

Open Research Online

The Open University's repository of research publications and other research outputs

Discovery of Novel Hub Genes Across Breast, Ovarian, and Prostate Cancers

Thesis

How to cite:

Chandler, Stephen (2022). Discovery of Novel Hub Genes Across Breast, Ovarian, and Prostate Cancers. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2022 Stephen Chandler



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.00014bce>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk



School of life, health, and chemical sciences

Discovery of novel hub genes across breast, ovarian, and prostate cancers

Stephen Chandler BSc, MSc

18th July 2022

Publications

Conference proceedings

Matheou, Chrysanthi; Chandler, Stephen; Crea, Francesco; Faria, Nuno; Powell, Jonathan; McDonald, Fraser; Rietdorf, Katja and Bootman, Martin D. (2016). Mibefradil; a T-type calcium channel blocker as a potential anti-cancer agent. In: 14th International Meeting of the European Calcium Society, 25-29 Sep 2016, Valladolid, Spain.

Chandler, Stephen; Crea, Francesco; Hirst, Mark; Cox, Angela and Rigas, Sushilaben (2017). Identification of putative functional genes in breast and other cancers with potentially shared aetiology. In: The Human Genome in Healthcare, 23-24 Nov 2017, The Royal Society, London, UK.

Chandler, Stephen; Crea, Francesco; Hirst, Mark; Cox, Angela and Rigas, Sushila (2018). Identification of candidate genes that may function in ER- breast cancer. In: BACR Conference: Response and Resistance in Cancer Therapy, 10-12 Sep 2018, University of Kent, Canterbury.

Chandler, Stephen; Gangadharannambiar, Priyadarsini; Crea, Francesco; Hirst, Mark; Cox, Angela and Rigas, Sushila (2019). Discovery of candidate hub genes in breast cancer. In: New Developments in Breast Cancer Research - From the Lab to the Clinic, 2019 BACR Special Conference, 9-11 Oct 2019, The Sage, Newcastle Gateshead, UK.

Chandler, Stephen; Gangadharannambiar, Priyadarsini; Crea, Francesco; Hirst, Mark; Cox, Angela and Rigas, Sushila (2019). Discovery of candidate long non-coding RNAs as biomarkers and treatment targets for prostate cancer. In: NEPC 2019. Treatment-emergent neuroendocrine prostate cancer: clinical relevance and molecular pathogenesis., 25 Jun 2019, The Open University, Milton Keynes, UK.

Journal Publications

Silvestri, Roberto; Pucci, Perla; Venalainen, Erik; Matheou, Chrysanthi; Mather, Rebecca; Chandler, Stephen; Aceto, Romina; Rigas, Sushila; Wang, Yuzhuo; Rietdorf, Katja; Bootman, Martin and Crea, Francesco (2019). T-type calcium channels drive the proliferation of androgen-receptor negative prostate cancer cells. *The Prostate*, 79(13) pp. 1580–1586.

Additional Publications

Chandler, Stephen; Rigas, Sushila (2017). Big data and Bioinformatics: Powerful tools for decoding DNA. Open learn: Tomorrows World. Available at: <https://www.open.edu/openlearn/science-maths-technology/biology/big-data-and-bioinformatics-powerful-tools-decoding-dna>

Acknowledgements

I would like to thank the Open university for the opportunity to develop and grow as a scientific researcher. My supervisors Dr Sushila Rigas, Dr Mark Hirst, Dr Francesco Crea, and Professor Angela Cox for their help, support, encouragement, and guidance throughout my PhD. This work could not have been completed without them.

To Dr Sushila Rigas I would like thank in particular. I consider myself incredibly lucky to have had such an encouraging supervisor who was never too busy to help. You gave me an opportunity to achieve something not many are given and for that I am incredibly grateful.

I would also like to particularly thank Priyadarsini Gangadharannambiar for her help with collecting the literature of the functions of the hub genes. It is my hope that she will continue in research in the future and will go on to discover many great things.

Lastly to Rebeccah Greenwood, who sat up with me in the many late nights of frustrated analysis and writing. Listened to my worries and didn't let me get disheartened or give up. There are no words I can write to express my gratitude and all I can offer are these.

Contents

Publications	ii
Conference proceedings	ii
Journal Publications	ii
Acknowledgements.....	iii
Table of Figures/Tables.....	vii
Abstract	xii
Chapter 1. Introduction	1
1.1 Breast, ovarian and prostate cancer incidence and mortality.....	5
1.2 Risk factors for breast, ovarian and prostate cancers	9
1.2.1 Ageing	9
1.2.2 Menopause and hormone replacement therapy (oestrogen).....	13
1.2.3 Obesity	13
1.2.4 Family history and hereditary risk	14
1.3 Classification of breast, ovarian and prostate cancers	16
1.3.1 Breast cancers	17
1.3.2 Ovarian cancers.....	24
1.3.3 Prostate cancers.....	30
1.4 Molecular pathogenesis of breast, ovarian and prostate cancers	35
1.4.1 Cancer-related pathways in breast, ovarian and prostate cancers	45
1.5 Similarities between breast, ovarian and prostate cancers	63
1.6 Bioinformatics for novel biomarker identification.....	66
1.6.1 Network analysis for cancer biomarker identification.....	66
1.6.2 Network analysis in cancer research	73
1.6.3 Hub genes	75
1.7 Literature summary, and research aims	81
1.7.1 Aims.....	82
1.7.2 Objectives.....	82
Chapter 2. Methodology.....	84
2.1 Overview of analyses and workflow	84
2.2 Curation of microarray data sets	86
2.2.1 Breast cancer data sets	87
2.2.2 Ovarian cancer data sets.....	92
2.2.3 Prostate cancer data sets.....	97
2.3 Quality control and pre-processing of data sets.....	100
2.4 Background correction, quantile normalisation and summarisation	101
2.5 Annotation of microarray platforms.....	103
2.6 Missing value imputation of microarray data sets	105

2.7	Cross-platform/data set batch correction	109
2.8	Identification of DEGs	117
2.9	Cross-cancer identification of DEGs.....	121
2.10	Survival analysis	125
2.11	Co-expression network analysis.....	129
2.11.1	Hub gene identification.....	131
2.11.2	GO analysis of co-expression networks	131
Chapter 3.	Results.....	132
3.1	Differentially expressed gene (DEG) analysis.....	132
3.1.1	Breast cancer histological DEGs	132
3.1.2	Ovarian tissue subtype analysis	136
3.1.3	Prostate cancer Gleason Grade group analysis	139
3.1.4	Breast cancer molecular subtype analysis	143
3.1.5	Ovarian cancer epithelial subtype analysis.....	148
3.1.6	Prostate cancer Gleason Score analysis.....	152
3.1.7	DEG analysis summary	158
3.2	Cross-cancer gene expression analysis	160
3.2.1	Histological cross-cancer group	161
3.2.2	Molecular cross-cancer group.....	163
3.2.3	Cross-cancer results summary	165
3.3	Novel gene identification.....	166
3.3.1	Novel cross-cancer DEGs in histological subtypes	168
3.3.2	Novel cross-cancer DEGs in molecular subtypes	169
3.3.3	Novel cross-cancer gene identification summary.....	171
3.4	Survival analysis of the three novel cross-cancer DEGs.....	172
3.4.1	Recurrence free survival (RFS)	172
3.4.2	Overall survival (OS).....	177
3.4.3	Survival analysis summary	182
3.5	Summary of DEG, cross-cancer, and novel/survival results	183
3.6	Network and hub gene analysis.....	185
3.6.1	Breast histological, ovarian tissue, and prostate Gleason grade networks.....	186
3.6.2	Breast molecular, ovarian epithelial, and prostate Gleason score networks.....	192
3.6.3	Summary of gene network and functional analyses.....	198
3.7	Functions and mechanisms of HMMR, CENPE, and STIL	200
3.7.1	HMMR – A HA receptor	200
3.7.2	CENPE – A chromosome separator.....	211
3.7.3	STIL – A centriole assembler	219
3.7.4	Combined mechanism of action	226

Chapter 4. Discussion.....	229
Aim 1: Develop a method and a workflow to integrate publicly available microarray gene expression data from breast, ovarian, and prostate cancer subtypes.	229
Aim 2: Use this approach to identify differentially expressed genes, and of those, further identify novel candidate hub genes common across the cancer subtypes.	233
Aim 3: Determine whether the novel candidate hub genes identified, may be useful as potential biomarkers and/or treatment targets for these cancers.....	234
<i>HMMR</i> gene	234
<i>CENPE</i> gene	237
<i>STIL</i> gene	239
<i>HMMR</i> , <i>CENPE</i> and <i>STIL</i> combined gene targeting.....	239
Potential impact.....	240
Chapter 5. Conclusions	243
Chapter 6. Future work.....	244
Appendices	245
References	371

Table of Figures/Tables

Figure 1. Cancer incidence (blue) and mortality (red) globally in 2020. (A) Women, (B) Men.	6
Figure 2 Cancer incidence and mortality different regions of the world.	8
Figure 3. Breast, Ovarian and Prostate cancer incidence with age.	11
Figure 4. Stages of ductal breast cancer progression.	18
Figure 5. Stages of lobular breast cancer progression.	20
Figure 6. Molecular breast cancer classifications.	21
Figure 7. Epithelial to mesenchymal transition (EMT).	23
Figure 8. Breast cancer molecular subtypes and proportion of overall survival over time.	24
Figure 9. Patient survival in ovarian cancer epithelial subtypes.	30
Figure 10. Prostate cancer Gleason grades and structure.	31
Figure 11. Prostate cancer Gleason grade groups and overall survival probability.	34
Figure 12. ER signalling pathways that result in altered transcription and gene expression.	47
Figure 13. PGR pathway.	51
Figure 14. PI3K/AKT pathway in cell proliferation.	56
Figure 15. The MAPK/ERK pathway.	58
Figure 16. Apoptosis pathways.	60
Figure 17. Flowchart depicting objectives and stages of the work to achieve the aims of the study.	85
Figure 18. Example of the normalisation steps performed for microarray data using robust median averaging. (a) Quantile normalisation of microarray, (b) Median polish summarisation.	102
Figure 19. Microarray probe intensity quality control.	103
Figure 20. Normalised root mean square error (NRMSE) equation.	107
Figure 21. PCA of breast, ovarian and prostate cancer data sets showing pre-cross-platform normalised samples for each of the data sets.	111
Figure 22. PCA of breast, ovarian and prostate cancer data sets showing post-cross-platform normalised samples for each of the data sets.	113
Figure 23. MDS of pre-ComBat batch-corrected gene expression matrices.	115
Figure 24. MDS of post-ComBat batch-corrected gene expression matrices.	116
Figure 25. Summary of the network construction and visualisation analysis.	129
Figure 26. Log ₂ Fold-change gene expression values for genes HMMR, CENPE, and STIL in breast histological, ovarian tissue, and prostate Gleason group subtypes.	168
Figure 27. The Log ₂ Fold-change gene expression values for genes HMMR, CENPE, and STIL in breast molecular, ovarian epithelial, and prostate Gleason score subtypes.	170
Figure 28. Kaplan Meier RFS survival curves in breast cancer primary tumour samples.	173
Figure 29. Kaplan Meier RFS survival curves in ovarian cancer primary tumour samples.	174
Figure 30. Kaplan Meier RFS survival curves in prostate cancer primary tumour samples.	175
Figure 31. Kaplan-Meier OS survival curves in breast cancer primary tumour samples.	178
Figure 32. Kaplan Meier OS survival curves in ovarian cancer primary tumour samples.	179
Figure 33. Kaplan Meier OS survival curves in prostate cancer primary tumour samples.	180
Figure 34. Co-expression networks for HMMR, CENPE, and STIL.	188
Figure 35. Co-expression networks for HMMR, CENPE, and STIL.	193
Figure 36. Proposed mechanism of HMMR regulated apicobasal polarity and microtubule formation.	203
Figure 37. Proposed mechanisms of action for HMMR in cell proliferation.	205
Figure 38. HMMR alterations to the tumour microenvironment.	208
Figure 39. HMMR promotes EMT phenotype via utilising TGF- β receptor and RTKs.	210
Figure 40. Summary of CENPE function in microtubule dynamics.	217
Figure 41. Centrosome duplication and genomic instability.	221
Figure 42. STIL promotes centriole formation in the S-G ₂ phase prior to G ₂ -M phase transition.	222

Figure 43. Proposed mechanism of STIL.	224
Figure 44. Summary of the proposed effect of hub genes overexpression as encoded proteins HMMR, CENPE, and STIL in the development and progression of breast, ovarian, and prostate cancer.....	228
Figure 45. Volcano plot of breast cancer histological subtype IDC tumours.....	245
Figure 46. Volcano plot of breast cancer histological subtype ILC tumours.....	246
Figure 47. Volcano plot of breast cancer histological subtype DCIS tumours.	247
Figure 48. Volcano plot of ovarian cancer tissue location ovarian tumours.	248
Figure 49. Volcano plot of ovarian cancer tissue location fallopian tumours.	249
Figure 50. Volcano plot of ovarian cancer tissue location peritoneum tumours.	250
Figure 51. Volcano plot of prostate cancer Gleason grade 1 tumours.	251
Figure 52. Volcano plot of prostate cancer Gleason grade 2 & 3 tumours.	252
Figure 53. Volcano plot of prostate cancer Gleason grade 4 tumours.	253
Figure 54. Volcano plot of prostate cancer Gleason grade 5 tumours.	254
Figure 55. Volcano plot of breast cancer molecular subtype normal like tumours.	255
Figure 56. Volcano plot of breast cancer molecular subtype luminal A tumours.	256
Figure 57. Volcano plot of breast cancer molecular subtype luminal B tumours.	257
Figure 58. Volcano plot of breast cancer molecular subtype HER2 tumours.....	258
Figure 59. Volcano plot of breast cancer molecular subtype basal tumours.	259
Figure 60. Volcano plot of ovarian cancer epithelial subtype serous tumours.	260
Figure 61. Volcano plot of ovarian cancer epithelial subtype endometrioid tumours.....	261
Figure 62. Volcano plot of ovarian cancer epithelial subtype mucinous tumours.	262
Figure 63. Volcano plot of ovarian cancer epithelial subtype clear cell tumours.....	263
Figure 64. Volcano plot of prostate cancer Gleason score 4 tumours.	264
Figure 65. Volcano plot of prostate cancer Gleason score 5 tumours.	265
Figure 66. Volcano plot of prostate cancer Gleason score 6 tumours.	266
Figure 67. Volcano plot of prostate cancer Gleason score 7 tumours.	267
Figure 68. Volcano plot of prostate cancer Gleason score 8 tumours.	268
Figure 69. Volcano plot of prostate cancer Gleason score 9 tumours.	269
Figure 70. IDC breast tumour networks for HMMR, CENPE, and STIL.....	321
Figure 71. ILC breast tumour networks for HMMR, CENPE, and STIL.....	322
Figure 72. DCIS breast tumour networks for HMMR, CENPE, and STIL.....	323
Figure 73. Ovarian tissue tumour networks for HMMR, CENPE, and STIL.	324
Figure 74. Fallopian tissue tumour networks for HMMR, CENPE, and STIL.	325
Figure 75. Peritoneum tissue tumour networks for HMMR, CENPE, and STIL.....	326
Figure 76. Gleason grade 1 tumour networks for HMMR, CENPE, and STIL.	327
Figure 77. Gleason grade 2 & 3 tumour networks for HMMR, CENPE, and STIL.....	328
Figure 78. Gleason grade 4 tumour networks for HMMR, CENPE, and STIL.	329
Figure 79. Gleason grade 5 tumour networks for HMMR, CENPE, and STIL.	330
Figure 80. Normal-like breast tumour networks for HMMR, CENPE, and STIL.	331
Figure 81. Luminal A breast tumour networks for HMMR, CENPE, and STIL.	332
Figure 82. Luminal B breast tumour networks for HMMR, CENPE, and STIL.	333
Figure 83. HER2 breast tumour networks for HMMR, CENPE, and STIL.....	334
Figure 84. Basal breast tumour networks for HMMR, CENPE, and STIL.....	335
Figure 85. Serous tumour networks for HMMR, CENPE, and STIL.....	336
Figure 86. Endometrioid tumour networks for HMMR, CENPE, and STIL.	337
Figure 87. Mucinous tumour networks for HMMR, CENPE, and STIL.....	338
Figure 88. Clear cell tumour networks for HMMR, CENPE, and STIL.....	339
Figure 89. Gleason score 5 tumour networks for HMMR, CENPE, and STIL.....	340
Figure 90. Gleason score 6 tumour networks for HMMR, CENPE, and STIL.....	341

Figure 91. Gleason score 7 tumour networks for HMMR, CENPE, and STIL.	342
Figure 92. Gleason score 8 tumour networks for HMMR, CENPE, and STIL.	343
Figure 93. Gleason score 9 tumour networks for HMMR, CENPE, and STIL.	344
Figure 94. Expanded network of HMMR in breast tumour samples.	345
Figure 95. Expanded network of HMMR in Ovarian tumour samples.	346
Figure 96. Expanded network of HMMR in Prostate tumour samples.	347
Figure 97. Expression of CDC20 and BUBR1 in breast histological, ovarian tissue, and prostate Gleason grade group subtypes.	364
Figure 98. Expression of CDC20 and BUBR1 in breast molecular, ovarian epithelial, and prostate Gleason score subtypes.	365
Figure 99. Kaplan-Meier OS survival curves of HSC in lung cancer primary tumour samples.	369
Figure 100. Kaplan-Meier OS survival curves of HSC in gastric cancer primary tumour samples.	370
Table 1. Histological breast cancer subtype classification.	17
Table 2. Current Gleason scores and their equivalent Gleason grade groups.	34
Table 3. Most common cancer-associated genes and the approximate number of known pathogenic mutations across breast, ovarian and prostate cancers.	36
Table 4. The three most common network types used to construct networks.	70
Table 5. Network centrality measures.	72
Table 6. A summary of papers published on hub genes.	79
Table 7. Breast cancer histological and molecular studies, including normal samples identified from the GEO database.	88
Table 8. Breast cancer histological subtypes and normal sample sizes.	91
Table 9. Breast cancer molecular subtypes and normal sample sizes.	92
Table 10. Ovarian cancer tissue and epithelial subtype studies, including normal samples identified from the GEO database.	93
Table 11. Ovarian cancer tissue subtype and normal sample sizes.	96
Table 12. Ovarian cancer epithelial subtype and normal sample sizes.	96
Table 13. Prostate cancer Gleason group and Gleason score studies, including normal samples identified in the GEO database.	98
Table 14. Prostate cancer Gleason score and normal samples.	99
Table 15. Prostate cancer Gleason score samples and their corresponding Gleason grades group.	100
Table 16. Affymetrix microarrays and their corresponding annotation packages.	104
Table 17. Microarray platforms probes and the number of unique target genes.	106
Table 18. The total number of missing values identified in each cancer expression matrix and the percentage of missing values identified for each of the cancer expression matrices.	109
Table 19. Differential expression analysis comparisons for breast cancer subtypes.	118
Table 20. Differential expression analysis comparisons for ovarian cancer subtypes.	119
Table 21: Differential expression analysis comparisons for prostate cancer subtypes.	120
Table 22. Cross-cancer comparisons of histological breast cancer subtypes, ovarian cancer tissue subtypes and prostate cancer Gleason groups.	122
Table 23. Cross-cancer comparisons of molecular breast cancer subtypes, ovarian cancer epithelial subtypes and prostate cancer Gleason scores.	123
Table 24. Primary patient survival data for breast, ovarian and prostate cancer.	127
Table 25. Summary numbers of DEGs identified in breast histological subtype analysis.	133
Table 26. The top 10 most significant DEGs identified in histological breast cancer subtype analysis.	134
Table 27. Summary numbers of DEGs identified in ovarian cancer tissue subtype analysis.	136
Table 28. The top 10 most significant DEGs identified in ovarian tissue location subtype analysis.	137

Table 29. Summary numbers of DEGs identified in prostate cancer Gleason grade subtype analysis.	139
Table 30. The top 10 most significant DEGs identified in prostate Gleason grade group subtype analysis.....	141
Table 31. Summary numbers of DEGs identified in breast cancer molecular subtype analysis.	144
Table 32. The top 10 most significant DEGs identified in breast cancer molecular subtype analysis.	146
Table 33. Summary numbers of DEGs identified in ovarian cancer epithelial subtype analysis.	149
Table 34. The top 10 most significant genes identified in ovarian cancer epithelial subtype analysis.....	150
Table 35. Summary numbers of DEGs identified in prostate cancer Gleason score subtype analysis.	153
Table 36. The top 10 most significant genes identified in prostate cancer Gleason score subtype analysis.....	155
Table 37. The 10 most highly significant cross-cancer DEGs in breast histological, ovarian tissue, and prostate Gleason group cancer subtypes.	162
Table 38. The 10 most highly significant cross-cancer DEGs in breast molecular, ovarian epithelial, and prostate Gleason score cancer subtypes.	164
Table 39. The log2FC and P-value of the three cross-cancer DEGs identified in the two cross-cancer analyses.....	167
Table 40. RFS P-values and hazard ratios for each of the three novel cross-cancer DEGs for breast, ovarian, and prostate cancers.	176
Table 41. OS P-values and hazard ratios for each of the three novel cross-cancer DEGs for breast, ovarian, and prostate cancers.	182
Table 42. Gene Ontology (GO) terms identified in breast histological, ovarian tissue location, and prostate Gleason grade for HMMR, CENPE, and STIL.....	189
Table 43. Gene Ontology (GO) terms identified in breast molecular, and ovarian epithelial, for HMMR, CENPE, and STIL.	193
Table 44. Gene Ontology (GO) terms identified in prostate Gleason scores for HMMR, CENPE, and STIL.	196
Table 45. Top 50 most significant DEGs in breast cancer IDC.....	270
Table 46. Top 50 most significant DEGs in breast cancer DCIS.....	272
Table 47. Top 50 most significant DEGs in breast cancer ILC.	274
Table 48. Top 50 most significant DEGs in ovarian cancer ovarian tissue.....	276
Table 49. Top 50 most significant DEGs in ovarian fallopian.....	278
Table 50. Top 50 most significant DEGs in ovarian peritoneum.....	280
Table 51. Top 50 most significant DEGs in Prostate Gleason grade group 1.....	282
Table 52. Top 50 most significant DEGs in prostate cancer Gleason grade group 2 & 3.....	284
Table 53. Top 50 most significant DEGs in prostate cancer Gleason grade group 4.	286
Table 54. Top 50 most significant DEGs in prostate cancer Gleason grade group 5.	288
Table 55. Top 50 most significant DEGs in breast cancer molecular normal like.	290
Table 56. Top 50 most significant DEGs in breast cancer luminal A.....	292
Table 57. Top 50 most significant DEGs in breast cancer luminal B.	294
Table 58. Top 50 most significant DEGs in breast molecular HER2.	296
Table 59. Top 50 most significant DEGs in breast molecular Basal.	298
Table 60. Top 50 most significant DEGs in ovarian cancer serous.....	300
Table 61. Top 50 most significant DEGs in ovarian cancer endometrioid.	302
Table 62. Top 50 most significant DEGs in ovarian cancer mucinous.....	304
Table 63. Top 50 most significant DEGs in ovarian cancer clear cell.	306
Table 64. Top 50 most significant DEGs in prostate Gleason score 4.....	308

Table 65. Top 50 most significant DEGs in prostate Gleason score 5.....	310
Table 66. Top 50 most significant DEGs in prostate Gleason score 6.....	312
Table 67. Top 50 most significant DEGs in prostate Gleason score 7.....	314
Table 68. Top 50 most significant DEGs in prostate Gleason score 8.....	316
Table 69. Top 50 most significant DEGs in prostate cancer Gleason score 9.	318
Table 70. HMMR, CENPE, and STIL expression in histological (and equivalent) subtype comparison.	320
Table 71. HMMR, CENPE, and STIL expression in molecular (and equivalent) subtype comparison.	320
Table 72. Expression of CD44 in cancer subtypes.....	348
Table 73. Expression of BRCA1 in cancer subtypes	350
Table 74. Expression of CXCR4 in cancer subtypes.....	352
Table 75. Expression of TGF- β in cancer subtypes.....	354
Table 76. Expression of HAS1 in cancer subtypes.....	356
Table 77. Expression of HAS2 in cancer subtypes.....	358
Table 78. Expression of TNF- α in cancer subtypes	360
Table 79. Expression of MMP9 in cancer subtypes	362
Table 80. Expression of CEP135 in cancer subtypes	366
Table 81. Expression of CENPJ in cancer subtypes	368

Abstract

Breast, ovarian, and prostate cancers are amongst the most prevalent diagnosed cancers annually. These cancers are classified into subtypes based on specific characteristics, however, there are some common underlying aetiological features such as hormone receptors. Few studies have investigated such common features. Here, breast, ovarian and prostate cancers were analysed to identify cross-cancer gene/s that could potentially be used as biomarkers and/or treatment targets. A novel method/workflow using R programming was developed to integrate publicly available microarray tissue (RNA) data to increase sensitivity/sample numbers.

Significantly upregulated differentially expressed genes (DEGs) that were identified, were compared between the three cancers using classifications identified in literature. Two analyses were conducted comparing subtypes: 1) histological breast cancer, ovarian cancer tissue location, and prostate cancer Gleason grade group; 2) breast cancer molecular, ovarian epithelial, and prostate cancer Gleason score. Three cross-cancer significantly upregulated DEGs, *HMMR*, *CENPE*, and *STIL* were identified in both analyses: 1) *HMMR* Log₂FC = 1.37 (P = 1.29E-10), *CENPE* Log₂FC = 1.12 (P = 6.62E-08), and *STIL* Log₂FC = 1.07 (P = 4.38E-08); 2) *HMMR* Log₂FC = 1.64 (P = 2.23E-04), *CENPE* Log₂FC = 1.42 (P = 5.82E-05), *STIL* Log₂FC = 1.35 (P = 7.82E-06).

HMMR, *CENPE*, and *STIL* were significantly associated with reduced survival times (using 95% confidence interval) in recurrence free survival (Hazard Ratio (HR) = 1.18 - 1.84, P = 1E-16 - 0.012), and in overall survival (HR = 1.29 - 2.10, P = 7.1E-03 - 5.2E-09).

Network analysis identified *HMMR*, *CENPE*, and *STIL* as hub genes. Gene ontology (GO) and literature search identified them as functioning in G₂/M phase cell cycle transition, indicating a shared mechanism/s of action. Therefore, the hub genes (*HMMR*, *CENPE*, *STIL*) termed 'HCS three-gene signature' were identified as potential biomarkers and treatment targets in breast, ovarian, or prostate cancers, regardless of subtype.

Chapter 1. Introduction

The number of cancer cases diagnosed each year is on the rise globally, and cancer is one of the leading causes of mortality. This brings with it new and continuing challenges for research institutions and clinicians that aim to alleviate increasing pressure on healthcare systems. The identification of new diagnostics as well as the development of improved treatments is the best method for reducing this pressure and for lowering cancer mortality. To achieve this goal, research efforts have been aimed at earlier cancer detection, improved classification and more efficient treatments. This strategy aims to provide personalised therapies and to increase survival rates in as large a proportion of patients as possible.

In this research project, three of the most prevalent and lethal malignancies will be investigated: breast, ovarian and prostate cancers. In females, breast cancer has the highest incidence with ~2.26 million new annual cases worldwide (Ferlay *et al.*, 2020). Male breast cancers are much rarer than female cases, accounting for less than 1% of all breast cancer cases globally (Yalaza, Inan and Bozer, 2016). In contrast to breast cancer, the incidence of ovarian cancer is much lower: this malignancy is the eighth most common female cancer with ~314,000 new cases annually. However, the mortality for ovarian cancer (684,996 mortalities per annum, 30% mortality rate) is double that of breast cancer (207,252 mortalities per annum, 66% mortality rate). In men, prostate cancer accounts for ~1.41 million new cases annually (375,304 annual deaths), and is ranked second in incidence among all males cancers (lung cancer being first) (Ferlay *et al.*, 2020).

Due to the problem of increasing cases of breast, ovarian and prostate cancer annually, the identification of new biomarkers and treatment targets is more vital than ever. These three cancers share cellular pathway alterations that promote cancer initiation, progression and treatment resistance (Elledge *et al.*, 1995; Fernandez-Cuesta *et al.*, 2011; Zhang *et al.*, 2016b; Chappell *et al.*, 2012). Biologically, these malignancies are characterised by the activation of hormone-dependent signalling pathways. Both the oestrogen receptor (ER) and androgen receptor (AR) promote carcinogenesis in these three cancers (Shen *et al.*, 2017; Elledge and

Osborne, 1997; Lafront *et al.*, 2020), albeit with different biological roles in each malignancy. This suggests that these cancers may harbour further similarities in their genetic and molecular aetiology. The identification of molecular similarities among these three cancers may enable the development of treatments and/or diagnostic methods that can be applied to more than one malignancy.

Previous research has focused on therapeutic targets and biomarkers that were specifically expressed by one cancer type (e.g. prostate cancer) or even sub-type (e.g. neuroendocrine prostate cancer). This has led to biomarkers and treatments that are employed in narrow clinical settings. Despite this, there is evidence that some current treatments are effective in multiple cancers or in several subtypes (groups of a cancer defined by specific characteristics). These cancers will usually share an underlying signature of genes that may be aberrantly expressed (either increased or decreased). For example, taxanes are used as a first-line chemotherapy treatment for ovarian cancers (Mikuła-Pietrasik *et al.*, 2019), most of which display a gene expression signature that predicts good response to this type of treatment. This predictive signature contains genes involved in cytoskeletal organisation and microtubule dynamics (Creekmore *et al.*, 2011). Recently, taxanes have been also used as a treatment for metastatic breast cancers (Gradishar, 2012), because of the identification of a similar gene expression signature in this malignancy (Dezsó *et al.*, 2014). Hence it is conceivable that gene expression signatures shared across different cancers could be used to identify patients more likely to benefit from a particular treatment.

Other cancer treatments are designed to treat specific cancer subtypes. For example, trastuzumab is used to treat human epidermal growth factor 2 positive (HER2+) breast cancers. Targeted therapies are effective, but limited to the cancer expressing the specific target. The identification of novel cross-cancer targets and biomarkers has been relatively unexplored. Cross-cancer targets may lead to more efficient diagnostic and treatment options; in this case, the therapeutic choice will not be guided by the cancer's histology, but by the gene expression profile of the specific cancer.

Exploration of this strategy will require the development of a novel method that can allow the discovery of genes expressed in multiple cancers and the selection of the most clinically-relevant genes. This could be achieved by analysing samples from multiple cancers or subtypes together. The benefit of this approach is that the genes are validated as potential biomarkers in multiple cancers simultaneously. This makes them far more useful than biomarkers (or targeted therapies) that are only relevant to one cancer or subtype alone. These genes can also be used for targeted therapies for multiple cancers following laboratory-based validations. This approach is likely to identify genes with high functional significance underlying multiple tumours or a signature of genes.

With the increasing amounts of publicly available transcriptomic data (microarray and RNA-Seq), the identification of novel biomarkers and therapeutic targets becomes more easily achievable. Curated data repositories such as the Gene Expression Omnibus (GEO) (Edgar, Domrachev and Lash, 2002), the Sequence Read Archive (SRA) (Leinonen *et al.*, 2011) and The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Research Network *et al.*, 2013) provide valuable resources for researchers to identify novel biomarkers and treatment targets. In order to analyse the data, knowledge in biology and computational languages such as R or Python is required together with access to high processing computing hardware. This is not always available in all research groups and can be a limiting factor. Various tools have been developed to allow analysis of these public data, many of which do not exploit the genetic data fully and limit the users to specific inputs and factors. Importantly, these data repositories bring with them new opportunities to increase sample sizes significantly by integrating multiple cancer studies. This allows the identification of novel genes that may be missed at lower sample size (Maleki *et al.*, 2019). Hence, the main advantage of integrating data sets is the increase in sensitivity and robustness of gene expression analyses. Very few methods or tools currently exist that are designed to achieve this.

This study focussed on breast, ovarian and prostate cancers as they are three of the most commonly diagnosed cancers both globally and in the UK, for men and women, and because there is evidence of aberrant genetic and molecular alterations shared by these tumours (a

shared aetiology). The role of new bioinformatics techniques in integrating cancer data and analysing gene expression data in a combination of new ways (e.g. considering cancer classifications with histological or molecular similarities for comparisons) to answer novel research questions. This allowed identification of novel genes as biomarkers and potential therapeutic targets in the three cancers.

1.1 Breast, ovarian and prostate cancer incidence and mortality

The incidence of breast, ovarian and prostate cancers is increasing globally. In 2018, breast, ovarian and prostate cancer global incidence was 2.09 million, 295,414 and 1.28 million cases, respectively. In 2020, these cases increased to 2.26 million, 313,959 and 1.41 million cases, respectively. Cancer mortality has also increased alongside incidence. In 2018, the mortality was: 626,679 (breast cancer), 184,799 (ovarian cancer) and 358,989 (prostate cancer). In 2020, annual mortality has increased to 684,996 (breast cancer), 207,252 (ovarian cancer) 375,304 (prostate cancer) (Sung *et al.*, 2021; Bray *et al.*, 2018).

The incidence and mortality of breast, ovarian and prostate cancer differs depending on geographical location. It also varies depending on a country's Human Development Index (HDI). HDI is the primary method used to categorise countries according to life expectancy, per capita income and education. The incidence of breast, ovarian and prostate cancer is highest for high/very high HDI countries. For breast cancer, the rate per 100,000 females is 55.9 for high/very high HDI countries, compared with only 29.7 for low/medium HDI countries (Figure 1 A). This trend is the same for ovarian cancer: the rate per 100,000 females is 7.1 for high/very high HDI countries, compared with 5.8 for low/medium HDI countries, (Figure 1 A). Similarly the incidence of prostate cancer is observed as 37.5 per 100,000 males in high/very high HDI countries and 11.3 per 100,000 in low/middle HDI countries (Figure 1 B).

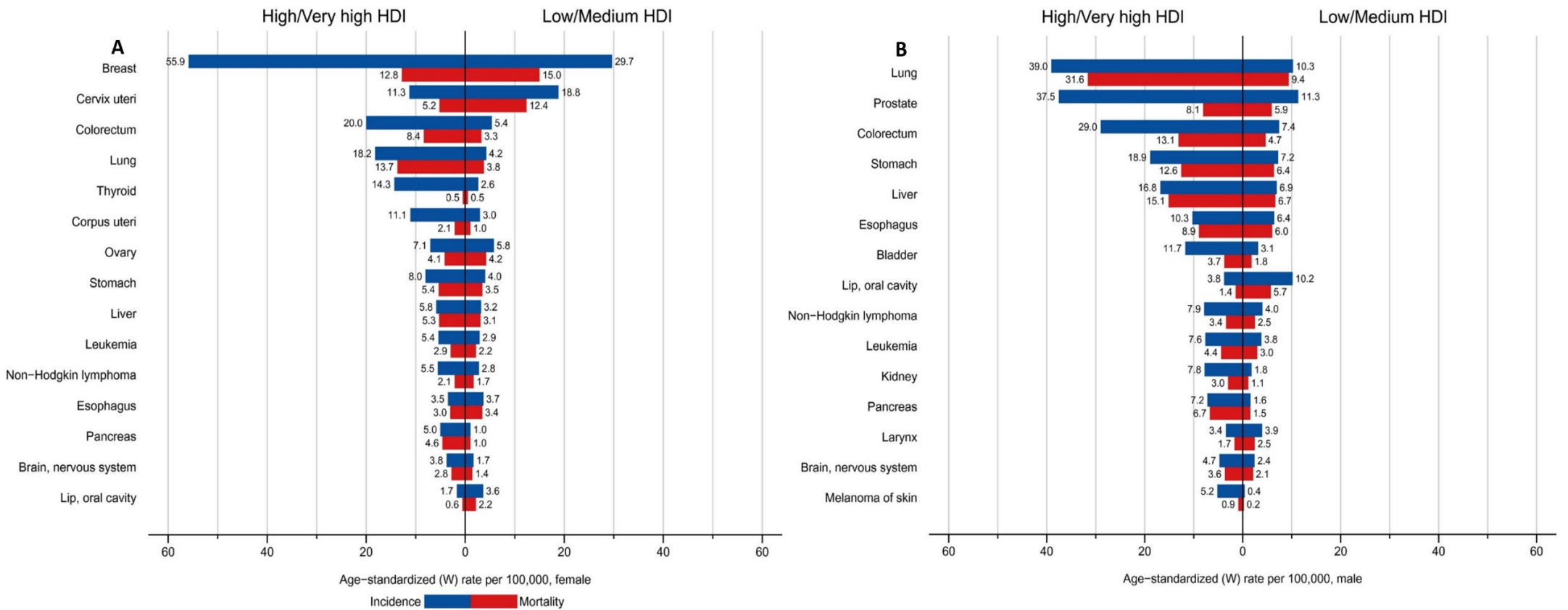


Figure 1. Cancer incidence (blue) and mortality (red) globally in 2020. (A) Women, (B) Men.

(A) In both high/very high HDI regions and low/medium HDI regions breast cancer is the most commonly diagnosed malignancy, occurring in 55.9 females per 100,000 in high/very high HDI and 29.7 females per 100,000 in low/medium HDI. Ovarian cancer has a lower incidence than breast cancer in both high/very high and low/medium HDI regions and occurs in 7.1 females per 100,000 in high/very high HDI and 5.8 females per 100,000 for low/medium HDI. (B) Prostate cancer is the second most commonly diagnosed cancer in males globally with 37.5 per 100,000 in high/very high HDI and 11.3 in low/medium HDI. From Sung et al., 2021.

HDI: Human Development Index.

Higher incidence is also observed in regions of high/very high HDI and lower incidence in regions of low/medium HDI. For example, in the Western European population (high HDI), incidence per 100,000 is 169.4 (breast cancer), 16 (ovarian cancer) 176.4 (prostate cancer) (Figure 2). In low/medium HDI regions such as Central Africa, incidence is only 19.9, 2.5 and 14.9 per 100,000 for breast, ovarian and prostate cancer respectively (Figure 2). This higher incidence rate in high/very high HDI countries may be due to the implementation of effective population screening and diagnosis initiatives, which are enabled by the presence of well-established healthcare systems leading to the population having a longer lifespan. The different incidence may also be due to the reduced life expectancy in low/middle HDI countries, compared to high/very high HDI countries; all these three cancers are more prevalent in older patients.

The total number of deaths caused by these cancers is higher in some high/very high HDI countries, compared with other countries. However, a higher mortality rate (deaths per cases) is observed in most low/medium HDI populations (Figure 2); this apparent paradox is probably due to reduced access to screening in low/middle HDI countries, which leads to diagnosis of cancers at later stages/grades. This effect is potentially exacerbated by reduced access to treatment options after diagnosis. However, with the continued improvements in developing countries (earlier detection and more diverse treatment options), it is expected that there will be a further increase in cancer incidence being recorded globally (Jemal *et al.*, 2011), and hopefully a reduction in the mortality rates.

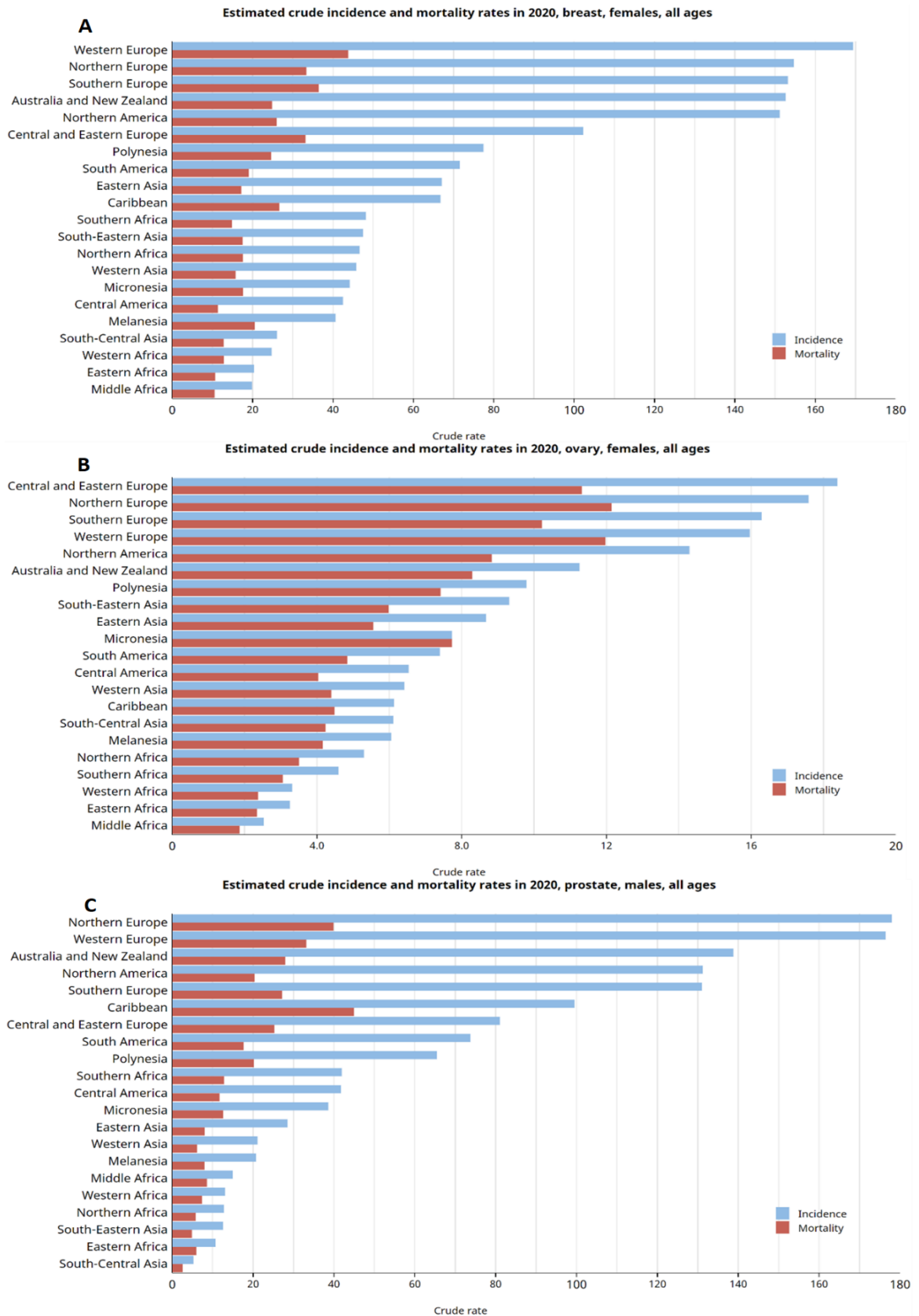


Figure 2 Cancer incidence and mortality different regions of the world.

Populations in high/very high HDI global regions (e.g. Europe) have higher crude mortality of breast, ovarian and prostate cancer with much higher incidence. In populations in low/medium HDI regions (e.g. parts of Africa and Asia), lower crude mortality is observed with much lower incidence. This results in lower mortality rate (death per cases) in high/very high HDI countries. From Sung et al., 2021.

HDI: Human Development Index.

