Loss y Ions: True
- Use Flanking Ions: True
- Weight of a Ions: 0

- Weight of b Ions: 1
- Weight of c Ions: 0
- Weight of x Ions: 0
- Weight of y Ions: 1
- Weight of z Ions: 0

```
4. Dynamic Modifications:
- Max. Equal Modifications Per Peptide:  3
- Max. Dynamic Modifications Per Peptide:  4
- 1. Dynamic Modification:  Oxidation / +15.995 Da (M)
- 2. Dynamic Modification:  Deamidated / +0.984 Da (N, Q)

5. Dynamic Modifications (peptide terminus):
- 1. N-Terminal Modification:  Gln->pyro-Glu / -17.027 Da (Q)

7. Static Modifications:
- 1. Static Modification:  Carbamidomethyl / +57.021 Da (C)

----------------------------------------------------------------
Processing node 4: Percolator
----------------------------------------------------------------
1. Target/Decoy Strategy:
- Target/Decoy Selection:  Concatenated
- Validation based on:  q-Value

2. Input Data:
- Maximum Delta Cn:  0.05
- Maximum Rank:  0

3. FDR Targets:
- Target FDR (Strict):  0.01
- Target FDR (Relaxed):  0.05

----------------------------------------------------------------
Processing node 5: Spectrum Confidence Filter
----------------------------------------------------------------
Filter Settings:
- Spectrum Confidence:  Worse Than High

----------------------------------------------------------------
Processing node 6: Minora Feature Detector
----------------------------------------------------------------
1. Peak & Feature Detection:
- Min. Trace Length:  5
- Max. ΔRT of Isotope Pattern Multiplets [min]:  0.2

2. Feature to ID Linking:
- PSM Confidence At Least:  High
```

Consensus Workflow Settings

Description: Result filtered for high confident peptides.
Workflow based on template: CWF_Basic_withCharts_Alex
Creation date: 3/18/2020 12:53:14 PM

```
Created with Discoverer version: 2.4.0.305


--------------------------------------------------------------
The workflow tree:
--------------------------------------------------------------

  |-(0) MSF Files
    |-(1) PSM Grouper
      |-(2) Peptide Validator
        |-(3) Peptide and Protein Filter
          |-(4) Protein Scorer
            |-(5) Protein Grouping
              |-(10) Peptide in Protein Annotation
            |-(6) Protein FDR Validator
          |-(8) Protein Annotation
          |-(9) Protein Marker
    |-(12) Feature Mapper
      |-(13) Precursor Ions Quantifier

Post-processing nodes:
------------------------------

  |-(7) Display Settings
  |-(11) Data Distributions

--------------------------------------------------------------
Processing node 0: MSF Files
--------------------------------------------------------------
1. Storage Settings:
- Spectra to Store:  Identified or Quantified
- Feature Traces to Store:  All

2. Merging of Identified Peptide and Proteins:
- Merge Mode:  Globally by Search Engine Type

3. FASTA Title Line Display:
- Reported FASTA Title Lines:  Best match
- Title Line Rule:  standard

4. PSM Filters:
- Maximum Delta Cn:  0.05
- Maximum Rank:  0
- Maximum Delta Mass:  0 ppm

--------------------------------------------------------------
Processing node 1: PSM Grouper
--------------------------------------------------------------
```

```
1. Peptide Group Modifications:
- Site Probability Threshold:  75

---------------------------------------------------------------
Processing node 2: Peptide Validator
---------------------------------------------------------------
1. General Validation Settings:
- Validation Mode:  Automatic (Control peptide level error rate if possible)
- Target FDR (Strict) for PSMs:  0.01
- Target FDR (Relaxed) for PSMs:  0.05
- Target FDR (Strict) for Peptides:  0.01
- Target FDR (Relaxed) for Peptides:  0.05

2. Specific Validation Settings:
- Validation Based on:  q-Value
- Target/Decoy Selection for PSM Level FDR Calculation Based on Score:
Automatic
- Reset Confidences for Nodes without Decoy Search (Fixed score thresholds):
False

---------------------------------------------------------------
Processing node 3: Peptide and Protein Filter
---------------------------------------------------------------
1. Peptide Filters:
- Peptide Confidence At Least:  High
- Keep Lower Confident PSMs:  False
- Minimum Peptide Length:  6
- Remove Peptides Without Protein Reference:  False

2. Protein Filters:
- Minimum Number of Peptide Sequences:  1
- Count Only Rank 1 Peptides:  False
- Count Peptides Only for Top Scored Protein:  False

---------------------------------------------------------------
Processing node 4: Protein Scorer
---------------------------------------------------------------
No parameters

---------------------------------------------------------------
Processing node 5: Protein Grouping
---------------------------------------------------------------
1. Protein Grouping:
- Apply strict parsimony principle:  True

---------------------------------------------------------------
Processing node 10: Peptide in Protein Annotation
---------------------------------------------------------------
```

```
1. Flanking Residues:
- Annotate Flanking Residues of the Peptide:  True
- Number Flanking Residues in Connection Tables:  1

2. Modifications in Peptide:
- Protein Modifications Reported:  Only for Master Proteins

3. Modifications in Protein:
- Modification Sites Reported:  All And Specific
- Minimum PSM Confidence:  High

- Report Only PTMs:  True

4. Positions in Protein:
- Protein Positions for Peptides:  Only for Master Proteins

----------------------------------------------------------------
Processing node 6: Protein FDR Validator
----------------------------------------------------------------
1. Confidence Thresholds:
- Target FDR (Strict):  0.01
- Target FDR (Relaxed):  0.05

----------------------------------------------------------------
Processing node 8: Protein Annotation
----------------------------------------------------------------
1. Annotation Aspects:
- 1. Aspect:  Biological Process
- 2. Aspect:  Cellular Component
- 3. Aspect:  Molecular Function
- 4. Aspect:  None
- 5. Aspect:  None
- 6. Aspect:  None

----------------------------------------------------------------
Processing node 9: Protein Marker
----------------------------------------------------------------
5. Annotate Species:
- As Species Map:  False
- As Species Names:  False


6. Mark Additional Entities:
- Annotation Groups:  False
- Pathway Groups:  False
- Modification Sites:  True
- Peptide Isoform Groups:  True
```

```
-----------------------------------------------------------------
Processing node 12: Feature Mapper
-----------------------------------------------------------------
1. Chromatographic Alignment:
- Perform RT Alignment:  True
- Maximum RT Shift [min]:  10
- Mass Tolerance:  10 ppm
- Parameter Tuning:  Coarse

2. Feature Linking and Mapping:
- RT Tolerance [min]:  0
- Mass Tolerance:  0 ppm
- Min. S/N Threshold:  5

-----------------------------------------------------------------
Processing node 13: Precursor Ions Quantifier
-----------------------------------------------------------------
1. General Quantification Settings:
- Peptides to Use:  Unique + Razor
- Consider Protein Groups for Peptide Uniqueness:  True
- Use Shared Quan Results:  True
- Reject Quan Results with Missing Channels:  False

2. Precursor Quantification:
- Precursor Abundance Based On:  Intensity
- Min. # Replicate Features [%]:  0

3. Normalization and Scaling:
- Normalization Mode:  None
- Scaling Mode:  None

4. Exclude Peptides from Protein Quantification:
- For Normalization:  Use All Peptides
- For Protein Roll-Up:  Use All Peptides
- For Pairwise Ratios:  Exclude Modified

5. Quan Rollup and Hypothesis Testing:
- Protein Abundance Calculation:  Summed Abundances
- N for Top N:  3
- Protein Ratio Calculation:  Pairwise Ratio Based
- Maximum Allowed Fold Change:  100
- Imputation Mode:  None
- Hypothesis Test:  t-test (Background Based)

6. Quan Ratio Distributions:
- 1st Fold Change Threshold:  2
- 2nd Fold Change Threshold:  4
- 3rd Fold Change Threshold:  6
- 4th Fold Change Threshold:  8
- 5th Fold Change Threshold:  10

-----------------------------------------------------------------
Processing node 7: Display Settings
-----------------------------------------------------------------
1. General:
- Filter Set:
        ###   Filter Set MasterProteinFilter.filterset contains the following filters:
        ###   Row Filter for TargetProtein:
        ###   Master is equal to Master

-----------------------------------------------------------------
Processing node 11: Data Distributions
-----------------------------------------------------------------
1. ID Distributions (Bottom-up):
- Peptides to Use:  Only unique peptides based on protein groups
```

Appendix 4

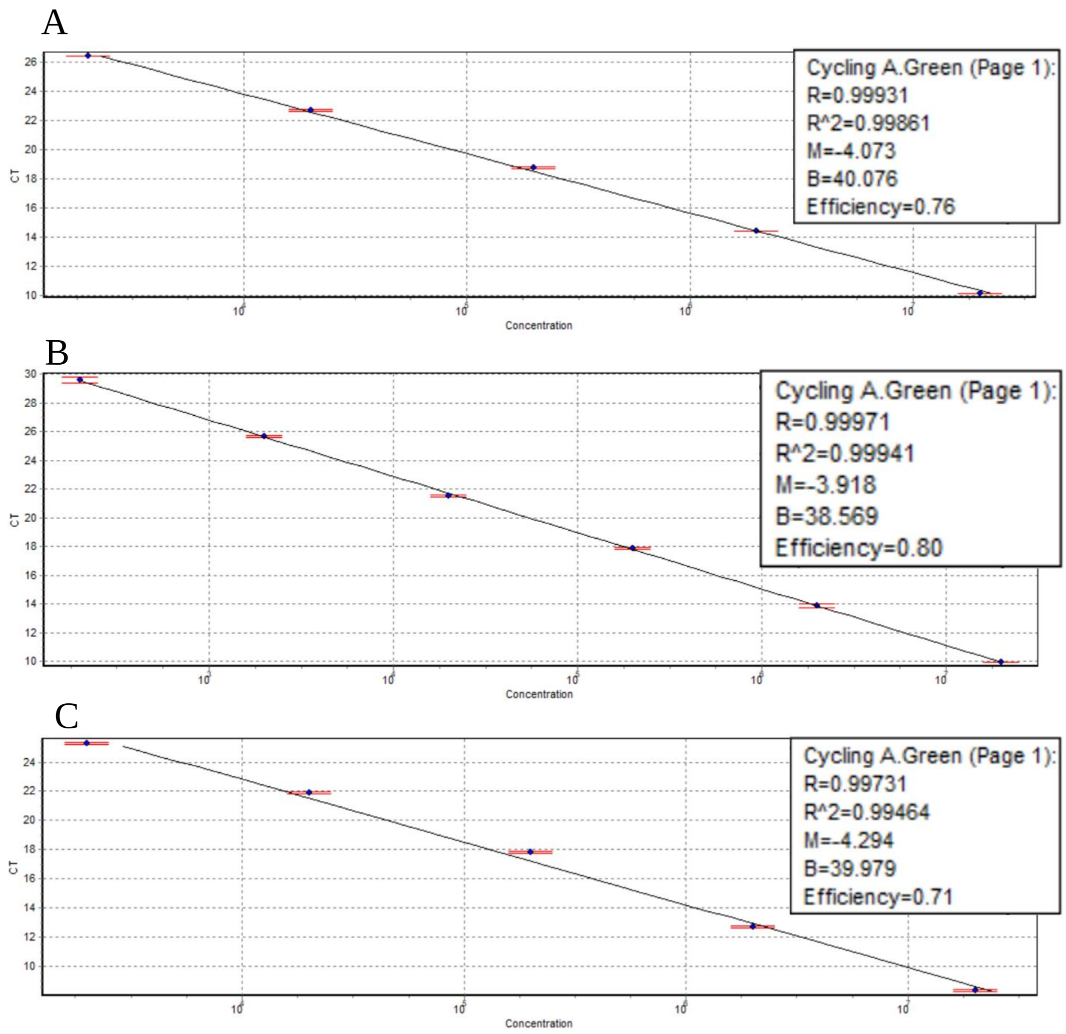


Figure A- 2 ***CircACVR2A* SYBR Green qRT-PCR standard curves**

CircACVR2A 2X SYBR Green qFT-PCR standard curves at (A) 54°C (B) 56°C (C) 58°C. PCR amplicon was used as DNA template for a 10-fold serial dilution up to 6 times. Concentration is copies per qRT-PCR reaction with 3 technical reps. Denaturation at 95°C for 15 seconds, annealing at above temperatures for 15 seconds, extension at 72°C for 15 seconds. Error bars are standard error of mean.

Appendix 5

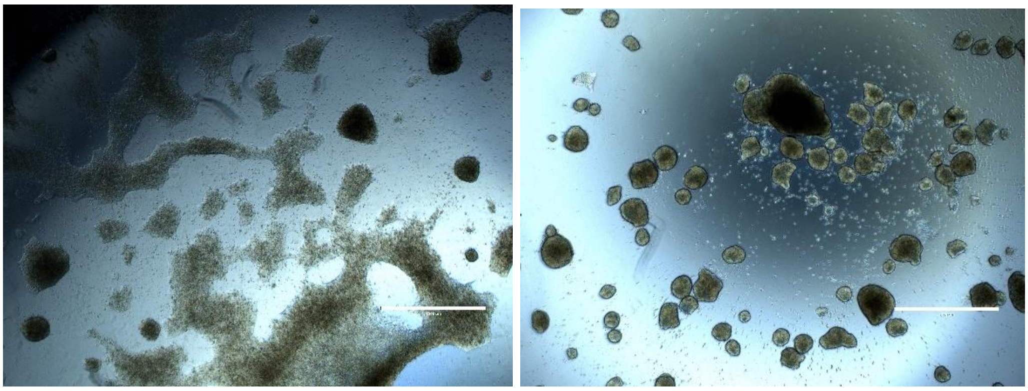


Figure A- 3 H1 *circACVR2A* shRNA knockdown induces EB formation

Appendix 6

Table A 2 Calculations for proportion of *circACVR2A* and *U1* transcripts in CCNH RIP compared to RIP input.

Delta Ct of *circACVR2A* and *U1* from CCNH FLAG IP and IgG IP were transformed into a percentage of the RNA levels from the input sample.

circACVR2A

	IP Ct reps	mean input	Ct diff	Delta Ct	Factor	final calc	mean	SD	Percent of input	Mean	SD
rep 1	33.81	26.44	7.37	132.16	5.00	660.80	581.84	96.39	0.15%	0.18%	0.03%
rep 2	33.31	26.44	6.87	94.89	5.00	474.43			0.21%		
rep 3	33.69	26.44	7.25	122.06	5.00	610.28			0.16%		

U1 snRNA

	IP Ct reps	mean input	Ct diff	Delta Ct	Factor	final calc	mean	SD	Percent of input	Mean	SD
rep 1	23.76	19.31	4.45	21.86	5.00	109.28	108.90	14.24	0.92%	0.93%	0.12%
rep 2	23.93	19.31	4.62	24.59	5.00	122.95			0.81%		
rep 3	23.55	19.31	4.24	18.90	5.00	94.48			1.06%		

References

Abdelmohsen, K. *et al.* (2017) ‘Identification of HuR target circular RNAs uncovers suppression of PABPN1 translation by CircPABPN1’, *RNA biology,* 14(3), pp. 361–369. doi:10.1080/15476286.2017.1279788.