1.2 Risk factors for breast, ovarian and prostate cancers

Breast, ovarian and prostate cancers are multifactorial diseases caused by a combination of biological and environmental factors. The onset and development of these cancers is influenced by factors such as age (Yancik *et al.*, 2001; Bechis, Carroll and Cooperberg, 2011), diet and obesity (Howe *et al.*, 1990; Lin, Aronson and Freedland, 2015), age of menopause (breast and ovarian cancer) (Key, Verkasalo and Banks, 2001; Moorman *et al.*, 2008), family history (Pharoah *et al.*, 1997; Negri *et al.*, 2003; Chen *et al.*, 2008) and exposure to environmental factors (such as carcinogens, UV radiation etc.).

Many of these risk factors will interact with one another, increasing the likelihood of cancer over time. For example, family history has been found to be associated with the onset of cancer at a younger age (Berkemeyer, Lemke and Hense, 2016; Brandt *et al.*, 2008). Family history of cancer is also strongly related to hereditary mutations. This is because hereditary mutations in cancer-promoting genes are commonly identified in those individuals with a family history of the disease (Feng *et al.*, 2018). For example, in individuals with first-degree relatives (parent, children and/or siblings) who had a diagnosis of breast or ovarian cancer before the age of 50, mutations in the breast cancer associated gene 1 (*BRCA1*) increased the risk of breast and ovarian cancer 1.2 and 1.6-fold respectively, when compared with individuals who had *BRCA1* mutation but no prior family history (Metcalf *et al.*, 2010).

1.2.1 Ageing

Age is one of the most significant risk factors for cancers in general and in the cancers discussed here. With increasing age there is an increase in cancer incidence (Thakkar, McCarthy and Villano, 2014). This trend has been observed in breast, ovarian and prostate cancer, with older age correlating with increased risk and incidence. In breast and ovarian cancer, the median age at diagnosis is 63. In prostate cancer, this is slightly higher at 67 years of age (Duggan *et al.*, 2016).

In women, the risk of breast cancer occurrence by the age of 50 is 2.4% (1 in 28 people); this rises to 4.09% (1 in 24) by the age of 70 (Howlader *et al.*, 2020). For men, the risk of prostate cancer

occurrence over the age of 65 is 60% (Rawla, 2019). Incidence of breast and prostate cancer increases up to the age range 65–74, with the majority of new cases being diagnosed in this age group (Figure 3). Unlike breast and prostate cancer, ovarian cancer has the highest incidence in the 55–64 age range, with 24.7% of all the cases occurring in this age range (Figure 3).

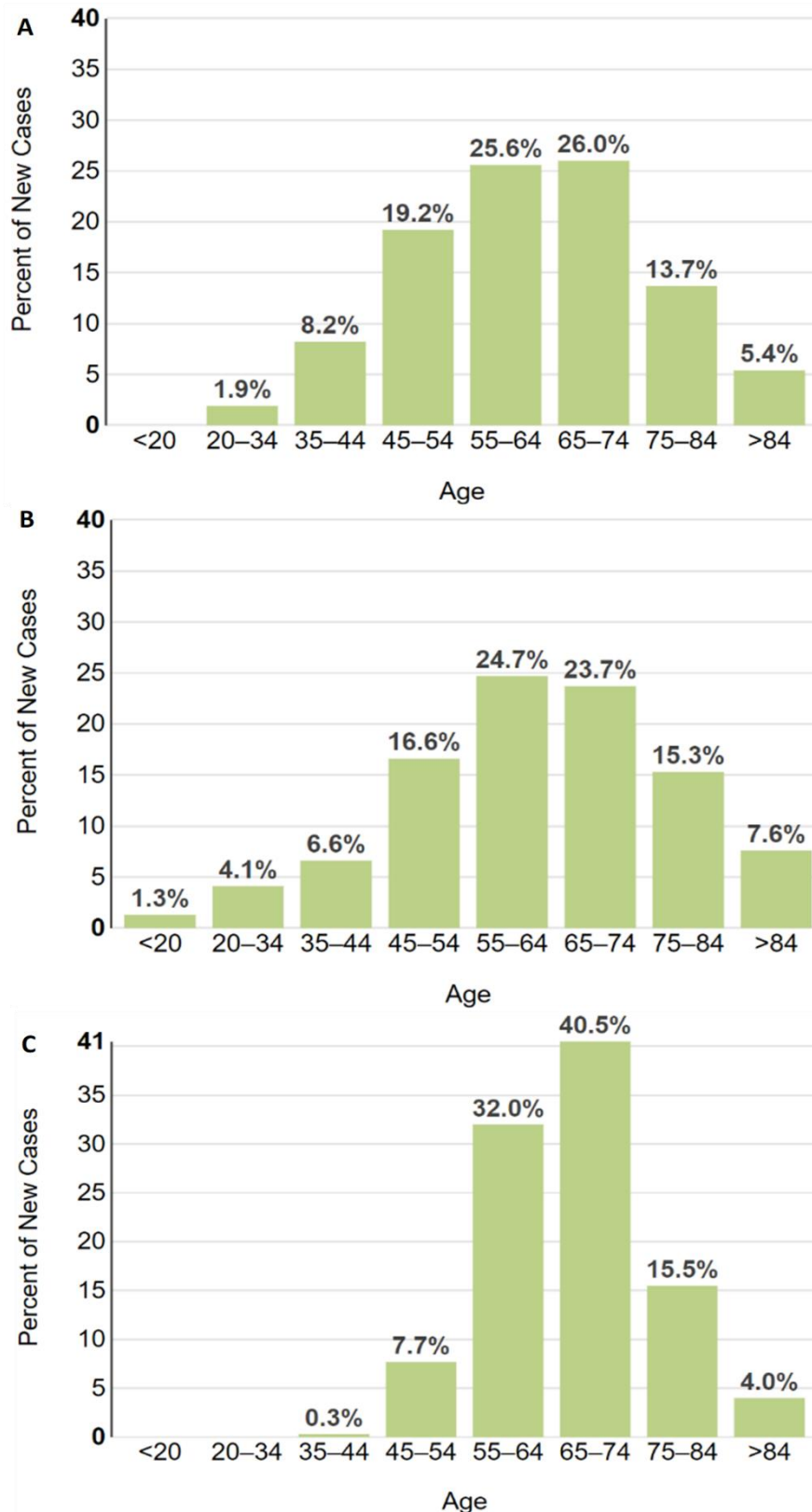


Figure 3. Breast, Ovarian and Prostate cancer incidence with age.

(A) Breast cancer incidence increases up to the age of 60+. Patients over the age of 60 account for 77% of new cases. New cases of breast cancer are observed less frequently before the age of 55. (B) Ovarian cancer incidence increases with age up to the age of 55. (C) Prostate cancer incidence increases with age up to the age of 65. There is lower incidence observed before the age of 55 for all three cancers. From National Cancer Institute., 2020, accessed February 2020.

Age is also related to family history of breast, ovarian and prostate cancer. Breast cancer has been found to have an earlier age of onset in individuals for whom a first-degree relative (mother or sister) is diagnosed with the disease, compared with those with no family history (Brandt *et al.*, 2010). This is also seen in prostate cancer, for which a first-degree relative (father or brother) being diagnosed with the disease corresponds to a younger age of onset (Ang *et al.*, 2020). In ovarian cancer, family history has been observed to increase incidence before the age of 60 (Zheng *et al.*, 2018).

As shown in Figure 3, the incidence of breast, ovarian and prostate cancer is less frequent before the age of 55. However, some subtypes of these three cancers have been observed more frequently in younger age groups, with others showing a greater number of observations in older age groups. For example, triple-negative breast cancers occur more frequently in younger adults (below the age of 40) (Bauer *et al.*, 2007; Tzikas, Nemes and Linderholm, 2020). The HER2+ breast cancer subtype is also observed frequently in the 70+ age group (Parise, Bauer and Caggiano, 2010). In prostate cancer, lower grade tumours (Gleason score <7) are observed at a younger age (below the age of 60) in those with a family history of the disease (Ang *et al.*, 2020; Pepe and Pennisi, 2015) indicating a connection between young people with cancer and the inheritance of a faulty gene. Subtypes of breast, ovarian and prostate cancers are described in more detail in section 1.3.

1.2.2 Menopause and hormone replacement therapy (oestrogen)

Exposure to the hormone oestrogen promotes the development of both ovarian and breast cancer. Hence, women with higher levels of oestrogens during their lifetime could have higher risk of developing these malignancies. Menopause involves the loss or reduction of circulating sex hormones such as oestrogen. Women who go into menopause earlier are exposed to less circulating oestrogen compared to women who go into menopause later. In keeping with this model, women who have undergone early onset menopause (before the age of 45) have a lower risk of breast cancer than premenopausal women of the same age (Monninkhof, van der Schouw and Peeters, 1999). However, earlier age of menopause is not a risk nor a protective factor for ovarian cancer (Schildkraut *et al.*, 2001).

Hormone replacement therapy (HRT) contains oestrogen to counteract the menopausal symptoms, which are partly due to reduction of oestrogen. HRT is commonly prescribed post-menopause, and has been found to increase breast and ovarian cancer risk (Schairer *et al.*, 2000; Ross *et al.*, 2000; Iversen *et al.*, 2018; Lacey *et al.*, 2002). HRT use has also been associated with increased risk of specific cancer subtypes, such as serous and endometrioid ovarian cancers (Beral *et al.*, 2015).

1.2.3 Obesity

Obesity, as measured by body mass index (BMI), has been found to be either a protective or a risk factor in breast, ovarian and prostate cancer. Postmenopausal women who have been identified as obese are at greater risk of breast and ovarian cancer than women of normal/healthy weight (Ewertz *et al.*, 2011; Beehler *et al.*, 2006). This is because the levels of oestrogen following menopause does not reduce as much as in healthy weight women due to the increased fat adipose tissue producing oestrogen. In contrast, premenopausal women with a BMI >30 have been observed to have a decreased risk of breast cancer (Schoemaker *et al.*, 2018). Similarly, in males, a BMI >30 confers a lower risk of prostate cancer (Vidal *et al.*, 2014). However, those who are diagnosed with prostate cancer with a BMI >30 are more likely to present with a more aggressive malignancy (Allott, Masko and Freedland, 2013). This is believed to be due to increased

levels of circulating IGF-1 identified in obese males where it can act similarly to the androgen receptor pathway in cases of low androgen availability and promote prostate cancer progression (Kaaks, Lukanova and Sommersberg, 2000; Freedland and Aronson, 2004).

Obesity as a risk factor is primarily due to changes in surrounding adipose tissue. Obesity has been observed to increase the number of adipocytes (hyperplasia) and cause them to enlarge (hypertrophy), thereby altering normal adipose tissue function (Fuster *et al.*, 2016). Obesity-driven changes in adipose tissue correspond to increased proinflammatory signals. This leads to faster growth and a higher grade of breast, ovarian and prostate cancer, and metastasis in all three cancers (Bousquenaud *et al.*, 2018; Arendt *et al.*, 2013; Wade and Kyprianou, 2019; Dai, Song and Di, 2020).

Obesity can also alter the levels of cancer-promoting hormones. Increases in both androgen and oestrogen hormones are associated with increased risk of cancer (Singletary and Gapstur, 2001). In breast and ovarian cancer, premenopausal oestrogen is produced by the ovaries; postmenopausal oestrogen is primarily produced by adipose tissue (Cleary and Grossmann, 2009). Changes to adipose tissue therefore increase the oestrogen availability to breast and ovarian tumours, promoting cancer initiation and progression (Cleary and Grossmann, 2009; Leitzmann *et al.*, 2009). Androgen levels are frequently decreased in obese males, but oestrogen levels have also been observed to increase. Oestrogen from adipose tissue can promote prostate cancer initiation and progression independent of androgen (Di Zazzo *et al.*, 2018). The role of hormones in the development and progression of breast, ovarian and prostate cancer is discussed in more detail later in the thesis (section 1.4.1).

1.2.4 Family history and hereditary risk

Family history was highlighted earlier in reference to the correlation between age and the incidence of breast, ovarian and prostate cancers (section 1.2.1). The relative risk (RR) of developing breast, ovarian or prostate cancers in individuals with family history of these diseases differs depending on the degree of the relative. The highest RR of breast and ovarian cancer has

been identified in those who have a history of first-degree relatives (sister, mother, daughter) with the disease (Pharoah *et al.*, 1997). This is also the case with regard to prostate cancer; those having first-degree relatives (father, son, brother) with the disease have been found to have a 2.3-fold increased risk of prostate cancer (Chen *et al.*, 2008).

A family history of breast, ovarian and prostate cancer that is observed in either first-, second- or third-degree relatives is frequently associated with hereditary mutations (Petrucci, Daly and Tuya, 2016). Hereditary breast and ovarian cancer (HBOC) is observed to have mutations in the *BRCA1* and *BRCA2* genes. This syndrome now includes prostate cancer, for which a *BRCA1/2* mutation has been identified if a family history is present. In families in which HBOC is present, there is an increased risk of breast, ovarian and also prostate cancer (Ruiz de Sabando *et al.*, 2019; Pilié *et al.*, 2017). These mutations in *BRCA1/2* are discussed in more detail in the following sections (section 1.4). In summary, heritable mutations in either *BRCA1* or *BRCA2* increase the risk of breast, ovarian and prostate cancer by 45%–65%, 20%–50% and 5%–10%, respectively (Hodgson and Turashvili, 2020; Giri and Beebe-Dimmer, 2016). This observation suggests that the pathogenesis of these three cancers shares some molecular mechanisms.

1.3 Classification of breast, ovarian and prostate cancers

Whilst breast, ovarian and prostate cancer share a number of similar characteristics, each carcinoma has its own diagnostic subtype classification system, which is based on either a histological or molecular profile. Many of these classifications are strongly associated with patients' prognosis. This makes them both highly important for diagnosis and of considerable relevance for clinicians who are deciding between different treatment options.

Histological classifications are primarily defined by the analysis of a tissue biopsy using microscopy. For example, in breast cancer, there are four main histological subtypes (discussed in more detail in the following subsections). These histological subtypes consider the location of the carcinoma, how different it is from normal tissue and if it has spread to the surrounding tissue (becoming invasive). Similar to breast cancer, ovarian and prostate cancer can be classified into different histological subtypes.

In contrast to histological subtype classifications, molecular classification systems show closer levels of similarity among different cancers. This is because the molecular classification is always based on the presence or absence of a particular tumour-associated protein. Such classifications are commonly achieved through immunohistochemical detection of receptors or other biomarkers. The results can then be used to guide optimal treatment options for the carcinoma of interest. For example, breast and ovarian carcinomas are commonly assessed for the presence of oestrogen receptor (ER) and/or the progesterone receptor (PR). In prostate cancer, the assessment is done for the androgen receptor (AR). More recently, it has been found that many of these receptors, which were initially believed to be exclusive to molecular subtypes of particular cancers, are also relevant for the molecular classifications of other cancers. One example is AR, which has been proposed for testing in breast cancers (Giovannelli *et al.*, 2018).

The most common classification systems used in breast, ovarian and prostate cancers are described in the next subsections, and their clinical relevance to prognosis is discussed.

1.3.1 Breast cancers

Histological categorisations of breast cancers

Breast histological subtypes can be split into two main groups: non-invasive and invasive. Non-invasive subtypes include ductal carcinoma *in situ* (DCIS) and lobular carcinoma *in situ* (LCIS).

Invasive subtypes include invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC) (Table 1).

Table 1. Histological breast cancer subtype classification.

Histological subtypes	Ductal	Lobular	
Preinvasive cancer	Ductal carcinoma in situ (DCIS) 80%	Lobular carcinoma in situ (LCIS) 20%	
	May spread through ducts and distort duct architecture.	Does not distort duct architecture.	
	1% progress to invasive cancer.	Same genetic abnormality as ILC with loss of E-cadherin.	
Invasive cancer	Invasive ductal carcinoma (IDC) 75%	Invasive lobular carcinoma (ILC) 15%	
	Usually from DCIS precursor.	Usually from LCIS precursor.	
	Causes fibrous response, producing a palpable mass on examination.	Minimal fibrous response, presents less often with palpable mass.	
	Metastasis through lymphatics and blood.		Metastasis through abdominal viscera to ovaries, uterus.
			Almost always ER+.

Histological subtypes are identified based on localisation within the tissue and on whether the carcinoma has invaded surrounding tissue (microinvasion). From Wong E., 2012; Duraker et al., 2020.

DCIS: ductal carcinoma in situ; LCIS: lobular carcinoma in situ; IDC: Invasive ductal carcinoma; ILC: invasive lobular carcinoma; GI: gastrointestinal; ER+: oestrogen receptor positive.

Breast carcinomas such as DCIS and LCIS are hormonally driven by oestrogen with the majority of breast carcinomas starting as hormone-dependent, i.e. requiring oestrogen for the tumour to grow (Saha Roy and Vadlamudi, 2012). Most malignancies are initially confined within the

basement membrane and are therefore called 'in situ'. Over time, undetected/untreated non-invasive breast carcinomas will progress to invasive breast carcinomas, either by accumulation of additional mutations, chromosomal alterations or via altered gene expression.

DCIS and IDC

DCIS is called 'ductal' because it originates from the ducts, i.e. tubes that move the milk from the breast to the nipple. DCIS commonly grows in the basement membrane-bound structures of breast tissue (Pinder *et al.*, 2010). Figure 4 shows the abnormal growth within the milk duct of the breast where the myoepithelial cells grow into a tumour. DCIS is the most commonly identified type of breast carcinoma being found in ~80% of all breast carcinomas; 79% of invasive breast carcinomas have been found to be IDC (Bombonati and Sgroi, 2011).

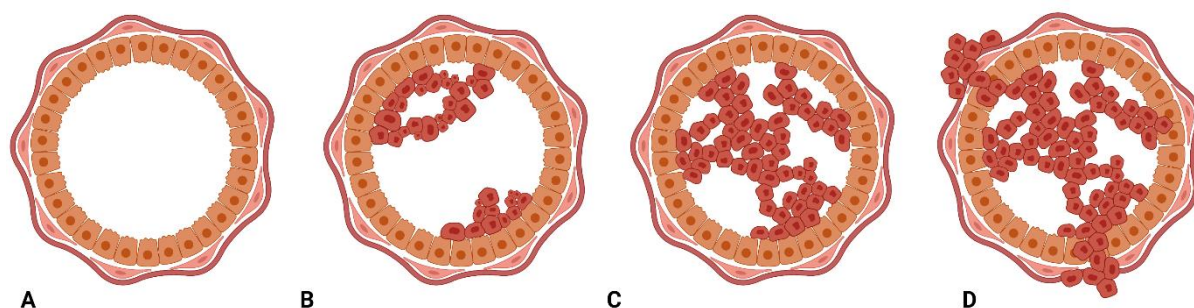


Figure 4. Stages of ductal breast cancer progression.

A) Normal breast duct consisting of outer layer of myoepithelial cells surrounding an interior of luminal cells. B) Atypical ductal hyperplasia is observed with abnormal growth within the duct but still considered a pre-cancerous lesion. C) Ductal carcinoma in situ the carcinoma is still located within the duct and therefore not invasive. D) Invasive ductal carcinoma which spreads outside the ducts into surrounding tissue (stroma).

Studies have suggested that there are two distinct forms of epithelial breast cell alterations that can progress from benign lesions to DCIS, and in some cases further to IDC. The first route is via flat epithelial atypia (FEA), which is found as an initial precursor for DCIS (Schnitt, 2003). It is considered a proliferative benign breast lesion and is associated with an increased risk of low-grade DCIS (Aulmann *et al.*, 2012). Alternatively, DCIS and IDC have been found to develop by atypical ductal hyperplasia (ADH) (Bombonati and Sgroi, 2011). ADH is observed as an intraductal proliferative lesion with similar features to low-grade DCIS (Figure 4) (Tozbikian *et al.*, 2017).

The genomic changes that drive the transition from DCIS to IDC occur partly through chromosomal alterations, which lead to changes in the normal cellular structure and function (*via* mutations and the accumulation of novel sequence variants). There have been several observed chromosomal changes in both low and high-grade IDC tumours compared with DCIS. In low-grade IDC tumours, the loss of chromosome region 16q and gains in 1q, 8q and 16p are frequently observed. In high-grade IDC, tumours show losses in 11q, 13q, 18q, 1p and 8p and gains in 8q, 17q, 20q and 16p (Gronwald *et al.*, 2005).

LCIS and ILC

Lobular carcinomas originate from the lobules, which are the milk-producing structures of the breast. LCIS grows as a linear sheet or row and it is often oestrogen receptor positive (ER+), progesterone receptor positive (PR+) and HER2+ (Sawyer *et al.*, 2014). LCIS has a lower incidence than DCIS with ~20% of breast carcinomas; ~10%–14% are ILC (Table 1). Studies of the progression of LCIS to ILC have found that atypical lobular hyperplasia (ALH) is a frequent precursor of LCIS (Bombonati and Sgroi, 2011). ALH is similar to ADH described previously in that it is a premalignant proliferative lesion. The difference between ALH and ADH is that ALH occurs in the lobules of the breast rather than the ducts.

Lobular carcinomas are frequently associated with loss of the *CDH1* gene which encodes E-cadherin, a cell adhesion junction protein. Women with *CDH1* mutations also have an increased risk of invasive breast cancer (~50%) (Dossus and Benusiglio, 2015). In comparison with LCIS, ILC has a chromosomal loss in 16q, in which *CDH1* is located (Bombonati and Sgroi, 2011). The loss of E-cadherin promotes invasiveness in ILC tumours, facilitating infiltration into the surrounding stroma (Sawyer *et al.*, 2014) (Figure 5).

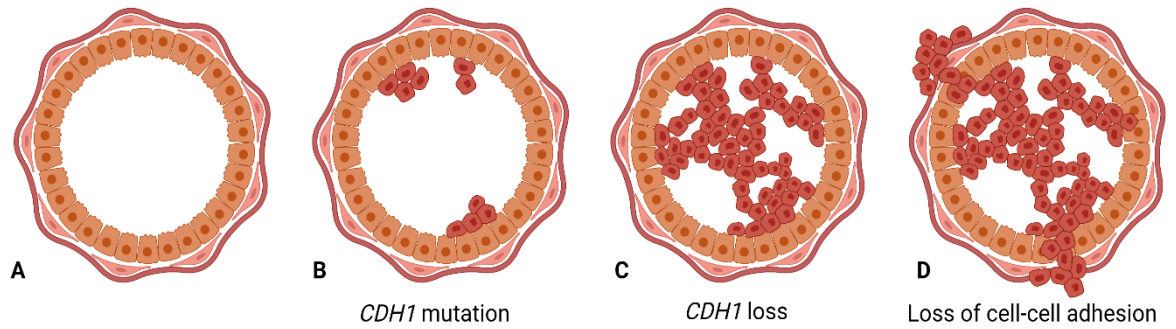


Figure 5. Stages of lobular breast cancer progression.

A) Normal breast lobule consisting of outer layer of myoepithelial cells surrounding an interior of luminal cells. B) atypical lobule hyperplasia is commonly observed with *CDH1* mutation. C) lobule carcinoma in-situ is usually classified with loss of *CDH1* and reduction in cell-cell adhesion but still contained within the lobule. D) invasive lobule carcinoma grow outside the lobule into the surrounding stromal tissue.

Molecular categorisation of breast carcinoma

The molecular classification of breast cancers differs from histological classification because molecular factors are taken into account. These molecular factors include hormone and growth receptor status and gene expression profiles. Since proliferative pathways in cells are activated via hormone receptor signalling, hormone receptor status provides valuable prognostic and therapeutic information. For example, hormone-driven breast cancers can be treated with hormone deprivation therapy (HDT), to reduce, slow or even stop the growth of the carcinoma. This therapy can also be combined with either surgery and/or chemotherapy to improve its efficacy.

Based on extensive transcriptomic profiling, breast cancers have been commonly subdivided into four main molecular subtypes: luminal A, luminal B, HER2+ and basal-like/triple negative (Figure 6) (Eliyatkin *et al.*, 2015; Pusztai *et al.*, 2006). The luminal subtypes are generally more differentiated. The HER2+ subtype is more aggressive than the luminal subtype and less differentiated. The basal type recapitulates the expression profile of undifferentiated cells.

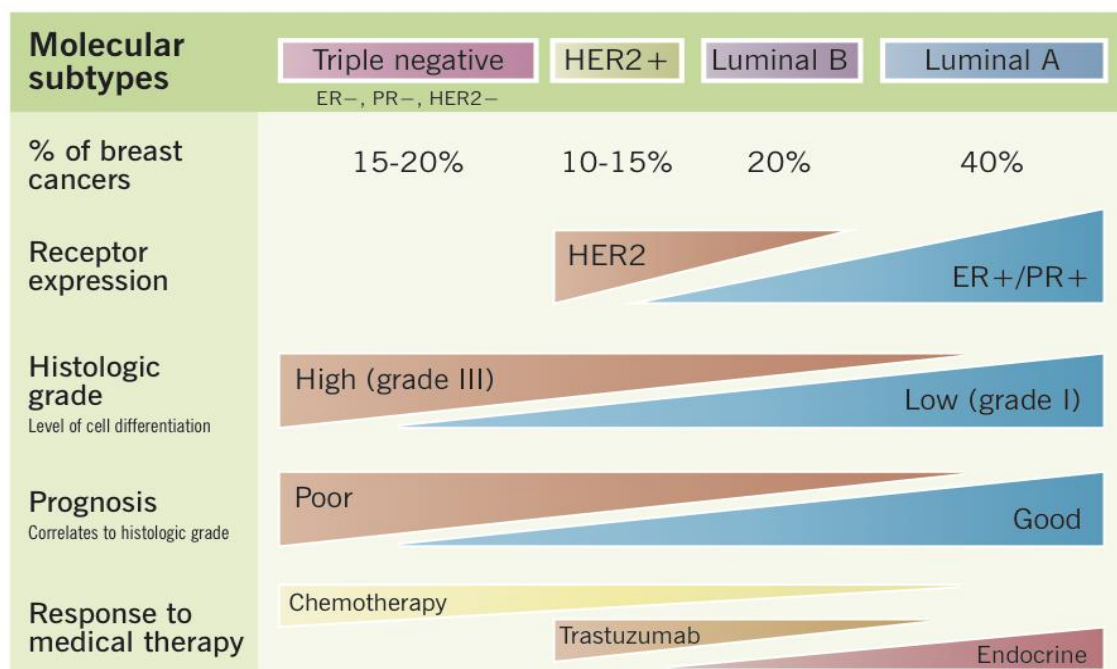


Figure 6. Molecular breast cancer classifications.

Luminal A is the most responsive to treatment, for example, to endocrine therapy. Commonly, luminal breast cancers are ER+ and PR+. HER2+ is commonly ER- and PR- but HER2+. Triple negative has the poorest prognosis and is hormone receptor negative. From Wong E., 2012.

ER-: oestrogen receptor negative; PR-: progesterone receptor negative; HER2-: human epidermal growth factor receptor 2 negative; HER2+: human epidermal growth factor receptor 2 positive; HER2: human epidermal growth factor receptor 2; ER+: oestrogen receptor positive; PR+: progesterone receptor positive.

Luminal subtypes

The luminal A subtype is defined as being ER+, PR+ and HER2-, and is identified in low-grade tumours (well- differentiated cells that resemble normal luminal cells). Incidence of the luminal A subtype has been found to be the greatest of all breast carcinomas, i.e. ~55.4% of primary breast tumours. The luminal B subtype is different to luminal A in that it can be a combination of ER+/-, PR+/- and HER2+/- cells, and is also found in low- and high-grade tumours. Incidence of luminal B has been observed as being ~11.8% of primary breast tumours (Ihemelandu *et al.*, 2007).

HER2+ breast cancer

The growth factor receptor HER2 is overexpressed in 25-30% of breast cancers (Jeong *et al.*, 2017). HER2+ breast cancer is considered more aggressive both in histological grading and prognosis, compared to luminal cancers (Menard *et al.*, 2001). The HER2+ subtype is also typically ER-, PR- and is predominantly high grade. HER2+ breast cancers account for ~11.6% of all breast carcinomas (Ihemelandu *et al.*, 2007). Current treatment for HER2+ breast cancer includes

trastuzumab (commercially called Herceptin), which is a monoclonal antibody targeting the HER2 receptor (Vu and Claret, 2012; Rubin and Yarden, 2001). This antibody prevents the activation overexpression of HER2 receptor signalling and the aberrant cell proliferation observed in HER2+ breast cancers. HER2 signalling also increases the metastatic potential of HER2 breast carcinomas, making them worse in terms of prognosis (Tan and Yu, 2007). Hence, trastuzumab may also have antimetastatic properties.

Triple negative and basal breast cancers

In rarer cases, the development of hormone-negative breast cancers can take place. These are considered the most aggressive of the breast cancer subtypes (Ovcaricek *et al.*, 2011). The basal-like subtype is sometimes referred to as triple negative due to a lack of ER, PR and HER2 expression; however, this description does not imply absent expression of any growth factor receptor. Although basal-like breast cancers have been found to be highly similar to triple negative in that they are hormone receptor negative (ER-, PR- and HER2-), they can also express the epidermal growth factor receptor (EGFR) (Eliyatkin *et al.*, 2015; Cheang *et al.*, 2008). Basal cancers frequently express some cytokeratins (CKs). These CKs include CK5, CK6, CK14, CK17 and vimentin (VIM) (Choo and Nielsen, 2010). Approximately 75% of triple negative cancers fall within basal-like phenotype breast cancers (BLBCs) and show early metastasis as well as rapid proliferation (Anders and Carey, 2009). Therefore, most basal-like cancers have an underlying triple negative phenotype. However, triple negative breast cancers do not necessarily have a basal-like phenotype, because they do not express the same set of cytokeratins (Eliyatkin *et al.*, 2015). Triple negative and basal-like subtypes are both observed more frequently in younger patients (before the age of 50) compared with other subtypes (Toft and Cryns, 2011; Dolle *et al.*, 2009). They are also more commonly found with underlying *BRCA1* gene mutations (Ressler, Mlineritsch and Greil, 2010).

Triple negative breast cancer accounts for ~21.2% of all breast cancer cases (Ihemelandu *et al.*, 2007) and for 10%–20% of all invasive breast cancers (Kumar and Aggarwal, 2016). These cancers are associated with the worst prognosis and highest mortality rate (Brouckaert *et al.*, 2012). Triple negative and basal-like cancers have low expression of the cyclin D1 gene (*CCND1*) and reduced expression of E-cadherin (Rakha *et al.*, 2007; Savagner, 2001; Li *et al.*, 2016). Loss of expression of cell adhesion proteins is associated with an increased epithelial to mesenchymal transition (EMT) (Kalluri and Weinberg, 2009) and a higher grade of tumour (Plasilova *et al.*, 2016). EMT is commonly observed in the progression of non-invasive breast carcinomas to invasive breast carcinomas. Initial downregulation of E-cadherin is frequently observed in cancer cells and is an important factor in EMT phenotype progression (Larue and Bellacosa, 2005; Kalluri and Weinberg, 2009). Hence the EMT phenotype of basal breast cancer enables their increased metastatic potential. The process of EMT is summarised in Figure 7.

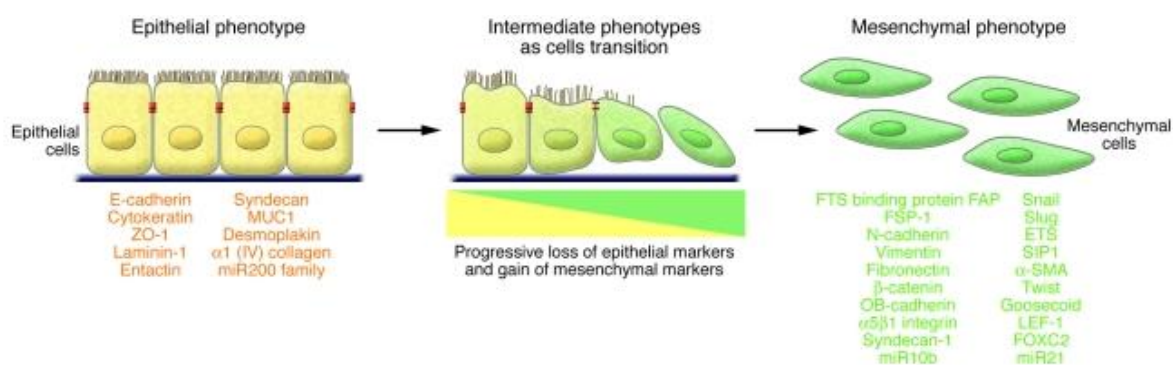


Figure 7. Epithelial to mesenchymal transition (EMT).

This figure shows the transformation of epithelial cell into a mesenchymal phenotype. A mesenchymal phenotype is associated with increased metastasis in certain breast cancers, for example, the triple negative subtype. Expression of focal adhesive genes such as E-cadherin is downregulated in a mesenchymal phenotype. Upregulation of genes such as fibronectin occurs in a mesenchymal phenotype. From Kalluri & Weinberg., 2009.

Breast cancer subtypes and prognosis

Breast cancer molecular subtypes are also useful for prognosis. This is an important factor for clinicians in guiding treatment options. Commonly, breast cancer hormone receptor-positive subtypes show better prognosis than hormone receptor-negative subtypes (Figure 8). As such, the luminal A subtype has been found to have the best prognosis in terms of survival time, followed

by luminal B and HER2. The worst prognosis is seen in triple negative and basal-like subtypes (Hennigs *et al.*, 2016; Howlader *et al.*, 2018; Fallahpour *et al.*, 2017; Cheang *et al.*, 2008).

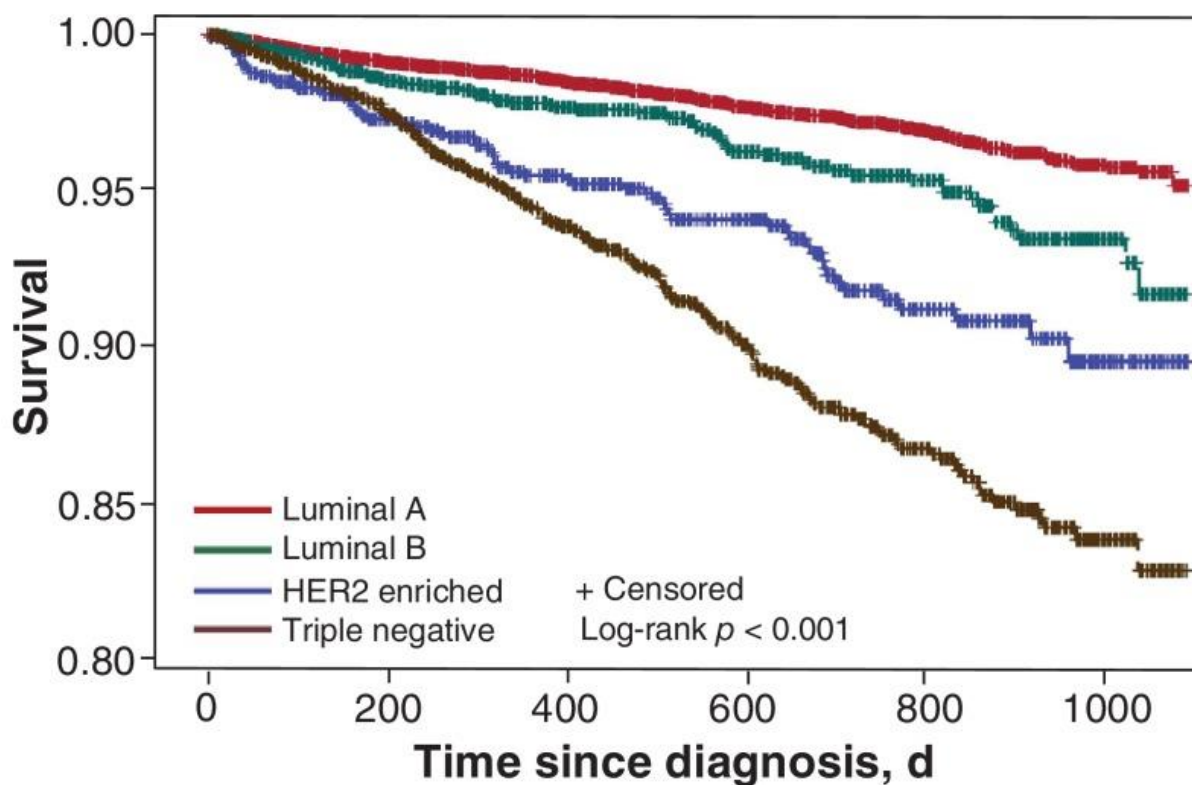


Figure 8. Breast cancer molecular subtypes and proportion of overall survival over time.

Luminal A and B show higher probability of survival compared with HER2; the lowest probability of survival is associated with the triple negative subtype. From Fallahpour *et al.*, 2017.

HER2: human epidermal growth receptor 2.

1.3.2 Ovarian cancers

Similar to breast cancer, a histological classification is used to identify the different subtypes of ovarian cancer. The ovarian cancer histological classifications have been further studied for receptor status such as ER and PR (used in breast cancer molecular classifications). However, despite this, there is no formal molecular classification system for ovarian cancers currently in clinical use. Hence the histological classification is still the most clinically relevant for ovarian cancers.

Additionally, ovarian cancers are sometimes classified by their tissue of origin, as many of these cancers migrate into ovarian tissue from elsewhere (for example, from the fallopian tubes) (Kurman and Shih, 2010). This site of origin theory means that ovarian cancers are sometimes classified according to the site of tissue origin rather than by histological subtype.

In this section, the classification of ovarian cancer tissue location subtypes and ovarian cancer histological subtypes are discussed.

Ovarian cancer classification based on tissue of location

There are three main subtypes based on the tissue location where the ovarian carcinoma was identified: ovarian, fallopian and peritoneum. The tissue of origin is important especially for determining metastatic potential. Many ovarian carcinomas may not originate in the ovaries at all; they actually develop in other tissue sites (such as fallopian or peritoneum) and migrate to the ovaries (Kurman and Shih, 2010). Consequently, a tissue location classification is now used. This classification is based on the location of the carcinoma at first diagnosis; for example, a fallopian cancer which migrated to and was discovered in the ovaries, would be labelled as ovarian tissue cancer. This classification system can be misleading due to the high migration potential of these cancers. For this reason an additional classification is used to separate cancers into epithelial subtypes.

Ovarian tissue cancers

The ovaries are the primary hormone-producing glands of the female reproductive system and produce oestrogen and progesterone. Ovarian tissue cancers are the most frequently observed subtype of ovarian cancers. This tissue subtype accounts for ~65% of identified ovarian carcinoma cases (Ehdaivand, 2012). Using receptor status, these are further classified into distinct histological ovarian cancer subtypes (discussed in more detail later). Ovarian cancers originate commonly in other tissue sites from which they migrate to the ovaries; only serous ovarian carcinomas have been observed to occur locally (Lengyel, 2010).

Peritoneum tissue cancers

The peritoneal cavity is part of the abdominal cavity that lines the surface of the ovaries. It is formed of two layers: the first is the visceral layer, which surrounds the intestines and other organs; the second is the parietal layer, which surrounds the abdominal cavity. Both layers are separated by peritoneal fluid.

Peritoneal tissue ovarian cancers are the second rarest of the ovarian tissue subtypes. A study conducted in the United States over a 10 year period found peritoneal cancer accounted for ~7 cases per million compared to ~120 cases per million for ovarian cancer (Goodman and Shvetsov, 2009). Tumours in the peritoneum tissues can metastasise to the ovaries due to the peritoneal fluid which can aid cancer cell migration (Lengyel, 2010).

Fallopian tissue cancers

The fallopian tubes connect the ovaries to the uterus, allowing the passage of eggs from the ovaries to the uterus. Fallopian tissue ovarian cancers are far rarer than other ovarian tissue subtypes; only ~4 cases per million (Goodman and Shvetsov, 2009). This is partially attributed to difficulties in diagnosis of ovarian cancers preoperatively (Rexhepi *et al.*, 2017). Many ovarian tissue cancers originate in the fallopian tubes and migrate to the ovaries, where they are diagnosed (George, Garcia and Slomovitz, 2016; Erickson, Conner and Landen, 2013). Fallopian tissue cancers are however, more frequently identified as serous carcinomas, specifically, high-grade serous carcinomas.

Histological subtype classification of ovarian cancers

Ovarian epithelial cancers occur in the surface epithelial cells (Kaku *et al.*, 2003); ~95% of ovarian cancers are epithelial ovarian cancers (Desai *et al.*, 2014). The four most common histological subtypes are serous, endometrioid, mucinous and clear cell. Classification of ovarian cancers is more frequently conducted using the epithelial subtypes rather than tissue subtype alone. This is because previous findings have shown that ovarian carcinomas have probably come from other tissue sites of origin and, therefore, it is difficult to identify their true sites of origin (George, Garcia and Slomovitz, 2016; Erickson, Conner and Landen, 2013; Kyo *et al.*, 2020). Because of this,

ovarian cancers that are classified by tissue location subtype are still graded and treated in the same manner. This has led to the tissue location subtypes being combined into one overall ovarian epithelial cancer classification system based on the histology and molecular characteristics (receptor status) of the cancer.

Serous

Serous ovarian cancer is characterised by the presence of papillary structures and marked cytologic atypia. This subtype is the most common ovarian cancer histological subtype and has been found to arise in all three tissue subtypes described previously (Lengyel, 2010); ~50% of identified ovarian cancers are of the serous subtype (Zhu *et al.*, 2006). Genetic analysis has found that high-grade serous ovarian cancer (HGSOC) of the ovaries is similar in nature to carcinomas found in fallopian and peritoneum tissue, suggesting the same source of origin (Labidi-Galy *et al.*, 2017; Kurman and Shih, 2010; Erickson, Conner and Landen, 2013; Meinhold-Heerlein *et al.*, 2016).

KRAS, *BRAF* and *ERBB2* gene mutations are commonly observed in both low-grade and high-grade serous cancers (Lengyel, 2010; Auner *et al.*, 2009; George, Garcia and Slomovitz, 2016). However, *TP53* gene mutations are rare in low-grade serous cancers, but frequent (~80%) in those that are high grade (Kurman and Shih, 2010; Erickson, Conner and Landen, 2013).

CK7 and CK8 have been identified as markers for serous cancer (Lengyel, 2010). These CKs have been found to differentiate metastatic potential, which is important in determining site of origin for ovarian cancer and prognosis. CK7 has been found to be significantly upregulated in serous ovarian cancer that has metastasised. Additionally, CK7 and CK20 are useful to distinguish between serous and mucinous carcinomas. In contrast to mucinous carcinomas, serous carcinomas were found to be CK20- and CK7+ (Kriplani and Patel, 2013; Cathro and Stoler, 2002).

Serous carcinomas also frequently harbour underlying *BRCA1* and *BRCA2* gene mutations, much like breast carcinomas. It has been observed that 15%–20% of serous carcinomas have germinal *BRCA1* and *BRCA2* mutations (George, Garcia and Slomovitz, 2016).

Endometrioid

Endometrioid ovarian carcinomas are usually derived from an endometriosis precursor when this tissue proliferates outside the uterine cavity or as an endometrioid adenofibroma tumour (fibrous benign tumour) (Ramalingam, 2016). It has been found that there is an increased number of cases of ovarian cancer (endometrioid subtype) in patients with underlying endometriosis (Vercellini *et al.*, 1993).