Abreu, R. de S. *et al.* (2009) ‘Global signatures of protein and mRNA expression levels’, *Molecular BioSystems,* 5(12), pp. 1512–1526. doi:10.1039/B908315D.

Abudayyeh, O.O. *et al.* (2016) ‘C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector’, *Science*, 353(6299), p. aaf5573. doi:10.1126/science.aaf5573.

Abudayyeh, O.O. *et al.* (2017) ‘RNA targeting with CRISPR–Cas13’, *Nature*, 550(7675), pp. 280–284. doi:10.1038/nature24049.

Aceto, N. *et al.* (2014) ‘Circulating Tumor Cell Clusters Are Oligoclonal Precursors of Breast Cancer Metastasis’, *Cell*, 158(5), pp. 1110–1122. doi:10.1016/j.cell.2014.07.013.

Acloque, H. *et al.* (2009) ‘Epithelial-mesenchymal transitions: the importance of changing cell state in development and disease’, *The Journal of Clinical Investigation,* 119(6), pp. 1438–1449. doi:10.1172/JCI38019.

Addgene: Protocol - pLKO.1 – TRC Cloning Vector (no date). Available at: https://www.addgene.org/protocols/plko/ (Accessed: 7 April 2020).

Aguilar-Fuentes, J. *et al.* (2006) ‘TFIIH trafficking and its nuclear assembly during early Drosophila embryo development’, *Journal of Cell Science*, 119(18), pp. 3866–3875. doi:10.1242/jcs.03150.

Ahmed, I. *et al.* (2016) ‘Altered expression pattern of circular RNAs in primary and metastatic sites of epithelial ovarian carcinoma’, *Oncotarget*, 7(24), pp. 36366–36381. doi:10.18632/oncotarget.8917.

Aiello, N.M. *et al.* (2017) ‘Upholding a role for EMT in pancreatic cancer metastasis’, *Nature*, 547(7661), pp. E7–E8. doi:10.1038/nature22963.

Aiello, N.M. *et al.* (2018) ‘EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration’, *Developmental Cell*, 45(6), pp. 681-695.e4. doi:10.1016/j.devcel.2018.05.027.

Aktaş, T. *et al.* (2017) ‘DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome’, *Nature*, 544(7648), pp. 115–119. doi:10.1038/nature21715.

Amatori, S., Persico, G. and Fanelli, M. (2017) ‘Real-time quantitative PCR array to study drug-induced changes of gene expression in tumor cell lines’, *Journal of Cancer Metastasis and Treatment*, 3, pp. 90–99. doi:10.20517/2394-4722.2017.22.

Amita, M. *et al.* (2013) ‘Complete and unidirectional conversion of human embryonic stem cells to trophoblast by BMP4’, *Proceedings of the National Academy of Sciences*, 110(13), pp. E1212–E1221. doi:10.1073/pnas.1303094110.

Antibodies, Proteins, Kits and Reagents for Life Science | Abcam (no date). Available at: https://www.abcam.com/ (Accessed: 19 October 2021).

Anton, R., Kestler, H.A. and Kühl, M. (2007) ‘Beta-catenin signaling contributes to stemness and regulates early differentiation in murine embryonic stem cells’, *FEBS letters*, 581(27), pp. 5247–5254. doi:10.1016/j.febslet.2007.10.012.

Arellano, M. and Moreno, S. (1997) ‘Regulation of CDK/cyclin complexes during the cell cycle’, *The International Journal of Biochemistry & Cell Biology*, 29(4), pp. 559–573. doi:10.1016/S1357-2725(96)00178-1.

Arnoux, V. *et al.* (2008) ‘Erk5 Controls Slug Expression and Keratinocyte Activation during Wound Healing’, *Molecular Biology of the Cell*, 19(11), pp. 4738–4749. doi:10.1091/mbc.e07-10-1078.

Ashwal-Fluss, R. *et al.* (2014) ‘circRNA Biogenesis Competes with Pre-mRNA Splicing’, *Molecular Cell*, 56(1), pp. 55–66. doi:10.1016/j.molcel.2014.08.019.

Athanasiadis, A., Rich, A. and Maas, S. (2004) ‘Widespread A-to-I RNA Editing of Alu-Containing mRNAs in the Human Transcriptome’, *PLOS Biology*, 2(12), p. e391. doi:10.1371/journal.pbio.0020391.

Attia, W.A. *et al.* (2014) ‘Evidence for self-maintaining pluripotent murine stem cells in embryoid bodies’, *Stem Cell Reviews and Reports*, 10(1), pp. 1–15. doi:10.1007/s12015-013-9472-1.

Aviner, R. *et al.* (2015) ‘Uncovering Hidden Layers of Cell Cycle Regulation through Integrative Multi-omic Analysis’, *PLOS Genetics*, 11(10), p. e1005554. doi:10.1371/journal.pgen.1005554.

Barrangou, R. *et al.* (2007) ‘CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes’, *Science*, 315(5819), pp. 1709–1712. doi:10.1126/science.1138140.

Begum, S. *et al.* (2018) ‘Novel tumour suppressive protein encoded by circular RNA, circ-SHPRH, in glioblastomas.’, *Oncogene* [Preprint]. doi:10.1038/s41388-018-0230-3.

Bertero, A. *et al.* (2015) ‘Activin/Nodal signaling and NANOG orchestrate human embryonic stem cell fate decisions by controlling the H3K4me3 chromatin mark’, *Genes & Development*, 29(7), pp. 702–717. doi:10.1101/gad.255984.114.

Beswick, R.W., Ambrose, H.E. and Wagner, S.D. (2006) ‘Nocodazole, a microtubule depolymerising agent, induces apoptosis of chronic lymphocytic leukaemia cells associated with changes in Bcl-2 phosphorylation and expression’, *Leukemia Research*, 30(4), pp. 427–436. doi:10.1016/j.leukres.2005.08.009.

Bjørlykke, Y. *et al.* (2019) *Reprogrammed Cells Display Distinct Proteomic Signatures Associated with Colony Morphology Variability*, *Stem Cells International*. Hindawi. doi:https://doi.org/10.1155/2019/8036035.

Black, D.L. (2003) ‘Mechanisms of Alternative Pre-Messenger RNA Splicing’, *Annual Review of Biochemistry*, 72(1), pp. 291–336. doi:10.1146/annurev.biochem.72.121801.161720.

Blajeski, A.L. *et al.* (2002) ‘G1 and G2 cell-cycle arrest following microtubule depolymerization in human breast cancer cells’, *The Journal of Clinical Investigation*, 110(1), pp. 91–99. doi:10.1172/JCI13275.

Blanco, M.J. *et al.* (2007) ‘Snail1a and Snail1b cooperate in the anterior migration of the axial mesendoderm in the zebrafish embryo’, *Development*, 134(22), pp. 4073–4081. doi:10.1242/dev.006858.

Bone, H.K. *et al.* (2011) ‘A novel chemically directed route for the generation of definitive endoderm from human embryonic stem cells based on inhibition of GSK-3’, *Journal of Cell Science*, 124(12), pp. 1992–2000. doi:10.1242/jcs.081679.

Bonneton, C., Sibarita, J.B. and Thiery, J.P. (1999) ‘Relationship between cell migration and cell cycle during the initiation of epithelial to fibroblastoid transition’, *Cell Motility and the Cytoskeleton*, 43(4), pp. 288–295. doi:10.1002/(SICI)1097-0169(1999)43:4<288::AID-CM2>3.0.CO;2-Y.

Bonomi, S. *et al.* (2013) ‘HnRNP A1 controls a splicing regulatory circuit promoting mesenchymal-to-epithelial transition’, *Nucleic Acids Research*, 41(18), pp. 8665–8679. doi:10.1093/nar/gkt579.

Boyd, N.L. *et al.* (2009) ‘Human Embryonic Stem Cell–Derived Mesoderm-like Epithelium Transitions to Mesenchymal Progenitor Cells’, *Tissue Engineering. Part A*, 15(8), pp. 1897–1907. doi:10.1089/ten.tea.2008.0351.

Brabletz, S. and Brabletz, T. (2010) ‘The ZEB/miR-200 feedback loop—a motor of cellular plasticity in development and cancer?’, *EMBO reports*, 11(9), pp. 670–677. doi:10.1038/embor.2010.117.

Brabletz, T. *et al.* (2001) ‘Variable β-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment’, *Proceedings of the National Academy of Sciences*, 98(18), pp. 10356–10361. doi:10.1073/pnas.171610498.

Brabletz, T. (2012) ‘To differentiate or not — routes towards metastasis’, *Nature Reviews Cancer*, 12(6), pp. 425–436. doi:10.1038/nrc3265.

Braeutigam, C. *et al.* (2014) 'The RNA-binding protein Rbfox2: an essential regulator of EMT-driven alternative splicing and a mediator of cellular invasion', *Oncogene*, 33(9), pp. 1082–1092. doi:10.1038/onc.2013.50.

Bronsert, P. *et al.* (2014) 'Cancer cell invasion and EMT marker expression: a three-dimensional study of the human cancer–host interface', *The Journal of Pathology*, 234(3), pp. 410–422. doi:10.1002/path.4416.

Brouns, S.J.J. *et al.* (2008) 'Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes', *Science*, 321(5891), pp. 960–964. doi:10.1126/science.1159689.

Brummelkamp, T.R., Bernards, R. and Agami, R. (2002) 'A System for Stable Expression of Short Interfering RNAs in Mammalian Cells', *Science*, 296(5567), pp. 550–553. doi:10.1126/science.1068999.

Cahan, P. *et al.* (2014) 'CellNet: Network Biology Applied to Stem Cell Engineering', *Cell*, 158(4), pp. 903–915. doi:10.1016/j.cell.2014.07.020.

Cai, Z. *et al.* (2021) 'VirusCircBase: a database of virus circular RNAs', *Briefings in Bioinformatics*, 22(2), pp. 2182–2190. doi:10.1093/bib/bbaa052.

Caixeiro, N.J. *et al.* (2014) 'Circulating tumour cells—a bona fide cause of metastatic cancer', *Cancer and Metastasis Reviews*, 33(2), pp. 747–756. doi:10.1007/s10555-014-9502-8.

Chambers, I. *et al.* (2007) 'Nanog safeguards pluripotency and mediates germline development', *Nature*, 450(7173), pp. 1230–1234. doi:10.1038/nature06403.

Chambers, S.M. *et al.* (2009) 'Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling', *Nature Biotechnology*, 27(3), pp. 275–280. doi:10.1038/nbt.1529.

Chan, D.N. *et al.* (2012) 'Ptk7 Marks the First Human Developmental EMT In Vitro', *PLoS ONE*. Edited by A.B. Pant, 7(11), p. e50432. doi:10.1371/journal.pone.0050432.

Chaudhary, R. *et al.* (no date) 'A Circular RNA from the MDM2 Locus Controls Cell Cycle Progression by Suppressing p53 Levels', *Molecular and Cellular Biology*, 40(9), pp. e00473-19. doi:10.1128/MCB.00473-19.

Chen, K.G. *et al.* (2012) 'Non-colony type monolayer culture of human embryonic stem cells', *Stem Cell Research*, 9(3), pp. 237–248. doi:10.1016/j.scr.2012.06.003.

Chen, N. *et al.* (2018) 'A novel FLI1 exonic circular RNA promotes metastasis in breast cancer by coordinately regulating TET1 and DNMT1', *Genome Biology*, 19(1), p. 218. doi:10.1186/s13059-018-1594-y.

Chen, P. *et al.* (2020) 'Circular RNA circCTNNA1 promotes colorectal cancer progression by sponging miR-149-5p and regulating FOXM1 expression', *Cell Death & Disease*, 11(7), pp. 1–12. doi:10.1038/s41419-020-02757-7.

Chen, Q. *et al.* (2018) ‘hsa_circ_0061140 Knockdown Reverses FOXM1-Mediated Cell Growth and Metastasis in Ovarian Cancer through miR-370 Sponge Activity’, *Molecular Therapy. Nucleic Acids,* 13, pp. 55–63. doi:10.1016/j.omtn.2018.08.010.

Chen, R.-X. *et al.* (2019) ‘N6-methyladenosine modification of circNSUN2 facilitates cytoplasmic export and stabilizes HMGA2 to promote colorectal liver metastasis’, *Nature Communications,* 10(1), p. 4695. doi:10.1038/s41467-019-12651-2.

Chen, S. *et al.* (2019) ‘Widespread and Functional RNA Circularization in Localized Prostate Cancer’, *Cell,* 176(4), pp. 831-843.e22. doi:10.1016/j.cell.2019.01.025.

Chen, W. *et al.* (2020) ‘Overexpression of circRNA circUCK2 Attenuates Cell Apoptosis in Cerebral Ischemia-Reperfusion Injury via miR-125b-5p/GDF11 Signaling’, *Molecular Therapy - Nucleic Acids,* 22, pp. 673–683. doi:10.1016/j.omtn.2020.09.032.

Chen, X. *et al.* (2018) ‘PRMT5 Circular RNA Promotes Metastasis of Urothelial Carcinoma of the Bladder through Sponging miR-30c to Induce Epithelial-Mesenchymal Transition’, *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research,* 24(24), pp. 6319–6330. doi:10.1158/1078-0432.CCR-18-1270.

Chen, Y. and Wang, X. (2020) ‘miRDB: an online database for prediction of functional microRNA targets’, *Nucleic Acids Research,* 48(D1), pp. D127–D131. doi:10.1093/nar/gkz757.

Chen, Y.G. *et al.* (2017) ‘Sensing Self and Foreign Circular RNAs by Intron Identity’, *Molecular Cell,* 67(2), pp. 228-238.e5. doi:10.1016/j.molcel.2017.05.022.

Chu, F.-Y., Haley, S.C. and Zidovska, A. (2017) ‘On the origin of shape fluctuations of the cell nucleus’, *Proceedings of the National Academy of Sciences,* 114(39), pp. 10338–10343. doi:10.1073/pnas.1702226114.

Chung, V.Y. *et al.* (2016) ‘GRHL2-miR-200-ZEB1 maintains the epithelial status of ovarian cancer through transcriptional regulation and histone modification’, *Scientific Reports,* 6(1), p. 19943. doi:10.1038/srep19943.

Cieply, B. *et al.* (2012) ‘Suppression of the Epithelial–Mesenchymal Transition by Grainyhead-like-2’, *Cancer Research,* 72(9), pp. 2440–2453. doi:10.1158/0008-5472.CAN-11-4038.

Cocquerelle, C. *et al.* (1993) ‘Mis-splicing yields circular RNA molecules’, *FASEB journal: official publication of the Federation of American Societies for Experimental Biology,* 7(1), pp. 155–160.

Cohen, S.J. *et al.* (2009) ‘Prognostic significance of circulating tumor cells in patients with metastatic colorectal cancer’, *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, 20(7), pp. 1223–1229. doi:10.1093/annonc/mdn786.

Colunga, T. *et al.* (2019) ‘Human Pluripotent Stem Cell-Derived Multipotent Vascular Progenitors of the Mesothelium Lineage Have Utility in Tissue Engineering and Repair’, *Cell reports*, 26(10), pp. 2566-2579.e10. doi:10.1016/j.celrep.2019.02.016.

Combes, A.N. *et al.* (2019) ‘Single-cell analysis reveals congruence between kidney organoids and human fetal kidney’, *Genome Medicine*, 11, p. 3. doi:10.1186/s13073-019-0615-0.

Compe, E. *et al.* (2019) ‘TFIIE orchestrates the recruitment of the TFIIH kinase module at promoter before release during transcription’, *Nature Communications*, 10(1), p. 2084. doi:10.1038/s41467-019-10131-1.

Conklin, J.F., Baker, J. and Sage, J. (2012) ‘The RB family is required for the self-renewal and survival of human embryonic stem cells’, *Nature Communications*, 3, p. 1244. doi:10.1038/ncomms2254.

Conn, S.J. *et al.* (2015) ‘The RNA Binding Protein Quaking Regulates Formation of circRNAs’, *Cell*, 160(6), pp. 1125–1134. doi:10.1016/j.cell.2015.02.014.

Conn, V.M. *et al.* (2017) ‘A circRNA from *SEPALLATA3* regulates splicing of its cognate mRNA through R-loop formation’, *Nature Plants*, 3(5), p. 17053. doi:10.1038/nplants.2017.53.

Cooper, D.A., Cortés-López, M. and Miura, P. (2018) ‘Genome-Wide circRNA Profiling from RNA-seq Data’, in Dieterich, C. and Papantonis, A. (eds) *Circular RNAs: Methods and Protocols*. New York, NY: Springer (Methods in Molecular Biology), pp. 27–41. doi:10.1007/978-1-4939-7562-4_3.

Cooper, S. (2003) ‘Reappraisal of serum starvation, the restriction point, G0, and G1 phase arrest points’, *The FASEB Journal*, 17(3), pp. 333–340. doi:10.1096/fj.02-0352rev.

Crooke, S.T. (2017) ‘Molecular Mechanisms of Antisense Oligonucleotides’, *Nucleic Acid Therapeutics*, 27(2), pp. 70–77. doi:10.1089/nat.2016.0656.

D’Ambra, E., Capauto, D. and Morlando, M. (2019) ‘Exploring the Regulatory Role of Circular RNAs in Neurodegenerative Disorders’, *International Journal of Molecular Sciences*, 20(21). doi:10.3390/ijms20215477.

D’Amour, K.A. *et al.* (2005) ‘Efficient differentiation of human embryonic stem cells to definitive endoderm’, *Nature Biotechnology*, 23(12), pp. 1534–1541. doi:10.1038/nbt1163.

Danan, M. *et al.* (2012) ‘Transcriptome-wide discovery of circular RNAs in Archaea’, *Nucleic Acids Research*, 40(7), pp. 3131–3142. doi:10.1093/nar/gkr1009.

Dang, R.-Y., Liu, F.-L. and Li, Y. (2017) ‘Circular RNA hsa_circ_0010729 regulates vascular endothelial cell proliferation and apoptosis by targeting the miR-186/HIF-1α axis’, *Biochemical and Biophysical Research Communications*, 490(2), pp. 104–110. doi:10.1016/j.bbrc.2017.05.164.

Davidson, E.H. and Erwin, D.H. (2006) ‘Gene regulatory networks and the evolution of animal body plans’, *Science (New York, N.Y.)*, 311(5762), pp. 796–800. doi:10.1126/science.1113832.

Davidson, K.C. *et al.* (2012) ‘Wnt/β-catenin signaling promotes differentiation, not self-renewal, of human embryonic stem cells and is repressed by Oct4’, *Proceedings of the National Academy of Sciences of the United States of America*, 109(12), pp. 4485–4490. doi:10.1073/pnas.1118777109.

Davis, F.M. *et al.* (2014) ‘Targeting EMT in cancer: opportunities for pharmacological intervention’, *Trends in Pharmacological Sciences*, 35(9), pp. 479–488. doi:10.1016/j.tips.2014.06.006.

De, D. *et al.* (2014) ‘Inhibition of master transcription factors in pluripotent cells induces early stage differentiation’, *Proceedings of the National Academy of Sciences*, 111(5), pp. 1778–1783. doi:10.1073/pnas.1323386111.

Devault, A. *et al.* (1995) ‘MAT1 ('menage à trois’) a new RING finger protein subunit stabilizing cyclin H-cdk7 complexes in starfish and Xenopus CAK.’, *The EMBO Journal*, 14(20), pp. 5027–5036.

Dhara, S.K. and Stice, S.L. (2008) ‘Neural differentiation of human embryonic stem cells’, *Journal of cellular biochemistry*, 105(3), pp. 633–640. doi:10.1002/jcb.21891.

Di Timoteo, G. *et al.* (2020) ‘Modulation of circRNA Metabolism by m6A Modification’, *Cell Reports*, 31(6), p. 107641. doi:10.1016/j.celrep.2020.107641.

Differentiation of Stem Cells | LSR | Bio-Rad (no date). Available at: https://www.bio-rad.com/en-au/applications-technologies/differentiation-stem-cells?ID=LUSR2L8UU (Accessed: 9 April 2020).

Ding, V.M.Y. *et al.* (2010) ‘FGF-2 modulates Wnt signaling in undifferentiated hESC and iPS cells through activated PI3-K/GSK3beta signaling’, *Journal of Cellular Physiology*, 225(2), pp. 417–428. doi:10.1002/jcp.22214.

Dobin, A. *et al.* (2013) ‘STAR: ultrafast universal RNA-seq aligner’, *Bioinformatics*, 29(1), pp. 15–21. doi:10.1093/bioinformatics/bts635.

Domanskyi, A. *et al.* (2014) ‘Transcription factors Foxa1 and Foxa2 are required for adult dopamine neurons maintenance’, *Frontiers in Cellular Neuroscience*, 8, p. 275. doi:10.3389/fncel.2014.00275.

Dong, W. *et al.* (2019) ‘Circular RNA ACVR2A suppresses bladder cancer cells proliferation and metastasis through miR-626/EYA4 axis’, *Molecular Cancer*, 18(1), p. 95. doi:10.1186/s12943-019-1025-z.

Dongre, A. and Weinberg, R.A. (2018) ‘New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer’, *Nature Reviews Molecular Cell Biology*, p. 1. doi:10.1038/s41580-018-0080-4.

Du, W.W. *et al.* (2016) ‘Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2’, *Nucleic Acids Research*, 44(6), pp. 2846–2858. doi:10.1093/nar/gkw027.

Du, William W. *et al.* (2017) ‘Foxo3 circular RNA promotes cardiac senescence by modulating multiple factors associated with stress and senescence responses’, *European Heart Journal*, 38(18), pp. 1402–1412. doi:10.1093/eurheartj/ehw001.

Du, William W *et al.* (2017) ‘Induction of tumor apoptosis through a circular RNA enhancing Foxo3 activity’, *Cell Death and Differentiation*, 24(2), pp. 357–370. doi:10.1038/cdd.2016.133.

DuBridge, R.B. *et al.* (1987) ‘Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system’, *Molecular and Cellular Biology*, 7(1), pp. 379–387. doi:10.1128/mcb.7.1.379-387.1987.

Eastham, A.M. *et al.* (2007) ‘Epithelial-Mesenchymal Transition Events during Human Embryonic Stem Cell Differentiation’, *Cancer Research*, 67(23), pp. 11254–11262. doi:10.1158/0008-5472.CAN-07-2253.

Ebbesen, K.K., Hansen, T.B. and Kjems, J. (2017) ‘Insights into circular RNA biology’, *RNA Biology*, 14(8), pp. 1035–1045. doi:10.1080/15476286.2016.1271524.

Eiselleova, L. *et al.* (2009) ‘A Complex Role for FGF-2 in Self-Renewal, Survival, and Adhesion of Human Embryonic Stem Cells’, *Stem Cells (Dayton, Ohio)*, 27(8), pp. 1847–1857. doi:10.1002/stem.128.

Elkabetz, Y. *et al.* (2008) ‘Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage’, *Genes & Development*, 22(2), pp. 152–165. doi:10.1101/gad.1616208.

Embryonic and Induced Pluripotent Stem Cell Differentiation Pathways & Lineage-specific Markers (no date) *www.rndsystems.com*. Available at: https://www.rndsystems.com/pathways/embryonic-induced-pluripotent-stem-cell-differentiation-pathways-lineage-specific-markers (Accessed: 5 May 2022).

Enkhbaatar, Z. *et al.* (2013) ‘KDM5B histone demethylase controls epithelial-mesenchymal transition of cancer cells by regulating the expression of the microRNA-200 family’, *Cell Cycle*, 12(13), pp. 2100–2112. doi:10.4161/cc.25142.

Enuka, Y. *et al.* (2016) ‘Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor’, *Nucleic Acids Research*, 44(3), pp. 1370–1383. doi:10.1093/nar/gkv1367.

Errichelli, L. *et al.* (2017) ‘FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons’, *Nature Communications*, 8, p. 14741. doi:10.1038/ncomms14741.

Fabian, M.R., Sonenberg, N. and Filipowicz, W. (2010) ‘Regulation of mRNA Translation and Stability by microRNAs’, *Annual Review of Biochemistry*, 79(1), pp. 351–379. doi:10.1146/annurev-biochem-060308-103103.

Fang, L. *et al.* (2019) ‘The circular RNA circ-Ccnb1 dissociates Ccnb1/Cdk1 complex suppressing cell invasion and tumorigenesis’, *Cancer Letters* [Preprint]. doi:10.1016/j.canlet.2019.05.036.

Fareh, M. *et al.* (2021) ‘Reprogrammed CRISPR-Cas13b suppresses SARS-CoV-2 replication and circumvents its mutational escape through mismatch tolerance’, *Nature Communications*, 12(1), p. 4270. doi:10.1038/s41467-021-24577-9.

Fei, T. *et al.* (2017) ‘Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing’, *Proceedings of the National Academy of Sciences*, p. 201617467. doi:10.1073/pnas.1617467114.

Feng, X. *et al.* (2012) ‘Protein Kinase C Mediated Extraembryonic Endoderm Differentiation of Human Embryonic Stem Cells’, *STEM CELLS*, 30(3), pp. 461–470. doi:10.1002/stem.1018.

Fidorra, J. *et al.* (1981) ‘Cellular and nuclear volume of human cells during the cell cycle’, *Radiation and Environmental Biophysics*, 19(3), pp. 205–214. doi:10.1007/BF01324188.

Filipczyk, A.A. *et al.* (2007) ‘Differentiation is coupled to changes in the cell cycle regulatory apparatus of human embryonic stem cells’, *Stem Cell Research*, 1(1), pp. 45–60. doi:10.1016/j.scr.2007.09.002.

Fischer, K.R. *et al.* (2015) ‘Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance’, *Nature*, 527(7579), pp. 472–476. doi:10.1038/nature15748.

Fischer, Y. *et al.* (2010) ‘NANOG Reporter Cell Lines Generated by Gene Targeting in Human Embryonic Stem Cells’, *PLOS ONE*, 5(9), p. e12533. doi:10.1371/journal.pone.0012533.

Francis, K.R. and Wei, L. (2010) ‘Human embryonic stem cell neural differentiation and enhanced cell survival promoted by hypoxic preconditioning’, *Cell Death & Disease*, 1(2), pp. e22–e22. doi:10.1038/cddis.2009.22.

Franz, H. *et al.* (2009) ‘Multimerization and H3K9me3 Binding Are Required for CDYL1b Heterochromatin Association’, *The Journal of Biological Chemistry*, 284(50), pp. 35049–35059. doi:10.1074/jbc.M109.052332.

Gao, D. *et al.* (2017) ‘Screening circular RNA related to chemotherapeutic resistance in breast cancer’, *Epigenomics*, 9(9), pp. 1175–1188. doi:10.2217/epi-2017-0055.

Gao, Z. *et al.* (2017) ‘Mutation of nucleotides around the +1 position of type 3 polymerase III promoters: The effect on transcriptional activity and start site usage’, *Transcription*, 8(5), pp. 275–287. doi:10.1080/21541264.2017.1322170.

Garikipati, V.N.S. *et al.* (2019) ‘Circular RNA CircFndc3b modulates cardiac repair after myocardial infarction via FUS/VEGF-A axis’, *Nature Communications*, 10(1), p. 4317. doi:10.1038/s41467-019-11777-7.

Gasimli, L. *et al.* (2014) ‘Changes in glycosaminoglycan structure on differentiation of human embryonic stem cells towards mesoderm and endoderm lineages’, *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1840(6), pp. 1993–2003. doi:10.1016/j.bbagen.2014.01.007.

Geng, H.-H. *et al.* (2016) ‘The Circular RNA Cdr1as Promotes Myocardial Infarction by Mediating the Regulation of miR-7a on Its Target Genes Expression’, *PLOS ONE*, 11(3), p. e0151753. doi:10.1371/journal.pone.0151753.

Geng, X. *et al.* (2020) ‘Circular RNA: biogenesis, degradation, functions and potential roles in mediating resistance to anticarcinogens’, *Epigenomics*, 12(3), pp. 267–283. doi:10.2217/epi-2019-0295.

Giard, D.J. *et al.* (1973) ‘In Vitro Cultivation of Human Tumors: Establishment of Cell Lines Derived From a Series of Solid Tumors2’, *JNCI: Journal of the National Cancer Institute*, 51(5), pp. 1417–1423. doi:10.1093/jnci/51.5.1417.

Gibbons, D.L. *et al.* (2009) ‘Contextual extracellular cues promote tumor cell EMT and metastasis by regulating miR-200 family expression’, *Genes & Development*, 23(18), pp. 2140–2151. doi:10.1101/gad.1820209.

Giral, H., Landmesser, U. and Kratzer, A. (2018) ‘Into the Wild: GWAS Exploration of Non-coding RNAs’, *Frontiers in Cardiovascular Medicine*, 5. doi:10.3389/fcvm.2018.00181.

Gordeeva, O. (2019) ‘TGFβ Family Signaling Pathways in Pluripotent and Teratocarcinoma Stem Cells’ Fate Decisions: Balancing Between Self-Renewal, Differentiation, and Cancer’, *Cells*, 8(12). doi:10.3390/cells8121500.

Görlach, A. and Holdenrieder, S. (2017) ‘Circular RNA maps paving the road to biomarker development?’, *Journal of Molecular Medicine*, 95(11), pp. 1137–1141. doi:10.1007/s00109-017-1603-8.