Endometrioid ovarian cancer accounts for ~10% of ovarian cancer cases and frequently harbours mutations in *PIK3CA*, *PTEN*, *KRAS* and *TP53* genes (Hollis *et al.*, 2020).

Clear cell

Clear cell ovarian carcinomas are characterized by clear papillary and tubulocystic pattern, and similarly to endometrioid carcinomas, are also associated with endometriosis (Ramalingam, 2016).

Clear cell cancer accounts for 5%–10% of ovarian cancer cases (Kaku *et al.*, 2003). Despite being treated with the same regimens used for serous ovarian cancers, this subtype has been found to be more resistant to platinum-based chemotherapy compared with its serous counterpart (Kuo *et al.*, 2009).

It has been observed that clear cell cancer displays a higher expression of annexin A4 protein compared with other subtypes such as serous carcinomas (Zhu *et al.*, 2006; Mogami *et al.*, 2013). Additionally, ANXA4 protein has been found to promote chemotherapy resistance in ovarian carcinomas (Mogami *et al.*, 2013).

Similar to endometrioid ovarian cancer, underlying mutations have also been observed in clear cell cancer. In a study of 69 clear cell tumours, ~33% had *PIK3CA* mutations. Other mutations observed in clear cell carcinoma are in *KRAS*, *TP53*, *PTEN* and *CTNNB1* (Kuo *et al.*, 2009).

Mucinous

Mucinous ovarian cancers are usually cystic containing high amounts of intracellular mucin (Babaier and Ghatage, 2020). Mucinous ovarian carcinomas are rarer than serous or endometrioid

types, comprising only ~5% of ovarian cancer cases (Brown and Frumovitz, 2014). They have also been observed to occur in younger women and tend to be of a lower grade (Ricci *et al.*, 2018).

CA72-4 is expressed to a high degree in mucinous subtypes compared with CA125, and is a potential biomarker for diagnosing this subtype. However, CA125 is observed more in serous carcinomas (Negishi *et al.*, 1993), and can also be used to distinguish between mucinous and serous types.

As with breast cancer and other ovarian cancers, *KRAS*-activating mutations are frequently identified in mucinous carcinomas in ~50% of cases (Lengyel, 2010; Ricci *et al.*, 2018). This is in contrast to *TP53* and *BRCA* mutations which are rare in mucinous ovarian carcinomas (Brown and Frumovitz, 2014).

Ovarian epithelial subtypes and prognosis

In ovarian epithelial subtypes, prognosis is best in endometrioid subtypes followed by serous subtypes (low-grade and high-grade). The worst prognosis is observed in clear cell and mucinous subtypes, with the mucinous subtype showing the lowest overall survival time, see Figure 9 (Winter *et al.*, 2007; Mackay *et al.*, 2010; Zhou *et al.*, 2018a). These observations correspond to findings with regard to receptor status in which endometrioid and serous cancers are frequently ER+ and PR+, and clear cell and mucinous subtypes are ER- and PR- (Chen *et al.*, 2017).

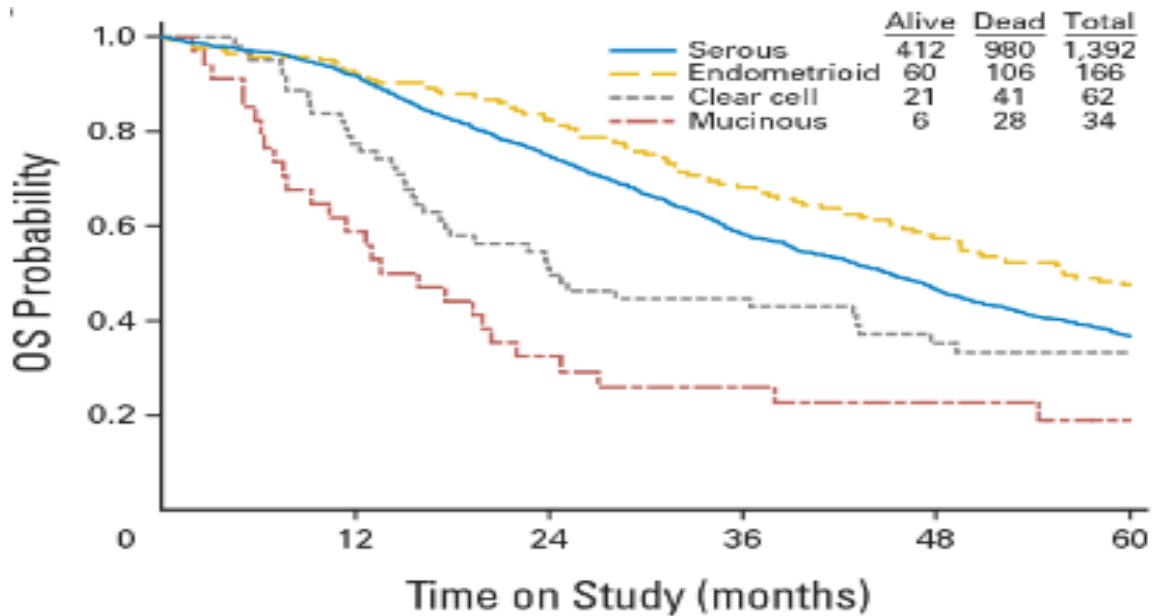


Figure 9. Patient survival in ovarian cancer epithelial subtypes.

Endometrioid and serous have the highest overall survival over time with clear cell and mucinous the lowest. From Winter et al., 2007.

OS: Overall Survival

1.3.3 Prostate cancers

Prostate carcinomas are classified primarily by the Gleason system, which is based on histological analysis of the cancer and on determining the level of differentiation from normal prostate tissue. More recently, a Gleason grade group system has been proposed to replace the older Gleason score. This is primarily aimed at providing a more consistent classification system based on the same histological information, while still matching back to the original Gleason score. It also has the benefit of creating a more uniform and consistent system for clinicians.

Gleason grading

Prostate cancers are classified based on a combined Gleason score. Here we refer to individual Gleason score as a Gleason grade for clarity. The individual Gleason grades are based on the structure/pattern of the tumour cell and glands in a specific group of cells. The more poorly differentiated the carcinoma (i.e. cells are less like normal cells), the higher the Gleason grade assigned (Figure 10).

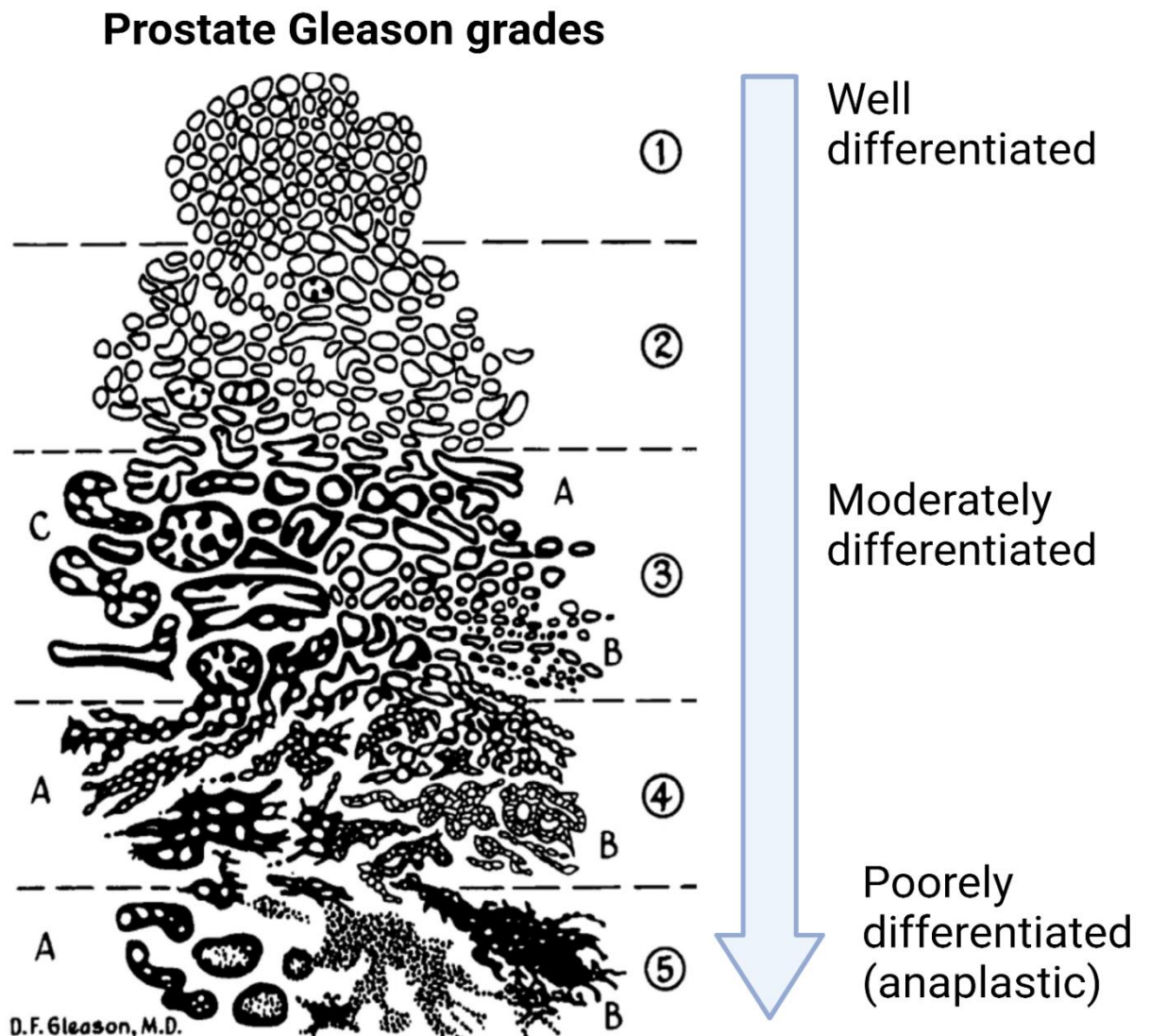


Figure 10. Prostate cancer Gleason grades and structure.

1) Gleason grade 1 are well differentiated, uniform and round glands separated by small stromal rims. 2) Gleason grade 2 are still well differentiated like grade score 1 but a larger stromal space is observed between glands and less rounded. 3) Gleason grade 3 are moderately differentiated, medium sized and infiltrate into the stroma. 4) Gleason grade 4 are poorly differentiated, infiltrative, and can be fused, cribriform or papillary in structure with few glands. 5) Gleason grade 5 are the most poorly differentiated Gleason group of infiltrative anaplastic sheets of adenocarcinoma cells with lack of occasional glands. From Humphrey., 2004.

Pathologists will examine different areas of the same sample, and identify different patterns in different areas (Figure 10). The level of differentiation is determined by assigning one Gleason grade to the most predominant pattern in a person's biopsy and another grade to the second most predominant. This is because prostate cancers are heterogenous and multifocal. The two grades will then be added together to obtain a combined Gleason score. For example, a prostate carcinoma in which the largest proportion is observed to be Gleason grade 4, and the smaller proportion Gleason grade 3 will be assigned a combined Gleason score of 7 (4 + 3).

The grades are on a scale between 3 and 5 and will form a combined score of anything up to 10, this being the most severe and have poorest prognosis. It is important to note that Gleason grades of 1 and 2 are no longer used, because they have been found to be identical to score 3.

Gleason grade 3 cancers are considered to be well differentiated; they have branching glands that form circular structures and are not fused (Iczkowski, 2019). Prostate cancers assigned to Gleason grade 4 have been identified as having fused and cribriform (multiple gaps between) glands and to form glandular structures that are less well-defined. Gleason 5 cancers appear as sheets of individual tumour cells. In some cases, they have also been found to contain necrotic cells (Gordetsky and Epstein, 2016).

In the following subsections, combined Gleason scores are discussed along with the newer Gleason grade groups. Both the Gleason score and Gleason grade are associated with prognosis.

Scores 4–7

Gleason score 6 cancers are one of the most common, being highly differentiated and more like normal prostate tissue. They are also relatively slow compared to cancers with a higher Gleason score (8 and above).

Prostate cancers with a Gleason score of 7 are considered moderately differentiated. Gleason scores of 7 or less are considered to be better prognostically. It is important to know what combination of Gleason score patterns form the majority of the tumour. For example, the treatment course for a 3 + 4 tumour (predominantly grade 3) will be different to a 4 + 3 tumour

(predominantly grade 4), despite both being considered Gleason score 7. Tumours with a Gleason score of 4 + 3 are considered more aggressive than those that are 3 + 4. This is because the larger proportion of the tumour belongs to the higher Gleason score (4). Consequently, the identification and classification of Gleason score 7 prostate cancers requires skilled pathologists for accurate determination. This is one of the main reasons for the adoption of the newer Gleason grade group classification.

Scores 8–10

Cancers with a Gleason score of 8 are less differentiated and considered anaplastic (lost specialised features). The glands are frequently fused and show a cribriform growth pattern. This is also observed frequently with cancers having a Gleason score of 9 or 10. In general, tumours with a Gleason score of between 8 and 10 are observed to proliferate rapidly and spread. These are considered the most aggressive and have the poorest prognosis.

Gleason grade group

The Gleason grade group is now being used as a substitute for the Gleason score. The primary reason for this is to provide a consistent grading system for clinicians, as well as to simplify the current 'two-score' Gleason system that directs treatment approaches. The new grading group system is shown in Table 2 below, along with the equivalent Gleason score. The most obvious benefit of using the Gleason grade groups, as opposed to the Gleason score is to differentiate between different Gleason score 7 malignancies: Group 3, which is mainly composed of individual Gleason score 4+3 cells and Group 2, which is mainly composed of individual Gleason score 3+4 cells. These two groups have substantially different prognosis. The Gleason grade group system has been found to provide more accurate prognostic discrimination for treatment (Kryvenko and Epstein, 2016; Epstein *et al.*, 2016). This has the benefit of preventing overtreatment of lower-grade prostate carcinomas. Overtreatment will cause unnecessary side effects and provide a poorer quality of life for the patient receiving treatment; achieving a balance between treatment choice and side effects is therefore important.

Gleason score	Gleason grade group
Score 6 (3 + 3)	Group 1
Score 7 (3 + 4)	Group 2
Score 7 (4 + 3)	Group 3
Score 8 (4 + 4)	Group 4
Score 9 (4 + 5, 5 + 4) and 10 (5 + 5)	Group 5

Table 2. Current Gleason scores and their equivalent Gleason grade groups.

It has been proposed that the older Gleason score classification system be replaced by the newer Gleason grade system.

Gleason grade groups are strong predictors of prognosis as shown in Figure 11, where Group 1 is associated with the best prognosis, whilst Group 5 with the worst.

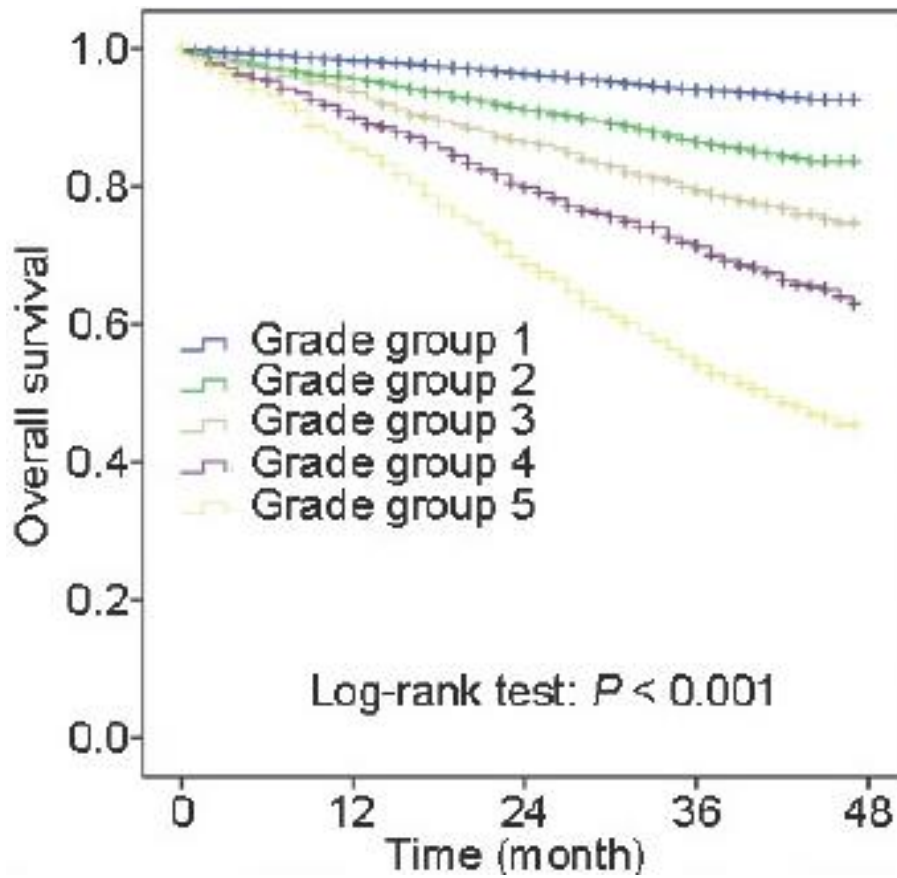


Figure 11. Prostate cancer Gleason grade groups and overall survival probability.

Higher Gleason grade groups show reduced overall survival times and poorer prognosis. From Chen et al., 2018.

1.4 Molecular pathogenesis of breast, ovarian and prostate cancers

With the identification and an increasing understanding of the risk factors that contribute to breast, ovarian and prostate cancer (and many other cancers), greater attention has been focused on the identification of the underlying molecular pathogenesis that promotes carcinogenesis. The molecular pathogenesis of breast, ovarian and prostate cancer is often associated with mutations in cancer-related genes and/or with risk factors described in section 1.2. One example is the *BRCA* genes that frequently harbour mutations in women that develop HBOC (section 1.2.4). Ultimately, mutations in such genes lead to alterations in the normal function of molecular pathways, leading to aberrant cell proliferation, aberrant migration or to other 'cancer hallmarks' (Hanahan and Weinberg, 2011). This can be either through loss-of-function or gain-of-function mutations. For example, mutations are frequently observed in many of the hormone receptor pathways. This can lead to overactivation of these pathways, thereby increasing rates of proliferation and to the activation of anti-apoptosis processes.

Breast, ovarian, and prostate cancers are often affected by the same risk factors. They have also been found to harbour mutations within the same genes. A summary of the genes and the number of known pathological mutations common across breast, ovarian and prostate cancers is shown in Table 3.

Table 3. Most common cancer-associated genes and the approximate number of known pathogenic mutations across breast, ovarian and prostate cancers.

Gene	Number of known mutations	Cancer	Reference
BRCA1	3430	Breast, ovarian, prostate	(Tsaousis <i>et al.</i> , 2019; Chen and Parmigiani, 2007)
BRCA2	3974	Breast, ovarian, prostate	(Tsaousis <i>et al.</i> , 2019; Chen and Parmigiani, 2007)
TP53	490	Breast, ovarian, prostate	(Tsaousis <i>et al.</i> , 2019; Gallardo-Alvarado <i>et al.</i> , 2019)
PALB2	724	Breast, ovarian, prostate	(Tsaousis <i>et al.</i> , 2019; Southey, Winship and Nguyen-Dumont, 2016)
ATM	1494	Breast, prostate	(Tsaousis <i>et al.</i> , 2019; Angèle <i>et al.</i> , 2004)
CHEK2	412	Breast, ovarian, prostate	(Tsaousis <i>et al.</i> , 2019; Hale, Weischer and Park, 2014; Dong <i>et al.</i> , 2003)

Mutations identified from the ClinVar database (Landrum *et al.*, 2018, accessed January 2018), and studies where the genes with pathogenic mutations have been identified in breast, ovarian, and prostate cancers.

Many genes harbour mutations that are usually inherited through the germline and confer a high relative risk of developing cancer these include *BRCA1*, *BRCA2* and *TP53*. Some genes have been found to commonly mutate in populations, but may require an additional ‘activator’ event or mutation in an additional gene for a cancer to develop. These genes for example include *PALB2*, *ATM*, and *CHEK2*.

In this section and the corresponding subsections, some of the most frequently identified mutated genes in breast, ovarian and prostate cancers are presented. Because of the considerable number of mutations that have been identified in these genes, only a few are described, along with their functional effect, where this is known. The role of the three main hormone receptor pathways (oestrogen, progesterone and androgen) in promoting breast, ovarian and prostate cancers is also discussed. The most frequently altered pathways in breast, ovarian and prostate cancers are described. Among all the cancer hallmarks, we have focussed on proliferation and anti-apoptosis mechanisms, which are essential prerequisites of carcinogenesis.

BRCA genes

The genes associated with breast cancer, specifically, *BRCA1* and *BRCA2*, are located in the 17q21.31 and 13q13.1 chromosome regions, respectively. They are both primarily involved in the DNA damage repair (DDR) pathway. When DNA damage occurs and a double strand break (DSB) is produced, this is detected, and the *BRCA1* and *BRCA2* proteins are recruited to the site to repair the DNA strands. *BRCA1* has multiple actions within the DDR pathway, for example, in activating checkpoint activation and DNA repair (Roy *et al.*, 2012). *BRCA2* is a key mediator in homologous recombination (HR) by recruiting *RAD51* to the site of repair (Roy *et al.*, 2012).

Mutations in these two genes account for ~5%–10% of breast cancers (Martin and Weber, 2000), and both have also been found to be the cause of hereditary ovarian and prostate carcinomas (Castro and Eeles, 2012; Neff, Senter and Salani, 2017). In a study of ~8,000 patients with either breast or ovarian cancer, the cumulative risk for breast cancer in those with *BRCA1* mutations was 65%, and 39% for ovarian cancer. For *BRCA2*, the cumulative risk was 45% and 11% for breast and ovarian cancers, respectively (Antoniou *et al.*, 2003). However, unlike breast and ovarian cancers,

BRCA1 mutations pose less of a risk of prostate cancer than *BRCA2* mutations (Nyberg *et al.*, 2020).

BRCA1/2 mutations

Mutations in *BRCA1* and *BRCA2* lead to dysfunctional and aberrant DNA repair. If sufficient DNA repair no longer occurs, further mutations and genomic instability can accumulate, contributing to the development and progression of cancer.

BRCA1 mutations alter the binding sites of the BRCA1 protein that binds to the ataxia telangiectasia mutated (ATM) protein. The mutations in *BRCA1*: K309T (c.926A>C), S632N (c.1897C>A), S1143F (c.3428C>T), Q1144H (c.3432G>T), Q1281P (c.3842A>C) and S1542C (c.4625C>G), have been found to alter the kinase motif of BRCA1 leading to aberrant functions in the BRCA1 protein. The S1542C mutation removes the phosphate on Ser1542 amino acid on the BRCA1 protein (Tram, Savas and Ozcelik, 2013). This phosphorylation site is important for the activation of BRCA1 by the ATM protein following the detection of DNA DSBs by ATM. Further to this the mutations S1143F, Q1144H and Q1281P all prevent ATM binding to Ser1143 and Ser1280 thereby altering BRCA1 function in single strand break (SSB) repair.

BRCA1 germline mutations c.287C>T and c.192T>C, as well as the deletion c.315del in the promoter region, alter *BRCA1* promoter activity (Burke *et al.*, 2018). Importantly, these mutations alter the binding of transcription factors FOXA1, E2F1 and MYC to the *BRCA1* promoter. For example, loss of FOXA1 binding to the *BRCA1* promoter has been found to suppress *BRCA1* expression (Gong *et al.*, 2015).

BRCA2 mutations have been found to occur primarily in the DNA binding domains (DBD) of the protein (Borg *et al.*, 2010). This reduces BRCA2 binding to the disordered and multifunctional protein (DSS1) in the C-terminus DNA binding domain (C-DBD) (Li *et al.*, 2006a; Mishra *et al.*, 2022). Loss of DSS1 binding reduces BRCA2 nuclear localisation (Jeyasekharan *et al.*, 2013) as well as binding of RAD51 to BRCA2 at the site of single strand DNA binding (ssDNA) of HR (Paul *et al.*, 2021). In breast cancer, the *BRCA2* mutation c.4585G>A was found to affect the binding site of

RAD51 to BRCA2 (Borg *et al.*, 2010). Loss of RAD51 binding to BRCA2 prevents RAD51 recruitment to the site of DSB repair (Prakash *et al.*, 2015). RAD51 is an important factor in HR of DSB repair, because it locates and binds the homologous sequence, thereby allowing correct repair and subsequent BRCA2 binding.

BRCA1 and *BRCA2* mutations occur in 8%–15% of ovarian cancer patients (Maistro *et al.*, 2016). A study of *BRCA1* in 158 ovarian cancer patients found that three mutations (c.2566dupC, c.3331_3334delCAAG and c.211A>G) were commonly identified (Cotrim *et al.*, 2019). In a separate study of 12 ovarian cancer patients, 17 *BRCA1* and 2 *BRCA2* mutations were found (Maistro *et al.*, 2016). *BRCA2* variants c.687T>A and c.7007G>A have been identified in HBOC as causing aberrant splicing of the *BRCA2* mRNA (Sanz *et al.*, 2010).

In prostate cancer, *BRCA1* mutations occur frequently as well. In a study of Ashkenazi Jewish men (979 with prostate cancer and 1,251 controls), three mutations commonly identified within the *BRCA1* gene (185delAG and 5382incC) and *BRCA2* gene (6174delT) were found to confer increased prostate cancer risk compared to wild-type. The *BRCA2* mutation 6174delT carriers had a three-fold greater risk of developing a higher Gleason score (7–10) prostate cancers. Additionally the *BRCA1* deletion mutation 185delAG also had a three-fold increased the overall risk of prostate cancer as well as the risk of developing a higher Gleason score (>7) tumours (Agalliu *et al.*, 2009).

PALB2 gene

The partner and localiser of BRCA2 (PALB2) is primarily involved in HR. HR forms part of the DSB pathway with BRCA1 and BRCA2. PALB2 acts as a binder for the BRCA complex of HR (BRCA1-PALB2-BRCA2-RAD51). In summary, DNA damage leads to BRCA1 recruitment to the site of damage, whereby PALB2 binds to BRCA1 and recruits BRCA2, then RAD51 is recruited to the site of damage to begin HR. The HR of DNA damage signalling (DDS) helps to maintain genomic stability in the accumulation of DNA lesions, which may impair correct DNA transcription (Jackson and Bartek, 2009).

PALB2 mutations

PALB2 mutations are frequent (Wu *et al.*, 2020; Wokołorczyk *et al.*, 2021) and associated with poorer prognosis in breast, ovarian and prostate cancers (Li *et al.*, 2017; Wokołorczyk *et al.*, 2021).

Mutations in *PALB2* occur in 40% of triple negative breast cancer subtypes (Tischkowitz and Xia, 2010) and confer a 2.3-fold increased risk of breast cancer (Rahman *et al.*, 2007). In a Finnish population, the *PALB2* mutation c.1592delT was found to cause truncation of the PALB2 protein. This led to decreased binding affinity of BRCA2 and, subsequently, inefficient DNA repair recruitment of BRCA2 and RAD51 (Hannele *et al.*, 2007). A separate study of triple negative breast cancer patients identified that three mutations (c.758dup, c.2390del, and c.3113+5G>C) caused truncation of PALB2, also preventing BRCA2 binding (Wong-Brown *et al.*, 2014). In a Spanish population consisting of families with a history of breast, pancreatic and ovarian cancer, two truncating mutations (c.2653T>A and c.3362del) were frequently identified. Both mutations are considered to be pathogenic due to the creation of stop codons that truncated the PALB2 protein, thereby preventing normal function (Blanco *et al.*, 2013).

PALB2 mutations increase the risk of ovarian cancer. A study of 1,421 ovarian cancer patients and 4,300 controls, mutations in *PALB2* were found to pose a four-fold increase in risk of developing ovarian cancer (Kotsopoulos *et al.*, 2017). *PALB2* mutations c.509_510del, c.172_175del and c.347ins, are frequent in ovarian cancers, leading to truncations in the PALB2 protein, as observed in breast cancer as well (Kluska *et al.*, 2017; Pauty *et al.*, 2017).

The truncating deletion mutation c.1592del, has been identified in both breast and prostate cancers (Obermeier *et al.*, 2016; Pakkanen *et al.*, 2009) and is more prevalent in aggressive breast and prostate cancers in both hereditary and sporadic cases (Pakkanen *et al.*, 2009).

TP53 gene

The *TP53* gene encodes the tumour-suppressor protein P53. Because of its role as a tumour suppressor, *TP53* is usually inactivated in the first steps towards the development of many cancers.

P53 suppresses cell proliferation in response to detection of DNA damage (e.g. single and double strand breaks), hypoxia and replicative stress (genomic instability). For example, when DNA damage is detected in either the G1 or G2/M phase of the cell cycle, P53 is phosphorylated by ATM, thereby preventing the MDM2-dependent ubiquitination of P53 (a mechanism that induces P53 degradation). P53 intracellular levels are therefore increased. P21 (encoded by the *CDKN1A* gene) is then phosphorylated by P53, inducing cell cycle arrest. During this time, P53 also phosphorylates BAX (encoded by the *BAX* gene also called *BCL2*), which causes the induction of downstream apoptosis.

TP53 mutations

TP53-inactivating mutations have been observed in nearly all cancers, including breast, ovarian and prostate (Olivier, Hollstein and Hainaut, 2010; Malkin, 2011; Ecker *et al.*, 2010). Loss of P53 activity leads to increased genomic instability in cancers, as well as their rapid proliferation.

In breast cancers, *TP53* is mutated in 17% of luminal A, 41% of luminal B, 50% of HER2 and 88% of basal-like cancers (Bertheau *et al.*, 2013). Luminal A and B breast cancers have predominantly A/T and C/G substitutions in *TP53*, whereas basal-like cancers have more complex mutations such as insertions and deletions (indels), resulting in complete loss of function (Dumay *et al.*, 2013). In a study in which *TP53* mutations were analysed in breast cancer patients, 90% of missense mutations were identified in the DNA binding domain (DBD) of *TP53*, which corresponded to poorer prognosis. Frameshift mutations also induced the loss of parts of the DBD domain, thereby removing correct DNA binding (Vegran *et al.*, 2013).

In ovarian cancer, *TP53* is mutated in ~96% of HGSOC cases (Cancer Genome Atlas Research Network *et al.*, 2013) and ~80% in all ovarian cancers, mainly in exons 5, 7 and 8, which are all

part of the DBD of *TP53* (Zhang *et al.*, 2016b; Garziera *et al.*, 2018). These mutations alter the DNA binding capacity of P53 protein, thus preventing it from binding to the transcription sites of target genes involved in apoptosis and cell cycle progression (Zhao *et al.*, 2000).

TP53 mutations in prostate cancer also promote tumour progression. Of 90 patients with prostate cancer, 36% were identified as having underlying *TP53* mutations (Ecke *et al.*, 2010). As with breast and ovarian cancer mutations, these mutations occurred in exons 7 and 8, in the DBD of P53. *TP53* mutations have been found to be positively associated with higher Gleason scores in prostate cancer (Sun *et al.*, 2019).

CHEK2 gene

The *CHEK2* (checkpoint kinase 2) gene encodes the CHK2 protein and is a key component of the cell cycle checkpoint activation, DDR and induction of apoptosis. Like *TP53*, the *CHEK2* gene is classed as a tumour suppressor gene. Similar to the *TP53* and *BRCA* genes, the CHK2 protein is also part of the DSB repair mechanism, and is directly activated by ATM. Once CHK2 has been phosphorylated, in turn, it phosphorylates CDC25 thereby promoting cell cycle arrest and activating P53.

CHEK2 mutations

A c.1100delC germinal mutation in exon 10 of *CHEK2* is frequently identified in European populations, causing increased risk of breast, lung and prostate cancer (Nevanlinna and Bartek, 2006; Huijts *et al.*, 2014). This mutation causes a truncation of the CHK2 protein, leading to reduced function. It can increase breast cancer risk two-fold (Muranen *et al.*, 2011) and is associated with later onset breast cancer in those with no family history of the disease (Nevanlinna and Bartek, 2006). In 237 *BRCA1/BRCA2* mutation negative patients with a family history of breast cancer, 11.4% were identified as carrying the c.1100delC mutation in *CHEK2* (Oldenburg *et al.*, 2003). In 507 *BRCA1* and *BRCA2* negative patients with a family history of breast cancer, 16 mutations were identified, including the c.1100delC in *CHEK2*. Within these *CHEK2* mutation positive individuals, the calculated odds ratio (OR) for breast cancer was 4.15 excluding the deleterious c.1100delC, which rose to 5.18 with the inclusion of the c.1100delC deleterious

mutation (Desrichard *et al.*, 2011). The c.1100delC mutation in prostate cancer confers an eight to nine-fold increase of developing prostate cancer (Nevanlinna and Bartek, 2006; Wang, Dai and Ye, 2015).

Within the forkhead-associated (FHA) domain of the CHK2 protein, the 470T>C missense mutation confers a 8.1% increased risk of breast cancer by age 70 (Nevanlinna and Bartek, 2006). Other missense mutations have also been identified in the FHA domain in prostate cancer; these, too, lead to cancer progression due to a reduction in the CHK2 protein activation. It was observed that the somatic missense mutation p.R1117G within the FHA domain caused loss of the normally conserved Arg117 residue, which is required for phosphopeptide binding (phosphorylation) (Wu *et al.*, 2006). This mutation can prevent activation of CHK2 via ATM phosphorylation (Ahn, Urist and Prives, 2004) when DNA damage is detected.

In a study of 15,397 patients with invasive ovarian cancer, mutations in the *CHEK2* gene were associated with increased risk of developing the disease. A germline intron mutation (g.28696732C>T) was identified as frequently expressed in the serous subtype, and had the strongest association with cancer risk of all the mutations identified (Lawrenson *et al.*, 2015).

ATM gene

The *ATM* gene encodes the ATM protein which is a serine/threonine kinase primarily involved in the initial activation of the DNA damage response/repair pathway. As with the *BRCA* and *TP53* genes, ATM also detects damage to the DNA. For example, when DSBs are detected, it initiates a phosphorylation cascade to stall cell cycle progression and induce apoptosis via P53. This is because ATM also controls PIDD (P53-induced protein with a death domain), which is an important driver of cell apoptosis (Shiloh and Ziv, 2013).

When DNA DSBs are detected, the DDR signalling pathway is activated. ATM detects DNA ends and chromatin and phosphorylates histone H2A.X variant histone (H2AX). This allows downstream binding of DNA repair proteins and chromatin remodelling proteins at the site of DNA breaks.

Additionally, ATM also activates BRCA1, P53 and CHK2 via phosphorylation to begin DNA repair (Maréchal and Zou, 2013).

ATM mutations

Mutations in *ATM* are associated with an increase in the rate of malignancies (driven by DNA lesions), due to inefficient DNA strand break repair (Roberts *et al.*, 2012).

ATM germinal mutations occur in 0.5%–1% of populations (Ahmed and Rahman, 2006). Mutations in *ATM* are associated with a 100-fold increased risk of epithelial cancers such as breast, ovarian and prostate cancer (Ahmed and Rahman, 2006; Rahman *et al.*, 2007; Angèle *et al.*, 2004). In a study across the exons of *ATM* in a German population, 46 *ATM* mutations were identified. The majority of these were nonsense mutations and frameshift indels, which led to a stop codon being created and the premature termination of the ATM protein (Sandoval *et al.*, 1999). In an analysis of 1,160 relatives, 169 of whom had *ATM* mutations, the relative risk of developing breast cancer was 2.23 higher; the relative risk of developing other cancer types was 2.05 higher in female carriers and 1.23 in male (Thompson *et al.*, 2005). A separate analysis of 443 hereditary breast cancers 9 germline *ATM* mutations resulted in premature truncation or exon skipping and 3 missense mutations were also identified (Renwick *et al.*, 2006).

In HBOC cases in which 270 individuals were sequenced, 137 different sequence variants were identified in *ATM* (Thorstenson *et al.*, 2003). Seven sequence variants promoted ataxia telangiectasia, which is an autosomal recessive disorder that is associated with an increase in susceptibility to cancer due to the loss of genome stability (Rothblum-Oviatt *et al.*, 2016). An L1420F *ATM* mutation was observed exclusively in HBOC cases and not in controls in a separate study (Thorstenson *et al.*, 2003).

A c.3161C>G germline mutation significantly increases the risk of developing prostate cancer as well as breast and ovarian cancer (Angèle *et al.*, 2004). This corresponds to previous findings in which HBOC was expanded to include prostate cancer. It had originally included only breast and

ovarian cancer. ATM-inactivating mutations occur more frequently in higher-grade prostate cancers (Kaur *et al.*, 2020).

1.4.1 Cancer-related pathways in breast, ovarian and prostate cancers

Hormone receptor pathways have been found to be important drivers of cancer. The most well-known hormone receptors in breast, ovarian and prostate cancers are ER, PR, HER2 and AR, which are important mediators of cell proliferation. These signalling pathways are commonly mutated and/or characterised by aberrant expression profiles in cancers; as such, they can also be used to classify breast cancers into molecular subtypes, ovarian cancers into epithelial subtypes and prostate cancers into histological cells subtypes. Because hormone receptors are transcription factors, they can cause aberrant expression of several downstream gene targets, leading to activation or inactivation of several downstream pathways. Many pathways will also induce the activation of further subsequent pathways through 'cross-talk communication', as well as by eliciting multiple cellular responses such as increased proliferation and/or migration and reduced apoptosis.

Oestrogen receptor (ER) pathway

Oestrogen is a steroid hormone that is important in the female reproductive system, especially during puberty in mammary gland development and maturation. Oestrogen is also present in males but at much lower levels, and it is important for prostate development (Prins and Korach, 2008). Oestrogen is part of a family of sex steroid hormones primarily secreted by the ovaries in females. In males, it is produced in the testes and other tissues, for example, adipose tissue. In males, however, testosterone is converted to oestrogen via aromatase activity (Cooke *et al.*, 2017).

Oestrogen has two main forms: oestrone (E1) and oestradiol (E2). The primary type of oestrogen formed is oestradiol (E2), which in its natural state is the 17 β -oestradiol ligand. 17 β -oestradiol is the primary initiator for the oestrogen receptor signalling pathway and binds to ER. Oestrone (E1) is produced from adipose tissue of the breast. Therefore, diet can lead to increased adipose tissue

and therefore increased oestrogen levels. Overexpression of 17 β -oestradiol has been identified as promoting breast cancer progression and invasion (Zheng *et al.*, 2011) through activation of other cell proliferation-related pathways such as PI3K/AKT (section 1.4.1).

The ER pathway has been identified as an important factor in the promotion of increased proliferation, metastasis and treatment-resistance in cancers. The pathway acts upon multiple mechanisms, which, in turn, affects transcriptional regulation. This promotes the expression of many proliferation-related genes and furthers cell cycle progression. Mutations in *ERs* are usually in the ligand binding domain (LBD), which allows aberrant ER pathway activation without the binding of the oestrogen hormone (such as ER- breast cancers). The K303 mutation has been found to prevent the binding of ER co-repressors (one being BRCA1), thereby enhancing ligand-independent activity. The Y537N and Y537S *ER* mutations allow for ER ligand-independent activity by removing regulation. This imitates oestrogen hormone binding activating ER (Harrod *et al.*, 2017). This process has been found to confer resistance to anti-oestrogen treatments such as tamoxifen (Fuqua, Gu and Rechoum, 2014). The Y537S mutation has been observed in both breast and ovarian cancers (Fuqua, Gu and Rechoum, 2014; Langdon *et al.*, 2020; Gaillard *et al.*, 2018).

The ER pathway is composed of two mechanisms of transcriptional regulation and is initiated via two main ERs: ER α and ER β (Figure 12).

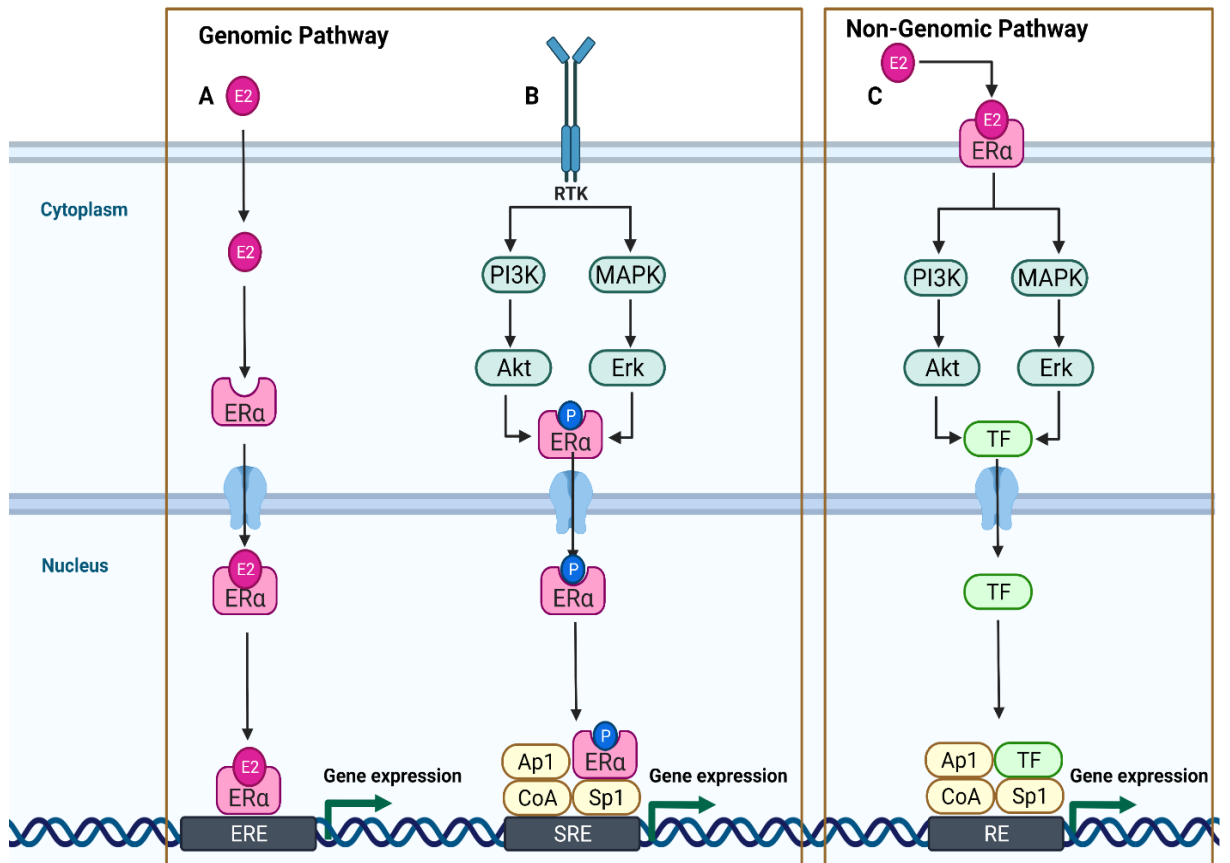


Figure 12. ER signalling pathways that result in altered transcription and gene expression.