Graham, F.L. *et al.* (1977) ‘Characteristics of a human cell line transformed by DNA from human adenovirus type 5’, *The Journal of General Virology*, 36(1), pp. 59–74. doi:10.1099/0022-1317-36-1-59.

Grande, M.T. *et al.* (2015) ‘Snail1-induced partial epithelial-to-mesenchymal transition drives renal fibrosis in mice and can be targeted to reverse established disease’, *Nature Medicine*, 21(9), pp. 989–997. doi:10.1038/nm.3901.

Greenburg, G. (1982) ‘Epithelia suspended in collagen gels can lose polarity and express characteristics of migrating mesenchymal cells’, *The Journal of Cell Biology*, 95(1), pp. 333–339. doi:10.1083/jcb.95.1.333.

Gregory, P.A. *et al.* (2008) ‘The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1’, *Nature Cell Biology*, 10(5), pp. 593–601. doi:10.1038/ncb1722.

Grigore, A.D. *et al.* (2016) ‘Tumor Budding: The Name is EMT. Partial EMT.’, *Journal of Clinical Medicine*, 5(5), p. 51. doi:10.3390/jcm5050051.

Gruner, H. *et al.* (2016) ‘CircRNA accumulation in the aging mouse brain’, *Scientific Reports*, 6(1). doi:10.1038/srep38907.

Guo, J.U. *et al.* (2014) ‘Expanded identification and characterization of mammalian circular RNAs’, *Genome Biology*, 15(7). doi:10.1186/s13059-014-0409-z.

Hamidi, S. *et al.* (2019) ‘Mesenchymal-Epithelial Transition Regulates Initiation of Pluripotency Exit before Gastrulation’, *bioRxiv* [Preprint]. doi:10.1101/655654.

Hansen, T.B. *et al.* (2013) ‘Natural RNA circles function as efficient microRNA sponges’, *Nature*, 495(7441), pp. 384–388. doi:10.1038/nature11993.

Hao, Y., Baker, D. and ten Dijke, P. (2019) ‘TGF-β-Mediated Epithelial-Mesenchymal Transition and Cancer Metastasis’, *International Journal of Molecular Sciences*, 20(11). doi:10.3390/ijms20112767.

Haussecker, D. and Proudfoot, N.J. (2005) ‘Dicer-Dependent Turnover of Intergenic Transcripts from the Human β-Globin Gene Cluster’, *Molecular and Cellular Biology*, 25(21), pp. 9724–9733. doi:10.1128/MCB.25.21.9724-9733.2005.

He, Q. *et al.* (2018) ‘circ-SHKBP1 Regulates the Angiogenesis of U87 Glioma-Exposed Endothelial Cells through miR-544a/FOXP1 and miR-379/FOXP2 Pathways’, *Molecular Therapy - Nucleic Acids*, 10, pp. 331–348. doi:10.1016/j.omtn.2017.12.014.

He, R. *et al.* (2017) ‘OCT4 supports extended LIF-independent self-renewal and maintenance of transcriptional and epigenetic networks in embryonic stem cells’, *Scientific Reports*, 7(1), pp. 1–19. doi:10.1038/s41598-017-16611-y.

Herranz, N. *et al.* (2008) ‘Polycomb Complex 2 Is Required for E-cadherin Repression by the Snail1 Transcription Factor’, *Molecular and Cellular Biology*, 28(15), pp. 4772–4781. doi:10.1128/MCB.00323-08.

Hoagland, M.B., Keller, E.B. and Zamecnik, P.C. (1956) ‘ENZYMATIC CARBOXYL ACTIVATION OF AMINO ACIDS’, *Journal of Biological Chemistry*, 218(1), pp. 345–358.

Hoffman, L.M. and Carpenter, M.K. (2005) ‘Characterization and culture of human embryonic stem cells’, *Nature Biotechnology*, 23(6), pp. 699–708. doi:10.1038/nbt1102.

Holdt, L.M. *et al.* (2016) ‘Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans’, *Nature Communications*, 7, p. 12429. doi:10.1038/ncomms12429.

Holmgren, G. *et al.* (2015) ‘Identification of stable reference genes in differentiating human pluripotent stem cells’, *Physiological Genomics*, 47(6), pp. 232–239. doi:10.1152/physiolgenomics.00130.2014.

Hong, T. *et al.* (2015) ‘An Ovol2-Zeb1 Mutual Inhibitory Circuit Governs Bidirectional and Multi-step Transition between Epithelial and Mesenchymal States’, *PLOS Computational Biology*, 11(11), p. e1004569. doi:10.1371/journal.pcbi.1004569.

Huang, C. *et al.* (2018) ‘A length-dependent evolutionarily conserved pathway controls nuclear export of circular RNAs’, *Genes & Development*, 32(9–10), pp. 639–644. doi:10.1101/gad.314856.118.

Huang, J. *et al.* (2019) ‘Identification of virus-encoded circular RNA’, *Virology*, 529, pp. 144–151. doi:10.1016/j.virol.2019.01.014.

Huang, R.Y.-J. *et al.* (2013) ‘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*, 4(11), pp. e915–e915. doi:10.1038/cddis.2013.442.

Huang, S. and Spector, D.L. (1992) ‘U1 and U2 small nuclear RNAs are present in nuclear speckles’, *Proceedings of the National Academy of Sciences*, 89(1), pp. 305–308. doi:10.1073/pnas.89.1.305.

Huang, Y. and Zhu, Q. (2021) ‘Why circRNAs are Abnormally Expressed in Tumor?—A Mini Review for the circRNA Biogenesis in Cancer’. Available at: https://www.preprints.org/manuscript/202107.0380/v1 (Accessed: 6 September 2021).

Huang, Z.-K. *et al.* (2018) 'Microarray Expression Profile of Circular RNAs in Peripheral Blood Mononuclear Cells from Active Tuberculosis Patients', *Cellular Physiology and Biochemistry*, 45(3), pp. 1230–1240. doi:10.1159/000487454.

Hynds, D.L. and Snow, D.M. (2001) 'Fibronectin and laminin elicit differential behaviors from SH-SY5Y growth cones contacting inhibitory chondroitin sulfate proteoglycans', *Journal of Neuroscience Research*, 66(4), pp. 630–642. doi:10.1002/jnr.10020.

Itskovitz-Eldor, J. *et al.* (2000) 'Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers.', *Molecular Medicine*, 6(2), pp. 88–95.

Ivanov, A. *et al.* (2015) 'Analysis of Intron Sequences Reveals Hallmarks of Circular RNA Biogenesis in Animals', *Cell Reports*, 10(2), pp. 170–177. doi:10.1016/j.celrep.2014.12.019.

Jafari Ghods, F. (2018) 'Circular RNA in Saliva', in Xiao, J. (ed.) *Circular RNAs: Biogenesis and Functions*. Singapore: Springer (Advances in Experimental Medicine and Biology), pp. 131–139. doi:10.1007/978-981-13-1426-1_11.

Jain, A.K. *et al.* (2012) 'p53 Regulates Cell Cycle and MicroRNAs to Promote Differentiation of Human Embryonic Stem Cells', *PLOS Biology*, 10(2), p. e1001268. doi:10.1371/journal.pbio.1001268.

James, D. *et al.* (2005) 'TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells', *Development (Cambridge, England)*, 132(6), pp. 1273–1282. doi:10.1242/dev.01706.

Jang, J. *et al.* (2019) 'Control over single-cell distribution of G1 lengths by WNT governs pluripotency', *PLOS Biology*, 17(9), p. e3000453. doi:10.1371/journal.pbio.3000453.

Jarlstad Olesen, M.T. and S. Kristensen, L. (2021) 'Circular RNAs as microRNA sponges: evidence and controversies', *Essays in Biochemistry* [Preprint], (EBC20200060). doi:10.1042/EBC20200060.

Javaid, S. *et al.* (2013) 'Dynamic Chromatin Modification Sustains Epithelial-Mesenchymal Transition following Inducible Expression of Snail-1', *Cell Reports*, 5(6), pp. 1679–1689. doi:10.1016/j.celrep.2013.11.034.

Jeck, W.R. *et al.* (2013) 'Circular RNAs are abundant, conserved, and associated with ALU repeats', *RNA*, 19(2), pp. 141–157. doi:10.1261/rna.035667.112.

Ji, Y. *et al.* (2016) 'Paeoniflorin suppresses TGF-β mediated epithelial-mesenchymal transition in pulmonary fibrosis through a Smad-dependent pathway', *Acta Pharmacologica Sinica*, 37(6), pp. 794–804. doi:10.1038/aps.2016.36.

Jiang, H. *et al.* (2012) ‘Transformation of Epithelial Ovarian Cancer Stemlike Cells into Mesenchymal Lineage via EMT Results in Cellular Heterogeneity and Supports Tumor Engraftment’, *Molecular Medicine*, 18(1), pp. 1197–1208. doi:10.2119/molmed.2012.00075.

Jie, M. *et al.* (2020) ‘CircMRPS35 suppresses gastric cancer progression via recruiting KAT7 to govern histone modification’, *Molecular Cancer*, 19(1). doi:10.1186/s12943-020-01160-2.

Johnson, D.G. and Walker, C.L. (1999) ‘Cyclins and Cell Cycle Checkpoints’, *Annual Review of Pharmacology and Toxicology*, 39(1), pp. 295–312. doi:10.1146/annurev.pharmtox.39.1.295.

Jolly, M.K., Mani, S.A. and Levine, H. (2018) ‘Hybrid epithelial/mesenchymal phenotype(s): The “fittest” for metastasis?’, *Biochimica Et Biophysica Acta. Reviews on Cancer*, 1870(2), pp. 151–157. doi:10.1016/j.bbcan.2018.07.001.

Kalluri, R. and Weinberg, R.A. (2009) ‘The basics of epithelial-mesenchymal transition’, *The Journal of Clinical Investigation*, 119(6), pp. 1420–1428. doi:10.1172/JCI39104.

Karedath, T. *et al.* (2019) ‘Silencing of ANKRD12 circRNA induces molecular and functional changes associated with invasive phenotypes’, *BMC Cancer*, 19(1), p. 565. doi:10.1186/s12885-019-5723-0.

Kasai, H. *et al.* (2005) ‘TGF-β1 induces human alveolar epithelial to mesenchymal cell transition (EMT)’, *Respiratory Research*, 6(1), p. 56. doi:10.1186/1465-9921-6-56.

Kelava, I. and Lancaster, M.A. (2016) ‘Stem Cell Models of Human Brain Development’, *Cell Stem Cell*, 18(6), pp. 736–748. doi:10.1016/j.stem.2016.05.022.

Khan, M.A.F. *et al.* (2016) ‘RBM20 Regulates Circular RNA Production From the Titin Gene’, *Circulation Research*, 119(9), pp. 996–1003. doi:10.1161/CIRCRESAHA.116.309568.

Khoo, B.L. *et al.* (2015) ‘Short-term expansion of breast circulating cancer cells predicts response to anti-cancer therapy’, *Oncotarget*, 6(17), pp. 15578–15593.

Khoo, M.L.M. *et al.* (2005) ‘Growth and Differentiation of Embryoid Bodies Derived from Human Embryonic Stem Cells: Effect of Glucose and Basic Fibroblast Growth Factor1’, *Biology of Reproduction*, 73(6), pp. 1147–1156. doi:10.1095/biolreprod.104.036673.

Kim, B.N. *et al.* (2020) ‘TGF-β induced EMT and stemness characteristics are associated with epigenetic regulation in lung cancer’, *Scientific Reports*, 10(1), p. 10597. doi:10.1038/s41598-020-67325-7.

Kim, J.H. *et al.* (2007) ‘Transtorming Growth Factor β1 Induces Epithelial-to-Mesenchymal Transition of A549 Cells’, *Journal of Korean Medical Science*, 22(5), pp. 898–904. doi:10.3346/jkms.2007.22.5.898.

Kim, J.M. *et al.* (2011) ‘Assessment of Differentiation Aspects by the Morphological Classification of Embryoid Bodies Derived from Human Embryonic Stem Cells’, *Stem Cells and Development*, 20(11), pp. 1925–1935. doi:10.1089/scd.2010.0476.

KO, L. *et al.* (1998) ‘p53 is phosphorylated by CDK7-cyclin H in a p36(MAT1)-dependent manner’, *Molecular and cellular biology*, 17, pp. 7220–9. doi:10.1128/MCB.17.12.7220.

Kong, D. *et al.* (2010) ‘Epithelial to Mesenchymal Transition Is Mechanistically Linked with Stem Cell Signatures in Prostate Cancer Cells’, *PLOS ONE*, 5(8), p. e12445. doi:10.1371/journal.pone.0012445.

Krebs, M.G. *et al.* (2011) ‘Evaluation and prognostic significance of circulating tumor cells in patients with non-small-cell lung cancer’, *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 29(12), pp. 1556–1563. doi:10.1200/JCO.2010.28.7045.

Krempler, A. *et al.* (2005) ‘Cyclin H is targeted to the nucleus by C-terminal nuclear localization sequences’, *Cellular and Molecular Life Sciences*, 62(12), pp. 1379–1387. doi:10.1007/s00018-005-5023-5.

Kriks, S. *et al.* (2011) ‘Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson’s disease’, *Nature*, 480(7378), pp. 547–551. doi:10.1038/nature10648.

Kukurba, K.R. and Montgomery, S.B. (2015) ‘RNA Sequencing and Analysis’, *Cold Spring Harbor Protocols*, 2015(11), p. pdb.top084970. doi:10.1101/pdb.top084970.

Lamouille, S., Xu, J. and Derynck, R. (2014) ‘Molecular mechanisms of epithelial–mesenchymal transition’, *Nature reviews. Molecular cell biology*, 15(3), pp. 178–196. doi:10.1038/nrm3758.

Langlois, M.-A. *et al.* (2005) ‘Cytoplasmic and Nuclear Retained DMPK mRNAs Are Targets for RNA Interference in Myotonic Dystrophy Cells *’, *Journal of Biological Chemistry*, 280(17), pp. 16949–16954. doi:10.1074/jbc.M501591200.

LaRoche-Johnston, F. *et al.* (2018) ‘Bacterial group II introns generate genetic diversity by circularization and trans-splicing from a population of intron-invaded mRNAs’, *PLoS Genetics*, 14(11). doi:10.1371/journal.pgen.1007792.

Larochelle, S. (2001) ‘T-loop phosphorylation stabilizes the CDK7-cyclin H-MAT1 complex in vivo and regulates its CTD kinase activity’, *The EMBO Journal*, 20(14), pp. 3749–3759. doi:10.1093/emboj/20.14.3749.

Lasda, E. and Parker, R. (2014) ‘Circular RNAs: diversity of form and function’, *RNA*, 20(12), pp. 1829–1842. doi:10.1261/rna.047126.114.

Leal-Esteban, L.C. and Fajas, L. (2020) ‘Cell cycle regulators in cancer cell metabolism’, *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1866(5), p. 165715. doi:10.1016/j.bbadis.2020.165715.