A) In the genomic pathway the ER (α and β) moves to the nucleus where it binds directly to the DNA. This can be either at EREs or SREs within the promoter regions of target genes, thereby altering gene expression. Oestrogen bound ER is able to bind to EREs. B) Alternatively non-oestrogen bound ER is phosphorylated via either activated PI3K/AKT or MAPK/ERK. Phosphorylated ER can then bind to SREs with other transcription factors (AP1, SP1, c-Fos and c-Jun). C) In the non-genomic pathway, oestrogen binds with membrane-bound ER to dimerise and activate PI3K and/or MAPK cascade signalling, which can alter transcription without direct binding of ER to DNA but instead through other transcription factors (AP1, SP1, c-Jun, c-Fos). Adapted from Musgrove R & Sutherland R., 2009; Barzi A, et al., 2013.

ER: oestrogen receptor; EREs: oestrogen response elements; SRE: serum response elements; RE: response element; RTK: receptor tyrosine kinase; MAPKs: mitogen-activated protein kinases; PKC: protein kinase C; PI3K: Phosphoinositide 3-kinases; AKT: AKT Serine/Threonine Kinase; ERK: Mitogen-Activated Protein Kinase 1.

The first of these mechanisms is genomic, where activated ER moves from the cytoplasm to the nucleus, where it binds to oestrogen response elements (EREs) in the DNA and acts on gene transcription. The second mechanism is nongenomic, and in this case, the activated ER does not bind directly to DNA to alter gene transcription; instead it causes the activation of other pathways, for example, mitogen-activated protein kinase (MAPK), which in turn trigger a signalling pathway that alters gene expression.

The genomic action of the ER pathway occurs primarily because ER α forms a dimer that can bind to EREs within the DNA (Figure 12). This can occur if ER is bound to a ligand (specifically oestrogen) or can be ligand independent. ERs can be phosphorylated by protein kinase pathways such as MAPK/extracellular signal-regulated kinase (ERK) or PI3K/AKT, whereby they can bind to EREs independent of ligand binding. EREs are found within promoter regions of their target genes and can regulate expression (Bjornstrom and Sjoberg, 2005). In the genomic mechanism, ERs can also bind to other transcription factors instead of EREs, such as serum response elements (SREs). This mechanism of genomic action occurs through other transcription factor response elements (Vrtacnik *et al.*, 2014). ERs bound to oestrogen translocate to the nucleus, forming a complex with transcription factors SP1 and AP1, which stimulates FOS and coactivators (CoAs) to alter gene expression (Elizabeth and Robert, 2009).

The nongenomic mechanism can be exerted without ER binding to DNA (either by EREs or transcription factors to SREs) (Figure 12). Upon binding of oestrogen to plasma membrane-bound ER, ER leads to the activation of protein kinase signalling cascades (such as PI3K/AKT and MAPK/ERK) (Vrtacnik *et al.*, 2014). Additionally, non-membrane bound ER, upon binding of oestrogen, allows the ER to bind to cellular SRC (c-Src) (Bjornstrom and Sjoberg, 2005) which can in turn activate receptor tyrosine kinases (RTKs) by phosphorylating tyrosine leading to signalling cascades (Belsches, Haskell and Parsons, 1997). ER can also interact with SRC kinase bound at RTKs (Cooper and Qian, 2008) to induce signalling cascades such as RAS.

Both RAS and SRC kinase are upstream activators of the PI3K/AKT and MAPK/ERK pathway. RTKs can also be activated through the plasma membrane-bound ER directly. ER α can activate the insulin growth factor receptor (IGFR) (Elizabeth and Robert, 2009; Bjornstrom and Sjoberg, 2005), which leads to increased MAPK and AKT signalling. Both the MAPK/ERK and PI3K/AKT pathways have been found to be important drivers of breast, ovarian and prostate cancers.

Overexpression and aberrant signalling of the ER pathway have been identified frequently in breast, ovarian and prostate cancers (Lee *et al.*, 2011; Lafront *et al.*, 2020; Langdon *et al.*, 2020). Increased expression of ER α in breast and ovarian cancers is associated with increased

proliferation (Andersen *et al.*, 2017; Xue *et al.*, 2019). Overexpression of ER α in prostate cancers is associated with a higher Gleason score and poorer prognosis (Di Zazzo *et al.*, 2018). This link was previously overlooked due to the limitations of prostate cancer models (Lafront *et al.*, 2020). ER α has been found to be expressed in the higher Gleason grade groups 4 and 5 (43% and 62% of cases). In CRPCs, ER α expression has been found to occur in 55.5% of cases and has also been associated with increased PR expression (a target gene of ER α) (Bonkhoff, 2018).

Loss or downregulation of ER α leads to increased expression of the EMT-related transcription factors *SNAIL* and *SLUG*, with the result that EMT promotes metastasis in ovarian cancers (Park *et al.*, 2008). In breast cancers, loss of ER α promotes activation of the SRC1 protein (steroid receptor coactivator-1), an ER coregulator that increases PEA3 (polyoma enhancer activator protein) activity initiating expression of transcription regulator *TWIST* (Chan *et al.*, 2015; Qin *et al.*, 2009). *PEA3* has previously been identified as an oncogene that has been found to be overexpressed in breast carcinomas (Gu *et al.*, 2011). *TWIST* protein promotes breast cancer metastasis via the destabilisation of chromosomes and the downregulation of *E-cadherin* (Vesuna, Bergman and Raman, 2017).

ER β has an antiproliferative effect in ovarian and breast cancers (Pinton, Nilsson and Moro, 2018; Schöler-Toprak *et al.*, 2018; Paruthiyil *et al.*, 2004). This is because it can act as an inhibitor of ER α and, as a result, reduced expression of *ER β* has been found to increase metastasis in DCIS and IDC breast cancers, which show a decrease in ER β in 21% of the cases (Skloris *et al.*, 2003). ER β is downregulated in high-grade prostate cancers (Signoretti and Loda, 2001). Further to this, the downregulation of ER β in prostate cancers has also been found to promote the EMT phenotype (Mak *et al.*, 2010). In prostate cancers, *ER β* has been found to be highly expressed, but lost in progression to CRPC (Bonkhoff, 2018). This is because ER β has been found to be a target of AR and, therefore, loss of AR leads to decreased ER β .

In conclusion, emerging evidence suggests that ER-dependent pathways are crucial mediators of all the three cancers studied in this research project.

Progesterone receptor (PGR) pathway

Progesterone is a steroid hormone expressed in epithelial cells. It is required for appropriate branching of the ductal system during female puberty and maturation of the mammary alveoli during pregnancy via alveologenesis (Conneely, Mulac-Jericevic and Arnett-Mansfield, 2007; Obr *et al.*, 2013; Obr and Edwards, 2012). This process is particularly important in the early development and then advancement of terminal ducts through the stroma in preparation for the maturation of secretory alveoli during pregnancy (Hinck and Silberstein, 2005). In the ovaries, progesterone is regulated by ERs. It also acts to control the menstrual cycle by reducing the expression of ERs in a feedback loop in the endometrium of the uterus (Reed and Carr, 2000). In males, progesterone is important for spermiogenesis and sperm maturation (Lue *et al.*, 2013). It has also been found that it can be converted to androstenedione, which can be further converted to testosterone (Miller and Flück, 2014).

There are two primary progesterone receptors: progesterone receptor A (PGR-A) and progesterone receptor B (PGR-B). As with ER α and ER β , PGR-A and PGR-B are able to control gene expression both genomically and non-genomically (Tan *et al.*, 2012). Both PGRs are cytoplasmic and migrate to the nucleus upon binding to progesterone.

PGR-A and *PGR-B* are expressed in similar levels in epithelial tissues; *PGR-B* has greater transcriptional regulation than *PGR-A* (Scarpin *et al.*, 2009). In cancers such as ovarian and breast, both PRs are aberrantly expressed. Either increased or decreased expression of *PGR-A* or *PGR-B* has been observed in breast cancer versus normal tissue (Bellance *et al.*, 2013; Khan *et al.*, 2012). These inconclusive results suggest that the relationship between *PGR* and breast carcinogenesis is complex and not completely elucidated. Similar to that of oestrogen binding to ERs, in the genomic pathway, progesterone binding to PGR-A/PGR-B leads to dimerization and binding to progesterone response elements (PREs), which alters gene transcription (Scarpin *et al.*, 2009) (Figure 13).

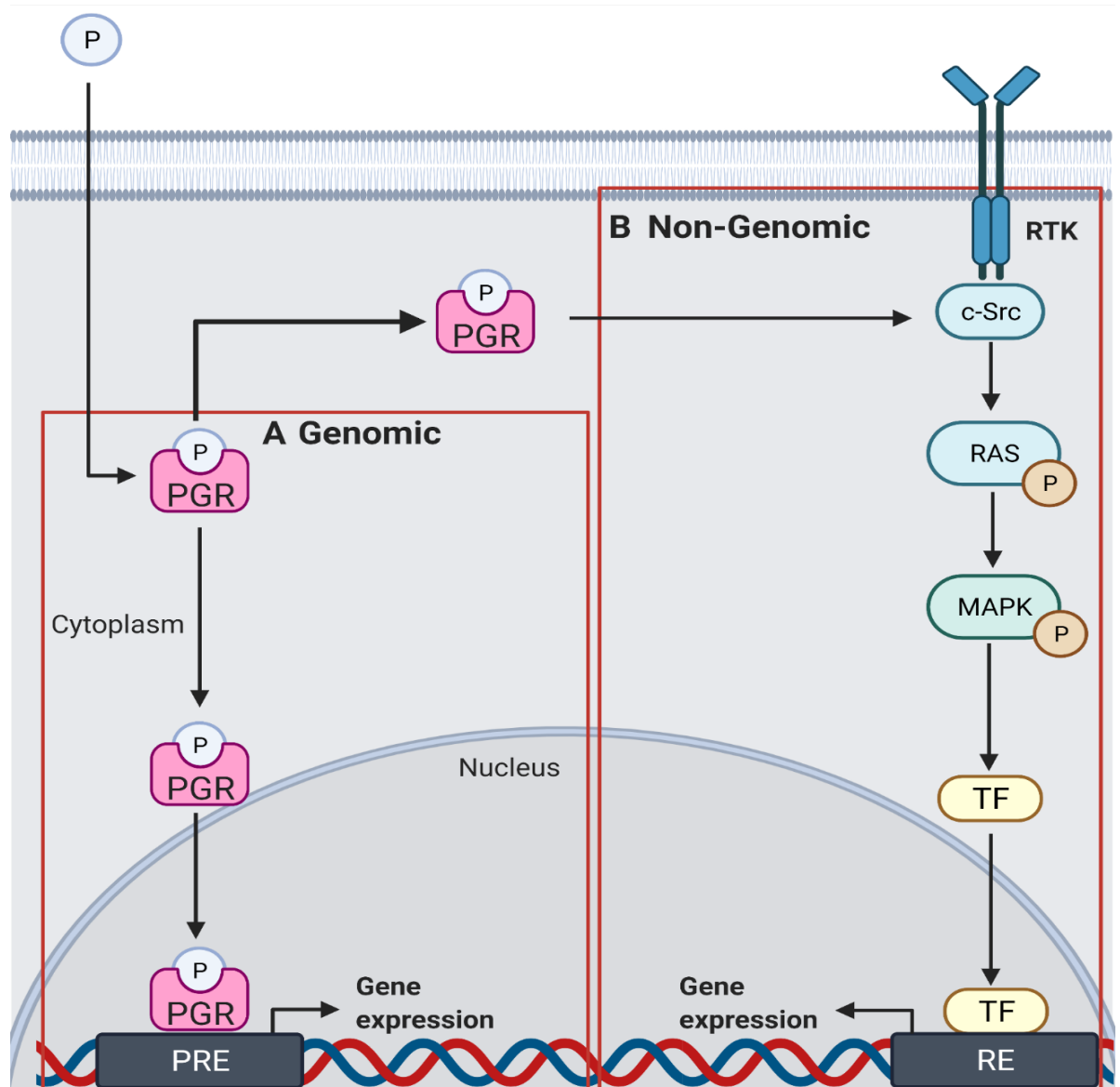


Figure 13. PGR pathway.

A) The genomic mechanism of the PGR pathway activates downstream gene expression via binding of progesterone-bound PGR to the PREs in DNA. This leads to changes in gene expression. B) In the non-genomic mechanism of the PGR pathway, the binding of progesterone with PGR leads to the activation of MAPK signalling pathway through binding to c-Src domains of RTKs. This modulates gene expression without direct binding of PGR to the DNA.

PGR: progesterone receptor; PREs: progesterone response elements; MAPK: mitogen-activated protein kinase; RAS: Rat sarcoma virus; RE: Response element.

Although PR-A and PR-B display a similar structure (PR-A has 164 fewer amino acids), they interact with different groups of coregulatory and transcription factors and, thus, have different effects on gene expression. PGR-A has also been found to act as a repressor of both PGR-B and ER α (Mulac-Jericevic *et al.*, 2000).

Both PGR-A and PGR-B are primarily expressed in accordance with ER α expression (Brisken, 2013). This is primarily due to *PGR* being a downstream target of ER signalling; as such, it increases expression (Grimm, Hartig and Edwards, 2016; Hisham *et al.*, 2015). Increased cellular proliferation from PR signalling (also ER) can speed up the transition from hyperplasia to IDC tumours (Grimm, Hartig and Edwards, 2016). Downregulation and loss of *PGR* within breast tumours is commonly observed as causing upregulation of MAPK, in which growth factors such as EGFR are able to promote proliferation and more aggressive breast cancer subtypes (Chandra *et al.*, 2013; Moerkens *et al.*, 2014; Lo, Hsu and Hung, 2006). In prostate cancers, PGR-B is significantly associated with poorer prognosis (Grindstad *et al.*, 2018). Progesterone and PGRs in ovarian cancers promote resistance to the anticancer drug Cisplatin treatment through PI3K/Akt pathway activation (Zhu *et al.*, 2013).

In conclusion, similar to the ER pathway, the PGR pathway is an important facilitator of the three cancers studied in this research project.

Androgen receptor (AR) pathway

AR is the primary hormone receptor for the prostate (Fujita and Nonomura, 2019). Androgens are the primary steroid hormones in males but are also present in lower concentrations in females. These androgens include testosterone. In hormone-sensitive prostate adenocarcinomas, high AR expression seems to be positively correlated with Gleason score and associated with worse prognosis (Hashmi *et al.*, 2019). AR is used as an additional factor to guide potential treatment options alongside the Gleason score. In AR+ hormone-dependant prostate cancers, testosterone binds with AR to induce cell proliferation pathways (Tan *et al.*, 2015). This is similar to the role of oestrogen and ERs in ovarian and breast cancers, activating similar pathways.

Testosterone is converted to dihydrotestosterone (DHT) by 5 α -reductase (Davey and Grossmann, 2016). This form of the hormone has a higher binding affinity to AR than testosterone. In hormone-dependent prostate cancers, androgen deprivation therapy (ADT) has been found to be beneficial in treating more advanced and metastatic cancer types (Perlmutter and Lepor, 2007). However, in some cases, prostate carcinomas can become androgen independent, and are

therefore termed castration-resistant prostate cancer (CRPC). CRPC is likely to occur in patients who are on ADT and is highly aggressive and prognostically poor (Tan *et al.*, 2015). In breast cancers, AR is commonly expressed in (60%–80%), and frequently in ER+ breast cancer (Kensler *et al.*, 2019). In triple negative breast cancer, AR activation inhibits apoptosis and increases cyclin D1 promoting proliferation (Zhu *et al.*, 2016). In ovarian cancers the AR has been associated with increased risk of ovarian cancers as well as inhibition of apoptosis also (Mizushima and Miyamoto, 2019). These results suggest a pathogenetic role of the AR in breast and ovarian cancers in addition to prostate cancer.

AR contains two zinc finger DNA binding domains that facilitate binding to androgen response elements (AREs), similar to the EREs of ER and the PREs of PR. As a result, as with ER in breast cancers, AR has both a genomic and nongenomic role in gene regulation.

Protein chaperones such as heat shock protein 90 (HSP90) keep AR in the cytoplasm. Upon DHT binding, AR dimerises and is released by HSPs; this leads to its translocation to the nucleus via a nuclear localisation signal (NLS) (Davey and Grossmann, 2016). Here, AR binds to an ARE, leading to transcription of target genes. Its target genes include *CDK1* and *2* (Stanbrough *et al.*, 2006; Gregory *et al.*, 1998), both of which encode proteins that are important in promoting cell cycle progression (Chen *et al.*, 2006).

AR's nongenomic role results in the activation of cell proliferation-related pathways. In prostate cancers, these proliferation pathways include ERK, AKT and MAPK (Chen *et al.*, 2006). Androgen binding to AR in the cytoplasm causes chaperones to release AR, which can then promote kinase signal cascades in these pathways by direct activation (e.g. without the activation of gene expression patterns in the nucleus).

In conclusion, similar to ER and PGR, the role of the AR is an important factor in the three cancers studied here. All three of these hormone receptors act in a similar manner to affect gene transcription via both a genomic or nongenomic mechanism. This is primarily through activation of proliferation pathways.

Proliferation pathways

Rapid proliferation is one of the most observed traits in all cancers. In particular, increasing rates of proliferation are associated with poorer prognosis in breast cancer (van Diest, van Der Wall and Baak, 2004). It has also been found that ovarian cancers have increased proliferation in patients with *BRCA* mutations (Ezzati *et al.*, 2014). This is most likely due to the loss of DNA repair mechanisms, which under normal conditions can activate apoptosis pathways when DNA damage occurs, thereby preventing proliferation. In prostate cancers, a 31-gene signature associated with increased proliferation was associated with reduced survival rates (Cuzick *et al.*, 2011).

The binding of hormones to ERs, PRs and ARs culminates in the increased expression of proliferation promoting genes. Proliferation is primarily mediated via cyclins and cyclin-dependent kinases, which promote advancement of the G1/S and G2/M phases of cell cycle transition. Both phases are downstream targets of ER, PR and AR. Additionally, these hormone receptors can increase proliferation by binding directly to G-proteins, which in turn activates G-proteins in G-protein-coupled receptors (GPCRs) such as RTKs. This activates proto-oncogene encoded tyrosine SRC (SRC), which is part of the initial kinase pathways signalling cascades (RAS and PI3K) (Saha, Dey and Nath, 2021).

Evading cell cycle checkpoints is one of the first steps in cancer. These checkpoints are important, because they are the stage at which cells are assessed for genomic stability and DNA damage. If DNA damage is detected, then cell cycle progression is paused until it is repaired; however, if this is not possible, then P53 (described earlier) induces apoptosis. Loss of P53-dependent cell cycle control halts apoptosis induction and allows cell cycle progression in both the G1/S and G2/M phases (Duronio and Xiong, 2013; Chen, 2016). As such P53 is frequently found to be mutated or downregulated in breast, ovarian and prostate cancers (Ozaki and Nakagawara, 2011; Chappell *et al.*, 2012).

Mutations that occur within the RTKs in proliferation pathways provide a link to the switch from paracrine to autocrine signalling cancers. These mutations are frequently observed in tumours

and promote an increased rate of proliferation (Obr and Edwards, 2012). This is because RTKs can activate proliferation pathways in the absence of ligand binding.

PI3K/AKT cancer mechanisms

In breast, ovarian and prostate cancers, the phosphoinositide 3-kinases (PI3K) / AKT serine-threonine protein kinase (AKT) pathway is overactive (Guille, Chaffanet and Birnbaum, 2013; Cheaib, Auguste and Leary, 2015; Chen *et al.*, 2016). This is due to underlying upregulation of RTKs in cancers. The most common of these receptors include IGFR, EGFR and ERBB2 (HER2). Activation of the PI3K pathway occurs by ligand binding to RTKs. This leads p85 and p110 to the receptor complex (forming PI3K). Once this has occurred, activated PI3K phosphorylates PIP2 to PIP3, thereby initiating phosphorylation of phosphoinositide-dependent kinase-1 (PDK1). This leads to AKT phosphorylation and prevents degradation of cyclin D1 (CCND1) by glycogen synthase kinase 3(GSK-3), thus promoting cell cycle progression (Figure 14). Following AKT phosphorylation, p21 is inhibited, which regulates cyclin D activity to the cytoplasm in the G1-S phase, thereby promoting progression (Chang *et al.*, 2003). Phosphorylated AKT can also inhibit apoptosis via BAD phosphorylation.

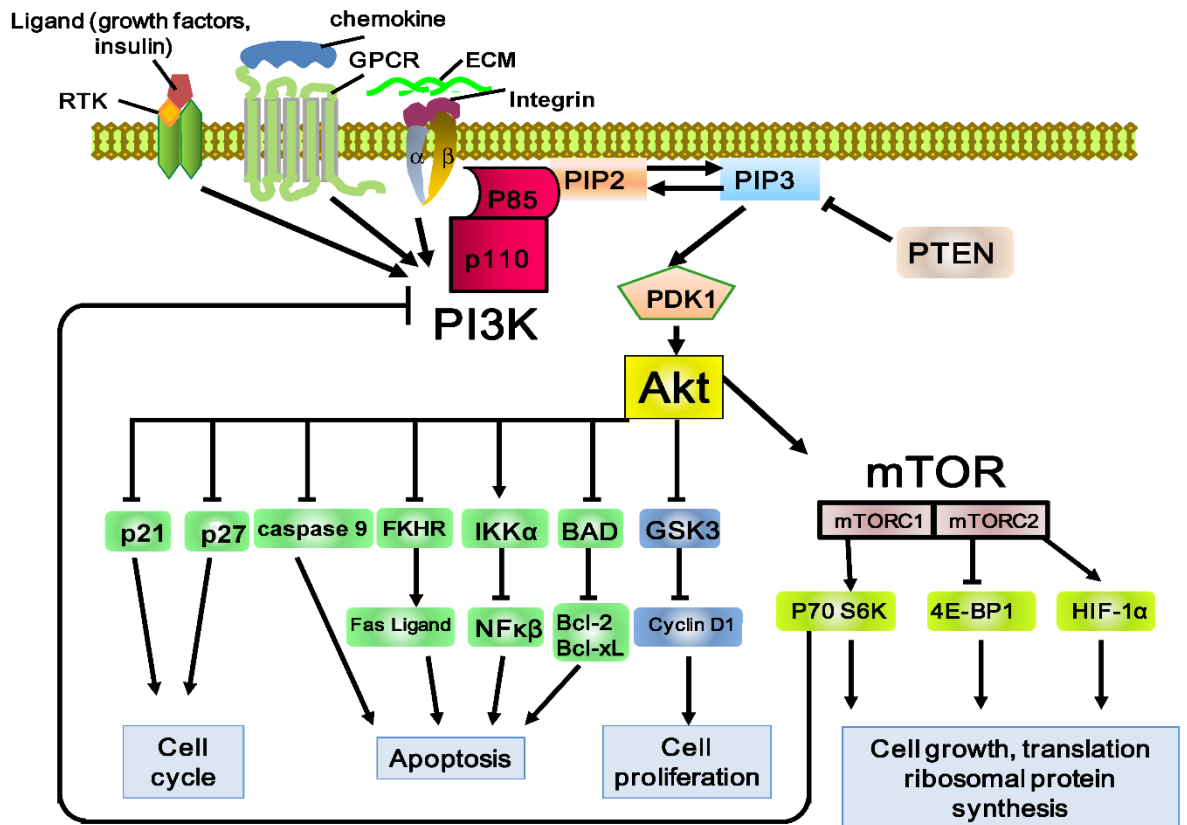


Figure 14. PI3K/AKT pathway in cell proliferation.

Growth factor receptors such as EGFR, IGFR, ER and ERBB2 can induce activation of the PI3K/AKT pathway, leading to cell proliferation and inhibition of apoptosis. PTEN is a potent suppressor of this pathway. The PTEN gene frequently mutates in cancers, leading to uncontrolled activation of the PI3K/AKT pathway. From Matsuoka & Yashiro., 2014.

RTK: receptor tyrosine kinase; GPCR: G-protein-coupled receptor; EGFR: Epidermal growth factor receptor; ERBB2: human epidermal growth factor receptor; IGFR: insulin-like growth factor receptor; ER: oestrogen receptor; PTEN: Phosphatase and tensin homolog.

Negative regulation of the PI3K/AKT pathway occurs primarily by the PTEN protein

dephosphorylating PIP3 to PIP2, which is why *PTEN* is classed as a tumour suppressor gene. The *PTEN* gene commonly mutates in cancers with aberrant activation of the PI3K pathway.

Activating mutations are common in *PIK3CA*, which encodes the P100 α protein, an essential component of the pathway described in Figure 14 (Miller *et al.*, 2011). An increased frequency of *PIK3CA* mutations is observed in hormone-positive breast cancers compared with hormone-negative cancers (Stemke-Hale *et al.*, 2008). *PIK3CA* mutations in ovarian and prostate cancers are associated with poor survival rates (Yuan and Cantley, 2008; Pearson *et al.*, 2018; Campbell *et al.*,

2004). In prostate cancers, *PIK3CA* mutations promote CRPC progression (Pearson *et al.*, 2018). In ovarian cancers, in a study of 97 ovarian clear cell carcinomas, 33% had mutated *PIK3CA* and increased AKT activation (Kuo *et al.*, 2009). *PIK3CA* mutations in cancer commonly occur within the kinase and helical domains (Li *et al.*, 2006b), where they enhance PI3K activity on PIP2 and PIP3 (Huang *et al.*, 2008). This removes the p85 regulatory subunit of the PIK3CA protein, which would normally prevent PIK3CA release from RTKs, thereby increasing cell proliferation (Papa and Pandolfi, 2019).

In conclusion, the PI3K/AKT pathway is activated in all the three cancers studied in this project; interestingly these cancers often share the same type of mutations in genes involved in this pathway.

The MAPK/ERK pathway

The MAPK/ERK pathway is involved in a large number of cellular processes including proliferation, apoptosis, differentiation and growth.

It can be activated by a large array of extracellular signals via RTKs (similar to the PI3K/AKT pathway) or GPCRs (Figure 15). One group of activating factors is the growth factors, for example, EGF; another is mitogens (Dhillon *et al.*, 2007). MAPK/ERK pathway activation occurs via the upstream RAS pathway. Growth factor ligand or mitogen binding to an RTK causes recruitment of SOS and GRB2 to the phosphorylated RTK. This then causes RAS–GDP recruitment to the RTK, whereby GDPase is phosphorylated to GTPase. RAS GTPase then activates RAF kinase, which starts a signal cascade similar to the PI3K pathway (Zhang and Liu, 2002). RAF then activates MEK1/MEK2 (also called MAPKK), which sequentially activates ERK1/2 (also called MAPK) (Guo *et al.*, 2020).

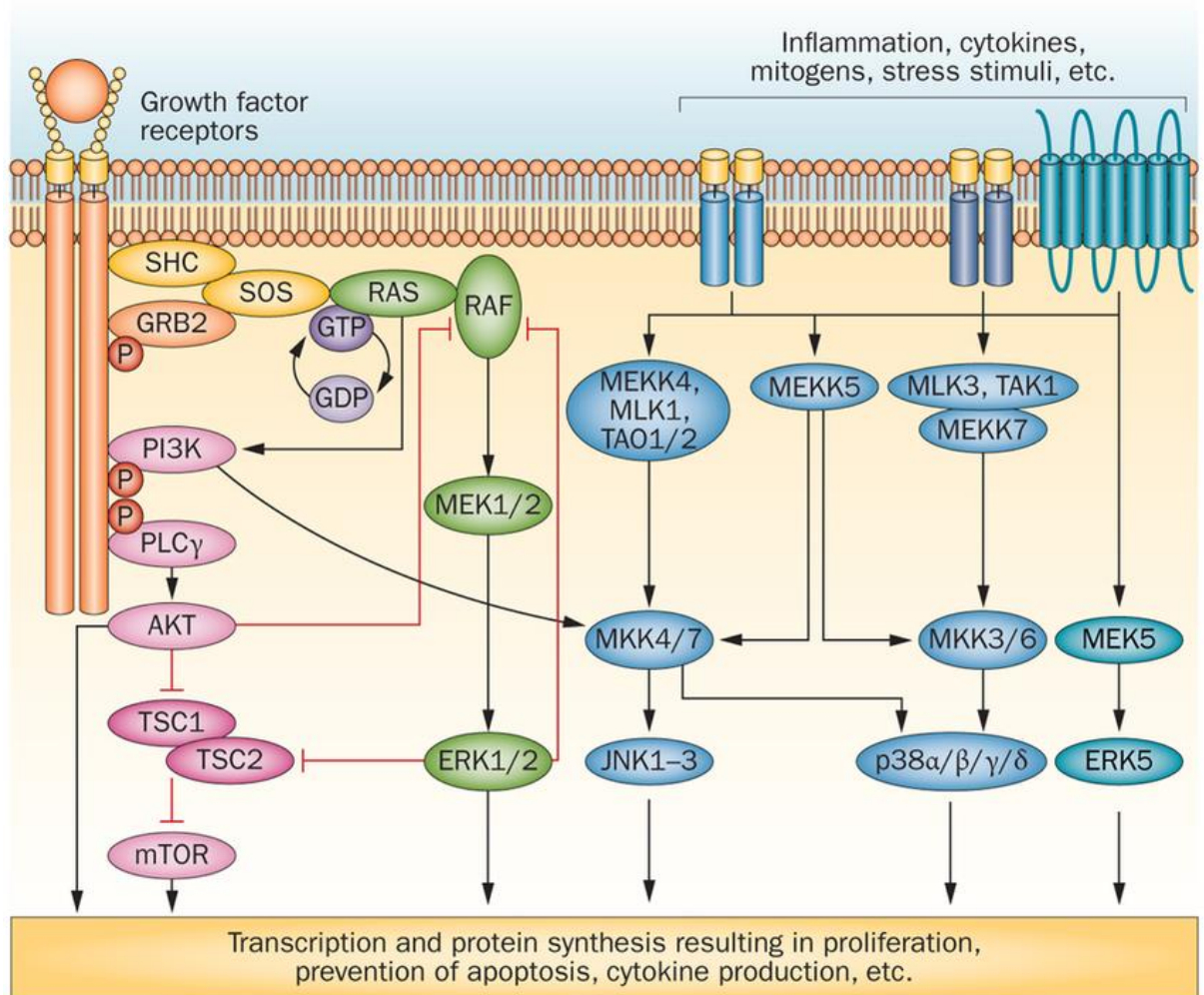


Figure 15. The MAPK/ERK pathway.

MAPK/ERK signalling begins when mitogens or growth factors bind to RTKs such as EGFR and IGFR receptors, activating them and inducing RAS recruitment. RAS begins a signalling cascade, whereby RAS phosphorylates RAF, in turn, phosphorylates MEK1/2, leading to ERK1/2 phosphorylation. From Zhao & Adjei., 2014.

MAPK: mitogen-activated protein kinase; ERK: extracellular signal-regulated kinase; RTK: receptor tyrosine kinase; EGFR: epidermal growth factor receptor; IGFR: insulin growth factor receptor.

The most observed somatic mutation in this pathway is the BRAF V600E. BRAF belongs to the RAF family kinases (Figure 15). The V600E mutation of *BRAF* has been observed in melanoma (Ascierto *et al.*, 2012), colon (Caputo *et al.*, 2019), lung (Alvarez and Otterson, 2019), ovarian (Tholander *et al.*, 2020) and breast cancers (Myers *et al.*, 2016). V600E refers to a valine substitution to glutamic acid at position 600 within the kinase domain. This mutation increases the signalling of the MAPK/ERK pathway by removing the glycine rich loop required for activation and regulation of BRAF by phosphorylation. Because of this BRAF V600E can activate MEK (MEK1/2) in the absence of RAS GTPase activation, which leads to the continual induction of aberrant proliferation (McCain, 2013). This mutation accounts for 95% of all *BRAF* mutations observed in ductal breast

cancers (Myers *et al.*, 2016). In a study of 230 breast cancers grouped as molecular subtypes, ~13% had *BRAF* V600E (Jung, Jung and Koo, 2016). It has also been observed in serous ovarian carcinomas, but to a much lesser degree than in breast cancers (Campos *et al.*, 2018). *BRAF* V600E mutations are more frequent in low-grade serous ovarian cancers (LGSOC) compared with HGSOC (Sieben *et al.*, 2004). Prostate cancers, however, have low rates of *BRAF* V600E mutations. In 43 prostate cancers, no *BRAF* mutations were identified (Köllermann *et al.*, 2010). In 200 prostate cancers, only four patients had *BRAF* V600E (Jafarian *et al.*, 2018). Therefore, in prostate cancers, *BRAF* mutations (and *KRAS*) do not seem to contribute to progression. However, the MAPK/ERK pathway is still commonly upregulated in prostate cancers and has been found to contribute to progression to CRPC (Nickols *et al.*, 2019).

Apoptosis pathways

In normal cellular function, DNA damage is normally repaired via the DDR pathways, requiring functional ATM, P53 and BRCA1/2 proteins, which frequently mutate in cancers (section 1.4). Loss of normal function of the DDR pathway due to mutations in one or more of these genes, can lead to cancer progression and tumorigenesis without sufficient genomic repair. This is the basis of genomic instability, which is frequently observed in cancers, including breast, ovarian and prostate cancer. Importantly, when irreversible DNA damage is detected, as in cancers through loss of genomic stability, a cellular stress signal is produced. This causes induction of apoptosis signalling (Figure 16), which occurs either by binding of a death receptor (DR) in the extrinsic pathway, or via mitochondria in the intrinsic pathway (Fulda and Debatin, 2006).

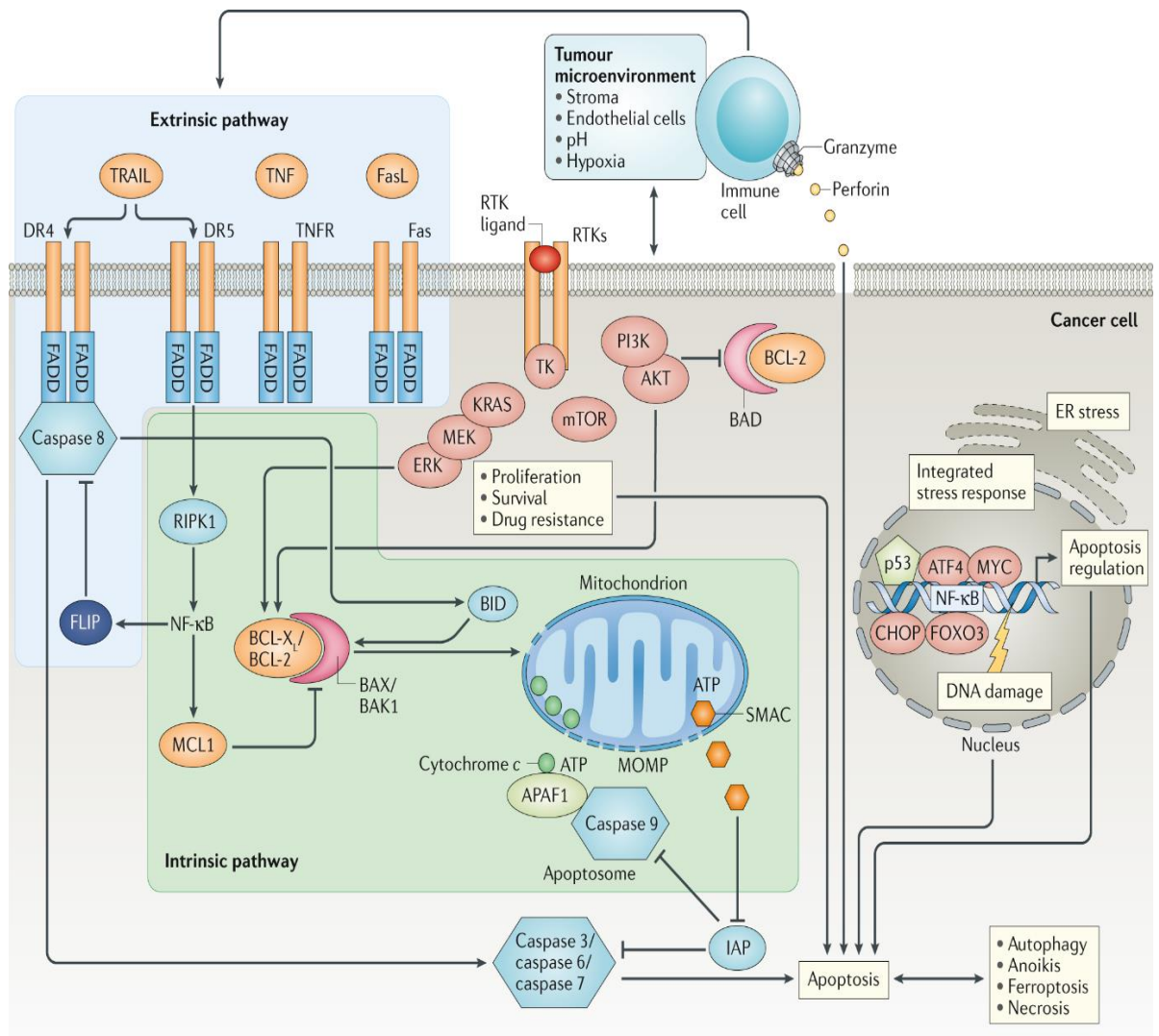


Figure 16. Apoptosis pathways.

The extrinsic pathway is activated through binding of a death receptor associated ligands such as TRAIL to death receptors 4/5, causing inactive procaspases 8 and 10 to become active caspases and subsequently activating caspase 3, 6 and 7. The intrinsic apoptosis pathway is induced through the release of cytochrome-c from the mitochondria. This leads to the formation of the apoptosome, causing procaspase 9 activation and apoptosis. The process also culminates in caspase 3 activation and apoptosis. From Carneiro & El-Deiry., 2020.

FADD: fas-associated death domain; DISC: death-inducing signalling complex.

Caspases are important downstream signalling components in both the intrinsic and extrinsic apoptosis pathways. Caspases facilitate the cleavage of a number of important substrates, for example, cytoskeletal components (caspase 1 and 3) and DNA-cleaving endonucleases (caspase 3) (Maravei *et al.*, 1997; Wolf *et al.*, 1999). Downregulation and loss of caspase 3 (*CASP3* gene) expression has been observed in breast, ovarian and prostate cancers (Devarajan *et al.*, 2002; Winter *et al.*, 2001). By altering caspase functions by mutations, which can also cause loss of

function (LOF) of genes (e.g. P53), apoptosis can be evaded in cancers. Many cancers are also able to inhibit apoptosis through the hyper-activation of proliferation pathways. For example the PI3K/AKT pathway is able to inhibit apoptosis to ensure increased proliferation through AKT phosphorylation of the BCL2- protein associated death promoter (BAD), thereby inhibiting apoptosis. AKT is also able to phosphorylate transcription factors that result in reduced expression of proapoptotic genes (Meng *et al.*, 2017).

Extrinsic apoptosis pathway and cancer

In normal cellular apoptosis, tumour necrosis factor (TNF) cytokines, such as the tumour necrosis factor apoptosis-inducing ligand (TRAIL), activate death receptors 4 and 5 (DR4/DR5). DR5 alters intracellular signalling pathways by inhibiting downstream BCL-2/BCL-X and BAX/BAK1 mitochondrial proteins which promotes apoptosis. These are also downstream targets of PI3K/AKT signalling acting to induce apoptosis.

The extrinsic pathway can also be initiated by the Fas ligand (FasL) binding to the Fas receptor or TNF binding to the TNF receptor (TNFR). All three pathways, involving ligands TRAIL (via DR4 and 5), FasL (via Fas) and TNF (via TNFR) initiate the formation of the proapoptotic death-inducing signalling complex (DISC) at the receptor through FAS-associated death domain (FADD) recruitment. DISC then recruits and activates procaspases 8 or 10 by the cleaving of aspartic acids, which in turn activate downstream caspases 3, 6 and 7. This leads to apoptotic cell death.

Downregulation of TRAIL-associated death receptors DR4 and DR5 prevents TRAIL-mediated apoptosis in breast cancer (Chandrasekaran *et al.*, 2014). In ovarian cancers downregulation of DR4 and DR5, is associated with poorer survival rates (Khaidar *et al.*, 2012), and in prostate cancers, it is associated with progression and a higher grade of tumour (Hernandez-Cueto *et al.*, 2014; Mittal *et al.*, 2015).

The *BCL2* (a mitochondrial antiapoptotic regulator) gene can be regulated by oestrogen and progesterone pathways. It has been determined that the *BCL2* gene is downregulated in P53-positive breast cancers, which are generally associated with better prognosis (Haldar *et al.*, 1994).

Interestingly, *BCL2* expression was higher with increased histological grade of breast cancers (Rostamizadeh *et al.*, 2013). However, it is worth noting that the loss of *BCL2* expression has also been related to poorer overall survival in invasive breast cancers (Hellemans *et al.*, 1995). In prostate cancer, *BCL2* has been identified as being upregulated, thereby inhibiting apoptosis. Additionally, *BCL2* promotes the progression from androgen-dependent to androgen-independent prostate cancers (Lin *et al.*, 2007; Raffo *et al.*, 1995).

Intrinsic apoptosis pathway and cancer

The intrinsic pathway is the main apoptosis pathway that is triggered in response to DNA damage or hypoxia. Following DNA damage detection by ATM, P53 initiates the intrinsic apoptosis pathway by promoting direct transcription expression of *PUMA* and *NOXA* (Aubrey *et al.*, 2018). Both *PUMA* and *NOXA* cause activation of proapoptotic BAX and BAK (Zhang, Li and Xu, 2013; Fulda and Debatin, 2006). This results in increased mitochondrial membrane permeability, causing the release of a small heme protein called cytochrome c from the mitochondria.

Cytochrome c protein plays a major part in the activation of caspases in the intrinsic pathway through apoptosome formation. The apoptosome consists of Cytochrome c, the apoptotic protease activating factor-1 (APAF-1) and procaspase 9 binding. Procaspase 9 cleavage by the apoptosome creates active caspase 9, in turn activating caspases 3, 6 and 7. In a study in which the mitochondrial DNA (mtDNA) genome was sequenced, variant m.6267G>A (causing Ala122Thr), reduced cytochrome c activity in breast, pancreatic, prostate and colon cancers inhibiting apoptosis (Gallardo *et al.*, 2006). Decreased cytochrome c gene (*CYCS*) expression leads to loss of formation of the apoptosome, thereby preventing downstream apoptosis and also cancer progression (Cai, Yang and Jones, 1998; Liu *et al.*, 2019c).

The intrinsic pathway can also be activated by paracrine signalling pathway involving the binding of growth factors (e.g. EGF) to receptor tyrosine kinase (RTK), can initiate apoptosis via KRAS/MEK/ERK or via PI3K/AKT both converging on BCLX/BCL-2 mitochondrial proteins in the intrinsic pathway leading to cytochrome-c release and apoptosis. This pathway is also where the switch occurs between proliferation and apoptosis where AKT phosphorylates both BCL-2

associated death promoter (BAD) and BAX preventing apoptosis through inhibiting cytochrome-c release (Meng *et al.*, 2017).