Lee, E.C.S. *et al.* (2019) ‘The roles of circular RNAs in human development and diseases’, *Biomedicine & Pharmacotherapy*, 111, pp. 198–208. doi:10.1016/j.biopha.2018.12.052.

Lee, J.M. *et al.* (2006) ‘The epithelial–mesenchymal transition: new insights in signaling, development, and disease’, *Journal of Cell Biology*, 172(7), pp. 973–981. doi:10.1083/jcb.200601018.

Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) ‘The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14’, *Cell*, 75(5), pp. 843–854.

Legnini, I. *et al.* (2017) ‘Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis’, *Molecular Cell*, 66(1), pp. 22-37.e9. doi:10.1016/j.molcel.2017.02.017.

Lei, W. *et al.* (2018) ‘Signature of circular RNAs in human induced pluripotent stem cells and derived cardiomyocytes’, *Stem Cell Research & Therapy*, 9(1), p. 56. doi:10.1186/s13287-018-0793-5.

Leopold, P.L., Vincent, J. and Wang, H. (2012) ‘A comparison of epithelial-to-mesenchymal transition and re-epithelialization’, *Seminars in Cancer Biology*, 22(5), pp. 471–483. doi:10.1016/j.semcancer.2012.07.003.

Leroy, P. and Mostov, K.E. (2007) ‘Slug Is Required for Cell Survival during Partial Epithelial-Mesenchymal Transition of HGF-induced Tubulogenesis’, *Molecular Biology of the Cell*, 18(5), pp. 1943–1952. doi:10.1091/mbc.e06-09-0823.

Levanon, E.Y. *et al.* (2004) ‘Systematic identification of abundant A-to-I editing sites in the human transcriptome’, *Nature Biotechnology*, 22(8), pp. 1001–1005. doi:10.1038/nbt996.

Li, J. *et al.* (2018) ‘An alternative splicing switch in FLNB promotes the mesenchymal cell state in human breast cancer’, *eLife*. Edited by D.L. Black and J.L. Manley, 7, p. e37184. doi:10.7554/eLife.37184.

Li, M., Li, Y. and Yu, M. (2020) ‘<p>CircRNA ZNF609 Knockdown Suppresses Cell Growth via Modulating miR-188/ELF2 Axis in Nasopharyngeal Carcinoma</p>’, *OncoTargets and Therapy*, 13, pp. 2399–2409. doi:10.2147/OTT.S234230.

Li, Q. *et al.* (2017) ‘A sequential EMT-MET mechanism drives the differentiation of human embryonic stem cells towards hepatocytes’, *Nature Communications*, 8, p. 15166. doi:10.1038/ncomms15166.

Li, R. *et al.* (2010) ‘A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts’, *Cell Stem Cell*, 7(1), pp. 51–63. doi:10.1016/j.stem.2010.04.014.

Li, W. and Kang, Y. (2016) ‘Probing the Fifty Shades of EMT in Metastasis’, *Trends in cancer*, 2(2), pp. 65–67. doi:10.1016/j.trecan.2016.01.001.

Li, X., Pei, D. and Zheng, H. (2014) ‘Transitions between epithelial and mesenchymal states during cell fate conversions’, *Protein & Cell*, 5(8), pp. 580–591. doi:10.1007/s13238-014-0064-x.

Li, Y. *et al.* (2015) ‘Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis’, *Cell Research*, 25(8), pp. 981–984. doi:10.1038/cr.2015.82.

Li, Y. *et al.* (2019) ‘Circular RNA circMTO1 suppresses bladder cancer metastasis by sponging miR-221 and inhibiting epithelial-to-mesenchymal transition’, *Biochemical and Biophysical Research Communications*, 508(4), pp. 991–996. doi:10.1016/j.bbrc.2018.12.046.

Li, Y.-H. *et al.* (2016) ‘Neuron-derived FGF10 ameliorates cerebral ischemia injury via inhibiting NF-κB-dependent neuroinflammation and activating PI3K/Akt survival signaling pathway in mice’, *Scientific Reports*, 6(1), p. 19869. doi:10.1038/srep19869.

Li, Z. *et al.* (2015) ‘Exon-intron circular RNAs regulate transcription in the nucleus’, *Nature Structural & Molecular Biology*, 22(3), pp. 256–264. doi:10.1038/nsmb.2959.

Liang, W.-C. *et al.* (2019) ‘Translation of the circular RNA circβ-catenin promotes liver cancer cell growth through activation of the Wnt pathway’, *Genome Biology*, 20(1), p. 84. doi:10.1186/s13059-019-1685-4.

Liao, Y., Smyth, G.K. and Shi, W. (2014) ‘featureCounts: an efficient general purpose program for assigning sequence reads to genomic features’, *Bioinformatics (Oxford, England)*, 30(7), pp. 923–930. doi:10.1093/bioinformatics/btt656.

Liao, Y.-J. *et al.* (2018) ‘Effects of Klf4 and c-Myc Knockdown on Pluripotency Maintenance in Porcine Induced Pluripotent Stem Cell’, *Cell Journal (Yakhteh)*, 19(4), pp. 640–646. doi:10.22074/cellj.2018.4428.

Lim, S. and Kaldis, P. (2013) ‘Cdks, cyclins and CKIs: roles beyond cell cycle regulation’, *Development*, 140(15), pp. 3079–3093. doi:10.1242/dev.091744.

Lin, Q. *et al.* (2018) 'Circular RNA circCDK13 suppresses cell proliferation, migration and invasion by modulating the JAK/STAT and PI3K/AKT pathways in liver cancer', *International Journal of Oncology*, 53(1), pp. 246–256.

Lin, Y. *et al.* (2010) 'The SNAG domain of Snail1 functions as a molecular hook for recruiting lysine-specific demethylase 1', *The EMBO Journal*, 29(11), pp. 1803–1816. doi:10.1038/emboj.2010.63.

Liu, C.-X. *et al.* (2019) 'Structure and Degradation of Circular RNAs Regulate PKR Activation in Innate Immunity', *Cell*, 177(4), pp. 865-880.e21. doi:10.1016/j.cell.2019.03.046.

Liu, D. *et al.* (2018) 'A Highly Efficient Strategy for Overexpressing circRNAs', in Dieterich, C. and Papantonis, A. (eds) *Circular RNAs: Methods and Protocols*. New York, NY: Springer (Methods in Molecular Biology), pp. 97–105. doi:10.1007/978-1-4939-7562-4_8.

Liu, J. *et al.* (2013) 'Suppression of SCARA5 by Snail1 is essential for EMT-associated cell migration of A549 cells', *Oncogenesis*, 2(9), pp. e73–e73. doi:10.1038/oncsis.2013.37.

Liu, L. *et al.* (2019) 'The cell cycle in stem cell proliferation, pluripotency and differentiation', *Nature Cell Biology*, 21(9), pp. 1060–1067. doi:10.1038/s41556-019-0384-4.

Liu, S. *et al.* (2017) 'Chromodomain Protein CDYL Acts as a Crotonyl-CoA Hydratase to Regulate Histone Crotonylation and Spermatogenesis', *Molecular Cell*, 67(5), pp. 853-866.e5. doi:10.1016/j.molcel.2017.07.011.

Liu, S.J. and Lim, D.A. (2018) 'Modulating the expression of long non-coding RNA s for functional studies', *EMBO reports*, 19(12). doi:10.15252/embr.201846955.

Liu, X. *et al.* (2020) 'Identification of mecciRNAs and their roles in the mitochondrial entry of proteins', *Science China. Life Sciences*, 63(10), pp. 1429–1449. doi:10.1007/s11427-020-1631-9.

Liu, X. *et al.* (2021) 'Genome-wide profiling of circular RNAs, alternative splicing, and R-loops in stem-differentiating xylem of Populus trichocarpa', *Journal of Integrative Plant Biology*, 63(7), pp. 1294–1308. doi:10.1111/jipb.13081.

Liu, Y. *et al.* (2012) 'One-Step Derivation of Mesenchymal Stem Cell (MSC)-Like Cells from Human Pluripotent Stem Cells on a Fibrillar Collagen Coating', *PLoS ONE*. Edited by C.J. Connon, 7(3), p. e33225. doi:10.1371/journal.pone.0033225.

Liu, Y. *et al.* (2018) 'Circular RNA-MTO1 suppresses breast cancer cell viability and reverses monastrol resistance through regulating the TRAF4/Eg5 axis', *International Journal of Oncology*, 53(4), pp. 1752–1762. doi:10.3892/ijo.2018.4485.

Liu, Y. *et al.* (2020) ‘Back-spliced RNA from retrotransposon binds to centromere and regulates centromeric chromatin loops in maize’, *PLOS Biology*, 18(1), p. e3000582. doi:10.1371/journal.pbio.3000582.

Long, K.R. and Huttner, W.B. (2019) ‘How the extracellular matrix shapes neural development’, *Open Biology*, 9(1), p. 180216. doi:10.1098/rsob.180216.

Love, M.I., Huber, W. and Anders, S. (2014) ‘Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2’, *Genome Biology*, 15(12), p. 550. doi:10.1186/s13059-014-0550-8.

Lovisa, S. *et al.* (2015) ‘Epithelial-to-mesenchymal transition induces cell cycle arrest and parenchymal damage in renal fibrosis’, *Nature Medicine*, 21(9), pp. 998–1009. doi:10.1038/nm.3902.

Lu, J. *et al.* (2006) ‘Defined culture conditions of human embryonic stem cells’, *Proceedings of the National Academy of Sciences of the United States of America*, 103(15), pp. 5688–5693. doi:10.1073/pnas.0601383103.

Lun, A.T.L., Chen, Y. and Smyth, G.K. (2016) ‘It’s DE-licious: A Recipe for Differential Expression Analyses of RNA-seq Experiments Using Quasi-Likelihood Methods in edgeR’, *Methods in Molecular Biology (Clifton, N.J.)*, 1418, pp. 391–416. doi:10.1007/978-1-4939-3578-9_19.

Ma, W. *et al.* (2008) ‘Cell-extracellular matrix interactions regulate neural differentiation of human embryonic stem cells’, *BMC Developmental Biology*, 8, p. 90. doi:10.1186/1471-213X-8-90.

Maheswaran, S. and Haber, D.A. (2015) ‘Transition loses its invasive edge’, *Nature*, 527(7579), pp. 452–453. doi:10.1038/nature16313.

Maier, T., Güell, M. and Serrano, L. (2009) ‘Correlation of mRNA and protein in complex biological samples’, *FEBS Letters*, 583(24), pp. 3966–3973. doi:10.1016/j.febslet.2009.10.036.

Major, M.L., Lepe, R. and Costa, R.H. (2004) ‘Forkhead Box M1B Transcriptional Activity Requires Binding of Cdk-Cyclin Complexes for Phosphorylation-Dependent Recruitment of p300/CBP Coactivators’, *Molecular and Cellular Biology*, 24(7), pp. 2649–2661. doi:10.1128/MCB.24.7.2649-2661.2004.

Mak, S.K. *et al.* (2012) *Small Molecules Greatly Improve Conversion of Human-Induced Pluripotent Stem Cells to the Neuronal Lineage*, *Stem Cells International*. Hindawi. doi:https://doi.org/10.1155/2012/140427.

Malumbres, M. and Barbacid, M. (2009) ‘Cell cycle, CDKs and cancer: a changing paradigm’, *Nature Reviews Cancer*, 9(3), pp. 153–166. doi:10.1038/nrc2602.

Mani, S.A. *et al.* (2008) ‘The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells’, *Cell*, 133(4), pp. 704–715. doi:10.1016/j.cell.2008.03.027.

Marco-Puche, G. *et al.* (2019) ‘RNA-Seq Perspectives to Improve Clinical Diagnosis’, *Frontiers in Genetics*, 10, p. 1152. doi:10.3389/fgene.2019.01152.

Mariani, J. *et al.* (2012) ‘Modeling human cortical development in vitro using induced pluripotent stem cells’, *Proceedings of the National Academy of Sciences*, 109(31), pp. 12770–12775. doi:10.1073/pnas.1202944109.

Mariani, J. *et al.* (2015) ‘FOXG1-dependent dysregulation of GABA/glutamate neuron differentiation in autism spectrum disorders’, *Cell*, 162(2), pp. 375–390. doi:10.1016/j.cell.2015.06.034.

Martin, B.L. and Kimelman, D. (2012) ‘Canonical Wnt Signaling Dynamically Controls Multiple Stem Cell Fate Decisions during Vertebrate Body Formation’, *Developmental cell*, 22(1), pp. 223–232. doi:10.1016/j.devcel.2011.11.001.

Martin, M. (2011) ‘Cutadapt removes adapter sequences from high-throughput sequencing reads’, *EMBnet.journal*, 17(1), pp. 10–12. doi:10.14806/ej.17.1.200.

Martin, R.M. *et al.* (2020) ‘Improving the safety of human pluripotent stem cell therapies using genome-edited orthogonal safeguards’, *Nature Communications*, 11(1), p. 2713. doi:10.1038/s41467-020-16455-7.

Martínez-Cerdeño, V., Noctor, S.C. and Kriegstein, A.R. (2006) ‘The Role of Intermediate Progenitor Cells in the Evolutionary Expansion of the Cerebral Cortex’, *Cerebral Cortex*, 16(suppl_1), pp. i152–i161. doi:10.1093/cercor/bhk017.

Matera, A.G. and Wang, Z. (2014) ‘A day in the life of the spliceosome’, *Nature reviews. Molecular cell biology*, 15(2), pp. 108–121. doi:10.1038/nrm3742.

Maurano, M.T. *et al.* (2012) ‘Systematic Localization of Common Disease-Associated Variation in Regulatory DNA’, *Science (New York, N.Y.)*, 337(6099), pp. 1190–1195. doi:10.1126/science.1222794.

McIntyre, G.J. and Fanning, G.C. (2006) ‘Design and cloning strategies for constructing shRNA expression vectors’, *BMC Biotechnology*, 6(1), p. 1. doi:10.1186/1472-6750-6-1.

Memczak, S. *et al.* (2013) ‘Circular RNAs are a large class of animal RNAs with regulatory potency’, *Nature*, 495(7441), pp. 333–338. doi:10.1038/nature11928.

Memczak, S. *et al.* (2015) ‘Identification and Characterization of Circular RNAs As a New Class of Putative Biomarkers in Human Blood’, *PLOS ONE*, 10(10), p. e0141214. doi:10.1371/journal.pone.0141214.

Meng, J. *et al.* (2018) ‘Twist1 Regulates Vimentin through Cul2 Circular RNA to Promote EMT in Hepatocellular Carcinoma’, *Cancer Research*, 78(15), pp. 4150–4162. doi:10.1158/0008-5472.CAN-17-3009.

Meng, X. *et al.* (2017) ‘Circular RNA: an emerging key player in RNA world’, *Briefings in Bioinformatics*, 18(4), pp. 547–557. doi:10.1093/bib/bbw045.

Mens, M.M.J. and Ghanbari, M. (2018) ‘Cell Cycle Regulation of Stem Cells by MicroRNAs’, *Stem Cell Reviews and Reports*, 14(3), pp. 309–322. doi:10.1007/s12015-018-9808-y.

Metzakopian, E. *et al.* (2012) ‘Genome-wide characterization of Foxa2 targets reveals upregulation of floor plate genes and repression of ventrolateral genes in midbrain dopaminergic progenitors’, *Development*, 139(14), pp. 2625–2634. doi:10.1242/dev.081034.

Mo, Y. *et al.* (2021) ‘Role of circRNAs in viral infection and their significance for diagnosis and treatment (Review)’, *International Journal of Molecular Medicine*, 47(5), p. 88. doi:10.3892/ijmm.2021.4921.

Mockenhaupt, S. *et al.* (2015) ‘Alleviation of off-target effects from vector-encoded shRNAs via codelivered RNA decoys’, *Proceedings of the National Academy of Sciences*, 112(30), pp. E4007–E4016. doi:10.1073/pnas.1510476112.

Molinari, M. (2000) ‘Cell cycle checkpoints and their inactivation in human cancer’, *Cell Proliferation*, 33(5), pp. 261–274. doi:10.1046/j.1365-2184.2000.00191.x.

Moore, C.B. *et al.* (2010) ‘Short Hairpin RNA (shRNA): Design, Delivery, and Assessment of Gene Knockdown’, *Methods in molecular biology (Clifton, N.J.)*, 629, pp. 141–158. doi:10.1007/978-1-60761-657-3_10.

Morris, K.V. *et al.* (2004) ‘Small Interfering RNA-Induced Transcriptional Gene Silencing in Human Cells’, *Science*, 305(5688), pp. 1289–1292. doi:10.1126/science.1101372.

Mossahebi-Mohammadi, M. *et al.* (2020) ‘FGF Signaling Pathway: A Key Regulator of Stem Cell Pluripotency’, *Frontiers in Cell and Developmental Biology*, 8. Available at: https://www.frontiersin.org/article/10.3389/fcell.2020.00079 (Accessed: 14 May 2022).

Nakaya, Y. and Sheng, G. (2008) ‘Epithelial to mesenchymal transition during gastrulation: an embryological view’, *Development, Growth & Differentiation*, 50(9), pp. 755–766. doi:10.1111/j.1440-169X.2008.01070.x.

Neumann, D.P., Goodall, G.J. and Gregory, P.A. (2018) ‘Regulation of splicing and circularisation of RNA in epithelial mesenchymal plasticity’, *Seminars in Cell & Developmental Biology*, 75, pp. 50–60. doi:10.1016/j.semcdb.2017.08.008.

Nieto, M.A. (2013) ‘Epithelial Plasticity: A Common Theme in Embryonic and Cancer Cells’, *Science*, 342(6159), pp. 1234850–1234850. doi:10.1126/science.1234850.

Nieto, M.A. *et al.* (2016) ‘EMT: 2016’, *Cell*, 166(1), pp. 21–45. doi:10.1016/j.cell.2016.06.028.

Nigro, J.M. *et al.* (1991) ‘Scrambled exons’, *Cell*, 64(3), pp. 607–613. doi:10.1016/0092-8674(91)90244-S.

Nilsen, T.W. and Graveley, B.R. (2010) ‘Expansion of the eukaryotic proteome by alternative splicing’, *Nature*, 463(7280), pp. 457–463. doi:10.1038/nature08909.

Ocaña, O.H. *et al.* (2012) ‘Metastatic Colonization Requires the Repression of the Epithelial-Mesenchymal Transition Inducer Prrx1’, *Cancer Cell*, 22(6), pp. 709–724. doi:10.1016/j.ccr.2012.10.012.

Oceguera-Yanez, F. *et al.* (2016) ‘Engineering the AAVS1 locus for consistent and scalable transgene expression in human iPSCs and their differentiated derivatives’, *Methods (San Diego, Calif.)*, 101, pp. 43–55. doi:10.1016/j.ymeth.2015.12.012.

Ogawa, K. *et al.* (2006) ‘Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells’, *Biochemical and Biophysical Research Communications*, 343(1), pp. 159–166. doi:10.1016/j.bbrc.2006.02.127.

O’Gorman, W. *et al.* (2005) ‘Analysis of U1 Small Nuclear RNA Interaction with Cyclin H *’, *Journal of Biological Chemistry*, 280(44), pp. 36920–36925. doi:10.1074/jbc.M505791200.

Oka, Y. *et al.* (2010) ‘293FT cells transduced with four transcription factors (OCT4, SOX2, NANOG, and LIN28) generate aberrant ES-like cells’, *Journal of Stem Cells & Regenerative Medicine*, 6(3), pp. 149–156.

O’Keefe, E.P. (2020) ‘siRNAs and shRNAs: Tools for Protein Knockdown by Gene Silencing’, *Materials and Methods* [Preprint]. Available at: /method/siRNAs-and-shRNAs-Tools-for-Protein-Knockdown-by-Gene-Silencing.html (Accessed: 4 April 2020).

Okita, K., Ichisaka, T. and Yamanaka, S. (2007) ‘Generation of germline-competent induced pluripotent stem cells’, *Nature*, 448(7151), pp. 313–317. doi:10.1038/nature05934.

Okoye, U., Malbon, C. and Wang, H. (2008) ‘Wnt and Frizzled RNA expression in human mesenchymal and embryonic (H7) stem cells’, *Journal of Molecular Signaling*, 3(0), p. Art. 16. doi:10.1186/1750-2187-3-16.

Özcan, A. *et al.* (2021) ‘Programmable RNA targeting with the single-protein CRISPR effector Cas7-11’, *Nature*, 597(7878), pp. 720–725. doi:10.1038/s41586-021-03886-5.

Pamudurti, N.R. *et al.* (2017) ‘Translation of CircRNAs’, *Molecular Cell*, 66(1), pp. 9-21.e7. doi:10.1016/j.molcel.2017.02.021.

Pan, Y. *et al.* (2018) ‘CircBA9.3 supports the survival of leukaemic cells by up-regulating c-ABL1 or BCR-ABL1 protein levels’, *Blood Cells, Molecules, and Diseases*, 73, pp. 38–44. doi:10.1016/j.bcmd.2018.09.002.

Panda, A.C. and Gorospe, M. (2018) 'Detection and Analysis of Circular RNAs by RT-PCR.', *Bio-protocol* [Preprint]. doi:10.21769/BioProtoc.2775.

Pandey, P.R. *et al.* (2019) 'RPAD (RNase R Treatment, Polyadenylation, and Poly(A)+ RNA Depletion) Method to Isolate Highly Pure Circular RNA', *Methods (San Diego, Calif.)*, 155, pp. 41–48. doi:10.1016/j.ymeth.2018.10.022.

Panina, Y. *et al.* (2018) 'Validation of Common Housekeeping Genes as Reference for qPCR Gene Expression Analysis During iPS Reprogramming Process', *Scientific Reports*, 8(1), p. 8716. doi:10.1038/s41598-018-26707-8.

Panina, Y., Germond, A. and Watanabe, T.M. (2020) 'Analysis of the stability of 70 housekeeping genes during iPS reprogramming', *Scientific Reports*, 10(1), p. 21711. doi:10.1038/s41598-020-78863-5.

Park, O.H. *et al.* (2019) 'Endoribonucleolytic Cleavage of m6A-Containing RNAs by RNase P/MRP Complex', *Molecular Cell*, 74(3), pp. 494-507.e8. doi:10.1016/j.molcel.2019.02.034.

Pastushenko, I. *et al.* (2018) 'Identification of the tumour transition states occurring during EMT', *Nature*, 556(7702), pp. 463–468. doi:10.1038/s41586-018-0040-3.

Patel, S.A. and Simon, M.C. (2010) 'Functional Analysis of the Cdk7·Cyclin H·Mat1 Complex in Mouse Embryonic Stem Cells and Embryos', *Journal of Biological Chemistry*, 285(20), pp. 15587–15598. doi:10.1074/jbc.M109.081687.

Pauklin, S. and Vallier, L. (2013) 'The Cell-Cycle State of Stem Cells Determines Cell Fate Propensity', *Cell*, 155(1), pp. 135–147. doi:10.1016/j.cell.2013.08.031.

Pauklin, S. and Vallier, L. (2015) 'Activin/Nodal signalling in stem cells', *Development*, 142(4), pp. 607–619. doi:10.1242/dev.091769.

Peinado, H. *et al.* (2004) 'Snail Mediates E-Cadherin Repression by the Recruitment of the Sin3A/Histone Deacetylase 1 (HDAC1)/HDAC2 Complex', *Molecular and Cellular Biology*, 24(1), pp. 306–319. doi:10.1128/MCB.24.1.306-319.2004.

Peng, C. *et al.* (2020) 'Cyclin H predicts the poor prognosis and promotes the proliferation of ovarian cancer', *Cancer Cell International*, 20(1), p. 316. doi:10.1186/s12935-020-01406-5.

Pillman, K.A. *et al.* (2018) 'miR-200/375 control epithelial plasticity-associated alternative splicing by repressing the RNA-binding protein Quaking', *The EMBO Journal*, p. e99016. doi:10.15252/embj.201899016.

Piwecka, M. *et al.* (2017) 'Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function', *Science*, 357(6357), p. eaam8526. doi:10.1126/science.aam8526.

Pristerà, A. *et al.* (2015) 'Transcription factors FOXA1 and FOXA2 maintain dopaminergic neuronal properties and control feeding behavior in adult mice', *Proceedings of the National Academy of Sciences*, 112(35), pp. E4929–E4938. doi:10.1073/pnas.1503911112.

Puisieux, A., Brabletz, T. and Caramel, J. (2014) 'Oncogenic roles of EMT-inducing transcription factors', *Nature Cell Biology*, 16(6), pp. 488–494. doi:10.1038/ncb2976.

Radley, A.H. *et al.* (2017) 'Assessment of engineered cells using CellNet and RNA-seq', *Nature Protocols*, 12(5), pp. 1089–1102. doi:10.1038/nprot.2017.022.

Rao, D.D. *et al.* (2009) 'Comparative assessment of siRNA and shRNA off target effects: what is slowing clinical development', *Cancer Gene Therapy*, 16(11), pp. 807–809. doi:10.1038/cgt.2009.53.

Reddy, A.S.N. *et al.* (2013) 'Complexity of the Alternative Splicing Landscape in Plants', *The Plant Cell*, 25(10), pp. 3657–3683. doi:10.1105/tpc.113.117523.

Reka, A.K. *et al.* (2011) 'Identifying Inhibitors of Epithelial-Mesenchymal Transition by Connectivity Map–Based Systems Approach', *Journal of Thoracic Oncology*, 6(11), pp. 1784–1792. doi:10.1097/JTO.0b013e31822adfb0.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics (Oxford, England)*, 26(1), pp. 139–140. doi:10.1093/bioinformatics/btp616.

Ros, X.B.-D. and Gu, S. (2016) 'Guidelines for the optimal design of miRNA-based shRNAs', *Methods (San Diego, Calif.)*, 103, pp. 157–166. doi:10.1016/j.ymeth.2016.04.003.

Rust, W.L., Sadasivam, A. and Dunn, N.R. (2006) 'Three-Dimensional Extracellular Matrix Stimulates Gastrulation-like Events in Human Embryoid Bodies', *Stem Cells and Development*, 15(6), pp. 889–904. doi:10.1089/scd.2006.15.889.

Rybak-Wolf, A. *et al.* (2015) 'Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed', *Molecular Cell*, 58(5), pp. 870–885. doi:10.1016/j.molcel.2015.03.027.

Sahara, S. and O'Leary, D.D.M. (2009) 'Fgf10 regulates transition period of cortical stem cell differentiation to radial glia controlling generation of neurons and basal progenitors', *Neuron*, 63(1), pp. 48–62. doi:10.1016/j.neuron.2009.06.006.

Sajini, A.A. *et al.* (2012) 'Loss of Oct4 expression during the development of murine embryoid bodies', *Developmental Biology*, 371(2), pp. 170–179. doi:10.1016/j.ydbio.2012.08.008.

Salzman, J. *et al.* (2012) ‘Circular RNAs Are the Predominant Transcript Isoform from Hundreds of Human Genes in Diverse Cell Types’, *PLOS ONE*, 7(2), p. e30733. doi:10.1371/journal.pone.0030733.

Salzman, J. *et al.* (2013) ‘Cell-Type Specific Features of Circular RNA Expression’, *PLOS Genetics*, 9(9), p. e1003777. doi:10.1371/journal.pgen.1003777.

Samavarchi-Tehrani, P. *et al.* (2010) ‘Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming’, *Cell Stem Cell*, 7(1), pp. 64–77. doi:10.1016/j.stem.2010.04.015.

Sang, Y. *et al.* (2019) ‘circRNA_0025202 Regulates Tamoxifen Sensitivity and Tumor Progression via Regulating the miR-182-5p/FOXO3a Axis in Breast Cancer’, *Molecular Therapy*, 27(9), pp. 1638–1652. doi:10.1016/j.ymthe.2019.05.011.

Sato, N. *et al.* (2004) ‘Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor’, *Nature Medicine*, 10(1), pp. 55–63. doi:10.1038/nm979.

Schliekelman, M.J. *et al.* (2015) ‘Molecular Portraits of Epithelial, Mesenchymal, and Hybrid States in Lung Adenocarcinoma and Their Relevance to Survival’, *Cancer Research*, 75(9), pp. 1789–1800. doi:10.1158/0008-5472.CAN-14-2535.

Schneider, E., Montenarh, M. and Wagner, P. (1998) ‘Regulation of CAK kinase activity by p53’, *Oncogene*, 17(21), pp. 2733–2741. doi:10.1038/sj.onc.1202504.

Sekar, S. and Liang, W.S. (2019) ‘Circular RNA expression and function in the brain’, *Non-coding RNA Research*, 4(1), pp. 23–29. doi:10.1016/j.ncrna.2019.01.001.

Shang, B.-Q. *et al.* (2019) ‘Functional roles of circular RNAs during epithelial-to-mesenchymal transition’, *Molecular Cancer*, 18(1), p. 138. doi:10.1186/s12943-019-1071-6.

Shang, J. *et al.* (2019) ‘CircPAN3 contributes to drug resistance in acute myeloid leukemia through regulation of autophagy’, *Leukemia Research*, 85, p. 106198. doi:10.1016/j.leukres.2019.106198.

Shang, Z. (2018) ‘Neural progenitor cells derived from human induced pluripotent stem cells.’ doi:10.17504/protocols.io.ntrdem6.

Shanmugapriya *et al.* (2018) ‘Functional Analysis of Circular RNAs’, in Xiao, J. (ed.) *Circular RNAs: Biogenesis and Functions*. Singapore: Springer (Advances in Experimental Medicine and Biology), pp. 95–105. doi:10.1007/978-981-13-1426-1_8.