1.5 Similarities between breast, ovarian and prostate cancers

Currently, most cancers are studied as individual diseases with independent aetiology. This is illustrated by the fact that they are classified into their own unique subtypes. However, cancers may be more similar than previously believed. For example, molecular breast cancer classifications use receptor status (ER, PR and HER2), and ovarian epithelial subtypes use both histology and receptor status (ER and PR) for classification. As a result, these subtypes share a similar underlying trait (ER and PR receptors). Similarly, prostate cancer is primarily classified based on histology (Gleason) but an important therapeutic consideration in this cancer regards the expression of another steroid hormone receptor (AR). Whilst early prostate cancers are invariably AR+, prolonged hormonal therapies can induce a subtype of CRPC that is AR-; this subtype is treated with a different set of drugs (e.g. platinum-based therapy) and is highly aggressive. Hence the hormone receptor-based classification is important in prostate cancer. Furthermore, the three hormone receptor pathways (AR, PR, ER) are often co-expressed and show a notable degree of molecular interaction in prostate, breast and ovarian cancers (Truong and Lange, 2018).

Many studies have found identical risk factors, pathways, receptors, genes and mutations across these three cancer types. These observations indicate that the same processes are occurring across multiple cancers. These factors can be referred to as cross-cancer factors. The most frequently observed cross-cancer risk factor is family history, which has led to an increased understanding that breast, ovarian and prostate cancers are more similar than previously thought. This has led to the grouping of some ovarian and breast cancers into HBOC, based on *BRCA1* and *BRCA2* mutations. The group now also includes prostate cancers, because some cancers have mutations of the same genes (*BRCA1/BRCA2*), as well as increased risk from the same factors (age, diet, etc.).

Studies that have been specifically aimed at identifying cross-cancer similarities are rare, but a few have pointed out that some cancer subtypes are highly similar. In a study comparing 545 women with either breast or ovarian cancer, the basal-like breast cancer and high grade serous ovarian cancer (HGSOC) subtypes had a very similar aetiology despite being derived from what are currently considered independent primary carcinomas (Begg *et al.*, 2017). It is important to note that this study looked at individuals who had developed two independent primary cancers and did not compare individuals with just one primary carcinoma (e.g. an individual who had breast cancer compared with a separate individual who had ovarian cancer). Therefore, it is likely that the germinal genetic alterations in the same person may have resulted in similar genes becoming aberrant and triggering the development of two separate malignancies. The Cancer Genome Atlas (TCGA) also highlighted previous similarities between breast (basal-like subtype) and ovarian cancers (HGSOC subtype). These were found to be gains in the same genomic regions (1q, 3q, 8q and 12p) as well as losses in the same regions (4q, 5q and 8p). The expression of genes in basal-like breast cancers was highly correlated with the gene expression of serous ovarian cancers. *BRCA1* and *BRCA2* inactivating mutations as well as *ATM* mutations were more common in basal-like breast cancers and serous ovarian cancers (The Cancer Genome Atlas Network *et al.*, 2012).

Such findings are not limited to similarities between breast and ovarian cancer. Prostate and breast cancers are similar in terms of the actions of their receptors in treatment resistance (Gail *et al.*, 2010). A pan-cancer study by Campbell *et al.*, 2020 looked at 38 tumour types, including those in breast, ovarian and prostate cancers. In comparing 2,658 tumour samples (across all 38 subtypes) and matching these with normal samples, the aim was to identify underlying mutations that were common drivers in cancer genomes. It was found that on average, cancers have four to five similar driver mutations. Driver mutations provide an advantage in terms of growth and survival of the tumours. However, it was also noted that in 5% of the cases, there were no driver mutations identified (Campbell *et al.*, 2020).

In addition to the identification of common mutations, which may act as drivers in multiple cancers, the identification of cross-cancer pathogenetic genes would potentially provide a more effective target for treatment. This is because treatments would be effective for multiple cancers. Identification of genes shared among cancers or subtypes could also enable the simultaneous treatment or diagnosis of both highly aggressive and less aggressive subtypes.

Many pathways cause the same downstream aberrant outcome that benefits cancer cells. For example, increased MAPK/AKT pathway activation leads to increased cell proliferation and reduced propensity to apoptosis, in all the three cancers studied here. Identifying the genes (and proteins) underpinning important pathways that are similarly conserved in multiple cancers (and subtypes), poses a novel and more efficient method of biomarker identification. As biomarkers, these genes highlight particularly important pathways relevant to the cancers that have been retained in order to provide a particular benefit. These biomarkers may also be possible treatment targets, especially if their function is a key factor for the stability and action of the pathway. For example, if a gene is important for multiple pathways, knocking out the gene or inhibiting the protein would remove and destabilise multiple mechanisms at once – ‘one hit, multiple outcomes’. If this was applied to a tumour cell, where there was an aberrant gene functioning in multiple pathways, such as enabling the switch between evading apoptosis and initiating cell proliferation, targeting that gene with a treatment might make it harder for the cancer cell to recover and survive, and apoptosis or necrosis pathways could be initiated to kill the cell/s. In conclusion using such ‘one hit’ genes as targets could be better than treating a cancer based on a specific subtype classification alone. It could be used for a number of subtypes or cancers regardless of classification, but the aberrant gene would need to be detected across the cancers. A targeted treatment based on such a biomarker or of a particular pathway, could be applied across multiple cancers. Therefore, the similarities between these three cancers supports the need for identifying genes that are up-regulated in more than one cancer type, to provide more efficient and effective treatment options.

1.6 Bioinformatics for novel biomarker identification

Over the past 10 years there has been an increase in the development of computer programs for analysing cancer data. This has been seen alongside an increasing amount of publicly available transcriptomic and high-throughput data. Importantly, this data can be reanalysed to answer new questions and find new biomarkers. They can be analysed using pre-existing tools; however, these can limit the user to the predefined functions of the program. In order to utilise this data fully, programming languages need to be used, for example, Python, C++, C# and R. This computational approach provides flexibility for identifying new biomarkers and/or treatment targets that are directly related to the research question.

Public databases such as Gene Expression Omnibus (GEO) (Edgar, Domrachev and Lash, 2002), Sequence Read Archive (SRA) (Leinonen *et al.*, 2011), ArrayExpress (Parkinson *et al.*, 2007) and TCGA (Cancer Genome Atlas Research *et al.*, 2013) are repositories for large amounts of transcriptomic data, which are derived from microarrays and less frequently from RNA-Seq. This is because microarray technology has been available and used for a longer time than RNA-Seq, and it is more cost effective. In contrast to microarrays however, RNA-Seq can be used to identify RNA isoforms, and is not limited to the probes used on a microarray platform.

New methodologies to identify novel biomarkers have been proposed. One approach is network analysis, which has become more popular in the past two years (2017 to 2019) and is useful to dissect the molecular pathogenesis of cancers and to identify new biomarkers.

1.6.1 Network analysis for cancer biomarker identification

With increasing amounts of high-throughput data being more publicly available and accessible, network analysis is becoming more frequently used in the biological sciences. Network analysis was originally derived from graph theory in mathematics to identify the relationship between pairwise objects. An object can be a gene (or a protein) for example, and in pairwise, one gene can be compared to another.

Network biology can be applied to analyse multiple genes in an environment, rather than using the single-gene approach, and there are many benefits of doing this. For example, determining if the expression of one gene is related to the expression of another gene, in a pair-wise comparison. The accessibility of more available data means more combinations of genes or proteins can be robustly analysed. This method is extremely useful for the mapping of highly complex data sets, which often contain the expression levels of thousands of genes. Network analysis allows the core features/genes of these data sets to be identified more clearly and concisely. Core features are those elements of a biological network that are important for the functioning of the overall network. For example, a core feature may be a gene that is important in connecting many other genes in the network. Many of the core features found in network analysis may therefore prove to be novel biomarkers across cancers or targets for treatment as they have the potential to be connected to multiple pathways. These important targets could otherwise be missed if studied alone, as they may not be among the most significantly upregulated or downregulated genes, or the patterns/combinations of genes may not be identified in the mass of data. Identification is carried out using the network topology (layout/organisation) to pinpoint overrepresented or recurring network motifs (recurring patterns/structures) within the larger network or across multiple networks from different conditions or states. For example, in comparing one gene at the centre of two networks, each from a different cancer subtype, it can be very difficult to visualize patterns or links to the most important genes; network topology helps in visualising the important links and patterns more clearly.

The basic structure of a network map consists of two objects. The first object is a node, which is usually represented as a circle, depicting a gene. The second object is an edge, representing a predicted interaction within the network, commonly represented as a line between two nodes/genes (pairwise interactions). Depending upon the data, the research aim, method, approach, or statistical test used, an 'interaction' could represent a correlation between the two genes, their protein products or their function. In this thesis, a statistical test of association between gene expression data was used to define interaction. This was to identify whether the

expression of one gene is associated with the expression of another, and whether expression levels are up- or down-regulated in different conditions (e.g. cancer vs normal tissue). This principle was used multiple times to analyse thousands of differentially expressed genes (DEGs). Therefore 'interaction' was used to indicate 'an association between DEGs', but not to identify the functions of the genes at this stage.

Nodes/genes and edges/interactions that cluster together form a network module, which is a group of nodes/genes that do not necessarily interact directly but have categories of functions in common. Because of this attribute, modules can be used in functional analyses to identify the potential functions of a group of associated genes. This is a useful determinant for finding potential functions of genes that may not have functional information available. The functions of such unknown genes can be predicted using a 'guilt by association' approach, which determines the function of a gene from the known functions of the other genes in the same module. The interacting DEGs can therefore be further grouped into functions/cell processes and into network modules. This provides biological information for sets of genes, which are more interpretable and applicable results when dealing with novel genes in particular and is an alternative to the traditional method of providing a list of the most significant genes that are up or down regulated in a study. This traditional method would only identify DEGs; genes where the expression is greater or lower than expected compared to the normal wild type gene. Such a list does not identify the interactions between these lists of genes or even further, any patterns/networks/modules and can lead to biased identification that prioritises one function over others or gives no clear insight as to the function of genes. Including genes that are downregulated in a module would provide a more complete and possibly more accurate set of functions.

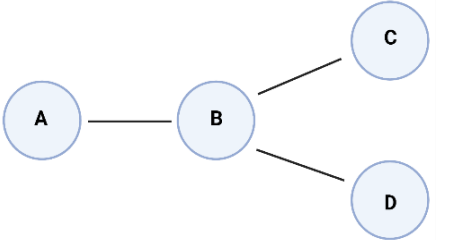
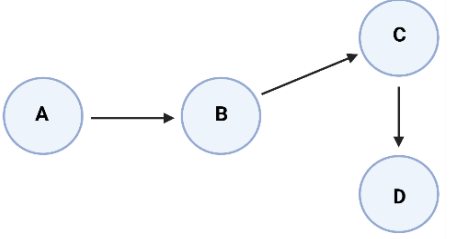
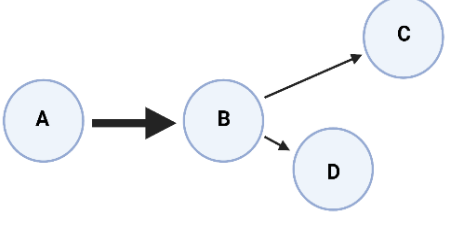
Importantly, network analysis also allows for the identification of subnetworks. These could be smaller, isolated networks that are independent from the larger network. They may also be nodes that are more connected to each other internally than they are to the larger network.

Subnetworks can represent isolated processes that may be important for the overall function of

the network, for example, gene regulation through the activation or modification of a transcription factor, which may then cause the expression or suppression of another gene important for the function of the network.

There are three main types of network that can be constructed: undirected, directed and weighted. Each of these can identify different associations and results. Table 4 summarises each of the network types and their interpretation.

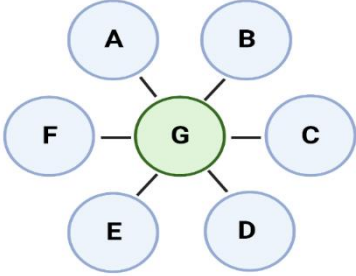
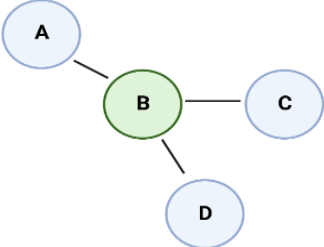
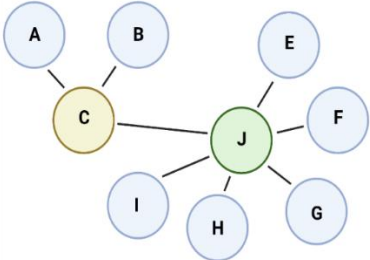
Table 4. The three most common network types used to construct networks.

Network type	Interpretation	Example
Undirected	These networks do not provide an implicit direction or sequence of interaction between the nodes. For example, nodes A and B interact but the interaction may be in either direction and does not define an order. In this thesis, it means that two DEGs (A and B) interact/are associated, and so on, as other genes are added (C and D), C and B interact, and D and B interact.	
Directed	These networks provide a direction of interaction between the nodes. For example, A interacts with B, then B interacts with C, then C interacts with D.	
Weighted	These networks can be either undirected or directed depending on the data analysed but in addition, the strength of association between two nodes is taken into account. For example, this can be depicted by shorter edges between two nodes, suggesting a higher weighted (stronger) interaction. This may also be represented as a thicker edge between the nodes rather than a thinner edge, the thicker edge also depicting a stronger association.	

Undirected networks do not provide the context in which the order of the nodes interact with one another via edges. Directed networks provide additional information as to how the order of the nodes interact. Weighted networks provide the 'strength' of an interaction. Both undirected and directed networks can also be weighted to provide additional strength of interaction. 'Strength of interaction' meaning a greater statistically significant association.

When analysing networks, the importance of a node (gene) and its associated edges is determined through 'centrality' measurements. There are multiple centrality measurements, the most common being 'closeness centrality', 'betweenness centrality' and 'degree centrality' (Table 5). In this thesis 'centrality' determines the patterns between nodes and how a network may be connected to another subnetwork. It identifies how the edges/interactions are closer between nodes/genes and identifies a node/gene that is mostly interacting to all other genes in the network. 'Closer' edges correspond to genes that are more significantly associated.

Table 5. Network centrality measures.

Centrality measurement	Interpretation	Example
Closeness centrality	Identifies a node that can be in the centre of a network due to how closely it interacts to other nodes. It determines the average length of the shortest edge between nodes, and identifies a node that interacts with all other nodes but has the shortest edges to all nodes. Therefore, the more central a node is in a network, the closer it is to all other nodes. In this example, node G has the highest closeness centrality in the network because it is closest to all the other nodes.	
Betweenness centrality	Identifies how nodes (blue) can be connected by a single node (green) by determining the number of the shortest paths that go through an edge. For example, node B lies on the shortest path to both nodes C and D from node A.	
Degree centrality	Determines a count of the number of edges that connect to a central node. These nodes are more likely to cause noticeable effects if lost, because they are important for the overall network connections and may mediate many functions. For example, node C has three edges, whereas node J has six and, therefore, has the highest degree of centrality in the network. Therefore node J is most important followed by node C in connecting all other nodes.	

Each method identifies different nodes in the network. Closeness centrality identifies nodes that have the shortest edges between the nodes in a network. Betweenness centrality specifically identifies the relationship/pathway between nodes. Degree centrality identifies crucial nodes, those which have more edges to more nodes; these nodes are highly interconnected within the network, i.e., they have a stronger association with other nodes.

The centrality measurement used in network analysis is an important factor to consider, because it can lead to different results from the same data set. This is particularly important for the identification of potential targets. There is no defined method of centrality measurement for cancer networks, and the method used would depend on what type of biomarker or target the study was aimed at identifying. For example, to identify a biomarker that may represent a certain functional pathway, closeness centrality may be appropriate, because it would identify genes linked within the pathway/network. It would identify which genes with high closeness centrality lie on the shortest distance between many nodes and, therefore, provide the majority of pathway activity (but not all). However, to identify which genes act as potential treatment targets, then degree centrality would be more appropriate. This is because the nodes identified by this method are the most highly connected to all other nodes and, therefore, more likely to have an effect on the overall network if altered, e.g. if you inhibited the crucial nodes/genes then you could affect the connected genes that may operate in other pathways.

1.6.2 Network analysis in cancer research

Network modelling is a promising approach to characterise cancer biology (Pe'er and Hacohen, 2011). In a similar manner to differential gene expression (DEG) analysis, differential network analysis is being utilised more (Ferreira *et al.*, 2021). It involves the identification of key network topological (structural) changes between test conditions. Network modules for example can be further annotated with relevant information to identify the known functions or novel functional changes between conditions (e.g. cancer vs normal tissue). Functional annotations, such as gene ontology (GO), provide categories of cellular pathways in which the genes function, both in a network and in a cancer subtype. Gene network modules are robust measurements in functional pathway analysis (Grimes, Potter and Datta, 2019). There are two main biological networks used in cancer research: gene–gene interaction (GGI) networks and protein–protein interaction (PPI) networks. Both of these network types are useful for the identification of novel biomarkers. Importantly, both GGI and PPI networks can be scale free, meaning that a small number of the nodes can be highly connected (high-degree centrality) to many other nodes. This feature enables

a network to be produced with only the nodes of importance viewed, taking out lots of other not so important genes. Such important nodes may be for example, a DNA repair gene (a central highly connected node) that recruits further genes (other nodes) or transcription factors (other nodes) to a promoter complex to alter transcription of a target gene (another central highly connected node) in GGI; or a protein that is the main scaffold for a complex in PPI. Targeting highly connected nodes should result in the collapse of the whole molecular network. This makes the highly connected nodes potential treatment targets and/or biomarkers.

GGI/GCN networks

GGI networks, which are also referred to as gene co-expression networks (GCNs), are biological networks that are aimed at the identification of biomarkers and their potential functions. They are often used to identify and map gene regulatory networks that control the expression of mRNA and proteins.

For example, in a GGI network analysis of 498 neuroblastoma tumour samples, 5 of the top 10 pathways identified were related to cell proliferation, whilst the other pathways were related to cell survival and apoptosis. Based on these results, it was proposed that underlying mutations in the genes may have altered the GGI resulting in metastatic growth (Grimes, Potter and Datta, 2019). In a comparison between a GGI network analysis and a DEG analysis of melanoma-infiltrating lymphocytes, seven DEGs were identified. The GGI network analysis found that many of the genes were significantly associated with modules that responded to chemokine ligands, and that the chemokine signalling pathway was downregulated in melanoma-infiltrating lymphocytes (Nacu *et al.*, 2007).

PPI networks

PPI networks are biological networks that show the relationship between the proteins of the cell(s) being analysed. Such networks are useful for identifying cell signalling pathways, determining putative functions of uncharacterised proteins or identifying the relationships between multiple protein complexes. The majority of PPI networks are scale free and are a useful tool to discover new biomarkers. In a study of 182 oesophageal adenocarcinomas (EACs), 27

proteins were important nodes and identified as biomarkers using multiple centrality measurements (betweenness, closeness, and degree centrality) (Rezaei-Tavirani *et al.*, 2017).

Caveats when using networks

The main limitation of network analysis is that it is highly dependent on the number of samples available. Networks require large sample sizes to produce accurate, robust and meaningful results. In the construction of biological networks, the larger the data set (often over 500 samples), the better the accuracy and reliability of the network (Liesecke *et al.*, 2019; Hevey, 2018). It has also been found that microarray data outperforms RNA-Seq data for network analysis, in terms of network robustness using identical sample sizes. Microarray derived networks also show more accurate Gene Ontology (GO) mapping than RNA-Seq, with more predicted interactions from microarray networks being observed for known GO interactions than RNA-Seq networks (Liesecke *et al.*, 2019).

Limitations in the functional annotation of networks must also be taken into account. Functional annotation can be biased towards more well-known and studied pathways (Charitou, Bryan and Lynn, 2016). This can make network modules that have many novel genes difficult to annotate reliably.

1.6.3 Hub genes

Networks can consist of multiple modules (often hundreds), and each module can also consist of hundreds or potentially thousands of genes. Not only can a network module provide an 'image' of the molecular environment being analysed (for example, a tumour or lesion), it can also provide information on which genes may be involved in a certain function. This requires the identification of key components of modules that may contribute to the overall function of the network.

Centrality analysis of GCNs is useful for the identification of key components of complex biological processes (Koschützki and Schreiber, 2008).

The key components of a network module are those areas where many genes are predicted to interact; therefore, these are considered to be 'hubs' (Goymer, 2008). Hubs can be created if a

single gene (or a protein in a PPI network) is involved in the actions of many other genes. For example, a transcription factor may interact with and regulate many different targets via an edge. Genes that have high connectivity (high-degree centrality) are termed hub genes (or hub proteins in a PPI network). Hub genes are key components in network modules because they play a pivotal role in the overall structure and function of the network. The way in which hub genes facilitate many of the interactions in a module and in the overall network means that their identification is becoming an important factor in cancer research. This is because they could potentially be used as reliable biomarkers and novel treatment targets.

Hub genes as biomarkers

Hub genes have been found to be useful biomarkers in cancer research. Identification of hub genes is commonly performed using weighted gene co-expression network analysis (WGCNA). These networks identify modules of co-expressed genes, which will normally correspond to a certain cellular pathway or pathways, using transcriptomic data. WGCNA hub gene analysis has been significantly correlated with phenotypic traits; for this reason, hub genes are often promising cancer biomarkers.

Hub gene biomarkers can also play a useful role as representatives of cellular pathways, molecular functions or processes. Instead of determining the expression levels of multiple genes for a pathway or process, a single hub gene can be used instead, making it a useful biomarker. This approach is more efficient; since hub genes have a high number of connected nodes (degree centrality), they will capture a large portion of the pathway or process interactions. This can provide more intelligible results from a larger and more complex gene co-expression network. By exploiting this characteristic of hub genes, scientists can develop a useful prognostic score. For example, in an analysis of breast cancer, four hub proteins (CENPL, ISG20L2, LSM4 and MRPL3) were identified. These four hub proteins (also termed hub genes) were found to be associated with poorer prognosis in patients with increased expression (Yin *et al.*, 2021). In a study of bladder cancer six genes were identified as hub genes which were also found to be significantly associated with reduced overall survival (Yan *et al.*, 2019). Hub genes can also be used to be biomarkers for

particular pathways. A study of three gynaecological cancers (cervical, endometrial and vulvar), seven hub genes were common amongst the three cancers and they were significantly associated with functioning in the PI3K/AKT pathway. This discovery led to these seven hub genes being proposed as diagnostic biomarkers and potential treatment targets for these cancers that are dependent on PI3K/AKT pathway activation (Liu *et al.*, 2019b).

Hub genes as treatment targets

Hub genes are also potential novel treatment targets. This is because they are highly important for the stability of the network due to their high-degree centrality. Targeting a single hub gene would affect a large proportion of connected genes and interactions, thereby amplifying the phenotypic effects of a therapy. This is in comparison to just targeting a single gene that is less interconnected with the functions of the network. A non-hub gene may be expressed in large numbers in a carcinoma, but knockdown or pharmacological inhibition of this gene may have no direct effect on the carcinoma's development or progression because the gene has only one or two interactions in the network. However, knockout or silencing of hub genes is more likely to result in the loss of network stability, leading for example to apoptosis, because many gene interactions are affected simultaneously.

The targeting of hub genes may prove to be far more effective than other treatments. Mutations occurring in cancer cells exposed to prolonged treatments (e.g. chemotherapy) can lead to the development of drug-resistant clones and to clinical relapse (D'Alterio *et al.*, 2020). This can allow for primary or secondary recurrence (e.g. metastasis), which has been observed in breast, ovarian and prostate cancers that have already been treated with chemotherapy (Rivera and Gomez, 2010; Pokhriyal *et al.*, 2019; Lohiya, Aragon-Ching and Sonpavde, 2016). Hub genes as treatment targets could allow for more direct and effective treatment. It would be expected that targeting one hub gene specifically in a pathway, would have a cascading destabilisation effect of changes affecting its multiple interacting genes, and triggering apoptosis through a DNA damage response. This will make it more difficult for a cancer cell to develop a resistance mechanism over a much shorter time period.

Because of the benefits of network construction from biological data, hub genes have been found in many different cancers. As a result, more effective treatment options, which can overcome current issues with potential resistance, could be offered. In addition, novel treatments for multiple cancers could be provided, because hub genes have been identified as functioning in more than one cancer, thus adding to their value as potential therapeutic targets. One example is the *CCNB1* hub gene, which has been identified in both pancreatic cancer (Zhou *et al.*, 2018b) and oesophageal squamous cell carcinoma (Yang *et al.*, 2019). In both studies, *CCNB1* was associated with reduced survival rates. In a separate pancreatic cancer study, the silencing of *CCNB1* reduced cell proliferation and promoted apoptosis (Zhang *et al.*, 2018).

A summary of the most recent hub gene papers published is shown in Table 6. Identical hub genes, for example, *CCNA2*, have been found in multiple cancers: pancreatic (Zhou *et al.*, 2018b); oesophageal squamous cell (Yang *et al.*, 2019); multiple gynaecological cancers (Liu *et al.*, 2019b); and breast cancer (Cai *et al.*, 2019a). *CDK1* is a hub gene that has also been found in multiple cancers: oesophageal (Yang *et al.*, 2019); and gynaecological (Liu *et al.*, 2019b). These results demonstrate that cancers often express the same hub genes to activate key phenotypes (e.g. evading apoptosis or promoting proliferation).

Table 6. A summary of papers published on hub genes.

Hub genes	Cancer	Clinical relevance	Reference
<i>CCNA2, CCNB1, CENPF, DLGAP5, KIF14, KIF23, NEK2, RACGAP1, TPX2, UBE2C</i>	Pancreatic cancer	Prognostic and diagnostic	(Zhou <i>et al.</i> , 2018b)
<i>CDK1, CCNB1, TOP2A, CCNB2, BUB1, CCNA2, NCAPG, AURKB, NDC80</i>	Oesophageal squamous cell cancer	Prognostic and diagnostic	(Yang <i>et al.</i> , 2019)
<i>COL4A1, VCAN, THBS2, TIMP1, COL1A2, SERPINH1, COL6A3</i>	Gastric cancer	Prognostic	(Li <i>et al.</i> , 2018)
<i>CDH11, COL3A1, COL6A3, COL5A1, AEBP1, COL1A2, NTM, COL11A1, THBS2, COL8A1, COL1A1, BGN, MMP2, PXDN, THY1, TGFB11</i>	Bladder cancer	Prognostic and diagnostic	(Di <i>et al.</i> , 2019)
<i>CASC5, CKAP2L, FAM83D, KIF18B, KIF23, SKA1, GINS1, CDCA5, MCM6</i>	Breast cancer	Prognostic and diagnostic	(Fu <i>et al.</i> , 2019)
<i>CCNB2, CDC20, CEP55, TOP2A, KIF20A, UBE2C</i>	Renal cancer	Prognostic and diagnostic	(Xiao <i>et al.</i> , 2019)
<i>PID1, ABCA6, TFPI, MPPED2, RP6KA6, MRO, RMDN1, ACACB, SLC4A4, TTC30A, RNF150, BCL2, CASC2, PRKCQ, SLC26A7, ITPR1</i>	Papillary thyroid cancer	Prognostic	(Liu <i>et al.</i> , 2020)

Hub genes	Cancer	Clinical relevance	Reference
<i>CCNA2, CDK1, CCND1, FGF2, IGF1, BCL2, VEGFA</i>	Gynecological cancers (cervical, endometrial and vulvar)	Diagnostic	(Liu <i>et al.</i> , 2019b)
<i>CA4, PCAM1, DNAJB4, AGER, GIMAP6, C10orf54, DOCK4, GOLM1, PFAH1B3</i>	Lung adenocarcinoma	Prognostic and diagnostic	(Yu <i>et al.</i> , 2020)
<i>TPX2, KIF2C, CDCA8, BUB1B, CCNA2</i>	Breast cancer	Prognostic and diagnostic	(Cai <i>et al.</i> , 2019a)

Identification of hub genes has been found to be a robust method of identifying many novel genes that can be used as biomarkers for prognosis or diagnostics in multiple cancers. One example being CCNA2 which has been identified as a prognostic and diagnostic hub gene in breast, gynecological and pancreatic cancer.

1.7 Literature summary, and research aims

The molecular pathogenesis of breast, ovarian and prostate cancers shows multiple similarities. The most frequently observed similarity is the utilisation of the same hormone receptors for cancer initiation and progression. However, these three cancers also exhibit hormone-independent subtypes in the most advanced phases of cancer progression. In addition, these cancers often carry similar mutations in proliferation and apoptosis related pathways. Ultimately, these mutations lead to the same changes and alterations in oncogenic and onco-suppressive pathways.

Despite the similarities observed, the classification of these three cancers is based on cancer-specific features. Very few studies have looked at comparing cancer subtypes to identify similarities that are conserved across different tissues. Previously, basal-like breast cancers have been found to be similar to serous ovarian cancers in terms of their aetiology (Begg *et al.*, 2017). It has also been found that when comparing three gynaecological cancers (cervical, endometrial and vulvar), important hub genes from network analysis were identified as being common to the pathogenesis of all three cancers (Liu *et al.*, 2019b); providing potential biomarkers. The novelty of comparing cancer subtypes and identifying hub genes can allow for more efficient identification of biomarkers and treatment targets. Although this process is primarily aimed at identifying genes or proteins that are important across cancers, it can also identify what is cancer specific. This is important in diagnosing cancer and distinguishing subtypes with specific biomarkers, and improve classification of tumours that may have features belonging to more than one classification. Examples of this are the normal-like and luminal A breast cancers.

There is now a large range of transcriptomic data that are publicly available. The majority of these are derived from microarrays. However, network analysis requires significant sample sizes in order to ensure robust network generation and hub gene identification. These individual data sets may not contain sufficient sample sizes to meet the requirements for network analysis. A potentially novel approach to this issue is to integrate suitable data sets, thus allowing for

increased sample sizes. Currently, there is no known standard integration method for combining microarray data sets.

Therefore, the research question for this thesis was:

Can hub genes common in breast, ovarian and prostate cancers be identified for use as potential biomarkers or treatment targets?

1.7.1 Aims

1. Develop a method/workflow to integrate publicly available microarray gene expression data from breast, ovarian, and prostate cancer subtypes.
2. Use this approach to identify differentially expressed genes (DEGs), and of those, further identify novel candidate hub genes common across the cancer subtypes.
3. Determine whether the novel candidate hub genes identified, may be useful as potential biomarkers and/or treatment targets for these cancers.

1.7.2 Objectives

To address Aim 1:

1. Develop an R workflow using publicly available microarray-based expression datasets of primary breast, ovarian, and prostate cancers to:
 - a. Perform quality control (QC) and standardise data.
 - b. Integrate and impute missing values accurately.
 - c. Compute differentially expressed gene (DEG) analysis via statistical analysis of cancers compared to normal samples.
 - d. Conduct Kaplan-Meier survival analysis on DEGs.
 - e. Construct gene co-expression networks and extract the DEGs of interest.
 - f. Perform Gene Ontology (GO) analysis on the extracted co-expression networks.

To address Aim 2:

2. Identify the most significant and altered DEGs commonly expressed in the three cancers.

3. Of the most upregulated and altered DEGs, identify those that are associated with patient survival outcomes using recurrence free survival (RFS) and overall survival (OS).

Downregulated DEGs were also identified but not taken forward, as there were too many genes to analyse.

4. Interpret the gene co-expression networks and GO cellular pathway analyses of the DEGs identified in objective 3, to identify candidate hub genes.

To address Aim 3:

5. Review the literature as to the potential mechanism of action of the novel candidate hub genes and determine whether they are involved in similar functions and may be useful as biomarkers or treatment targets.

Chapter 2. Methodology

2.1 Overview of analyses and workflow

In order to answer the research question, the first stage of this study (aim 1) required the collection of suitable microarray gene expression data from public databases. The data was then pre-processed, and quality control checks were performed (section 2.3) where poor quality data sets were removed.

Differential gene (RNA) expression analysis was conducted using a Bayesian modified t-test comparing cancer subtype with corresponding normal tissue (aim 2). Significantly upregulated DEGs were compared across breast, ovarian and prostate cancers using the current classification systems (section 1.3). These classifications were: histological and molecular for breast cancers, tissue location and epithelial ovarian subtypes for ovarian cancers, and Gleason grade group and Gleason score for prostate cancers.

The DEGs that were identified as being significantly upregulated across (and common in) each of the three cancers were termed cross-cancer genes, and the analysis was referred to as cross-cancer analysis. These cross-cancer genes were further analysed to determine if the level of gene expression was associated with patient survival in breast, ovarian and prostate cancers (aims 2 and 3).

The term cross-cancer is used as only three cancers were investigated and compared, although, cross-cancer can be considered a specific derivative of pan-cancer. It is termed cross-cancer rather than pan-cancer as mutations are not also analysed or identified. Therefore, the cross-cancer analysis conducted here is not considered a 'true' pan-cancer analysis. Future work could include identification of mutations/variants, which is more in keeping with a pan-cancer analysis.

Genes that were significantly associated with reduced survival were further analysed via network analysis. This was to determine if any of the cross-cancer genes were hub genes and could, therefore, serve as potential biomarkers or treatment targets (aim 3). The analysis also allowed

the potential identification of the function of these hub genes in the cancer subtypes. A flowchart depicting the objectives and stages of the work is shown in Figure 17.

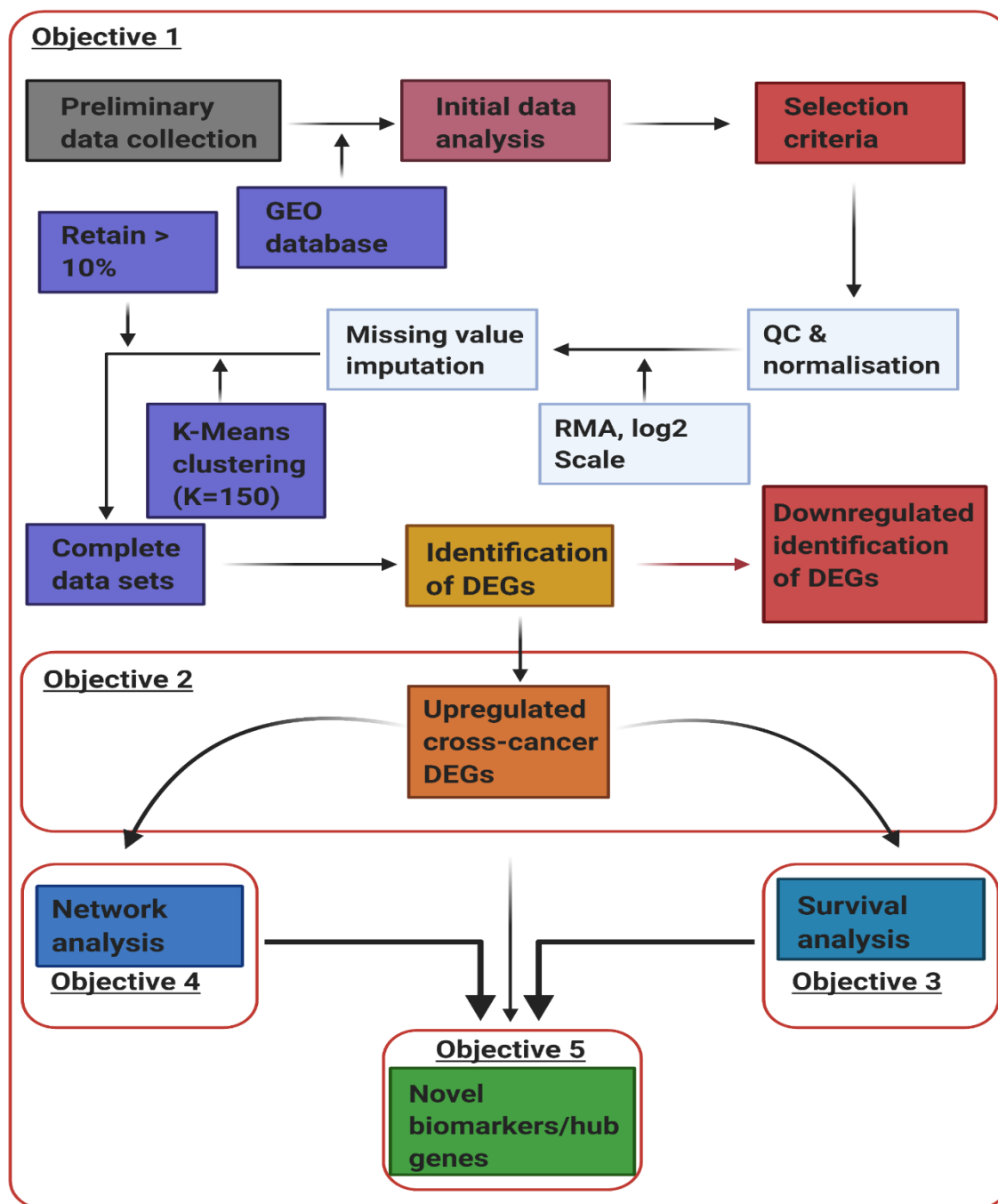


Figure 17. Flowchart depicting objectives and stages of the work to achieve the aims of the study.

Each of the stages of analysis is described in more detail in the following sections (sections 2.2-2.11). In summary, data sets were collected from the GEO database (Edgar, Domrachev and Lash, 2002). The datasets underwent quality control checks, and were then pre-processed to identify DEGs. In a comparison of the three cancers and subtypes, DEGs that were found in all three were selected for further analysis. Survival analysis was conducted to determine the relevance of significant DEGs for patient survival (section 2.10), and network analysis was carried out to identify the functions of candidate DEGs (section 2.11).

DEGs: differential gene expressions; GEO: Gene Expression Omnibus.

Downregulated genes were identified in this workflow, however, they were not taken forward and analysed further. This was due to the large number of both upregulated and downregulated genes identified in the DEG analyses. The cross-cancer analysis therefore focused on the upregulated genes only.

2.2 Curation of microarray data sets

As per aim 1 (objective 1) (Figure 17), publicly available tissue microarray data sets containing RNA expression levels of genes for breast, ovarian and prostate cancer studies were collected from the National Centre for Biotechnology Information (NCBI) GEO database (Edgar, Domrachev and Lash, 2002; Barrett *et al.*, 2013), which was selected because of its use of the MIAME (minimum information about microarray experiments) submission format. This format has been set as the standard for microarray experiment curation to ensure reproducibility and validation of analyses in microarray experiments (Brazma *et al.*, 2001). It requires submission of data to the NCBI as follows:

- raw unprocessed/non-normalised data (in CEL file format).
- processed data (normalised).
- sample annotation describing suitable treatment (i.e. dosage).
- experimental design information that accurately identifies which samples belong to which raw data file.
- sufficient array annotation for probes, either in the form of pre-annotated genes, or probe sequences for reannotation.
- information on how the raw data files were processed to produce the final normalised data.

The raw unprocessed/non-normalised data sets for relevant studies were downloaded (in CEL format) using the GEOQuery package (Davis and Meltzer, 2007) in R (R Core Team, 2017). This

allowed for consistent quality control assessment and downstream processing (background correction and normalisation) of the data (section 2.4).

An initial preliminary analysis was conducted prior to identifying DEGs in aim 2 (objectives 1 and 2). This was to assess the capabilities and limitations of each of the different microarray platform providers.

Following initial testing and preliminary analyses, three main microarray platform manufacturers were identified: Affymetrix, Agilent and Illumina. Illumina bead chip arrays were excluded because of inconsistencies found with probe measurements across the different array versions (whole genome versions 2 to 4). These microarrays made normalisation and comparison unreliable in comparison with the other microarray platforms. The Agilent arrays were also excluded because of difficulties with identifying and labelling probes that were not publicly available. As such, it was not possible to confirm the annotation of probe sequences to genes. The remaining arrays were from Affymetrix and, therefore, inclusion criteria for the data set were developed as follows:

- 1) Affymetrix platform – U95 onwards.
- 2) samples consisting of primary tumour tissue and/or normal tissue.
- 3) samples consisting of at least one relevant breast, ovarian or prostate subtype.
- 4) sample metadata with sufficient information to identify samples and subtypes.

2.2.1 Breast cancer data sets

For primary breast cancers, a total of 2,070 samples from across 24 studies were identified. The studies and platforms that met the selection criteria are shown in Table 7 for both histological and molecular classified data sets.

Table 7. Breast cancer histological and molecular studies, including normal samples identified from the GEO database.

	GEO identifier (study)	Total no. samples/applicable samples	Microarray platform	Study reference
Histological				
1	GSE21422	19	Affymetrix Human Genome U133 Plus 2.0 Array	(Kretschmer <i>et al.</i> , 2011)
2	GSE5764	30	Affymetrix Human Genome U133 Plus 2.0 Array	(Turashvili <i>et al.</i> , 2007)
3	GSE20194	278/253	Affymetrix Human Genome U133A Array	(Shi <i>et al.</i> , 2010; Wang <i>et al.</i> , 2014)
4	GSE65194	178/164	Affymetrix Human Genome U133 Plus 2.0 Array	(Maire <i>et al.</i> , 2013; Wang <i>et al.</i> , 2014)
5	GSE14548	66	Affymetrix Human X3P Array	(Ma <i>et al.</i> , 2009; Shi <i>et al.</i> , 2010)
6	GSE88770	117	Affymetrix Human Genome U133 Plus 2.0 Array	(Metzger-Filho <i>et al.</i> , 2013; Maire <i>et al.</i> , 2013)
7	GSE41194	50/26	Affymetrix Human Genome U95 Version 2 Array	(Lee <i>et al.</i> , 2012)
8	GSE41196	14	Affymetrix Human Genome U95 Version 2 Array	(Lee <i>et al.</i> , 2012; Metzger-Filho <i>et al.</i> , 2013)
9	GSE41197	18	Affymetrix Human Genome U95 Version 2 Array	(Lee <i>et al.</i> , 2012; Lehmann <i>et al.</i> , 2011)
10	GSE41198	22	Affymetrix Human U133 X3P Array	(Lee <i>et al.</i> , 2012)

	GEO identifier (study)	Total no. samples/applicable samples	Microarray platform	Study reference
Histological				
11	GSE41227	22	Affymetrix Human U133 X3P Array	(Lee <i>et al.</i> , 2012)
12	GSE76124	198	Affymetrix Human Genome U133 Plus 2.0 Array	(Burstein <i>et al.</i> , 2015; Lee <i>et al.</i> , 2012)
13	GSE9574	29	Affymetrix Human Genome U133A Array	(Tripathi <i>et al.</i> , 2008; Lee <i>et al.</i> , 2012)
14	GSE10780	185	Affymetrix Human Genome U133 Plus 2.0 Array	(Chen <i>et al.</i> , 2010; Burstein <i>et al.</i> , 2015)
15	GSE76274	67/62	Affymetrix Human Genome U133 Plus 2.0 Array	(Burstein <i>et al.</i> , 2015)
Total no. histological samples		1,224		
	GEO identifier (study)	Total no. samples	Microarray platform	Study reference
Molecular				
16	GSE31448	357/294	Affymetrix Human Genome U133 Plus 2.0 Array	(Sabatier <i>et al.</i> , 2011; Chen <i>et al.</i> , 2010)
17	GSE65194	178/164	Affymetrix Human Genome U133 Plus 2.0 Array	(Maire <i>et al.</i> , 2013)
18	GSE29431	66/53	Affymetrix Human Genome U133 Plus 2.0 Array	(Cuadros <i>et al.</i> , 2011; Sabatier <i>et al.</i> , 2011)

	GEO identifier (study)	Total no. samples	Microarray platform	Study reference
Molecular				
19	GSE1456	318/139	Affymetrix Human Genome U133A Array	(Hall <i>et al.</i> , 2006; Maire <i>et al.</i> , 2013)
20	GSE20711	90	Affymetrix Human Genome U133 Plus 2.0 Array	(Dedeurwaerder <i>et al.</i> , 2011; Cuadros <i>et al.</i> , 2011)
21	GSE43837	38	Affymetrix Human X3P Array	(McMullin <i>et al.</i> , 2014; Hall <i>et al.</i> , 2006)
22	GSE12763	30/12	Affymetrix Human Genome U133 Plus 2.0 Array	(Stinson <i>et al.</i> , 2011; McMullin <i>et al.</i> , 2014)
23	GSE3744	47/27	Affymetrix Human Genome U133 Plus 2.0 Array	(Richardson <i>et al.</i> , 2006; Burstein <i>et al.</i> , 2015)
24	GSE9574	29	Affymetrix Human Genome U133A Array	(Tripathi <i>et al.</i> , 2008; Stinson <i>et al.</i> , 2011)
Total no. molecular samples		846		
Total no. breast cancer samples		2,070		

All studies are from Affymetrix platforms. Both the total number of samples in each dataset and the number of relevant samples taken forward for analysis are shown. For example the breast histological dataset GSE20194 contained 278 samples, however 25 of these were not found to meet the selection criteria. This was because they could not be assigned to a specific subtype. As such the 25 samples were excluded from analysis meaning that only 253 remained.

GEO: Gene Expression Omnibus.

Breast cancer histological subtypes

Of the total 2,070 breast cancer samples, 1,224 were histologically graded into three subtypes and normal tissue: IDC (741), ILC (147), DCIS (99) and normal tissue (236). The LCIS subtype was excluded from subsequent analysis because of insufficient sample size. Table 8 shows the sample sizes and subtypes of each of the breast cancer data sets from the GEO database analysed in this thesis.

Table 8. Breast cancer histological subtypes and normal sample sizes.

Histological subtypes	No. of samples
Normal	236
IDC	741
ILC	148
DCIS	99
LCIS	excluded – low sample size (2)
Total	1,224

Invasive subtypes are IDC and ILC. Non-invasive subtypes are DCIS and LCIS. The LCIS subtype was excluded because of insufficient sample size.

IDC: invasive ductal carcinoma; ILC: invasive lobular carcinoma; DCIS: ductal carcinoma in situ; LCIS: lobular carcinoma in situ.

Breast cancer molecular subtypes

For the molecular assigned breast cancers, a total of 846 samples with sufficient classification were identified from the GEO database. From these samples, a total of five subtypes were analysed: normal-like (37), luminal A (181), luminal B (124), HER2 (187), and basal (226). Normal samples were also identified (91). Table 9 shows the sample sizes for each of the molecular breast cancer classifications from the GEO database.

Table 9. Breast cancer molecular subtypes and normal sample sizes.

Molecular subtypes	No. of samples
Normal	91
Normal-like	37
Luminal A	181
Luminal B	124
HER2	187
Basal	226
Total	846

2.2.2 Ovarian cancer data sets

A total of 1,782 ovarian cancer samples that contained sufficient information to identify subtypes were identified across 28 studies. The studies and platforms that met the data selection criteria are shown in Table 10.

Table 10. Ovarian cancer tissue and epithelial subtype studies, including normal samples identified from the GEO database.

	GEO identifier	Total no. samples	Microarray platform	Study reference
Tissue subtype				
1	GSE9891	285	Affymetrix Human Genome U133 Plus 2.0 Array	(Tothill <i>et al.</i> , 2008)
2	GSE26712	195	Affymetrix Human Genome U133A Array	(Bonome <i>et al.</i> , 2008)
3	GSE20565	140/96	Affymetrix Human Genome U133 Plus 2.0 Array	(Meyniel <i>et al.</i> , 2010)
4	GSE6008	103	Affymetrix Human Genome U133A Array	(Hendrix <i>et al.</i> , 2006)
5	GSE63885	101	Affymetrix Human Genome U133 Plus 2.0 Array	(Lisowska <i>et al.</i> , 2014)
6	GSE14764	80	Affymetrix Human Genome U133A Array	(Denkert <i>et al.</i> , 2009)
7	GSE18520	63	Affymetrix Human Genome U133 Plus 2.0 Array	(Mok <i>et al.</i> , 2009)
8	GSE44104	60	Affymetrix Human Genome U133 Plus 2.0 Array	(Wu <i>et al.</i> , 2014)
9	GSE10971	37	Affymetrix Human Genome U133 Plus 2.0 Array	(Tone <i>et al.</i> , 2008)
10	GSE51373	28	Affymetrix Human Genome U133 Plus 2.0 Array	(Koti <i>et al.</i> , 2013)
11	GSE14407	24	Affymetrix Human Genome U133 Plus 2.0 Array	(Bowen <i>et al.</i> , 2009)
12	GSE14001	23	Affymetrix Human Genome U133 Plus 2.0 Array	(Tung <i>et al.</i> , 2009)
13	GSE54388	22	Affymetrix Human Genome U133 Plus 2.0 Array	(Yeung <i>et al.</i> , 2017)

	GEO identifier (study)	Total no. samples	Microarray platform	Study reference
Tissue subtype				
14	GSE55512	12	Affymetrix Human Genome U133 Plus 2.0 Array	(Abiko <i>et al.</i> , 2015)
15	GSE36668	12	Affymetrix Human Genome U133 Plus 2.0 Array	(Elgaaen <i>et al.</i> , 2012)
16	GSE28044	24	Affymetrix Human Genome U133 Plus 2.0 Array	(George <i>et al.</i> , 2011)
	Total no. tissue location samples	1,165		
	GEO identifier (study)	Total no. samples	Microarray platform	Study reference
Epithelial				
17	GSE63885	101/94	Affymetrix Human Genome U133 Plus 2.0 Array	(Lisowska <i>et al.</i> , 2014)
18	GSE14764	80/77	Affymetrix Human Genome U133A Array	(Denkert <i>et al.</i> , 2009)
19	GSE18520	63	Affymetrix Human Genome U133 Plus 2.0 Array	(Mok <i>et al.</i> , 2009)
20	GSE44104	60	Affymetrix Human Genome U133 Plus 2.0 Array	(Wu <i>et al.</i> , 2014)
21	GSE10971	37	Affymetrix Human Genome U133 Plus 2.0 Array	(Tone <i>et al.</i> , 2008)
22	GSE14407	24	Affymetrix Human Genome U133 Plus 2.0 Array	(Bowen <i>et al.</i> , 2009)
23	GSE54388	22	Affymetrix Human Genome U133 Plus 2.0 Array	(Yeung <i>et al.</i> , 2017)

	GEO identifier (study)	Total no. samples	Microarray platform	Study reference
Epithelial				
24	GSE55512	12	Affymetrix Human Genome U133 Plus 2.0 Array	(Abiko <i>et al.</i> , 2015)
25	GSE36668	12	Affymetrix Human Genome U133 Plus 2.0 Array	(Elgaaen <i>et al.</i> , 2012)
26	GSE14001	23	Affymetrix Human Genome U133 Plus 2.0 Array	(Tung <i>et al.</i> , 2009)
27	GSE20565	140/90	Affymetrix Human Genome U133 Plus 2.0 Array	(Meyniel <i>et al.</i> , 2010)
28	GSE6008	103	Affymetrix Human Genome U133A Array	(Hendrix <i>et al.</i> , 2006)
	Total no. ovarian epithelial samples	617		
	Total no. ovarian cancer samples	1,782		

All studies are from Affymetrix platforms. Both the total number of samples in each dataset and the number of relevant samples taken forward for analysis are shown.

GEO: Gene Expression Omnibus.

Ovarian cancer tissue subtypes

In total, 1,165 ovarian cancer tissue subtype and normal samples were identified, the predominant tissue subtype being ovarian. The ovarian cancer tissue subtypes were as follows: ovarian (1,038), fallopian (32), peritoneum (34) and normal samples (61) (Table 11).

Table 11. Ovarian cancer tissue subtype and normal sample sizes.

Tissue subtype	No. of samples
Normal	61
Ovarian	1,038
Fallopian	32
Peritoneum	34
Total	1,165

Ovarian cancer epithelial subtypes

For the ovarian epithelial subtypes, 617 samples across 12 studies were identified in the GEO database. The subtypes were as follows: serous (413), endometrioid (74), mucinous (30), clear cell (39) and normal samples (61) (Table 12).

Table 12. Ovarian cancer epithelial subtype and normal sample sizes.

Epithelial subtype	No. of samples
Normal	61
Serous	413
Endometrioid	74
Mucinous	30
Clear cell	39
Total	617

2.2.3 Prostate cancer data sets

A total of 358 samples were identified from 8 prostate cancer studies in the GEO database. The studies that met the inclusion criteria are shown in Table 13 below.

Table 13. Prostate cancer Gleason group and Gleason score studies, including normal samples identified in the GEO database

	GEO identifier	Total no. samples	Microarray platform	Study reference
1	GSE2109	83	Affymetrix Human Genome U133 Plus 2.0 Array	(Intgen, 2005)
2	GSE6604	18	Affymetrix Human Genome U95 Version 2 Array	(Yu <i>et al.</i> , 2004)
3	GSE6606	65	Affymetrix Human Genome U95 Version 2 Array	(Yu <i>et al.</i> , 2004)
4	GSE6608	63	Affymetrix Human Genome U95 Version 2 Array	(Yu <i>et al.</i> , 2004)
5	GSE7055	57	Affymetrix Human Genome U133 Plus 2.0 Array	(Prueitt <i>et al.</i> , 2008)
6	GSE12630	14	Affymetrix Human Genome U133A Array	(Monzon <i>et al.</i> , 2009)
7	GSE45016	14	Affymetrix Human Genome U133 Plus 2.0 Array	(Satake <i>et al.</i> , 2010)
8	GSE46602	50	Affymetrix Human Genome U133 Plus 2.0 Array	(Mortensen <i>et al.</i> , 2015)
	Total no. prostate cancer samples	364		

All studies are from Affymetrix platforms. Both the total number of samples in each dataset and the number of relevant samples taken forward for analysis are shown.

GEO: Gene Expression Omnibus.

Prostate cancer Gleason scores

For prostate cancer, the primary tumour subtypes consisted of Gleason scores that were measured using the combined Gleason score (Table 14).

Table 14. Prostate cancer Gleason score and normal samples.

Gleason score subtypes	No. of samples
Normal	96
Score 4	8
Score 5	9
Score 6	68
Score 7	133
Score 8	13
Score 9	37
Total	364

Prostate cancer Gleason grade group

As discussed in the introduction (section 1.3.3), Gleason scores are now being reclassified under a Gleason grade group. Therefore, the prostate cancer Gleason score samples were reclassified into Gleason grade groups and are shown in Table 15. Gleason scores of 7 were assigned to a combined group of 2 and 3. This was because the overall Gleason score was only reported in the data sets and, therefore, it was impossible to differentiate between grade 2 (Gleason score 3+4) and grade 3 (Gleason score 4+3).

Table 15. Prostate cancer Gleason score samples and their corresponding Gleason grades group.

Gleason score subtypes	Gleason grade group subtypes	No. of samples
Normal	Normal	96
Score 4	Group 1	85
Score 5		
Score 6		
Score 7	Groups 2 and 3	133
Score 8	Group 4	13
Score 9	Group 5	37
Total		364

2.3 Quality control and pre-processing of data sets

Following the identification and curation of suitable data sets that met the microarray data selection criteria outlined in section 2.2, each data set was individually analysed for quality control. Microarray data sets were pre-processed in R version 3.4.1 (R Core Team, 2017) and Bioconductor version 3.9 (Huber *et al.*, 2015) using the affy package (Gautier *et al.*, 2004). Quality control checks were also carried out using the affy package. Each of the microarray samples were visually inspected for errors with the array chips used via probe intensity plots. These were generated both before and after normalisation (section 2.4) using box plots, MA plots and principal component analysis (PCA) (section 2.7). This allowed for the identification of erroneous probes and whether these were localised to one specific sample in the data set.

The identification of outliers and duplicates in each data set was undertaken using PCA and multidimensional scaling (MDS) of samples in each data set. No outliers that required exclusion were detected. The resulting sample metadata (phenodata) were compared to check if any duplicated samples were present in the data sets. Further to this, if samples were found to have identical probe expression in all probes, they were considered to be duplicates and removed; if duplicates were found, and the metadata could not be relied upon to identify the 'true'

phenotype i.e. if a sample was assigned to two subtypes then they were also removed. In total 433 samples were removed which were either duplicated or had insufficient sample data (subtype classification).

2.4 Background correction, quantile normalisation and summarisation

Following quality control, each microarray data set was background corrected, quantile normalised and summarised. The same method was applied to all data sets to ensure consistency in reducing technical effects, which can be the result of several things for example: preparation of samples in different wet labs at different intervals (commonly termed batch effects); differences in sequence hybridisation; and biased probe fluorescence.

Robust multiarray averaging (RMA) has been designed for normalising Affymetrix microarray data sets (Irizarry *et al.*, 2003). This was applied using the affy package (Gautier *et al.*, 2004).

To summarise, the RMA method first background corrects the array data (each array individually) in order to remove technical effects caused by the fluorescence of probes on the arrays themselves. During background correction, the signal intensity of each probe is adjusted according to the fluorescence of the surrounding probes (signal). This is necessary because probes that provide greater intensity can influence the surrounding probes, thereby leading to erroneous readings. The procedure helps to resolve technical noise and artifacts when reading the image of the probes.

Following background correction, normalisation of microarray data to remove batch effects that cause variations between the arrays of a study was carried out. These variations may be due to differences in, for example, preparation and hybridisation, and differences in DNA/RNA concentrations. The procedure uses information from all the arrays (the expression matrix) in the experiment, making their distributions identical and comparable. The data are first quantile normalised by ranking each row based on the probe measurement and imputing the mean average, which is calculated across each row. Then, the data are reordered back to their original positions. After quantile normalisation, summarisation by the RMA algorithm is carried out using

median polish average, whereby the residual of each probe is equal to 0. Finally, the data is log₂ transformed. A summary of the RMA normalisation steps is shown in Figure 18.

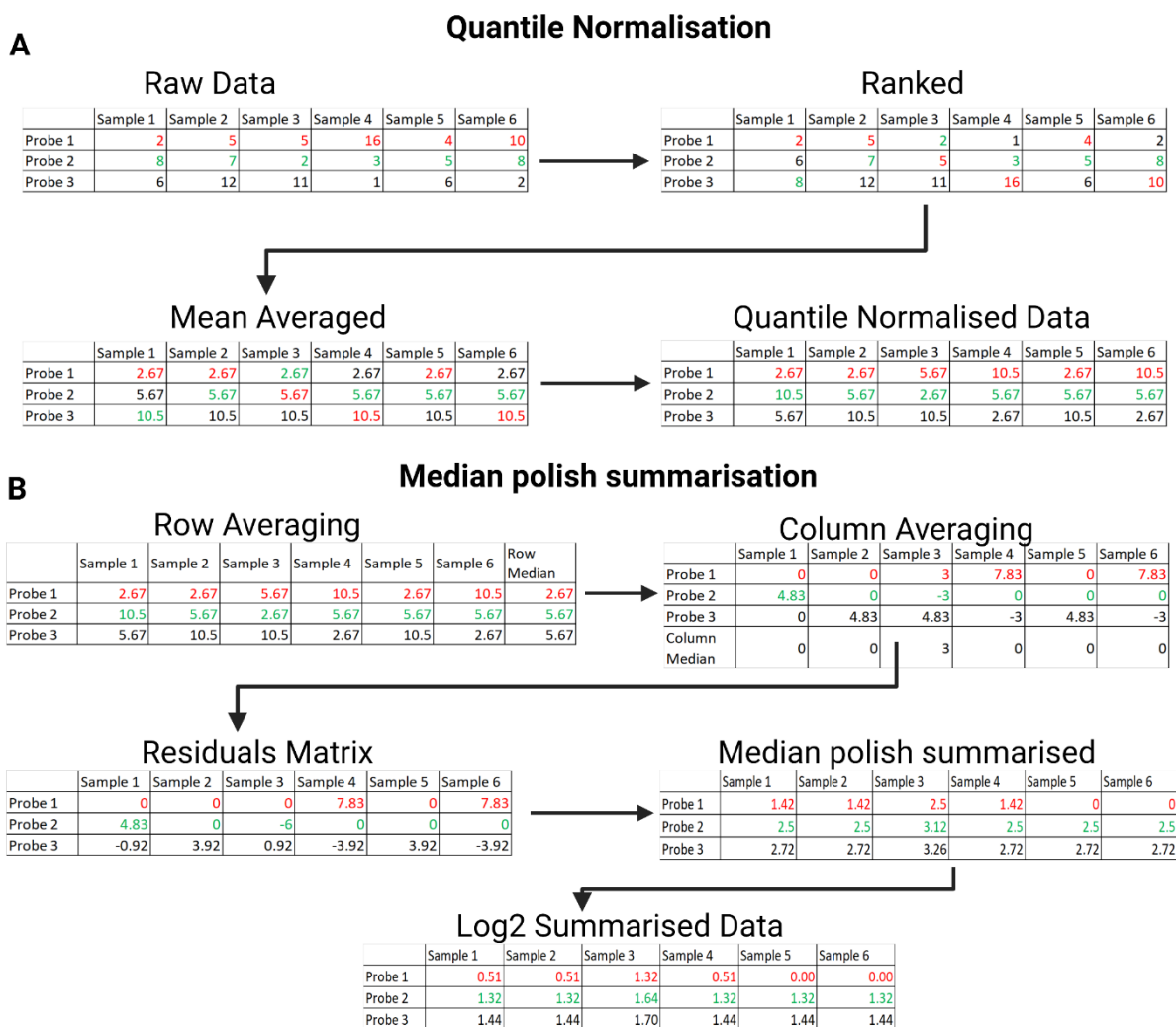


Figure 18. Example of the normalisation steps performed for microarray data using robust median averaging. (a) Quantile normalisation of microarray, (b) Median polish summarisation.

(a) Quantile normalisation of data removes variation within the expression data caused by batch effects. The background corrected data is first rank ordered based on the probe expressions for each sample. The mean average is calculated for each row and then imputed. Finally, the probe expressions are reordered back into their original places in the expression matrix.

(b) Median polish summarisation calculates the average of each probe (row) and sample (column) and subtracts this; the step is repeated until all residuals are equal to 0. This produces a residual matrix that is subtracted from the original expression matrix (the quantile normalised data) and log₂ transformed.

In addition, the normalised microarray expression data was quality checked to ensure suitable normalisation of samples. This was done using probe intensity box plots for each data set to identify any samples that required removal or were potential outliers. After normalisation, it was

not necessary to exclude any data sets or samples. An example probe intensity plot from the breast cancer study GSE65194 (a histological subtype study) is shown in Figure 19.

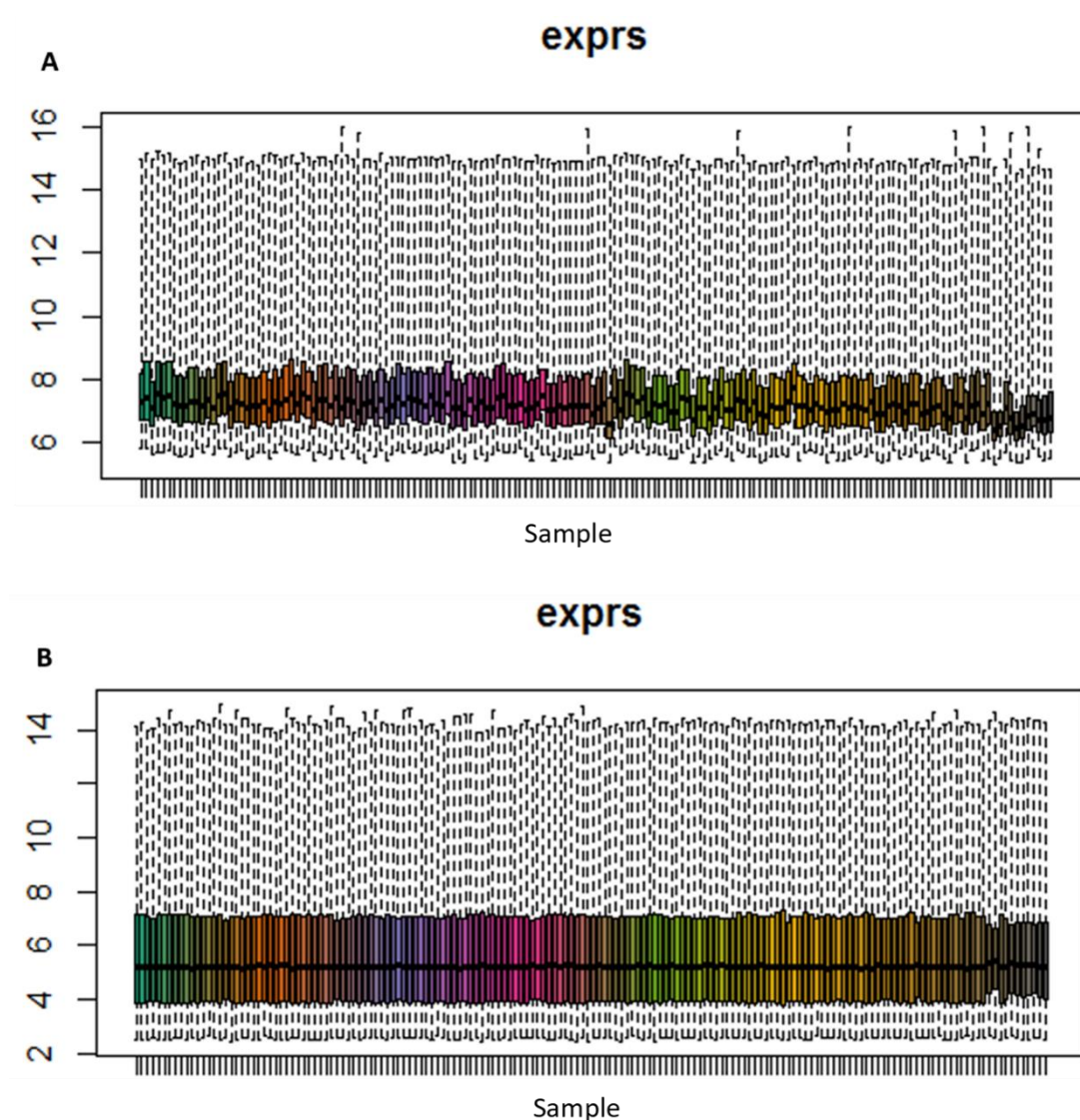


Figure 19. Microarray probe intensity quality control.

An example of the microarray quality control check of the probe intensity both prior and post background correction and normalisation. A) Prior to background correction and normalisation the samples of the microarrays show highly varied expression levels across the probes. B) Post background correction and normalisation the samples show more uniform expression across the probes.

2.5 Annotation of microarray platforms

Initially, probes were annotated using the appropriate annotation database (.db) package from Bioconductor. The annotation database packages allow simplified matching of probes to genes

using the AnnotationDbi package in R (Pagès *et al.*, 2018). They are also updated biannually with the latest probe information available from the supplier and the target gene. As a result, they have become used more widely for the mapping of microarray probe IDs to genes. These packages and their corresponding microarray platforms are shown in Table 16.

Table 16. Affymetrix microarrays and their corresponding annotation packages.

Array platform	Annotation package	Reference
HGU133A Plus 2.0	hgu133plus.db	(Carlson, 2016c)
HGU133A	hgu133a.db	(Carlson, 2016b)
Human X3P Array	u133x3p.db	(Carlson, 2016d)
HGU95 Version 2 Array	hgu95av2.bd	(Carlson, 2016a)

Each package contains probe IDs and their target gene according to manufacturer specifications.

The probe sequences for each array platform were reannotated to confirm that they corresponded to the correct target gene as stated by the manufacturer in the appropriate annotation package. This was necessary because each probe is designed to cover a certain portion of the target gene, and on average there are 11–20 probes for a gene (a probeset). With updates to the human reference genome, these probes may correspond to a different target than the original design of the probe.

To confirm the probe target sequences, the complete list of probe sequences (in FASTA format) for each of the array platforms was downloaded from Affymetrix (Thermo Fisher Scientific, 2017). Each of the probe target sequences was then compared with BLAST using the most current human reference genome (GRCh38.p12) at the time to confirm the correct target gene in accordance with the manufacturer’s information.

Genes were annotated further using the ensemble database with the BioMart package (Steffen *et al.*, 2009). This additional annotation allowed for the identification of probes that corresponded to known protein coding genes, miRNAs and lncRNAs, etc.

2.6 Missing value imputation of microarray data sets

Following quality control and normalisation, the breast, ovarian and prostate cancer data sets were combined into a gene expression matrix per cancer. For example, all the individual breast cancer data sets for molecular subtypes were combined into one larger expression matrix, meaning that the 9 studies that made up the breast cancer molecular subtype analysis were combined, making an expression matrix of 846 samples. The process was repeated separately for the breast cancer histological subtypes, ovarian cancer tissue subtypes, ovarian cancer epithelial subtypes, prostate cancer Gleason score subtypes, and prostate cancer Gleason grade group subtypes data sets, forming six individual gene expression matrices for missing value imputation.

As highlighted in the previous section (section 2.5), each array has probes that are designed to cover a range of genes. As such, they also have specific annotation packages designed to annotate these probes appropriately. However, each of the arrays identified in this study were designed at different times; the HGU95 arrays being older than the HGU133 arrays. The U95 array utilises build 95 of the UniGene database (Pontius, Wagner and Schuler, 2002) which was released 1999. The U133 utilises build 133 of the UniGene database and was released 2001. As a result, the arrays measure a different number of genes. For example, the HGU95 version 2 (HGU95 version 2) was designed with 12,626 probes, corresponding to 9,200 unique genes. The newer HGU133 Plus 2 arrays were designed with 54,675 probes covering 23,517 genes. The total number of probes and genes represented on each of the arrays used in this study is shown in Table 17.

Table 17. Microarray platforms probes and the number of unique target genes.

Array platform	Number of probes	Number of unique genes
HGU95 version 2	12,626	9,200
HGU 133A	22,283	13,512
HGU 133A X3P	61,359	20,994
HGU 133A Plus 2	54,675	23,517

An increasing number of genes are represented in newer arrays where older versions of the UniGene database are used.

The HGU133A plus covers the largest proportion of unique genes whereas the older HGU95 covers the least.

Because of the differences in the arrays, there were a number of genes that were not present consistently across all of them and, as a result, these genes were absent from some data sets. Rather than limit the analysis to just the genes present on all the arrays, thereby limiting analyses to a maximum of 9,200 genes (the limit of HGU95A), imputation of the values for missing genes was conducted to retain as many genes as possible.

In order to determine the accuracy of the different imputation methods, a subset of complete data was extracted from each of the three cancer expression data sets. This data included only the genes that had no missing entries in all samples. For each of the three data sets, a range of missing values was generated at random to determine the maximum threshold at which accurate imputation could be applied. The range of missing values was chosen as 1%, 2%, 5%, 10%, 15%, 20%, 50% and 80% of the total data set size (number of samples multiplied by number of probes).

Five methods were selected for missing value imputation: mean row (gene) value, K-nearest neighbour (KNN), local least squares (LLS), Bayesian principal component analysis (BPCA) and singular value decomposition (SVD). The mean expression for each gene (row) was calculated and the value imputed. KNN was calculated using the impute package in R (Hastie *et al.*, 2018), and LLS, BPCA and SVD were calculated using the pcaMethods package in R (Stacklies, Redestig and Wright, 2018). The accuracy of the imputed values was compared with their original (non-missing) value using the normalised root mean square error (NRMSE). This procedure, whereby the

imputed value is compared with the original value, was developed specifically for assessing the accuracy of methods for imputing missing values in gene expression data (Figure 20) (Oba *et al.*, 2003).

$$\text{NRMSE} = \sqrt{\frac{\text{mean}(X^{\text{true}} - X^{\text{imp}})^2}{\text{var}(X^{\text{true}})}}$$

Figure 20. Normalised root mean square error (NRMSE) equation.

The closer the NRMSE is to 0 the more accurate the imputation. From Stekhoven & Bühlmann., 2012.

The NRMSE for each imputation method was calculated using the hydroGOF package in R (Zambrano-Bigiarini, 2018). In summary, the closer the NRMSE value is to 0 the more accurate the imputation method (because the value is closer to the original) (Stekhoven and Bühlmann, 2012). Following the initial simulation analysis to determine the limits of accuracy for each imputation method, it was found that each method was most accurate with genes of 10% or lower missing values. As a result, genes with more than 10% missing values were removed from imputation of the three gene expression data sets. Of the imputation methods tested, LLS was found to provide the most accurate imputation using K=150. The K-value was tested using simulated data generated from multiple primary data sets. A K-value of 150 means that 150 genes that are most similar to the gene being imputed are used for imputation. Additionally, the LLS algorithm will prioritise the gene expression that is present in other samples of the same subtype over other genes. For example, if the expression of *BRCA1* was not present in a luminal A sample (most likely because of a different array version), the other luminal A samples in which the *BRCA1* expression is available would be prioritised over other samples and subtypes. Therefore, a K-value of 150 will first identify samples in data sets in which the gene is measured and expression values are available, and from this it will identify what the expression should be in the samples in which it is missing. From this initial imputation, the expression value is then adjusted according to the other genes that are most similarly expressed, in order to correct for sample-specific variances in expression. The total number of remaining genes after missing value imputation for each gene

expression matrix is shown in Table 18. The prostate cancer gene expression matrix had the fewest remaining genes (8,902). This is probably because nearly half of the samples (146) were derived from a HGU95 array (Table 13). This array, which is curated with 9,200 unique genes, is the most restrictive of the microarray platforms. Increasing the sample size of prostate cancer Gleason samples in this study using one of the newer microarray platforms (HGU133 or later) would be expected to increase the number of genes available because there would be more samples from which to impute missing values. For example, in the breast cancer molecular gene expression matrix, 12,952 unique genes remained after missing value imputation. The majority of these samples were from HGU133A Plus 2 (23,517 unique genes) arrays, the most restricted array being HGU133A (13,512 unique genes). Only 168 samples were from the HGU133A array, compared with 640 from the HGU133A Plus 2 array. This led to a larger proportion of genes being retained. Also the genes that were lost from the imputation were potentially those that are no longer applicable from older arrays and therefore not covered on the newer platforms; where probes for certain genes have been updated and confer to either a different gene or redundant sequence.

Table 18. The total number of missing values identified in each cancer expression matrix and the percentage of missing values identified for each of the cancer expression matrices.

Gene expression matrix	Total data set size (samples x genes)	No. of missing values	Percentage of missing values (%)	Total no. of remaining genes per sample
Breast cancer histological samples	43,841,133	9,702,715	22	7,169
Breast cancer molecular samples	24,880,806	2,005,291	8	12,952
Ovarian cancer tissue location	25,362,402	3,018,330	12	12,993
Ovarian cancer epithelial	14,789,490	1,461,255	10	12,993
Prostate cancer Gleason score	7,528,740	2,466,460	32.8	8,902
Prostate cancer Gleason group	7,528,740	2,466,460	32.8	8,902

The total dataset size was determined by the number of samples multiplied by the number of genes. The breast cancer molecular subtype expression matrix had the least number of missing values (8%). The prostate cancer expression matrix had the largest number of missing values (32.8%). This led to a large number of genes being removed from analysis. The number of remaining genes were those that had less than 10% missing values. In samples that had missing values for genes the values were imputed allowing the gene to be retained. The genes remaining after missing value imputation were taken forward for DEG analysis.

2.7 Cross-platform/data set batch correction

Cross-platform correction was conducted for each gene expression matrix for each type of cancer (each cancer expression matrix) to remove technical batch effects caused by the combination of multiple microarray platforms and studies. Cross-platform batch effects were identified using principal component analysis (PCA) and multidimensional scaling (MDS) for each gene expression

matrix (data set). Samples were found to cluster distinctively together by study, showing study-based batch effects even for studies in which the same microarray platform was used (Figure 21 A-C).

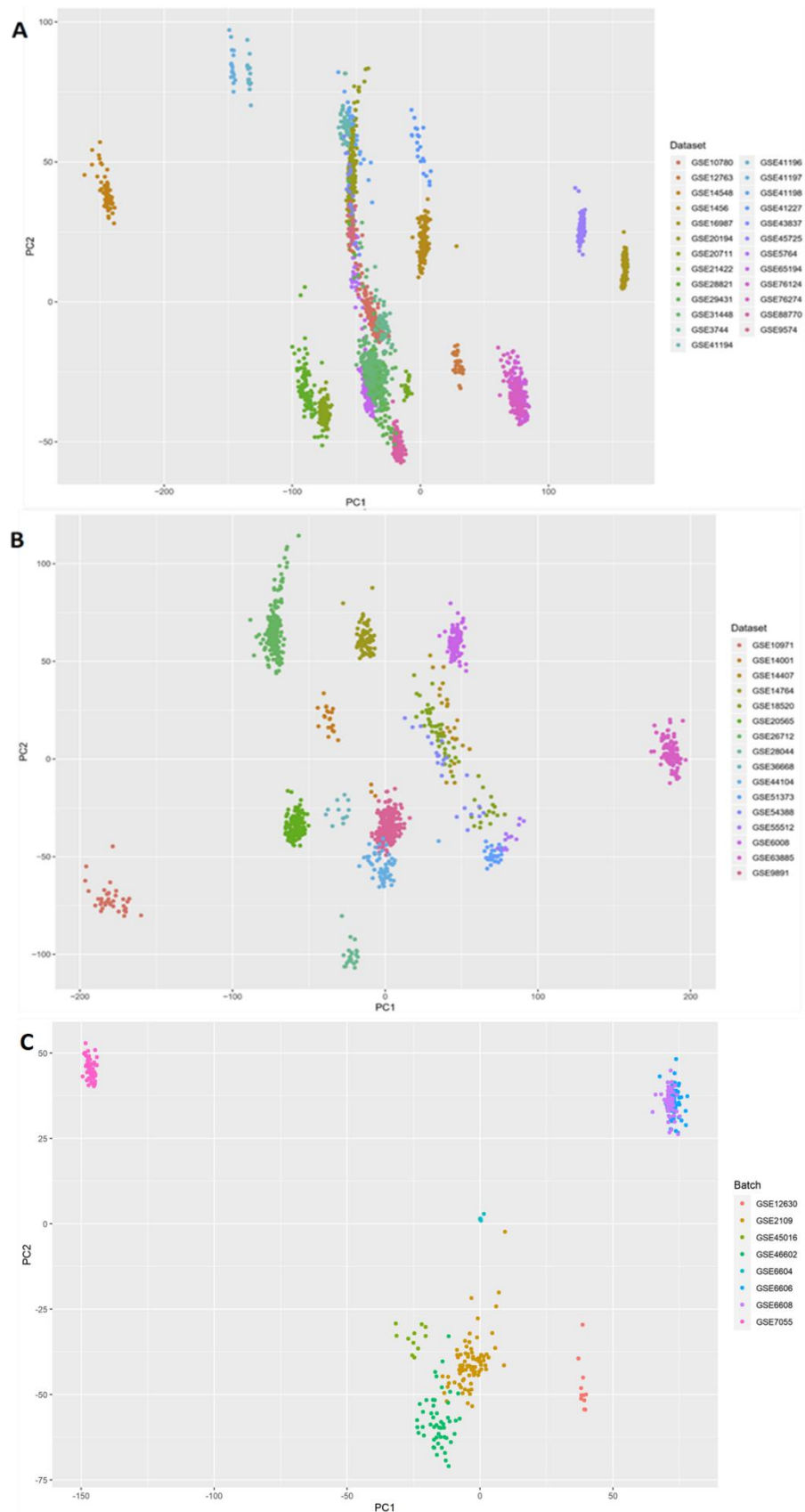


Figure 21. PCA of breast, ovarian and prostate cancer data sets showing pre-cross-platform normalised samples for each of the data sets.

(A) Breast cancer, (B) Ovarian cancer, (C) Prostate cancer. The samples show distinct clustering around samples from the same study, indicating a study-based batch effect.

PCA: principal component analysis.

Cross-platform batch correction was conducted using the ComBat algorithm from the SVA package (Leek, 2017) in R. For batch correction, sample types (cancer subtype or normal) were modelled as covariates into the correction algorithm in order to retain biological differences. PCA plots post ComBat batch correction for datasets are shown in Figure 22 A-C.

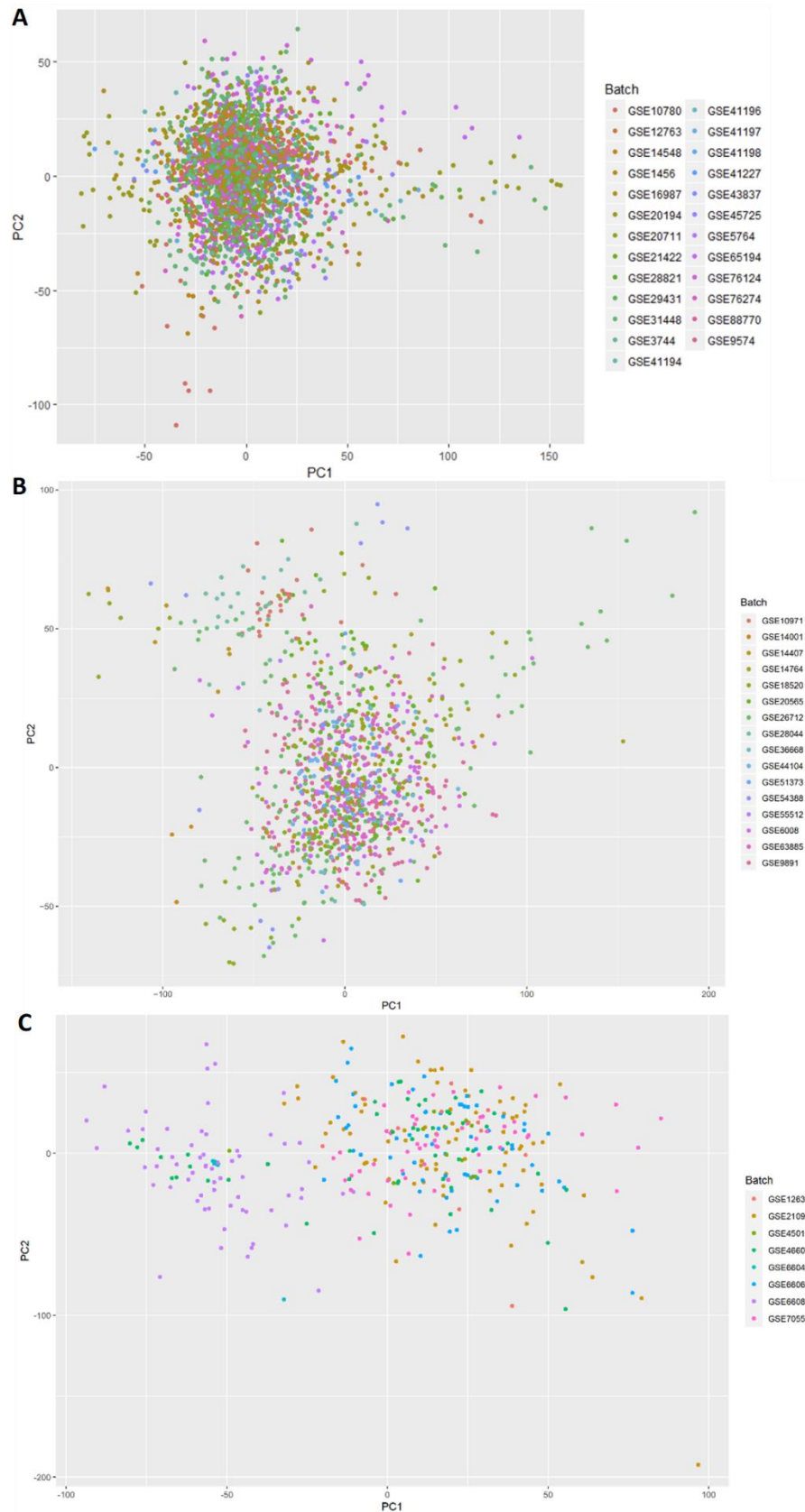


Figure 22. PCA of breast, ovarian and prostate cancer data sets showing post-cross-platform normalised samples for each of the data sets.

(A) Breast cancer, (B) Ovarian cancer, (C) Prostate cancer. The samples show no distinct clustering around samples from the same study, indicating a reduction/removal of study-based batch effect.

PCA: principal component analysis.

MDS plots prior to platform correction are shown in Figure 23 A-C. After ComBat batch correction, the data were then reassessed, again using MDS, to determine whether technical batch effects had been removed and biological variance had been retained (Figure 24 A-C).

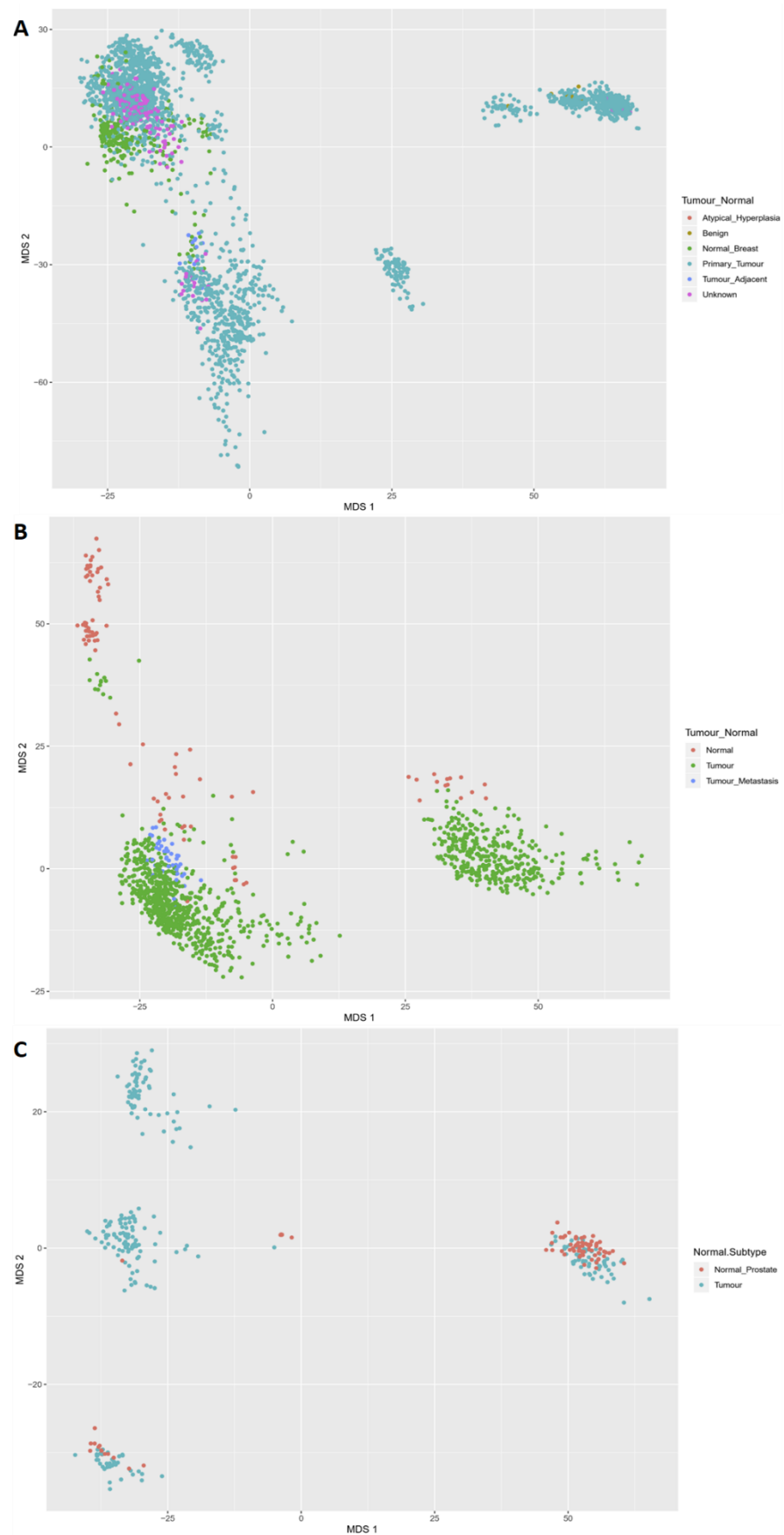


Figure 23. MDS of pre-ComBat batch-corrected gene expression matrices.