Shao, F. *et al.* (2018) ‘Circular RNA Signature Predicts Gemcitabine Resistance of Pancreatic Ductal Adenocarcinoma’, *Frontiers in Pharmacology*, 9, p. 584. doi:10.3389/fphar.2018.00584.

Sharma, S. and Rao, A. (2009) ‘RNAi screening: tips and techniques’, *Nature immunology*, 10(8), pp. 799–804. doi:10.1038/ni0809-799.

Sheng, P., Flood, K.A. and Xie, M. (2020) ‘Short Hairpin RNAs for Strand-Specific Small Interfering RNA Production’, *Frontiers in Bioengineering and Biotechnology*, 8, p. 940. doi:10.3389/fbioe.2020.00940.

Shi, Y. *et al.* (2012) ‘Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses’, *Nature Neuroscience*, 15(3), pp. 477–486. doi:10.1038/nn.3041.

Shi, Z. *et al.* (2017) ‘The circular RNA ciRS-7 promotes APP and BACE1 degradation in an NF-κB-dependent manner’, *The FEBS Journal*, 284(7), pp. 1096–1109. doi:10.1111/febs.14045.

Shintani, Y. *et al.* (2011) ‘Epithelial to Mesenchymal Transition Is a Determinant of Sensitivity to Chemoradiotherapy in Non-Small Cell Lung Cancer’, *The Annals of Thoracic Surgery*, 92(5), pp. 1794–1804. doi:10.1016/j.athoracsur.2011.07.032.

Siemens, H. *et al.* (2011) ‘miR-34 and SNAIL form a double-negative feedback loop to regulate epithelial-mesenchymal transitions’, *Cell Cycle (Georgetown, Tex.)*, 10(24), pp. 4256–4271. doi:10.4161/cc.10.24.18552.

Sokol, S.Y. (2011) ‘Maintaining embryonic stem cell pluripotency with Wnt signaling’, *Development (Cambridge, England)*, 138(20), pp. 4341–4350. doi:10.1242/dev.066209.

Song, Y.-Z. and Li, J.-F. (2018) ‘Circular RNA hsa_circ_0001564 regulates osteosarcoma proliferation and apoptosis by acting miRNA sponge’, *Biochemical and Biophysical Research Communications*, 495(3), pp. 2369–2375. doi:10.1016/j.bbrc.2017.12.050.

Stagsted, L.V. *et al.* (2019) ‘Noncoding AUG circRNAs constitute an abundant and conserved subclass of circles’, *Life Science Alliance*, 2(3), p. e201900398. doi:10.26508/lsa.201900398.

Stemmler, M.P. *et al.* (2019) ‘Non-redundant functions of EMT transcription factors’, *Nature Cell Biology*, 21(1), p. 102. doi:10.1038/s41556-018-0196-y.

Strauss, R. *et al.* (2011) ‘Analysis of Epithelial and Mesenchymal Markers in Ovarian Cancer Reveals Phenotypic Heterogeneity and Plasticity’, *PLoS ONE*. Edited by D. Gullberg, 6(1), p. e16186. doi:10.1371/journal.pone.0016186.

Sun, X. *et al.* (2016) ‘Integrative analysis of Arabidopsis thaliana transcriptomics reveals intuitive splicing mechanism for circular RNA’, *FEBS Letters*, 590(20), pp. 3510–3516. doi:10.1002/1873-3468.12440.

Sun, Y. *et al.* (2017) ‘A Novel Regulatory Mechanism of Smooth Muscle α-Actin Expression by NRG-1/circACTA2/miR-548f-5p Axis’, *Circulation research*, 121(6), pp. 628–635. doi:10.1161/CIRCRESAHA.117.311441.

Syed, N.H. *et al.* (2012) ‘Alternative splicing in plants – coming of age’, *Trends in Plant Science*, 17(10), pp. 616–623. doi:10.1016/j.tplants.2012.06.001.

Tam, W.L. and Weinberg, R.A. (2013) ‘The epigenetics of epithelial-mesenchymal plasticity in cancer’, *Nature Medicine*, 19(11), pp. 1438–1449. doi:10.1038/nm.3336.

Tan, T.Z. *et al.* (2014) ‘Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients’, *EMBO Molecular Medicine*, 6(10), pp. 1279–1293. doi:10.15252/emmm.201404208.

Tang, C. *et al.* (2020) ‘m6A-dependent biogenesis of circular RNAs in male germ cells’, *Cell Research*, 30(3), pp. 211–228. doi:10.1038/s41422-020-0279-8.

Tang, T. *et al.* (2021) ‘Programmable System of Cas13-Mediated RNA Modification and Its Biological and Biomedical Applications’, *Frontiers in Cell and Developmental Biology*, 9, p. 1806. doi:10.3389/fcell.2021.677587.

Taxman, D.J. *et al.* (2006) ‘Criteria for effective design, construction, and gene knockdown by shRNA vectors’, *BMC Biotechnology*, 6(1), p. 7. doi:10.1186/1472-6750-6-7.

Teo, A.K.K. *et al.* (2012) ‘Activin and BMP4 synergistically promote formation of definitive endoderm in human embryonic stem cells’, *Stem Cells (Dayton, Ohio)*, 30(4), pp. 631–642. doi:10.1002/stem.1022.

Teplova, M. *et al.* (2013) ‘Structure–function studies of STAR family Quaking proteins bound to their in vivo RNA target sites’, *Genes & Development*, 27(8), pp. 928–940. doi:10.1101/gad.216531.113.

Thiery, J.P. (2002) ‘Epithelial–mesenchymal transitions in tumour progression’, *Nature Reviews Cancer*, 2(6), pp. 442–454. doi:10.1038/nrc822.

Thiery, J.P. *et al.* (2009) ‘Epithelial-Mesenchymal Transitions in Development and Disease’, *Cell*, 139(5), pp. 871–890. doi:10.1016/j.cell.2009.11.007.

Thomson, J.A. *et al.* (1998) ‘Embryonic stem cell lines derived from human blastocysts’, *Science (New York, N.Y.)*, 282(5391), pp. 1145–1147. doi:10.1126/science.282.5391.1145.

Thomson, S. *et al.* (2008) ‘Kinase switching in mesenchymal-like non-small cell lung cancer lines contributes to EGFR inhibitor resistance through pathway redundancy’, *Clinical & Experimental Metastasis*, 25(8), pp. 843–854. doi:10.1007/s10585-008-9200-4.

Thorvaldsdóttir, H., Robinson, J.T. and Mesirov, J.P. (2013) 'Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration', *Briefings in Bioinformatics*, 14(2), pp. 178–192. doi:10.1093/bib/bbs017.

Ting, D.T. *et al.* (2014) 'Single-Cell RNA Sequencing Identifies Extracellular Matrix Gene Expression by Pancreatic Circulating Tumor Cells', *Cell Reports*, 8(6), pp. 1905–1918. doi:10.1016/j.celrep.2014.08.029.

Tirino, V. *et al.* (2013) 'TGF- β 1 exposure induces epithelial to mesenchymal transition both in CSCs and non-CSCs of the A549 cell line, leading to an increase of migration ability in the CD133 + A549 cell fraction', *Cell Death & Disease*, 4(5), pp. e620–e620. doi:10.1038/cddis.2013.144.

Tng, P.Y.L. *et al.* (2020) 'Cas13b-dependent and Cas13b-independent RNA knockdown of viral sequences in mosquito cells following guide RNA expression', *Communications Biology*, 3(1), pp. 1–9. doi:10.1038/s42003-020-01142-6.

Tomizawa, J. *et al.* (1981) 'Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA.', *Proceedings of the National Academy of Sciences of the United States of America*, 78(3), pp. 1421–1425.

Tsai, J.H. *et al.* (2012) 'Spatiotemporal Regulation of Epithelial-Mesenchymal Transition Is Essential for Squamous Cell Carcinoma Metastasis', *Cancer Cell*, 22(6), pp. 725–736. doi:10.1016/j.ccr.2012.09.022.

Ullmann, U. *et al.* (2007) 'Epithelial–mesenchymal transition process in human embryonic stem cells cultured in feeder-free conditions', *MHR: Basic science of reproductive medicine*, 13(1), pp. 21–32. doi:10.1093/molehr/gal091.

Vallier, L. *et al.* (2009) 'Activin/Nodal signalling maintains pluripotency by controlling Nanog expression', *Development*, 136(8), pp. 1339–1349. doi:10.1242/dev.033951.

Vallier, L., Alexander, M. and Pedersen, R.A. (2005) 'Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells', *Journal of Cell Science*, 118(19), pp. 4495–4509. doi:10.1242/jcs.02553.

Vallier, L., Reynolds, D. and Pedersen, R.A. (2004) 'Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway', *Developmental Biology*, 275(2), pp. 403–421. doi:10.1016/j.ydbio.2004.08.031.

Vausort, M. *et al.* (2016) 'Myocardial Infarction-Associated Circular RNA Predicting Left Ventricular Dysfunction', *Journal of the American College of Cardiology*, 68(11), pp. 1247–1248. doi:10.1016/j.jacc.2016.06.040.

Vazin, T. and Freed, W.J. (2010) 'Human embryonic stem cells: Derivation, culture, and differentiation: A review', *Restorative neurology and neuroscience*, 28(4), pp. 589–603. doi:10.3233/RNN-2010-0543.

Vega, S. (2004) ‘Snail blocks the cell cycle and confers resistance to cell death’, *Genes & Development*, 18(10), pp. 1131–1143. doi:10.1101/gad.294104.

Verduci, L. *et al.* (2021) ‘CircRNAs: role in human diseases and potential use as biomarkers’, *Cell Death & Disease*, 12(5), pp. 1–12. doi:10.1038/s41419-021-03743-3.

Vermeulen, K., Van Bockstaele, D.R. and Berneman, Z.N. (2003) ‘The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer’, *Cell Proliferation*, 36(3), pp. 131–149. doi:10.1046/j.1365-2184.2003.00266.x.

Vogel, C. and Marcotte, E.M. (2012) ‘Insights into the regulation of protein abundance from proteomic and transcriptomic analyses’, *Nature Reviews Genetics*, 13(4), pp. 227–232. doi:10.1038/nrg3185.

Vossaert, L. *et al.* (2013) ‘Reference loci for RT-qPCR analysis of differentiating human embryonic stem cells’, *BMC Molecular Biology*, 14(1), p. 21. doi:10.1186/1471-2199-14-21.

Wamaitha, S.E. *et al.* (2020) ‘IGF1-mediated human embryonic stem cell self-renewal recapitulates the embryonic niche’, *Nature Communications*, 11(1), p. 764. doi:10.1038/s41467-020-14629-x.

Wang, E.T. *et al.* (2008) ‘Alternative isoform regulation in human tissue transcriptomes’, *Nature*, 456(7221), pp. 470–476. doi:10.1038/nature07509.

Wang, L. *et al.* (2018) ‘Circular RNA hsa_circ_0008305 (circPTK2) inhibits TGF-β-induced epithelial-mesenchymal transition and metastasis by controlling TIF1γ in non-small cell lung cancer’, *Molecular Cancer*, 17(1), p. 140. doi:10.1186/s12943-018-0889-7.

Wang, L. *et al.* (2019) ‘Circular RNA circRHOT1 promotes hepatocellular carcinoma progression by initiation of NR2F6 expression’, *Molecular Cancer*, 18(1). doi:10.1186/s12943-019-1046-7.

Wang, M. *et al.* (2020) ‘Cyclin-dependent kinase 7 inhibitors in cancer therapy’, *Future Medicinal Chemistry*, 12(9), pp. 813–833. doi:10.4155/fmc-2019-0334.

Wang, P.L. *et al.* (2014) ‘Circular RNA Is Expressed across the Eukaryotic Tree of Life’, *PLOS ONE*, 9(3), p. e90859. doi:10.1371/journal.pone.0090859.

Wang, W.-L. *et al.* (2015) ‘Slug is temporally regulated by cyclin E in cell cycle and controls genome stability’, *Oncogene*, 34(9), pp. 1116–1125. doi:10.1038/onc.2014.58.

Wang, X. *et al.* (2018) ‘Increased circular RNA hsa_circ_0012673 acts as a sponge of miR-22 to promote lung adenocarcinoma proliferation’, *Biochemical and Biophysical Research Communications* [Preprint]. doi:10.1016/j.bbrc.2018.01.126.

Wang, X. and Zhou, Q. (2020) ‘Circular RNA ACVR2A Inhibits NSCLC Progression by Acting as a Molecular Sponge of miR-483-3p to Modulate Chemerin Expression’, in *C110. LUNG CANCER TREATMENT: IMMUNOTHERAPY AND BEYOND. American Thoracic Society 2020 International Conference, May 15-20, 2020 - Philadelphia, PA*, American Thoracic Society, pp. A6187–A6187. doi:10.1164/ajrccm-conference.2020.201.1_MeetingAbstracts.A6187.

Warzecha, C.C. *et al.* (2009) ‘ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing’, *Molecular Cell*, 33(5), pp. 591–601. doi:10.1016/j.molcel.2009.01.025.

Warzecha, C.C. *et al.* (2010) ‘An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition’, *The EMBO journal*, 29(19), pp. 3286–3300. doi:10.1038/emboj.2010.195.

Warzecha, C.C. and Carstens, R.P. (2012) ‘Complex changes in alternative pre-mRNA splicing play a central role in the epithelial-to-mesenchymal transition (EMT)’, *Seminars in Cancer Biology*, 22(5), pp. 417–427. doi:10.1016/j.semcancer.2012.04.003.

Watabe, T. and Miyazono, K. (2009) ‘Roles of TGF-β family signaling in stem cell renewal and differentiation’, *Cell Research*, 19(1), pp. 103–115. doi:10.1038/cr.2008.323.

Wattanapanitch, M. *et al.* (2014) ‘Dual Small-Molecule Targeting of SMAD Signaling Stimulates Human Induced Pluripotent Stem Cells toward Neural Lineages’, *PLoS ONE*, 9(9). doi:10.1371/journal.pone.0106952.

Wei, J.-X. *et al.* (2009) ‘Both Strands of siRNA Have Potential to Guide Posttranscriptional Gene Silencing in Mammalian Cells’, *PLoS ONE*, 4(4). doi:10.1371/journal.pone.0005382.

Weigelt, C.M. *et al.* (2020) ‘An Insulin-Sensitive Circular RNA that Regulates Lifespan in Drosophila’, *Molecular Cell*, 79(2), pp. 268-279.e5. doi:10.1016/j.molcel.2020.06.011.

Woźniak, G. *et al.* (2016) ‘Cell-cycle gene expression analysis using real time PCR in locally advanced squamous-cell head and neck cancer’, *Advances in Medical Sciences*, 61(2), pp. 293–299. doi:10.1016/j.advms.2016.03.009.

Wu, N. *et al.* (2019) ‘Translation of yes-associated protein (YAP) was antagonized by its circular RNA via suppressing the assembly of the translation initiation machinery’, *Cell Death & Differentiation*, 26(12), pp. 2758–2773. doi:10.1038/s41418-019-0337-2.