(A) Breast cancer, (B) Ovarian cancer, (C) Prostate cancer. Prior to ComBat batch correction, the cancer subtypes were observed to be clustered according to the underlying study.

MDS: multidimensional scaling

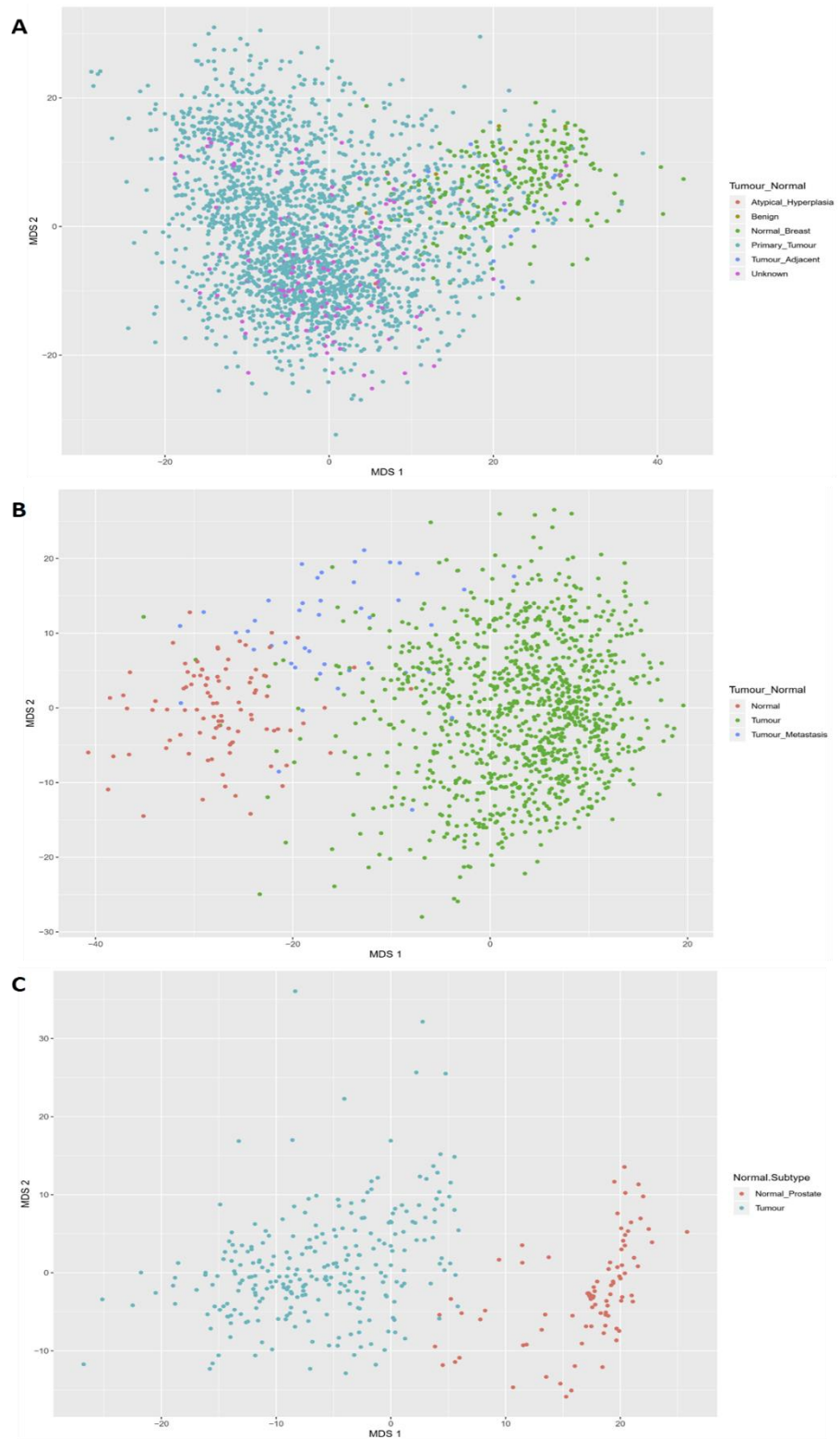


Figure 24. MDS of post-ComBat batch-corrected gene expression matrices.

(A) Breast cancer, (B) Ovarian cancer, (C) Prostate cancer. After ComBat batch correction, cancer samples were observed to be clustered according to subtype, suggesting biological variance was retained.

MDS: multidimensional scaling.

2.8 Identification of DEGs

Differential gene expression analysis was conducted using the LIMMA package in R (Ritchie *et al.*, 2015). An empirical Bayes modified t-test was used to compare cancer subtypes with normal tissue samples. For example, the breast cancer luminal A subtype was compared with normal breast tissue, ovarian serous cancer compared to normal ovarian tissue, and the prostate cancer Gleason score 7 was compared with normal prostate tissue samples. The analysis was necessary to determine which genes had significantly altered expression (differentially expressed) in the tumour samples compared with its corresponding normal tissue. The LIMMA empirical Bayes t-test overrides the base R program p-value cut-off, which is set to a default of $P=1e^{-16}$. This provided increased accuracy of reporting p-values that may otherwise be confounding. In an initial analysis of individual breast, ovarian and prostate cancer data sets, p-values were commonly found to be lower than $P=1e^{-16}$. This has also been observed in other microarray cancer studies, in which DEGs identified were found to be below $P=1e^{-16}$, and in some cases below $P=1e^{-25}$ (Long *et al.*, 2019; Zhao, Erwin and Xue, 2018; Zhang *et al.*, 2019; Tang *et al.*, 2020). Reporting the actual p-value here (rather than grouping most significant genes to $P=1e^{-16}$) was important to distinguish the most significantly differentially expressed genes.

The comparisons for each normal and carcinoma subtype for the three different cancers are shown in Table 19 (breast cancer), Table 20 (ovarian cancer) and Table 21 (prostate cancer). All subtypes were compared with corresponding normal samples to identify the DEGs. These were then ranked according to P-value significance with a $P \leq 0.01$ cut-off.

Table 19. Differential expression analysis comparisons for breast cancer subtypes.

Subtype	Subtype comparison								
	Histological				Molecular				
	Normal	IDC	ILC	DCIS	Normal-like	Luminal A	Luminal B	HER2	Basal
IDC		-							
ILC			-						
DCIS				-					
Normal-like					-				
Luminal A						-			
Luminal B							-		
HER2								-	
Basal									-

Blue shows the comparison of subtypes. All subtypes were compared with normal breast tissue samples to identify DEGs for further analysis. Additionally, breast cancer subtypes were compared with each other. For example, normal-like breast cancer tissue samples were compared with luminal A. Those in red are comparisons not conducted, either because they had been conducted previously or because they were the same subtypes.

IDC: invasive ductal carcinoma; ILC: invasive lobular carcinoma; DCIS: ductal carcinoma in situ; DEGs: differential gene expressions.

Table 20. Differential expression analysis comparisons for ovarian cancer subtypes.

Subtype	Subtype comparison							
	Histological				Histological with molecular status			
	Normal	Ovarian	Peritoneum	Fallopian	Serous	Endometrioid	Mucinous	Clear cell
Ovarian		-						
Peritoneum			-					
Fallopian				-				
Serous					-			
Endometrioid						-		
Mucinous							-	
Clear cell								-

Blue shows the comparison of subtypes. All subtypes were compared with normal ovarian tissue samples to identify DEGs for further analysis. Additionally, ovarian cancer subtypes were compared with each other. Those in red are comparisons not conducted, either because they had been conducted previously or because they were the same subtypes.

DEGs: differential gene expressions.

Table 21: Differential expression analysis comparisons for prostate cancer subtypes.

Subtype	Subtype comparison									
	Histological						Histological			
	Normal	Score 5	Score 6	Score 7	Score 8	Score 9	Group 1	Groups 2 and 3	Group 4	Group 5
Score 5		-								
Score 6			-							
Score 7				-						
Score 8					-					
Score 9						-				
Group 1							-			
Groups 2 and 3								-		
Group 4									-	
Group 5										-

Blue shows the comparison of subtypes. All subtypes were compared with normal prostate tissue samples to identify DEGs for further analysis. Additionally, prostate cancer subtypes were compared with each other. Those in red are comparisons not conducted, either because they had been conducted previously or because they were the same subtypes.

DEGs: differential gene expressions.

Following DEG identification, a Bonferroni correction was applied to correct for type 1 errors (detecting false positives) due to multiple testing across the gene expressions. This provided an adjusted p-value for further analysis. The adjusted p-values and log₂ fold changes (Log₂ FC) for each of the DEGs for each of the cancer subtypes were visualised using volcano plots with the ggplot2 package in R (Valero-Mora, 2010). These plots are shown in appendices Figure 45-69.

2.9 Cross-cancer identification of DEGs

The significant DEGs that were identified from each of the cancer subtype analyses (subtype compared to normal) were further cross-compared to identify if they were significant in other cancer subtypes. In the previous DEG analysis (section 2.8) subtypes were compared to normal, here the DEGs identified were directly compared between cancer subtypes. For example, the results from DEG analysis of breast cancer histological subtypes (IDC, DCIS, etc.) were compared with ovarian tissue subtypes (ovarian, fallopian, etc.).

The breast cancer histological subtypes (IDC, ILC and DCIS) were cross-compared with ovarian tissue subtypes (ovarian, fallopian and peritoneum) and prostate cancer Gleason grade groups (grade groups 1–5) (see Table 22). The breast cancer molecular subtypes (normal-like, luminal A, luminal B, HER2, and basal), were compared with ovarian cancer epithelial subtypes (serous, endometrioid, mucinous and clear cell) and prostate cancer Gleason scores (4–9) (see Table 23).

Table 22. Cross-cancer comparisons of histological breast cancer subtypes, ovarian cancer tissue subtypes and prostate cancer Gleason groups.

Subtype	Subtype comparison									
	Breast			Ovarian			Prostate			
	IDC	ILC	DCIS	Ovarian	Peritoneum	Fallopian	Group 1	Groups 2 and 3	Group 4	Group 5
IDC	-									
ILC		-								
DCIS			-							
Ovarian				-						
Peritoneum					-					
Fallopian						-				
Group 1							-			
Groups 2 and 3								-		
Group 4									-	
Group 5										-

Blue shows the comparison of subtypes. All subtypes were compared with each other, for example, all the significant DEGs identified in breast cancer histological subtypes were compared with the significant DEGs identified in ovarian cancer epithelial tissue subtypes and prostate cancer Gleason grade groups. Those in red are comparisons not conducted, either because they had been conducted previously or because they were the same subtypes.

IDC: invasive ductal carcinoma; ILC: invasive lobular carcinoma; DCIS: ductal carcinoma in situ; DEGs: differential gene expressions.

Table 23. Cross-cancer comparisons of molecular breast cancer subtypes, ovarian cancer epithelial subtypes and prostate cancer Gleason scores.

Subtype	Subtype comparison													
	Breast					Ovarian				Prostate				
	Normal-like	Luminal A	Luminal B	HER2	Basal	Serous	Endometrioid	Mucinous	Clear cell	Score 5	Score 6	Score 7	Score 8	Score 9
Normal-like	-													
Luminal A		-												
Luminal B			-											
HER2				-										
Basal					-									
Serous						-								
Endometrioid							-							
Mucinous								-						
Clear cell									-					
Score 5										-				
Score 6											-			
Score 7												-		
Score 8													-	
Score 9														-

Blue shows the comparison of subtypes. All subtypes were compared with each other, for example, all the significant DEGs identified in breast cancer molecular subtypes were compared with the significant DEGs identified in ovarian cancer epithelial subtypes and prostate cancer Gleason scores. Those in red are comparisons not conducted, either because they had been conducted previously or because they were the same subtypes.

A directed comparison analysis was used in which those genes that were significantly upregulated were then compared across other cancer subtypes. This analysis identified genes that were similarly upregulated across cancers and subtypes. Molecular breast cancer subtypes were compared with ovarian cancer epithelial subtypes based on previous studies highlighted in the literature review (section 1.5). They were also cross-compared with prostate cancer Gleason scores, because this classification is more differentiated. This approach was used instead of making a comparison with Gleason grade group, because the clustering of samples under one group classifier is less diverse, therefore making it a broader classification in cases in which Gleason grade groups consist of multiple Gleason scores. Gleason grade group is potentially more appropriate when compared with ovarian cancer tissue subtypes and histological breast cancer subtypes, which also consist of multiple underlying groups that are considered separate in other classification systems. For example, the histological subtypes of breast cancer commonly consist of luminal A or B, HER2, etc., which can be classified as either IDC or DCIS. Similarly, with regard to ovarian cancer tissue subtypes, these can consist of mucinous and serous epithelial subtypes together. Thus, a Gleason grade group consisting of multiple Gleason scores, for example, group 1 (Gleason score 6 or below), is potentially more appropriate. Therefore, the histological breast cancer subtypes, ovarian cancer tissue subtypes and prostate cancer Gleason groups were compared. The cross-cancer comparisons were conducted in R, and the 'conserved' cross-cancer genes found to be significantly upregulated in the comparison of all the subtypes were extracted for further analysis to determine if they were novel. To identify the novel cross-cancer genes, the genes were selected using the following criteria:

- significantly upregulated DEGs
- novel cross-cancer DEGs that are also associated with reduced survival
 - must be novel in at least 2 of the 3 cancers and the majority of subtypes

The validation of the DEG analysis and cross-cancer analysis will be shown in the results.

2.10 Survival analysis

To determine whether the significant cross-cancer conserved DEGs associated with poorer survival times, primary patient overall survival (OS) and recurrence-free survival (RFS) data for breast, ovarian and prostate cancers were collected from the GEO database (Edgar, Domrachev and Lash, 2002).

Overall survival (OS) was selected as it incorporates the time from diagnosis/treatment. This can be continuous as patient survival can be longer than the study/trial period. This classification includes local as well as secondary metastasis. Recurrence free survival (RFS) was chosen as it is the time from completion of primary treatment until symptoms are observed. This is usually encapsulating the effectiveness of treatments. Symptoms can include local or secondary metastasis.

These data sets were different from those used previously in the DEG analysis as they had associated survival data (OS and/or RFS) (section 2.8), and are shown in Table 24. Kaplan–Meier curves for OS and RFS for genes of interest were generated in R. If known for the data set, treatment data were also compared with the DEGs identified to determine whether potential resistance in patients may occur and affect survival. The significance of gene expression for patient survival time was determined using log-rank tests, and Cox proportion hazards ratio (HR) calculated using the survival package in R (Therneau, 2020). The log-rank test determined whether there was a significant difference in patient survival times between low and high expression of the gene of interest. The Cox proportion hazard ratio determined the effect of other independent variables on survival times, giving an overall value (and range) of the effect of the gene expression on survival. The log-rank test and Cox proportion hazard ratio results were considered significant if $P \leq 0.05$. Kaplan–Meier curves were then visualised using the Survminer package in R (Kassambara, Kosinski and Biecek, 2019).

To validate the survival plots, a combined Kaplan-Meier curve for genes of interest was created for breast and ovarian cancers for OS and RFS survival data using the Kaplan-Meier plotter web tool

(Lánczky and Gyórfy, 2021). The web tool was useful as a validation tool because it used the same statistical tests and similar visualisation but the majority of the datasets in the web tool were different from those used in DEG and survival analyses. The added benefit of using KM-plotter as a validation was that there were more samples available increasing robustness. The web tool did not contain prostate cancer data so this limited the validation to only breast and ovarian cancers. For prostate cancer, the SurvExpress (Aguirre-Gamboa *et al.*, 2013) database was used to analyse the three genes in OS, but for this, RFS data was not available.

Table 24. Primary patient survival data for breast, ovarian and prostate cancer.

GEO identifier	Survival data	Treatment	Total no. samples	Microarray platform	Study reference	Present in KM-plotter
Breast cancer						
GSE20685	OS and RFS	CMF and CAF	327	Affymetrix Human Genome U133 Plus 2.0 Array	(Kao <i>et al.</i> , 2011)	Yes
GSE1456	OS and RFS	Tamoxifen, CMF	159	Affymetrix Human Genome U133A Array	(Pawitan <i>et al.</i> , 2005)	No
GSE17705	RFS	Tamoxifen	196	Affymetrix Human Genome U133A Array	(Symmans <i>et al.</i> , 2010)	No
GSE21653	RFS	Adjuvant (tamoxifen)	266	Affymetrix Human Genome U133 Plus 2.0 Array	(Sabatier <i>et al.</i> , 2011)	Yes
Ovarian cancer						
GSE32062	OS and RFS	Taxane and platinum	270	Agilent-014850 Whole Human Genome Microarray 4x44K	(Yoshihara <i>et al.</i> , 2012)	No
GSE32063	OS and RFS	Taxane and platinum	40	Agilent-014850 Whole Human Genome Microarray 4x44K	(Yoshihara <i>et al.</i> , 2012)	No

GEO identifier	Survival data	Treatment	Total no. samples	Microarray platform	Study reference	Present in KM-plotter
GSE17260	OS and RFS	Taxane and platinum	110	Agilent-014850 Whole Human Genome Microarray 4x44K	(Yoshihara <i>et al.</i> , 2012; Yoshihara <i>et al.</i> , 2010)	No
GSE14764	OS	Carboplatin and paclitaxel	80	Affymetrix Human Genome U133A Array	(Denkert <i>et al.</i> , 2009)	Yes
GSE26712	OS	Untreated	195	Affymetrix Human Genome U133A Array	(Bonome <i>et al.</i> , 2008)	Yes
Prostate cancer						
GSE16560	OS	Untreated	281	Human 6k Transcriptionally Informative Gene Panel for DASL	(Sboner <i>et al.</i> , 2010)	No
GSE10645	OS	RRP	596	DASL Human Cancer Panel	(Nakagawa <i>et al.</i> , 2008)	No
GSE116918	OS and RFS	RRP	248	Almac Diagnostics Prostate Disease Specific Array	(Jain <i>et al.</i> , 2018)	No

Overall survival and recurrence-free survival data sets were collected from the GEO database. The GEO datasets that were also in KM-plotter are identified.

GEO: Gene Expression Omnibus; OS: overall survival; RFS: recurrence-free survival; DMFS: distant metastasis-free survival; RRP: radical retropubic prostatectomy.

2.11 Co-expression network analysis

A summary of the steps used for network construction and visualisation steps are shown in Figure 25. Each step of network analysis is described in more detail in the following section and subsections.

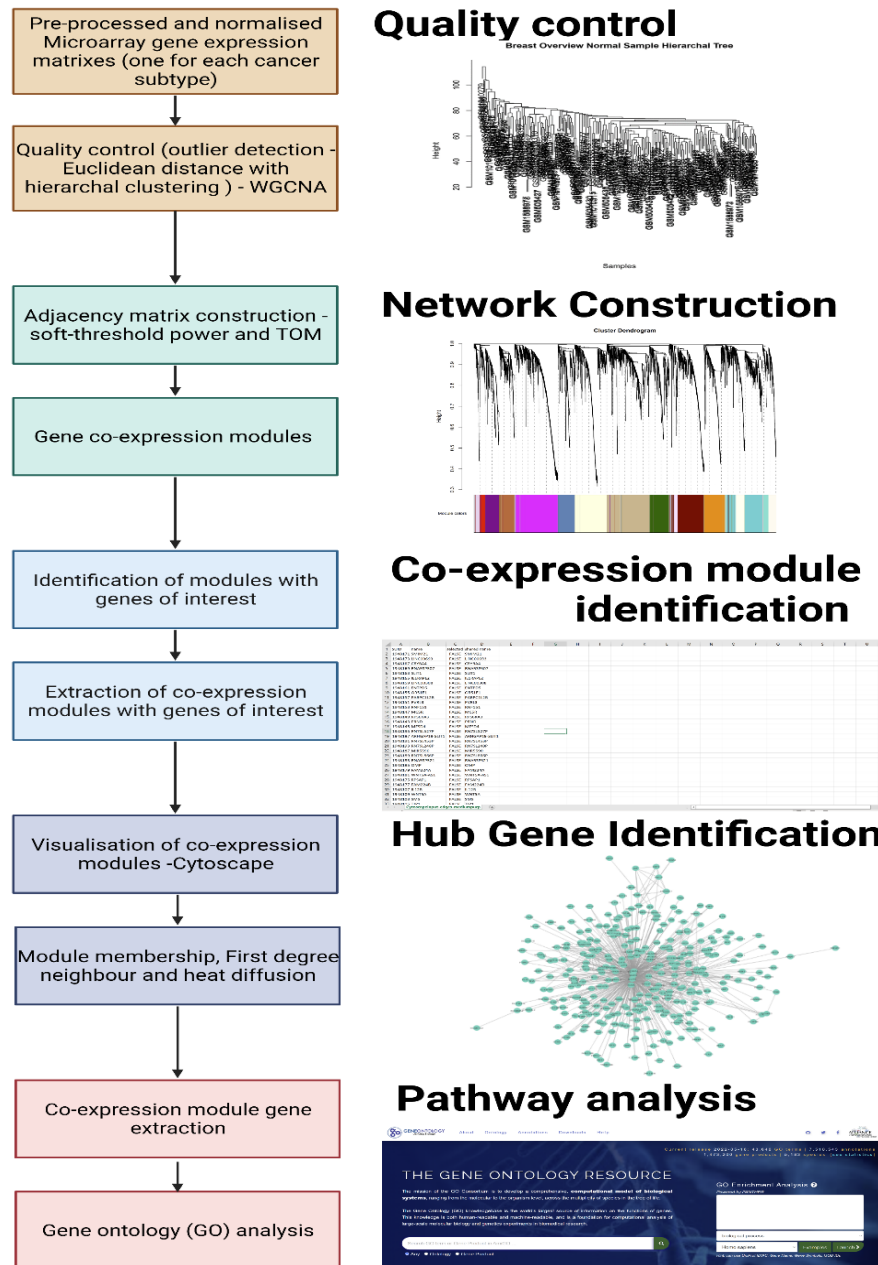


Figure 25. Summary of the network construction and visualisation analysis.

There were five main steps to the network analysis. First quality control was conducted to identify and remove any erroneous samples. Networks were then constructed for each of the cancer subtypes to produce gene co-expression modules. The co-expression modules containing genes of interest were extracted for further hub gene identification. Genes of interest that were also hub genes were then taken forward for pathway analysis. This also included the other genes from the same co-expression module to help determine potential role/roles of the hub genes.

Samples were checked for outliers using hierarchal clustering. No samples were found to be outliers in any of the three gene expression matrices (breast, ovarian and prostate cancers). Each of the three cancer expression matrices was separated further. Normal samples and primary tumours were extracted from the six gene expression matrices created for imputation and integration of microarray data sets earlier (sections 2.6 and 2.7) into smaller subtype-specific gene expression matrices and analysed separately. Weighted gene co-expression networks were constructed for each cancer subtype using the WGCNA package in R (Langfelder & Horvath, 2008). Parameter optimisation (power threshold, minimum module size and merge cut height) for network construction was tested prior to gene module construction. For the normal sample networks, a power threshold of two was chosen. This was also used for breast cancer subtypes (histological and molecular) and ovarian cancer subtypes (tissue location and epithelial subtype). However, for prostate cancer subtypes (Gleason grades and Gleason scores), a power threshold of three was chosen, because there were smaller prostate cancer sample sizes for each subtype. The minimum module size was set at 30, because it was found this produced sufficient minimal module size for functional analysis. The merge cut height of 0.2 was retained as the default value based on the hierarchal clustering of samples. A Pearson's correlation was conducted to determine the co-expressed genes within each of these. Following this, an adjacency matrix was created and the topological overlap measure (TOM) to cluster genes into a co-expression matrix. The gene network co-expression modules were examined in order to identify those with candidate genes that had earlier been identified from the cross-cancer comparisons and survival analysis; the modules in which these were identified were extracted for functional analysis using Gene Ontology (GO) (Michael *et al.*, 2000; The Gene Ontology Consortium, 2018) functional pathway analysis.

The co-expression modules extracted for each subtype and candidate gene were visualised in Cytoscape 3.7.1 (Shannon *et al.*, 2003). These were weighted GGI networks that were filtered prior to module construction using thresholding restriction to remove weak predicted interactions. This procedure was undertaken to reduce the co-expression module sizes to the

most highly weighted and strongest predicted interactions between genes. The resulting images are then clearer as they focus on the most functional predicted interactions operating in the gene network rather than encompassing all interactions in one image (reducing noise).

2.11.1 Hub gene identification

Following importation of the gene network modules into Cytoscape 3.7.1 (Shannon *et al.*, 2003), the networks constructed for each candidate gene were analysed to determine whether they contained potential hub/driver genes within their network modules and, therefore, could potentially serve as novel biomarkers or targets for treatment. The candidate genes were considered hub genes if they possessed a high degree of connectivity/centrality (a high number of predicted interactions with other co-expressed genes) within their assigned co-expression modules and also had a gene module membership >0.8 (Pearson correlation) (Liu *et al.*, 2019a).

For the candidate genes of interest, the first-degree neighbours were selected in order to reduce the network sizes down to the genes for which there was a strongly predicted association and a directly predicted interaction; these genes have the highest weighted interactions. This step removes weaker interactions that may have passed the previous threshold criteria for extraction and, therefore, represent predicted interactions that may be purely due to chance.

Gene networks were then filtered further using Cytoscape's heat diffusion algorithm. This was run on each of the candidate genes in order to match the first-degree neighbours with the highest connectivity within the network to the candidate genes. The top 25 most highly ranked genes were extracted for visualising the network.

2.11.2 GO analysis of co-expression networks

Genes from the modules with candidate genes were analysed using GO (Michael *et al.*, 2000; The Gene Ontology Consortium, 2018) to identify which biological pathways these co-expression modules were significantly ($P \leq 0.05$) involved in and determine their potential function in the cancers and subtypes under consideration. A Benjamini–Hochberg false discovery correction was applied to reduce type 1 error (detecting false positives).

Chapter 3. Results

In this chapter, the results of the differentially expressed gene (DEG) analysis are presented; firstly for each of the individual subtypes of breast, ovarian and prostate cancers and secondly across breast, ovarian and prostate cancers to identify cross-cancer DEGs with common changes in expression. The novel cross-cancer DEGs are then identified and their relation to survival is discussed. Finally, the novel cross-cancer DEGs were analysed to determine whether they were potential hub genes and therefore potentially useful as diagnostic or prognostic markers and/or treatment targets.

3.1 Differentially expressed gene (DEG) analysis

The DEG analysis identified thousands of differentially expressed genes in the different cancers and subtypes analysed. This provided an opportunity to in part verify the integration method through the identification of known cancer genes. Here the PAM50 gene list from breast cancer was used as potential confirmation. The choice of this gene list allowed for the integrated breast cancer data (specifically molecular) to be compared and identify if any abnormal values were present. This was done on the top 50 DEGs from each subtype and shown in appendices Table 45-69. The PAM50 genes were also looked at in the ovarian and prostate cancer subtypes for further validation in these cancers. Additionally DEGs that were the most significantly downregulated or upregulated are briefly discussed in regards to findings in cancers to also aid with additional validation.

The DEG lists were ranked primarily based on their significance due to the number of DEGs identified. Whilst this is an arbitrary ranking, it is important to note that if ranked on log₂ fold change (log₂FC), a different list would be obtained.

3.1.1 Breast cancer histological DEGs

The analysis of breast cancer histological subtypes utilised 15 datasets with 741 IDC samples, 148 ILC, and 99 DCIS. The number of significant DEGs (upregulated and downregulated) identified in

breast cancer histological subtypes are shown in Table 25. The lowest number of significant DEGs was observed in the non-invasive IDC subtype.

Table 25. Summary numbers of DEGs identified in breast histological subtype analysis.

Subtype	Total number of significant DEGs	Upregulated	Downregulated
IDC	8,700	4,911	3,789
ILC	7,244	4,265	2,979
DCIS	6,850	4,052	2,798

The total number of DEGs identified in breast cancer histological subtypes IDC, ILC, and DCIS. The number of upregulated and downregulated DEGs are also shown for each subtype.

IDC: Invasive Ductal Carcinoma, ILC: Invasive Lobular Carcinoma, DCIS: Ductal Carcinoma in-situ.

The 10 most highly significant DEGs, i.e. those ranked by the highest significance adjusted p-value in each of the three breast cancer histological subtypes (IDC, ILC and DCIS) are shown in Table 26. Notably, three of the classification and prognostic PAM50 genes *RRM2*, *MELK*, and *CCNB1* were identified in the top 50 most significant DEGs as shown in appendices Table 45-47. The PAM50 *RRM2* gene was observed across all histological subtypes within the top 10 significant DEGs. It was found to be most significantly upregulated DEG in both IDC and ILC samples and the second highest after *COL10A1* in DCIS. *MELK* and *CCNB1* were both observed in the top 50 DEGs of IDC (appendices Table 45). With *MELK* also being observed in the top 50 DCIS DEGs (appendices Table 46).

Table 26. The top 10 most significant DEGs identified in histological breast cancer subtype analysis.

DEG	Ensembl identifier	Adjusted P-Value	Log2 Fold Change
IDC			
<i>OXTR</i>	ENSG00000180914	1.59E-291	-3.38
<i>FXVD1</i>	ENSG00000266964	2.59E-285	-1.43
<i>TNXA</i>	ENSG00000248290	1.12E-274	-1.57
<i>INHBA</i>	ENSG00000122641	1.69E-251	1.76
<i>VEGFD</i>	ENSG00000165197	1.42E-250	-2.42
<i>RRM2 (PAM50)</i>	ENSG00000171848	1.32E-249	3.02
<i>CNN1</i>	ENSG00000130176	4.77E-238	-1.97
<i>ZWINT</i>	ENSG00000122952	2.00E-227	2.13
<i>MME</i>	ENSG00000196549	2.56E-222	-2.16
<i>NR3C2</i>	ENSG00000151623	7.04E-217	-1.73
DCIS			
<i>OXTR</i>	ENSG00000180914	2.17E-166	-3.25
<i>TNXA</i>	ENSG00000248290	2.56E-156	-1.51
<i>INHBA</i>	ENSG00000122641	6.05E-146	1.73
<i>FXVD1</i>	ENSG00000266964	8.41E-146	-1.28
<i>VEGFD</i>	ENSG00000165197	1.30E-136	-2.29
<i>CNN1</i>	ENSG00000130176	9.78E-132	-1.89
<i>RRM2 (PAM50)</i>	ENSG00000171848	3.54E-125	2.73
<i>COL10A1</i>	ENSG00000123500	2.14E-124	3.49
<i>WIF1</i>	ENSG00000156076	6.30E-121	-3.63
<i>ZWINT</i>	ENSG00000122952	5.93E-120	2.00
ILC			

DEG	Ensembl identifier	Adjusted P-Value	Log2 Fold Change
<i>RRM2</i> (PAM50)	ENSG00000171848	8.10E-122	3.05
<i>OXTR</i>	ENSG00000180914	1.04E-121	-3.05
<i>FXYD1</i>	ENSG00000266964	3.95E-118	-1.28
<i>TNXA</i>	ENSG00000248290	2.96E-112	-1.41
<i>VEGFD</i>	ENSG00000165197	7.88E-110	-2.28
<i>INHBA</i>	ENSG00000122641	2.92E-100	1.58
<i>ZWINT</i>	ENSG00000122952	1.17E-95	1.99
<i>MME</i>	ENSG00000196549	6.20E-91	-1.98
<i>HLF</i>	ENSG00000108924	2.45E-90	-1.67
<i>HOXA4</i>	ENSG00000197576	1.77E-82	-0.86

Primary IDC, ILC, and DCIS tumours were compared to normal breast samples individually. The DEGs are ordered by the highest significance adjusted p-value. DEGs highlighted in red were downregulated, and in green were upregulated. Known PAM50 genes are shown in brackets.

IDC: Invasive Ductal Carcinoma, ILC: Invasive Lobular Carcinoma, DCIS: Ductal Carcinoma in-situ.

In the non-invasive, DCIS, subtype *COL10A1* was a DEG that was not identified in the other subtypes in the top 10 most significant DEGs. The *COL10A1* gene has been observed in many breast cancer molecular subtypes and associated with poor prognosis and survival (reduced recurrence free survival) (Zhang *et al.*, 2020a). *COL10A1* was also significantly upregulated in IDC and ILC subtypes but was not significant enough to be included in the top 10 most highly significant DEGs list here in Table 26, but was observed in the top 50 genes (appendices Table 45 and Table 47). *COL10A* has not been identified as being differentially expressed in ILC and IDC prior to this study, but has been observed in DCIS tumours with high risk of developing into IDC via ECM remodelling (Giussani *et al.*, 2015). This is an interesting result as it helps in the validation of the integration methodology used here due to identifying a gene previously known specific to the DCIS subtype.

The *OXTR* gene was found to be the most significantly downregulated DEG in all breast cancer histological subtypes, and this is the first study to identify this. The only other study that has described changes in *OXTR* expression is in colon cancer, where downregulation was also identified. Normal expression of *OXTR* suppresses metastasis (Ma *et al.*, 2019).

3.1.2 Ovarian tissue subtype analysis

The ovarian tissue location subtype analysis included 16 datasets with 1,038 samples from primary ovarian tissue, 32 from fallopian tissue, and 34 from peritoneal tissue. The number of significant DEGs (upregulated and downregulated) identified in ovarian cancer tissue subtypes are shown in Table 27.

Table 27. Summary numbers of DEGs identified in ovarian cancer tissue subtype analysis.

Subtype	Total number of significant DEGs	Upregulated	Downregulated
Ovarian	10,539	6,398	4,141
Fallopian	4,275	2,290	1,985
Peritoneum	8,139	4,754	3,385

The total number of DEGs identified in ovarian cancer tissue subtypes ovarian, fallopian, and peritoneum. The number of upregulated and downregulated DEGs are also shown for each subtype.

The 10 most highly significant DEGs identified in ovarian cancer tissue subtypes are shown in Table 28. Notably, within the top 10 (and 50, appendix Table 48-50) significantly identified DEGs one PAM50 DEG, *PTTG1*, was upregulated. *PTTG1* was identified specifically in fallopian tissue samples as being significantly upregulated ($\log_2FC = 2.77$, $P=1.84E-14$). Though this was not seen in the top 50 DEGs for ovarian and peritoneum tissue subtypes, *PTTG1* was significantly upregulated in ovarian tissues samples ($\log_2FC = 2.23$, $P= 6.36E-100$) and peritoneum ($\log_2FC = 2.27$, $P= 1.84E-33$).

Table 28. The top 10 most significant DEGs identified in ovarian tissue location subtype analysis.

DEG	Ensembl identifier	Adjusted P-Value	Log2 Fold Change
Ovarian			
<i>HAND2-AS1</i>	ENSG00000237125	2.82E-285	-2.36
<i>C21orf62</i>	ENSG00000205929	4.52E-262	-1.81
<i>CLEC4M</i>	ENSG00000104938	1.27E-213	-2.20
<i>PDE8B</i>	ENSG00000113231	4.99E-200	-2.14
<i>FAM153B</i>	ENSG00000182230	7.00E-196	-3.16
<i>PRG4</i>	ENSG00000116690	1.88E-193	-2.52
<i>CLDN15</i>	ENSG00000106404	5.07E-188	-1.78
<i>BNC1</i>	ENSG00000169594	1.84E-185	-2.80
<i>KDR</i>	ENSG00000128052	2.38E-183	-2.46
<i>REEP1</i>	ENSG00000068615	5.75E-173	-3.25
Fallopian			
<i>C21orf62</i>	ENSG00000205929	4.15E-39	-1.95
<i>HAND2-AS1</i>	ENSG00000237125	3.44E-35	-2.24
<i>CLEC4M</i>	ENSG00000104938	2.16E-30	-2.42
<i>BNC1</i>	ENSG00000169594	1.81E-26	-3.18
<i>FAM153B</i>	ENSG00000182230	6.29E-26	-3.41
<i>PDE8B</i>	ENSG00000113231	9.60E-26	-2.27
<i>ABCA8</i>	ENSG00000141338	1.27E-24	-3.15
<i>PRG4</i>	ENSG00000116690	3.64E-24	-2.64
<i>CLDN15</i>	ENSG00000106404	6.72E-23	-1.85
<i>AOX1</i>	ENSG00000138356	4.04E-22	-3.20
Peritoneum			

DEG	Ensembl identifier	Adjusted P-Value	Log2 Fold Change
<i>C21orf62</i>	ENSG00000205929	6.26E-105	-1.82
<i>HAND2-AS1</i>	ENSG00000237125	5.18E-97	-2.11
<i>CLEC4M</i>	ENSG00000104938	1.57E-84	-2.27
<i>FAM153B</i>	ENSG00000182230	9.72E-80	-3.35
<i>PDE8B</i>	ENSG00000113231	9.80E-73	-2.13
<i>LOC100507387</i>	-	8.20E-69	-1.72
<i>CLDN15</i>	ENSG00000106404	1.23E-68	-1.79
<i>GPRASP1</i>	ENSG00000198932	9.96E-68	-3.06
<i>PRG4</i>	ENSG00000116690	2.29E-66	-2.43
<i>CD24</i>	ENSG00000272398	9.89E-65	3.87

Primary ovarian tissue, fallopian tissue, and peritoneum tissue tumours were compared to normal tissue samples from the same location individually. The DEGs are ordered by the highest significance adjusted p-value. DEGs highlighted in red were downregulated, and in green were upregulated. Known PAM50 genes are shown in brackets.

The majority of genes observed in the top 50 were found to be downregulated with the noticeable exception of *CD24*. This gene was significantly upregulated in ovarian ($\log_2FC = 3.51$, $P=1.98E-155$), fallopian ($\log_2FC = 3.52$, $P=1.3E-16$), and peritoneum ($\log_2FC = 3.87$, $P=9.89E-65$) tissues. This gene has been observed to be an important prognostic biomarker in ovarian cancers previously, with reduced survival times observed with increased expression and increased metastasis (Tarhriz *et al.*, 2019).

The same genes were identified to be significantly upregulated and downregulated in ovarian, fallopian and peritoneum samples. However, the exact level of expression and significance varied slightly from tissue subtype to subtype. For example, the lncRNA *HAND2-AS1*, which has been found recently to act as a tumour suppressor in ovarian cancers (Gokulnath *et al.*, 2020) was observed to be highly significantly downregulated in all three tissue subtypes (Table 28).

However, the expression change was lowest in ovarian tissue samples ($\log_2FC = -2.36$) and highest in peritoneum tissue samples ($\log_2FC = -2.11$). This highlights a key point discussed in section 1.5

that ovarian carcinomas can develop from cells that have migrated from the fallopian or peritoneum to the ovaries.

3.1.3 Prostate cancer Gleason Grade group analysis

As described in section 2.2.3, prostate cancer samples were classified by Gleason score and re-classified into Gleason grade groups. This analysis included 8 datasets with 77 samples from Gleason grade 1, 133 from Gleason grade 2 & 3, 13 from Gleason grade 4, and 37 from Gleason grade 5. The total number of significant DEGs identified in prostate cancer Gleason grade groups is shown in Table 29.

Table 29. Summary numbers of DEGs identified in prostate cancer Gleason grade subtype analysis.

Subtype	Total number of significant DEGs	Upregulated	Downregulated
Gleason grade group 1	6,630	3,949	2,681
Gleason grade group 2 & 3	7,039	4,196	2,843
Gleason grade group 4	4,992	3,219	1,773
Gleason grade group 5	5,977	3,637	2,340

The total number of DEGs identified in prostate cancer Gleason grade groups 1, 2&3, 4, and 5. The number of upregulated and downregulated DEGs are also shown for each subtype.

Notably, three of the genes identified in prostate Gleason grade analysis were known PAM50 genes. These were *MKI67*, *BIRC5* and *RRM2* genes and are shown in the top 50 DEGs in appendices Table 51-54. The log₂FC expression of *MKI67* was significantly upregulated with the expression in Gleason grade group 1 of 0.81 (P=2.10E-63), Gleason grade group 2 & 3 being 0.87 (P=1.74E-81), 0.92 (P=1.37E-27) for Gleason Grade group 4, and 0.81 (P=5.80E-44) for Gleason grade 5. The log₂FC expression of *BIRC5* was also significantly upregulated across the Gleason grade groups being 0.90 (P=1.45E-57), 1.00 (P=3.56E-78), 1.02 (P=5.05E-24), and 0.9 (P= 4.06E-39) in Gleason grade groups 1 to 5 (group 2 & 3 are combined) respectively. *RRM2* was only identified

to be significantly expressed in Gleason grade groups 2 & 3 ($\log_2FC = 2.27$, $P=2.09E-72$) and Gleason grade group 4 ($\log_2FC = 2.50$, $P=7.5E-25$).

Many of the top 10 and top 50 genes identified as being differentially expressed in each of the Gleason grade groups were found to have similar levels of change, as is shown in Table 30.

Interestingly, within the top 10 genes identified, the *SUPT3H* gene was found to be the most significantly downregulated gene across all the Gleason grade groups. The *SUPT3H* gene has not been identified in prostate cancers prior to this study but has been observed in ovarian cancer where its downregulated expression was associated with reduced progression free survival (PFS) (Chen *et al.*, 2020).

Table 30. The top 10 most significant DEGs identified in prostate Gleason grade group subtype analysis.