Wu, Z. *et al.* (2020) ‘Mitochondrial Genome-Derived circRNA mc-COX2 Functions as an Oncogene in Chronic Lymphocytic Leukemia’, *Molecular Therapy. Nucleic Acids*, 20, pp. 801–811. doi:10.1016/j.omtn.2020.04.017.

Xia, S. *et al.* (2017) ‘Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes’, *Briefings in Bioinformatics*, 18(6), pp. 984–992. doi:10.1093/bib/bbw081.

Xu, C. *et al.* (2021) ‘Programmable RNA editing with compact CRISPR–Cas13 systems from uncultivated microbes’, *Nature Methods*, 18(5), pp. 499–506. doi:10.1038/s41592-021-01124-4.

Xu, C., Yu, Y. and Ding, F. (2018) ‘Microarray analysis of circular RNA expression profiles associated with gemcitabine resistance in pancreatic cancer cells’, *Oncology Reports*, 40(1), pp. 395–404. doi:10.3892/or.2018.6450.

Xu, D. *et al.* (2020) ‘A CRISPR/Cas13-based approach demonstrates biological relevance of vlinc class of long non-coding RNAs in anticancer drug response’, *Scientific Reports*, 10(1), p. 1794. doi:10.1038/s41598-020-58104-5.

Xu, H. *et al.* (2015) ‘The circular RNA Cdr1as, via miR-7 and its targets, regulates insulin transcription and secretion in islet cells’, *Scientific Reports*, 5(1), p. 12453. doi:10.1038/srep12453.

Xu, N. *et al.* (2018) ‘Profiles and Bioinformatics Analysis of Differentially Expressed Circrnas in Taxol-Resistant Non-Small Cell Lung Cancer Cells’, *Cellular Physiology and Biochemistry*, 48(5), pp. 2046–2060. doi:10.1159/000492543.

Xu, T. *et al.* (2017) ‘Circular RNA expression profiles and features in human tissues: a study using RNA-seq data’, *BMC Genomics*, 18(Suppl 6), p. 680. doi:10.1186/s12864-017-4029-3.

Yan, N. *et al.* (2017) ‘Circular RNA profile indicates circular RNA VRK1 is negatively related with breast cancer stem cells’, *Oncotarget*, 8(56), pp. 95704–95718. doi:10.18632/oncotarget.21183.

Yang, C. *et al.* (2018) ‘Circular RNA circ-ITCH inhibits bladder cancer progression by sponging miR-17/miR-224 and regulating p21, PTEN expression’, *Molecular Cancer*, 17(1), p. 19. doi:10.1186/s12943-018-0771-7.

Yang, F. *et al.* (2018) ‘Cis-acting circ-CTNNB1 promotes β-catenin signaling and cancer progression via DDX3-mediated transactivation of YY1’, *Cancer Research*, p. canres.1559.2018. doi:10.1158/0008-5472.CAN-18-1559.

Yang, F. *et al.* (2019) ‘Circ-HuR suppresses HuR expression and gastric cancer progression by inhibiting CNBP transactivation’, *Molecular Cancer*, 18(1), p. 158. doi:10.1186/s12943-019-1094-z.

Yang, L., Fu, J. and Zhou, Y. (2018) ‘Circular RNAs and Their Emerging Roles in Immune Regulation’, *Frontiers in Immunology*, 9. doi:10.3389/fimmu.2018.02977.

Yang, Q. *et al.* (2017) 'A circular RNA promotes tumorigenesis by inducing c-myc nuclear translocation', *Cell Death and Differentiation*, 24(9), pp. 1609–1620. doi:10.1038/cdd.2017.86.

Yang, Y. *et al.* (2016) 'Determination of a Comprehensive Alternative Splicing Regulatory Network and Combinatorial Regulation by Key Factors during the Epithelial-to-Mesenchymal Transition', *Molecular and Cellular Biology*, 36(11), pp. 1704–1719. doi:10.1128/MCB.00019-16.

Yang, Y. *et al.* (2018) 'Novel Role of FBXW7 Circular RNA in Repressing Glioma Tumorigenesis', *JNCI: Journal of the National Cancer Institute*, 110(3), pp. 304–315. doi:10.1093/jnci/djx166.

Yang, Y. *et al.* (2017) 'Extensive translation of circular RNAs driven by N6-methyladenosine', *Cell Research*, 27(5), pp. 626–641. doi:10.1038/cr.2017.31.

Yang, Z. *et al.* (2019) 'Dysregulation of p53-RBM25-mediated circAMOTL1L biogenesis contributes to prostate cancer progression through the circAMOTL1L-miR-193a-5p-Pcdha pathway', *Oncogene*, 38(14), pp. 2516–2532. doi:10.1038/s41388-018-0602-8.

Yao, D., Dai, C. and Peng, S. (2011) 'Mechanism of the Mesenchymal–Epithelial Transition and Its Relationship with Metastatic Tumor Formation', *Molecular Cancer Research*, 9(12), pp. 1608–1620. doi:10.1158/1541-7786.MCR-10-0568.

Yao, S. *et al.* (2006) 'Long-term self-renewal and directed differentiation of human embryonic stem cells in chemically defined conditions', *Proceedings of the National Academy of Sciences of the United States of America*, 103(18), pp. 6907–6912. doi:10.1073/pnas.0602280103.

Ye, X. *et al.* (2017) 'Upholding a role for EMT in breast cancer metastasis', *Nature*, 547(7661), pp. E1–E3. doi:10.1038/nature22816.

Yu, C.-Y. *et al.* (2017) 'The circular RNA circBIRC6 participates in the molecular circuitry controlling human pluripotency', *Nature Communications*, 8(1), p. 1149. doi:10.1038/s41467-017-01216-w.

Yu, C.-Y. and Kuo, H.-C. (2019) 'The emerging roles and functions of circular RNAs and their generation', *Journal of Biomedical Science*, 26. doi:10.1186/s12929-019-0523-z.

Yu, M. *et al.* (2013) 'Circulating Breast Tumor Cells Exhibit Dynamic Changes in Epithelial and Mesenchymal Composition', *Science (New York, N.Y.)*, 339(6119), pp. 580–584. doi:10.1126/science.1228522.

Yu, X. *et al.* (2020) 'The circular RNA circMAST1 promotes hepatocellular carcinoma cell proliferation and migration by sponging miR-1299 and regulating CTNND1 expression', *Cell Death & Disease*, 11(5), p. 340. doi:10.1038/s41419-020-2532-y.

Zafarana, G. *et al.* (2009) ‘Specific Knockdown of OCT4 in Human Embryonic Stem Cells by Inducible Short Hairpin RNA Interference’, *Stem Cells (Dayton, Ohio)*, 27(4), pp. 776–782. doi:10.1002/stem.5.

Zahedi, A. *et al.* (2018) ‘Epithelial-to-mesenchymal transition of A549 lung cancer cells exposed to electronic cigarettes’, *Lung Cancer (Amsterdam, Netherlands)*, 122, pp. 224–233. doi:10.1016/j.lungcan.2018.06.010.

Zakrzewski, W. *et al.* (2019) ‘Stem cells: past, present, and future’, *Stem Cell Research & Therapy*, 10(1), p. 68. doi:10.1186/s13287-019-1165-5.

Zeisberg, M. and Neilson, E.G. (2009) ‘Biomarkers for epithelial-mesenchymal transitions’, *The Journal of Clinical Investigation*, 119(6), pp. 1429–1437. doi:10.1172/JCI36183.

Zeisberg, M., Shah, A.A. and Kalluri, R. (2005) ‘Bone Morphogenic Protein-7 Induces Mesenchymal to Epithelial Transition in Adult Renal Fibroblasts and Facilitates Regeneration of Injured Kidney’, *Journal of Biological Chemistry*, 280(9), pp. 8094–8100. doi:10.1074/jbc.M413102200.

Zeng, K. *et al.* (2018) ‘The pro-metastasis effect of circANKS1B in breast cancer’, *Molecular Cancer*, 17(1). doi:10.1186/s12943-018-0914-x.

Zhang, H. *et al.* (2017) ‘Increased circular RNA UBAP2 acts as a sponge of miR-143 to promote osteosarcoma progression’, *Oncotarget*, 8(37), pp. 61687–61697. doi:10.18632/oncotarget.18671.

Zhang, J. *et al.* (2019) ‘Circular RNA profiling provides insights into their subcellular distribution and molecular characteristics in HepG2 cells’, *RNA Biology* [Preprint]. doi:10.1080/15476286.2019.1565284.

Zhang, J. *et al.* (2020) ‘Accurate quantification of circular RNAs identifies extensive circular isoform switching events’, *Nature Communications*, 11(1), p. 90. doi:10.1038/s41467-019-13840-9.

Zhang, M. *et al.* (2018) ‘A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis’, *Oncogene*, 37(13), pp. 1805–1814. doi:10.1038/s41388-017-0019-9.

Zhang, N. *et al.* (2018) ‘Circ5379-6, a circular form of tumor suppressor PPARα, participates in the inhibition of hepatocellular carcinoma tumorigenesis and metastasis’, *American Journal of Translational Research*, 10(11), pp. 3493–3503.

Zhang, S.-C. *et al.* (2001) ‘In vitro differentiation of transplantable neural precursors from human embryonic stem cells’, *Nature Biotechnology*, 19(12), pp. 1129–1133. doi:10.1038/nbt1201-1129.

Zhang, X. *et al.* (2018) ‘circSMAD2 inhibits the epithelial–mesenchymal transition by targeting miR-629 in hepatocellular carcinoma’, *OncoTargets and therapy*, 11, pp. 2853–2863. doi:10.2147/OTT.S158008.

Zhang, X. *et al.* (2019) ‘Circular RNA Vav3 sponges gga-miR-375 to promote epithelial-mesenchymal transition’, *RNA biology* [Preprint]. doi:10.1080/15476286.2018.1564462.

Zhang, X.-O. *et al.* (2014) ‘Complementary Sequence-Mediated Exon Circularization’, *Cell*, 159(1), pp. 134–147. doi:10.1016/j.cell.2014.09.001.

Zhang, X.-O. *et al.* (2016) ‘Diverse alternative back-splicing and alternative splicing landscape of circular RNAs’, *Genome Research*, 26(9), pp. 1277–1287. doi:10.1101/gr.202895.115.

Zhang, Y. *et al.* (2013) ‘Circular Intronic Long Noncoding RNAs’, *Molecular Cell*, 51(6), pp. 792–806. doi:10.1016/j.molcel.2013.08.017.

Zhang, Y. *et al.* (2016) ‘The Biogenesis of Nascent Circular RNAs’, *Cell Reports*, 15(3), pp. 611–624. doi:10.1016/j.celrep.2016.03.058.

Zhang, Y. *et al.* (2021) ‘Optimized RNA-targeting CRISPR/Cas13d technology outperforms shRNA in identifying functional circRNAs’, *Genome Biology*, 22(1), p. 41. doi:10.1186/s13059-021-02263-9.

Zhang, Z., Yang, T. and Xiao, J. (2018) ‘Circular RNAs: Promising Biomarkers for Human Diseases’, *EBioMedicine*, 34, pp. 267–274. doi:10.1016/j.ebiom.2018.07.036.

Zhao, J. *et al.* (2019) ‘Transforming activity of an oncoprotein-encoding circular RNA from human papillomavirus’, *Nature Communications*, 10. doi:10.1038/s41467-019-10246-5.

Zhao, M. and Qu, H. (2020) ‘circVAR database: genome-wide archive of genetic variants for human circular RNAs’, *BMC Genomics*, 21(1), p. 750. doi:10.1186/s12864-020-07172-y.

Zhao, Q. *et al.* (2020) ‘Targeting Mitochondria-Located circRNA SCAR Alleviates NASH via Reducing mROS Output’, *Cell*, 183(1), pp. 76-93.e22. doi:10.1016/j.cell.2020.08.009.

Zhao, S. *et al.* (2014) ‘Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells’, *PLOS ONE*, 9(1), p. e78644. doi:10.1371/journal.pone.0078644.

Zhao, Y. *et al.* (2016) ‘Deficiency in the Ubiquitin Conjugating Enzyme UBE2A in Alzheimer’s Disease (AD) is Linked to Deficits in a Natural Circular miRNA-7 Sponge (circRNA; ciRS-7)’, *Genes*, 7(12), p. 116. doi:10.3390/genes7120116.

Zhao, Z. *et al.* (2017) ‘Hsa_circ_0054633 in peripheral blood can be used as a diagnostic biomarker of pre-diabetes and type 2 diabetes mellitus’, *Acta Diabetologica*, 54(3), pp. 237–245. doi:10.1007/s00592-016-0943-0.

Zheng, Q. *et al.* (2016) ‘Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs’, *Nature Communications*, 7, p. 11215. doi:10.1038/ncomms11215.

Zheng, X. *et al.* (2015) ‘Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer’, *Nature*, 527(7579), pp. 525–530. doi:10.1038/nature16064.

Zheng, X. *et al.* (2019) ‘A novel protein encoded by a circular RNA circPPP1R12A promotes tumor pathogenesis and metastasis of colon cancer via Hippo-YAP signaling’, *Molecular Cancer*, 18(1), p. 47. doi:10.1186/s12943-019-1010-6.

Zhou, B.P. *et al.* (2004) ‘Dual regulation of Snail by GSK-3β-mediated phosphorylation in control of epithelial–mesenchymal transition’, *Nature Cell Biology*, 6(10), pp. 931–940. doi:10.1038/ncb1173.

Zhou, C. *et al.* (2017) ‘Genome-wide maps of m6A circRNAs identify widespread and cell-type-specific methylation patterns that are distinct from mRNAs’, *Cell reports*, 20(9), pp. 2262–2276. doi:10.1016/j.celrep.2017.08.027.

Zhou, L.-H. *et al.* (2018) ‘CircRNA_0023642 promotes migration and invasion of gastric cancer cells by regulating EMT’, *European Review for Medical and Pharmacological Sciences*, 22(8), pp. 2297–2303. doi:10.26355/eurrev_201804_14818.

Zhou, W.-Y. *et al.* (2020) ‘Circular RNA: metabolism, functions and interactions with proteins’, *Molecular Cancer*, 19(1), pp. 1–19. doi:10.1186/s12943-020-01286-3.

Zhou, Y. *et al.* (2019) ‘Circular RNA hsa_circ_0004015 regulates the proliferation, invasion, and TKI drug resistance of non-small cell lung cancer by miR-1183/PDPK1 signaling pathway’, *Biochemical and Biophysical Research Communications*, 508(2), pp. 527–535. doi:10.1016/j.bbrc.2018.11.157.

Zhu, Y.-J. *et al.* (2019) ‘Circular RNAs negatively regulate cancer stem cells by physically binding FMRP against CCAR1 complex in hepatocellular carcinoma’, *Theranostics*, 9(12), pp. 3526–3540. doi:10.7150/thno.32796.