DEG	Ensembl identifier	Adjusted P-Value	Log2 Fold Change
Grade group 1			
<i>SUPT3H</i>	ENSG00000196284	9.03E-121	-1.78
<i>VSNL1</i>	ENSG00000163032	1.27E-111	-2.09
<i>TRAF3IP2</i>	ENSG00000056972	4.65E-100	-1.47
<i>LAMB3</i>	ENSG00000196878	2.89E-95	-2.91
<i>ZNF711</i>	ENSG00000147180	1.14E-92	-2.00
<i>GGA2</i>	ENSG00000103365	1.78E-88	-0.83
<i>TP63</i>	ENSG00000073282	6.07E-83	-1.86
<i>SLC14A1</i>	ENSG00000141469	1.03E-82	-4.04
<i>LUZP2</i>	ENSG00000187398	1.83E-82	2.67
<i>ALOX12P2</i>	ENSG00000262943	4.74E-79	-1.35
Grade Group 2 & 3			
<i>SUPT3H</i>	ENSG00000196284	4.03E-142	-1.86
<i>VSNL1</i>	ENSG00000163032	1.12E-131	-2.17
<i>TRAF3IP2</i>	ENSG00000056972	5.18E-117	-1.50
<i>LAMB3</i>	ENSG00000196878	1.38E-108	-2.89
<i>ZNF711</i>	ENSG00000147180	7.57E-108	-2.02
<i>GGA2</i>	ENSG00000103365	5.66E-103	-0.84
<i>SLC14A1</i>	ENSG00000141469	2.69E-101	-4.25
<i>LUZP2</i>	ENSG00000187398	7.22E-98	2.73
<i>TP63</i>	ENSG00000073282	2.12E-97	-1.89
<i>ALOX12P2</i>	ENSG00000262943	4.46E-96	-1.40
Grade Group 4			

DEG	Ensembl identifier	Adjusted P-Value	Log2 Fold Change
<i>SUPT3H</i>	ENSG00000196284	8.91E-60	-1.95
<i>VSNL1</i>	ENSG00000163032	4.23E-54	-2.30
<i>TRAF3IP2</i>	ENSG00000056972	5.17E-52	-1.74
<i>LAMB3</i>	ENSG00000196878	2.52E-46	-3.33
<i>GGA2</i>	ENSG00000103365	2.21E-38	-0.90
<i>LUZP2</i>	ENSG00000187398	1.19E-36	2.98
<i>TP63</i>	ENSG00000073282	1.72E-35	-2.02
<i>HOXC4</i>	ENSG00000198353	3.84E-35	2.04
<i>ZNF711</i>	ENSG00000147180	1.47E-34	-1.94
<i>SLC14A1</i>	ENSG00000141469	1.47E-34	-4.32
Grade Group 5			
<i>SUPT3H</i>	ENSG00000196284	5.95E-96	-1.86
<i>VSNL1</i>	ENSG00000163032	1.02E-83	-2.10
<i>TRAF3IP2</i>	ENSG00000056972	2.62E-72	-1.46
<i>ZNF711</i>	ENSG00000147180	1.70E-68	-2.03
<i>LAMB3</i>	ENSG00000196878	8.27E-66	-2.80
<i>LUZP2</i>	ENSG00000187398	2.38E-61	2.76
<i>ALOX12P2</i>	ENSG00000262943	1.03E-60	-1.43
<i>GGA2</i>	ENSG00000103365	1.34E-60	-0.80
<i>SLC14A1</i>	ENSG00000141469	1.12E-57	-3.98
<i>PDPN</i>	ENSG00000162493	1.95E-57	-0.90

Primary prostate cancer tumour graded by Gleason score was converted to equivalent Gleason grade group and compared to normal prostate tissue. The DEGs are ordered by the highest significance adjusted p-value. DEGs highlighted in red were downregulated, and in green were upregulated. Known PAM50 genes are shown in brackets.

The *LUZP2* gene was found to be a consistently upregulated gene in the top 10 genes across all Gleason grade groups (Table 30). Previously, changes in *LUZP2* expression have been associated

with hormone positive prostate cancers where they were observed to be downregulated in the development of castration resistant prostate cancer (CRPC) (Zhao *et al.*, 2016). The upregulated expression of *LUZP2* has been associated with reduced overall survival (OS) in gliomas and pancreatic cancer (Li *et al.*, 2020; Chen *et al.*, 2021).

In addition to *LUZP2* upregulation, the *HOXC4* gene was also observed in the top 10 genes in Gleason grade group 4 (Table 30). However, *HOXC4* was also significantly upregulated in the other Gleason grade groups analysed. In Gleason grade 1 the log fold change (log2FC) was 1.76 (P=6.77E-76), Gleason grade 2&3 1.86 (P=2.10E-94), in Gleason grade 4 2.04 (P=3.84E-35), and Gleason grade 5 1.83 (P=5.42E-57). The disrupted expression of *HOXC4* has previously been observed in prostate cancer with higher expression associated with increased aggressiveness through increased rates of proliferation (Luo and Farnham, 2020).

3.1.4 Breast cancer molecular subtype analysis

The differential gene expression analysis (DEGs) of breast cancer molecular subtypes included 9 datasets with 37 samples from normal-like, 181 from luminal A, 124 from luminal B, 187 from HER2, and 226 from basal subtypes. The total number of significant DEGs (upregulated and downregulated DEGs) identified in breast cancer molecular subtypes is shown in Table 31.

Table 31. Summary numbers of DEGs identified in breast cancer molecular subtype analysis.

Subtype	Total number of significant DEGs	Upregulated	Downregulated
Normal-like	6,702	2,818	3,884
Luminal A	8,635	4,000	4,635
Luminal B	8,524	4,349	4,175
HER2	8,835	4,216	4,619
Basal	9,061	4,575	4,486

The total number of DEGs identified in breast cancer molecular subtypes normal-like, luminal A, luminal B, HER2, and Basal. The number of upregulated and downregulated DEGs are also shown for each subtype.

The lowest number of significant DEGs was identified in normal-like breast cancer. These are ER+ and PR+, similar to luminal A carcinomas, but are more similar to normal breast growth with low levels of proliferation than the luminal A types. Because of this, they are difficult to classify as their own group because they are often mistaken for luminal A. The identification of a lower number of significant DEGs would be expected as these samples are considered to be more similar to normal breast tissue and would therefore have the least differentiated phenotype of the molecular subtypes.

The 10 most significant genes are shown in Table 32 and the top 50 gene lists in appendices Table 55-59. Four known PAM50 genes were identified in the top 50 gene lists of molecular subtypes: *RRM2*, *MELK*, *CEP55*, and *UBE2C*. The *RRM2* gene was only identified in the HER2 positive subtype ($\log_2FC = 4.01$, $P=7.64E-87$) and the basal subtype ($\log_2FC = 4.23$, $P=5.271E-95$) being significantly upregulated in both. Secondly, the *MELK*, *CEP55*, and *UBE2C* gene was only identified in the basal subtype ($\log_2FC = 3.74$, $P=9.14E-94$; $\log_2FC = 3.90$, $P=3.16E-92$; $\log_2FC = 3.36$, $P=1.23E-90$).

Recently, the increased expression of *RRM2* has been found to be observed in other HER2 positive and basal like with an association with significantly shorter recurrence free survival (RFS) (Abdel-Rahman, Mahfouz and Habashy, 2022). The normal-like, luminal A, and luminal B subtypes did

show significant expression of *RRM2* which are also frequently ER+ and PR+. Previously, it has also been seen that both ER+ and/or PR+ expressing breast cancer samples have been found to be negatively correlated with *RRM2* expression and associated with poorer survival (RFS and OS) (Chen *et al.*, 2019) coinciding with findings in this study.

Table 32. The top 10 most significant DEGs identified in breast cancer molecular subtype analysis.

DEG	Ensembl identifier	Adjusted P-Value	Log2 Fold Change
Normal-Like			
<i>CA4</i>	ENSG00000167434	2.06E-107	-2.10
<i>LHCGR</i>	ENSG00000138039	8.08E-98	-1.07
<i>GLYAT</i>	ENSG00000149124	3.90E-79	-1.51
<i>KANK3</i>	ENSG00000186994	9.73E-79	-1.74
<i>GPD1</i>	ENSG00000167588	8.46E-70	-1.68
<i>NPR1</i>	ENSG00000169418	2.19E-67	-1.38
<i>MYOM1</i>	ENSG00000101605	6.48E-60	-1.43
<i>LEP</i>	ENSG00000174697	7.43E-59	-3.68
<i>TIMP4</i>	ENSG00000157150	8.90E-59	-3.18
<i>CIDEC</i>	ENSG00000187288	2.00E-57	-2.61
Luminal A			
<i>CA4</i>	ENSG00000167434	4.87E-158	-2.33
<i>KANK3</i>	ENSG00000186994	5.30E-135	-2.10
<i>NPR1</i>	ENSG00000169418	2.04E-129	-1.77
<i>LHCGR</i>	ENSG00000138039	9.84E-125	-1.08
<i>GPD1</i>	ENSG00000167588	1.92E-113	-1.95
<i>MYOM1</i>	ENSG00000101605	5.86E-102	-1.70
<i>GLYAT</i>	ENSG00000149124	1.38E-97	-1.48
<i>TNXA</i>	ENSG00000248290	1.08E-96	-2.29
<i>LEP</i>	ENSG00000174697	1.08E-96	-4.26
<i>VEGFD</i>	ENSG00000165197	1.13E-96	-3.26
Luminal B			

DEG	Ensembl identifier	Adjusted P-Value	Log2 Fold Change
<i>CA4</i>	ENSG00000167434	3.92E-154	-2.39
<i>KANK3</i>	ENSG00000186994	3.26E-127	-2.10
<i>NPR1</i>	ENSG00000169418	6.45E-120	-1.75
<i>LHCGR</i>	ENSG00000138039	5.08E-118	-1.09
<i>GPD1</i>	ENSG00000167588	2.04E-113	-2.02
<i>SIK2</i>	ENSG00000170145	1.15E-108	-1.36
<i>FXD1</i>	ENSG00000266964	9.04E-108	-2.38
<i>TNXA</i>	ENSG00000248290	2.24E-103	-2.47
<i>TNXB</i>	ENSG00000168477	7.15E-102	-1.27
<i>CLDN5</i>	ENSG00000184113	2.17E-101	-2.54
HER2			
<i>CA4</i>	ENSG00000167434	5.34E-167	-2.42
<i>KANK3</i>	ENSG00000186994	2.01E-145	-2.20
<i>LHCGR</i>	ENSG00000138039	6.75E-131	-1.11
<i>NPR1</i>	ENSG00000169418	2.03E-127	-1.75
<i>GPD1</i>	ENSG00000167588	9.39E-125	-2.06
<i>ACACB</i>	ENSG00000076555	3.53E-117	-2.46
<i>CLDN5</i>	ENSG00000184113	4.46E-112	-2.59
<i>GLYAT</i>	ENSG00000149124	6.52E-111	-1.60
<i>MYOM1</i>	ENSG00000101605	1.31E-110	-1.78
<i>CLEC3B</i>	ENSG00000163815	4.93E-110	-3.81
Basal			
<i>CA4</i>	ENSG00000167434	5.56E-163	-2.38
<i>KANK3</i>	ENSG00000186994	9.10E-145	-2.19

DEG	Ensembl identifier	Adjusted P-Value	Log2 Fold Change
<i>ACACB</i>	ENSG00000076555	4.22E-127	-2.59
<i>LHCGR</i>	ENSG00000138039	2.84E-126	-1.09
<i>GPD1</i>	ENSG00000167588	9.99E-122	-2.03
<i>CLEC3B</i>	ENSG00000163815	1.51E-120	-4.03
<i>CLDN5</i>	ENSG00000184113	9.80E-116	-2.64
<i>FXD1</i>	ENSG00000266964	2.13E-115	-2.38
<i>SEMA3G</i>	ENSG00000010319	1.18E-113	-3.59
<i>NPR1</i>	ENSG00000169418	1.34E-113	-1.62

Primary breast cancer tumours were classified into molecular subtypes and compared to equivalent normal breast tissue. The genes are ordered by the highest significance adjusted p-value. DEGs highlighted in red were downregulated, and in green were upregulated. Known PAM50 genes are shown in brackets

HER2: Human epidermal growth factor receptor

3.1.5 Ovarian cancer epithelial subtype analysis

The analysis of ovarian epithelial subtypes included 12 datasets with 413 samples from the serous subtype, 74 from the endometrioid subtype, 30 from the mucinous subtype, and 39 from the clear cell subtype. The total significant DEGs identified are shown in Table 33.

Table 33. Summary numbers of DEGs identified in ovarian cancer epithelial subtype analysis.

Subtype	Total number of significant DEGs	Upregulated	Downregulated
Serous	9,686	5,662	4,024
Endometrioid	8,266	4,758	3,508
Mucinous	7,888	4,575	3,313
Clear cell	8,561	5,046	3,515

The total number of DEGs identified in ovarian cancer epithelial subtypes serous, endometrioid, mucinous, and clear cell.

The number of upregulated and downregulated DEGs are also shown for each subtype.

The 10 most highly significant DEGs are shown in Table 34. Five known PAM50 genes were identified in at least one of the subtypes top 50 DEG lists in appendices Table 60-63. *PTTG1* was significantly upregulated in serous subtype ($\log_2FC = 2.51$, $P=8.62E-74$) as well as *CENPF* ($\log_2FC = 1.70$, $P=3.32E-71$), and *CEP55* ($\log_2FC = 3.23$, $P=3.53E-71$). In clear cell *CCNE1* was significantly upregulated ($\log_2FC = 2.16$, $P=1.56E-45$) and *RRM2* ($\log_2FC = 3.52$, $P= 1.77E-43$). Notably there were no PAM50 genes observed in the top 50 DEGs lists for endometrioid and mucinous subtypes. This does not mean that they are not present and an expanded search through the full lists would likely identify them as significant DEGs. However, the aim of these results was to take a portion of the total DEGs and validate without introducing potential bias.

Table 34. The top 10 most significant genes identified in ovarian cancer epithelial subtype analysis.

DEG	Ensembl identifier	Adjusted P-Value	Log2 Fold Change
Serous			
<i>HAND2-AS1</i>	ENSG00000237125	9.12E-185	-2.68
<i>C21orf62</i>	ENSG00000205929	1.41E-124	-1.71
<i>LOC100507387</i>	-	1.66E-122	-1.87
<i>GPRASP1</i>	ENSG00000198932	2.98E-117	-3.23
<i>PDE8B</i>	ENSG00000113231	2.59E-115	-2.16
<i>ABCA8</i>	ENSG00000141338	8.69E-100	-2.37
<i>CLEC4M</i>	ENSG00000104938	3.65E-99	-2.05
<i>ALDH1A2</i>	ENSG00000128918	2.66E-96	-2.53
<i>RADX</i>	ENSG00000147231	8.84E-96	-1.16
<i>WNT2B</i>	ENSG00000134245	1.48E-95	-2.16
Endometrioid			
<i>HAND2-AS1</i>	ENSG00000237125	2.86E-134	-2.62
<i>LOC100507387</i>	-	5.42E-90	-1.92
<i>C21orf62</i>	ENSG00000205929	8.53E-84	-1.65
<i>FAM153B</i>	ENSG00000182230	3.90E-75	-2.95
<i>ALDH1A2</i>	ENSG00000128918	6.86E-75	-2.73
<i>GPRASP1</i>	ENSG00000198932	2.52E-74	-3.02
<i>CLEC4M</i>	ENSG00000104938	1.48E-69	-2.07
<i>RADX</i>	ENSG00000147231	1.10E-66	-1.17
<i>AOX1</i>	ENSG00000138356	8.85E-66	-2.94
<i>CLDN15</i>	ENSG00000106404	1.18E-65	-1.66
Mucinous			

DEG	Ensembl identifier	Adjusted P-Value	Log2 Fold Change
<i>HAND2-AS1</i>	ENSG00000237125	1.33E-96	-2.68
<i>LOC100507387</i>	-	6.78E-61	-1.94
<i>FAM153B</i>	ENSG00000182230	4.80E-54	-3.13
<i>C21orf62</i>	ENSG00000205929	3.16E-53	-1.62
<i>PDE8B</i>	ENSG00000113231	5.79E-53	-2.15
<i>INAVA</i>	ENSG00000163362	3.38E-50	4.17
<i>ALDH1A2</i>	ENSG00000128918	4.65E-47	-2.67
<i>LGALS4</i>	ENSG00000171747	6.85E-46	3.66
<i>CELF2</i>	ENSG00000048740	1.36E-45	-1.94
<i>TFF2</i>	ENSG00000160181	7.92E-45	2.91
Clear Cell			
<i>HAND2-AS1</i>	ENSG00000237125	5.49E-108	-2.66
<i>C21orf62</i>	ENSG00000205929	1.39E-61	-1.62
<i>GPRASP1</i>	ENSG00000198932	2.43E-61	-3.19
<i>LOC100507387</i>	-	1.99E-58	-1.73
<i>LBP</i>	ENSG00000129988	2.11E-58	2.66
<i>HAVCR1</i>	ENSG00000113249	2.33E-56	2.07
<i>REEP1</i>	ENSG00000068615	8.45E-56	-3.69
<i>CD24</i>	ENSG00000272398	8.45E-56	4.08
<i>ALDH1A2</i>	ENSG00000128918	1.86E-55	-2.69
<i>PDE8B</i>	ENSG00000113231	2.55E-55	-2.02

Primary ovarian cancer tumours were classified into epithelial subtypes and compared to equivalent normal ovarian tissue. The DEGs are ordered by the highest significance adjusted p-value. DEGs highlighted in red were downregulated, and in green were upregulated. Known PAM50 genes are shown in brackets.

The *INAVA* gene was found to be the most significantly upregulated gene in the mucinous subtype (log2FC = 4.17, P=3.38E-50) and also significantly upregulated in the clear cell subtype (log2FC =

3.67, $P=5.67E-47$). It was not significantly differentially expressed in the serous or endometrioid subtypes ($P<0.001$). The upregulated *INAVA* expression has been observed to promote increased aggressiveness through the downstream upregulation of matrix metalloproteinase 9 (*MMP9*) in papillary thyroid cancer (PTC) (Guan *et al.*, 2018). *INAVA* has also been found to promote cell migration in ovarian cancers (Zhao *et al.*, 2020) and this is most likely through the same mechanism affecting cell adhesion. However, this is the first study to identify specific epithelial subtypes. *HAND2-AS1* was found to be the most significantly downregulated gene across the epithelial subtypes as identified also in the ovarian tissue subtype analysis (section 3.1.2).

3.1.6 Prostate cancer Gleason Score analysis

Gleason scores and Gleason grade groups were kept as separate DEG analyses as each classification was used in a separate cross-cancer analysis (section 2.9). The analysis of prostate Gleason scores included 8 datasets with 8 samples from the Gleason score 4, 9 from Gleason score 5, 68 from Gleason score 6, 133 from Gleason score 7, 13 from Gleason score 8, and 37 from Gleason score 9. The number of significant DEGs identified in prostate Gleason score analysis are shown in Table 35.

Table 35. Summary numbers of DEGs identified in prostate cancer Gleason score subtype analysis.

Subtype	Total number of significant DEGs	Upregulated	Downregulated
Gleason score 4	2,643	1,616	1,027
Gleason score 5	3,016	1,653	1,363
Gleason score 6	6,523	3,930	2,593
Gleason score 7	7,047	4,198	2,849
Gleason score 8	5,008	3,229	1,779
Gleason score 9	6,067	3,650	2,417

The total number of DEGs identified in prostate cancer Gleason scores 4, 5, 6, 7, 8, and 9. The number of upregulated and downregulated DEGs are also shown for each subtype.

The 10 most significant genes are shown in Table 36 and 50 most significant in appendices Table 64-69. Relatively few genes were identified in the Gleason score 4 samples compared to the others. This is most likely due to smaller samples sizes potentially limiting the power of the analysis.

Four PAM50 DEGs were identified in prostate Gleason scores: *MKI67*, *CENPF*, *BIRC5*, and *RRM2*.

The PAM50 gene *MKI67* was identified as significantly upregulated in all Gleason scores with its highest expression in Gleason score 4 ($\log_2FC = 1.40$, $P=3.55E-11$). The *MKI67* expression was similar in other Gleason scores; Gleason score 5 ($\log_2FC = 0.67$, $P=4.69E-12$), Gleason score 6 ($\log_2FC = 0.81$, $P=1.10E-60$), Gleason score 7 ($\log_2FC = 0.86$, $P=5.20E-83$), Gleason score 8 ($\log_2FC = 0.92$, $P=2.42E-28$), and Gleason score 9 ($\log_2FC = 0.81$, $P=6.38E-46$). In other cancers, increased expression of *MKI67* has been identified in triple negative breast cancers and in gastric cancer (GC) as a marker for proliferation (Xiong *et al.*, 2019; Tan *et al.*, 2019). This coincided also with both reduced overall survival (OS) and disease free survival (DFS) (Xiong *et al.*, 2019). *MKI67* has been identified in prostate cancers in which increased *MKI67* expression was associated with higher Gleason scores (8-10) (Green *et al.*, 2016; Hammarsten *et al.*, 2019) and subsequently

reduced recurrence free survival (RFS) and OS (Tretiakova *et al.*, 2016). This is the first study to identify significantly upregulated *MKI67* expression in Gleason scores 4, 5, and 6.

Table 36. The top 10 most significant genes identified in prostate cancer Gleason score subtype analysis.

DEG	Ensembl identifier	Adjusted P-Value	Log2 Fold Change
Gleason score 4			
<i>MOG</i>	ENSG00000204655	5.19E-12	0.96
<i>CENPF (PAM50)</i>	ENSG00000117724	2.01E-11	1.65
<i>TRAF3IP2</i>	ENSG00000056972	2.01E-11	-1.80
<i>MKI67 (PAM50)</i>	ENSG00000148773	3.55E-11	1.40
<i>AURKA</i>	ENSG00000087586	1.24E-10	1.51
<i>SUPT3H</i>	ENSG00000196284	2.56E-10	-1.73
<i>ZNF711</i>	ENSG00000147180	7.08E-10	-2.43
<i>SNAPC4</i>	ENSG00000165684	8.75E-09	1.37
<i>IHH</i>	ENSG00000163501	9.07E-09	1.35
<i>MYEF2</i>	ENSG00000104177	1.28E-08	1.02
Gleason score 5			
<i>SUPT3H</i>	ENSG00000196284	2.16E-39	-1.75
<i>VSNL1</i>	ENSG00000163032	1.31E-36	-2.12
<i>ZNF711</i>	ENSG00000147180	1.78E-27	-2.02
<i>LAMB3</i>	ENSG00000196878	6.25E-26	-2.78
<i>LUZP2</i>	ENSG00000187398	1.11E-25	2.86
<i>GGA2</i>	ENSG00000103365	2.09E-24	-0.82
<i>CDC14B</i>	ENSG00000081377	8.45E-24	-1.15
<i>ALOX12P2</i>	ENSG00000262943	8.45E-24	-1.43
<i>ITGB6</i>	ENSG00000115221	1.01E-22	-1.45
<i>ACSBG1</i>	ENSG00000103740	2.01E-21	-0.72
Gleason score 6			

DEG	Ensembl identifier	Adjusted P-Value	Log2 Fold Change
<i>SUPT3H</i>	ENSG00000196284	1.56E-114	-1.78
<i>VSNL1</i>	ENSG00000163032	3.30E-105	-2.08
<i>TRAF3IP2</i>	ENSG00000056972	7.46E-98	-1.50
<i>LAMB3</i>	ENSG00000196878	5.45E-90	-2.92
<i>ZNF711</i>	ENSG00000147180	2.11E-86	-1.98
<i>GGA2</i>	ENSG00000103365	1.70E-82	-0.83
<i>SLC14A1</i>	ENSG00000141469	2.20E-78	-4.06
<i>TP63</i>	ENSG00000073282	4.43E-78	-1.87
<i>LUZP2</i>	ENSG00000187398	2.66E-75	2.60
<i>ALOX12P2</i>	ENSG00000262943	1.83E-73	-1.34
Gleason score 7			
<i>SUPT3H</i>	ENSG00000196284	1.96E-141	-1.86
<i>VSNL1</i>	ENSG00000163032	5.64E-131	-2.17
<i>TRAF3IP2</i>	ENSG00000056972	3.59E-118	-1.50
<i>LAMB3</i>	ENSG00000196878	5.75E-108	-2.89
<i>ZNF711</i>	ENSG00000147180	1.39E-107	-2.02
<i>GGA2</i>	ENSG00000103365	2.54E-102	-0.84
<i>SLC14A1</i>	ENSG00000141469	4.31E-101	-4.25
<i>LUZP2</i>	ENSG00000187398	9.27E-98	2.73
<i>TP63</i>	ENSG00000073282	6.19E-97	-1.89
<i>ALOX12P2</i>	ENSG00000262943	1.00E-95	-1.40
Gleason score 8			
<i>SUPT3H</i>	ENSG00000196284	1.65E-59	-1.95
<i>VSNL1</i>	ENSG00000163032	8.16E-54	-2.30

DEG	Ensembl identifier	Adjusted P-Value	Log2 Fold Change
<i>TRAF3IP2</i>	ENSG00000056972	6.99E-53	-1.74
<i>LAMB3</i>	ENSG00000196878	4.68E-46	-3.33
<i>GGA2</i>	ENSG00000103365	4.02E-38	-0.90
<i>LUZP2</i>	ENSG00000187398	1.12E-36	2.98
<i>TP63</i>	ENSG00000073282	2.49E-35	-2.02
<i>HOXC4</i>	ENSG00000198353	4.83E-35	2.04
<i>ZNF711</i>	ENSG00000147180	1.39E-34	-1.94
<i>SLC14A1</i>	ENSG00000141469	1.39E-34	-4.32
Gleason score 9			
<i>SUPT3H</i>	ENSG00000196284	4.18E-97	-1.87
<i>VSNL1</i>	ENSG00000163032	7.96E-85	-2.11
<i>TRAF3IP2</i>	ENSG00000056972	5.60E-75	-1.47
<i>ZNF711</i>	ENSG00000147180	1.02E-69	-2.04
<i>LAMB3</i>	ENSG00000196878	5.35E-67	-2.82
<i>LUZP2</i>	ENSG00000187398	2.34E-64	2.83
<i>GGA2</i>	ENSG00000103365	2.16E-62	-0.82
<i>ALOX12P2</i>	ENSG00000262943	7.76E-62	-1.44
<i>SLC14A1</i>	ENSG00000141469	2.73E-59	-4.02
<i>PDPN</i>	ENSG00000162493	1.30E-58	-0.91

Primary prostate cancer tumours were classified into Gleason score subtypes and compared to normal prostate tissue.

The DEGs are ordered by the highest significance adjusted p-value. DEGs highlighted in red were downregulated, and in green were upregulated. Known PAM50 genes are shown in brackets.

In addition to *MKI76*, the *CENPF* gene was found to have a higher level of upregulation in Gleason score 4 samples than in samples with higher Gleason scores; Gleason score 4 (log2FC = 1.65, P=2.01E-11), Gleason score 5 (log2FC =0.62, P=3.06E-8), Gleason score 6 (log2FC = 0.79, P=8.50E-49), Gleason score 7 (log2FC = 0.87, P=6.77E-70), Gleason score 8 (log2FC = 0.94, P=6.18E-23) and

Gleason score 9 ($\log_2FC=0.84$, $P=5.12E-39$). Expression of *CENPF* has previously been found to be associated with higher Gleason score prostate cancer (Göbel *et al.*, 2018). In an analysis of the expression of *CENPF* it was found to significantly increase from Gleason score 6 to 8, similarly to findings here. Contrary to those findings the highest expression of *CENPF* was observed in the lower Gleason score 4 group. However, the previous study did not include prostate tumours of below Gleason score 6 and therefore limited in these lower Gleason score groups.

As was observed with the Gleason grade group analysis (section 3.1.3), both the *LUZP2* and *SUPT3H* genes were found to be similarly expressed. The identification of these two genes across the Gleason score analysis supports the use of the newer Gleason grade group classification being implemented by clinicians in that it appears to capture the same information whilst reducing the overall number of classification groups.

3.1.7 DEG analysis summary

The analysis of differentially expressed genes (DEGs), presented above, has identified a number of significantly upregulated and downregulated genes. Some of these genes have been identified in previous studies, for example PAM50 genes (e.g. *RRM2*), whereas others are novel. The importance of identifying these PAM50 genes helps to validate the integration methodology used here. However, a number of novel findings were identified such as the identification of PAM50 genes being similarly expressed in ovarian and prostate cancers to breast cancers. Some DEGs were identified in previously known subtypes but they were also found to be novel in additional subtypes. For example, *COL10A* has been previously identified as significantly upregulated in DCIS breast cancer subtypes, here it was identified as significantly upregulated in IDC and ILC, which is novel. The novel approach taken in this study, through the pooling/integration approach, resulted in an increased sample size and increased sensitivity.

Many of the significant DEGs described above were notably not subtype specific, but instead are often observed to be significantly altered in multiple subtypes in several cancers (albeit at differing levels of expression). This potentially provides initial evidence that some of these cancers

may be more similarly related in their molecular pathogenesis than what has been previously described in studies.

Though it was not the aim of the DEG analysis, it is also important to note that the finding of changes in expression levels of genes (significant DEGs) such as *MME* in breast cancer histological subtypes or *CENPF* in prostate cancer Gleason score subtypes (where the expression level increases with Gleason score) could also be used to help with more robust classification. For example, the significant DEGs identified here that are conserved across cancer subtypes could be used to create a supervised clustering method to classify tumour samples based on the expression of these genes similarly to the PAM50 genes used in breast cancer currently. Sets of genes from each of these cancers identified could be developed further to classify a tumour regardless of knowledge of its source (i.e. from breast, ovarian, or prostate). Additionally, the genes that are identified further as subtype specific could be used for more robust classification. This is particularly useful for cancer subtypes that are difficult to classify such as the normal-like and luminal A breast cancer subtypes which are often difficult to distinguish due to their phenotype similarities. This is also particularly useful for ovarian and prostate carcinomas as there are no current classifications based on this approach. Whereas for breast cancer there is currently the PAM50 gene list used to aid classification.

Finally, it is important to note that one limitation of this analysis was the focus on only the 10 most disrupted genes. As the number of genes identified for many of the DEG comparisons ranged in the hundreds to thousands, it was not practical to analyse all these genes individually, although the data generated would be available should this be a suitable area for investigation. This focus on the most significantly disrupted genes here was because the results of this DEG analysis were primarily intended to facilitate the identification of cross-cancer targetable genes using the thousands of significant DEGs for comparison. This analysis is presented in the following section.

3.2 Cross-cancer gene expression analysis

Following the identification of significantly differentially expressed genes as detailed in section 3.1, the DEGs from each cancer and their subtypes were cross compared. This was to identify similarly upregulated and downregulated genes working on the expectation that these may undertake similar functions in the different breast, ovarian, and prostate cancers and their subtypes.

Prior to all comparisons, the upregulated and downregulated genes were separated into individual lists. These were then compared across cancers and subtype classifications. For example, the DEG analysis of Luminal A breast cancer identified 8,635 genes of interest, with 4,000 being upregulated and 4,635 downregulated. The upregulated genes were extracted into a separate list and compared to the upregulated gene lists from ovarian tissue samples and prostate Gleason grade groups.

This approach to cross-cancer comparison was chosen based upon evidence described in section 1.5. The comparisons made were based upon breast cancer (histological), ovarian (tissue location), and prostate (Gleason grade) as shown in section 2.9. Comparisons for molecular breast cancer, ovarian epithelial, and prostate Gleason score are shown in section 2.9. The underlying evidence is that basal breast cancers and serous carcinomas share similar phenotypic traits (Begg *et al.*, 2017), therefore providing evidence in which to base and compare the molecular breast cancer subtypes with ovarian cancer epithelial subtypes. These were further compared to prostate Gleason score. This was because breast histological classifications, ovarian tissue subtype, and prostate Gleason grade group are all primarily classified using microscopy imaging. Therefore, it was logical to compare these groups. This left the remaining Gleason scores to be compared to molecular breast and ovarian epithelial subtypes. This is by no means a comprehensive set of comparisons, but due to the novelty of this approach, it was determined to be the most appropriate initial comparison.

The upregulated genes identified to be cross-cancer related and potentially functional were extracted for subsequent survival analysis. Following the cross-cancer subtype comparison the number of genes was greatly reduced (down from thousands to a couple of hundred for both comparison analyses). This reduced number of genes was more appropriate for survival analysis. Findings from the two comparisons are detailed in the following sections.

3.2.1 Histological cross-cancer group

Differentially expressed genes identified in the breast histological (IDC, ILC, and DCIS), ovarian tissue (Ovarian, Peritoneum, and Fallopian), and Gleason grade groups (Group 1–5) were compared. A total of 395 genes were significantly upregulated across all of the histological breast cancer subtypes, ovarian tissue location subtypes, and prostate Gleason grade subtypes. These upregulated genes were taken forward for survival analysis (section 3.4). It was also found that 248 genes were downregulated across the subtypes. The 10 most significantly upregulated and downregulated genes are shown in Table 37.

Table 37. The 10 most highly significant cross-cancer DEGs in breast histological, ovarian tissue, and prostate Gleason group cancer subtypes.

Order	DEG	Ensembl identifier	Average log2 FC	Average P-Value
Upregulated				
1	<i>RRM2</i> (PAM50)	ENSG00000171848	2.51	8.76E-11
2	<i>COMP</i>	ENSG00000105664	1.89	2.46E-04
3	<i>CKS2</i>	ENSG00000123975	1.84	3.11E-08
4	<i>TPX2</i>	ENSG00000088325	1.82	2.53E-13
5	<i>BUB1B</i>	ENSG00000156970	1.76	1.50E-14
6	<i>TOP2A</i>	ENSG00000131747	1.71	1.69E-14
7	<i>MELK</i> (PAM50)	ENSG00000165304	1.70	4.70E-11
8	<i>ISG15</i>	ENSG00000187608	1.62	3.73E-06
9	<i>ZWINT</i>	ENSG00000122952	1.57	5.66E-06
10	<i>DLGAP5</i>	ENSG00000126787	1.53	3.02E-10
Downregulated				
1	<i>CHRD1</i>	ENSG00000101938	-1.93	3.72E-07
2	<i>MEIS2</i>	ENSG00000134138	-1.93	1.01E-10
3	<i>GPRASP1</i>	ENSG00000198932	-1.67	1.04E-07
4	<i>AOX1</i>	ENSG00000138356	-1.62	9.48E-06
5	<i>EFEMP1</i>	ENSG00000115380	-1.61	1.65E-09
6	<i>CAV1</i>	ENSG00000105974	-1.58	7.94E-04
7	<i>AMIGO2</i>	ENSG00000139211	-1.53	2.75E-04
8	<i>ANXA1</i>	ENSG00000135046	-1.48	1.35E-05
9	<i>TGFBR3</i>	ENSG00000069702	-1.47	1.51E-04
10	<i>ANG</i>	ENSG00000214274	-1.42	3.92E-06

DEGs were ranked based on the log2FC. The average log2FC was calculated across the cancers as well as the average P-values. Known PAM50 genes are also shown in brackets.

Of the upregulated genes, the *RRM2* gene was found to be the most highly upregulated in terms of fold change. As described in section 3.1, *RRM2* is a known PAM50 gene and has been observed to be disrupted in breast, ovarian and prostate cancers prior (Mazzu *et al.*, 2020; Zhan *et al.*, 2021; Abdel-Rahman, Mahfouz and Habashy, 2022). The *COMP* gene has also been observed in breast and prostate cancers previously and is associated with reduced recurrence free survival (RFS) via the expression of an epithelial-mesenchymal phenotype increasing metastasis (Englund *et al.*, 2017; Papadakos *et al.*, 2019). This, however, is the first time that it has been observed in ovarian cancer tissues. The *CKS2* gene was also identified to be upregulated across all subtypes. In earlier reports, *CKS2* was reported in metastatic breast cancer and metastatic prostate cancer (Gottardo *et al.*, 2020; Xu, Wang and Xu, 2019) and its expression was associated with increased proliferation in prostate cancer (Lan *et al.*, 2008). This is the first study in which its elevated expression in primary prostate tumours (Gleason grade groups) has been observed.

The commonly downregulated genes identified here, such as *CHRD1* and *TGFBR3*, have been identified in other cancers with tumour suppressive roles. In addition, the downregulation of the *MEIS2* gene has been found to promote resistance to oxaliplatin (Wang *et al.*, 2019b). This chemotherapy drug is used in the treatment of ovarian and prostate cancers (Bogliolo *et al.*, 2015; Zhou *et al.*, 2017) and metastatic breast cancers (Delpeuch *et al.*, 2011).

3.2.2 Molecular cross-cancer group

DEGs from breast molecular (Normal like, Luminal A, Luminal B, HER2, and Basal), ovarian epithelial (Serous, Endometrioid, Mucinous, and Clear Cell), and prostate Gleason scores (4-9) data sets were cross-cancer compared. Like the previous histological subtype level equivalent analysis, both the commonly upregulated and downregulated genes were identified. There was a total of 75 upregulated genes identified across the subtypes compared. For the downregulated genes 71, were identified. The 10 most significantly upregulated and downregulated genes are shown in Table 38. Similar to the previous histological subtype comparison, *RRM2*, *MELK* as well as *PTTG1* were found to be commonly elevated in breast molecular, ovarian epithelial, and prostate Gleason score subtypes.

Table 38. The 10 most highly significant cross-cancer DEGs in breast molecular, ovarian epithelial, and prostate Gleason score cancer subtypes.

Order	DEG	Ensembl identifier	Average log2 FC	Average P-Value
Upregulated				
1	<i>RRM2</i> (PAM50)	ENSG00000171848	2.86	5.07E-06
2	<i>TPX2</i>	ENSG00000088325	2.07	6.35E-05
3	<i>MELK</i>	ENSG00000165304	2.05	9.85E-06
4	<i>BUB1B</i>	ENSG00000156970	2.04	4.75E-05
5	<i>TOP2A</i>	NSG00000131747	1.92	2.89E-06
6	<i>UBE2C</i>	ENSG00000175063	1.83	1.75E-04
7	<i>EZH2</i>	ENSG00000106462	1.79	2.63E-09
8	<i>CCNB1</i> (PAM50)	ENSG00000134057	1.78	2.40E-04
9	<i>PTTG1</i> (PAM50)	ENSG00000164611	1.72	4.95E-05
10	<i>KIF11</i>	ENSG00000138160	1.71	6.36E-04
Downregulated				
1	<i>TGFBR3</i>	ENSG00000069702	-1.96	3.38E-04
2	<i>MEIS2</i>	ENSG00000134138	-1.83	1.10E-04
3	<i>EFEMP1</i>	ENSG00000115380	-1.78	8.38E-05
4	<i>AOX1</i>	ENSG00000138356	-1.69	1.09E-05
5	<i>PDK4</i>	ENSG00000004799	-1.68	1.26E-04
6	<i>DPT</i>	ENSG00000143196	-1.62	2.57E-04
7	<i>ANXA1</i>	ENSG00000135046	-1.61	1.02E-04
8	<i>ZNF204P</i>	ENSG00000204789	-1.53	1.68E-05
9	<i>ALDH1A2</i>	ENSG00000128918	-1.46	2.10E-04
10	<i>ID4</i>	ENSG00000172201	-1.32	3.19E-04

DEGs were ranked based on the log2FC. The average log2FC was calculated across the cancers as well as the average P-values. Known Pam50 genes are also shown in brackets.

The *TGFBR3* gene was found to be the most downregulated gene in regard to log₂FC. In breast cancers, the loss of *TGFBR3* has been reported to be associated with disease progression and reduced RFS (Dong *et al.*, 2007). This is also the case in ovarian cancer, where loss of *TGFBR3* expression was associated with reduced invasiveness and cell motility (Hempel *et al.*, 2007). In prostate cancer, the loss of *TGFBR3* has also been reported with advanced tumour stage (Turley *et al.*, 2007).

In the top 10 lists, in the previous histological comparison, the *AOX1* gene was one of the most significantly downregulated (log₂FC -1.63) genes across subtypes. In the molecular comparisons it had a similar level of expression (log₂FC -1.69). *AOX1* gene expression has also been found to be downregulated in colon cancer (Zhang *et al.*, 2020b) but not reported in breast, ovarian, or prostate cancer previously. Downregulation in *AOX1* gene expression is therefore a novel finding in this study.

The DEGs that were also found to be subtype specific were also collected for future research to identify DEGs that can classify subtypes similar to the PAM50 gene set but which would be particularly beneficial in ovarian and prostate cancers where more biomarkers are needed. The upregulated DEGs were extracted and taken forward for survival analysis.

3.2.3 Cross-cancer results summary

The significant DEGs from section 3.1 were cross compared in two separate comparison analyses comparing cancer subtypes. This was to determine whether changes in gene expression were conserved between cancers and, if so, to what level this was. This approach to generating a set of candidate genes involved across cancers had the advantage that it reduced the number of initial genes from the DEG analysis from thousands to a couple of hundred. This makes downstream analysis considerably less complex. This approach also meant that genes for downstream survival analysis were selected in an unbiased way. Rather than selecting a group of genes from the different subtypes that met a specific role or functional criteria, genes were selected based only on their potentially shared role across the cancers, irrespective of their function.

The identification of genes with conserved changes in expression across the cancer subtypes as highlighted previously (whether upregulated or downregulated) are both of equal importance despite emphasis in this thesis being on the upregulated genes. This approach is especially useful to identify genes that are subtype specific as these can be used to create new subtype specific classifications.

Whilst many of the top 10 genes (both upregulated and downregulated) identified here were previously shown to be altered in at least one of the cancers prior to this study, the identification of these genes validates the method of integrating multiple datasets.

In order to narrow down the list of upregulated genes of interest, a literature survey in conjunction with survival analyses was performed as described in the following sections.

3.3 Novel gene identification

The significantly upregulated DEGs identified that were also significantly upregulated in the histological and/or molecular cross-cancer classification analyses (section 3.2) were collected. This left, for the first comparison (breast histological, ovarian tissue location, and prostate Gleason grade group), 395 genes and 75 for the second comparisons (molecular breast, ovarian epithelial, and prostate Gleason scores) to compare to previous literature findings for these cancer types.

In order to identify novel genes of interest for the three cancers, only the top 50 genes ranked by average P-value from each of the two cross-cancer analyses (100 genes in total) were analysed in the scientific literature. The focus was on an examination of genes that are upregulated for several reasons. First, upregulated genes are potentially far more effective targets for downstream knockdown/silencing studies to understand more about their function. This could be employed in future studies by using siRNAs; for example, to silence the gene and measure potential functional effects. Second, downregulated genes are more likely to have tumour suppressor roles and, therefore, identification of their function is likely to be less informative for cancer progression beyond early-stage tumorigenesis.

The results from the histological and molecular (and corresponding ovarian and prostate subtypes) cross-cancer analyses complemented each other in refining the genes of interest more concisely. Three cross-cancer DEGs that were novel and expressed at almost twice the level as seen in normal tissue were identified; these are detailed in Table 39.

Table 39. The log₂FC and P-value of the three cross-cancer DEGs identified in the two cross-cancer analyses.

DEG	Log ₂ FC		Adjusted P-Value	
	Histological	Molecular	Histological	Molecular
<i>HMMR</i>	1.37	1.68	1.29E-10	2.09E-04
<i>CENPE</i>	1.12	1.45	6.62E-08	5.45E-05
<i>STIL</i>	1.07	1.38	4.38E-08	7.34E-06

The three cross-cancer DEGs; *HMMR*, *CENPE*, and *STIL* were found to be significantly upregulated in the histological breast, ovarian tissue location, and prostate Gleason grade group comparison. This was also observed in the molecular breast, ovarian epithelial, and prostate Gleason score comparison.

It is important to note that this does not mean that these three genes are not known in any other cancers. For example, more recently (within the last 12 months) hyaluronan mediated motility receptor (*HMMR*) has been identified in other cancers such as hepatocellular (Lu *et al.*, 2020) and lung cancer (Meng *et al.*, 2020). However, this is the first study to identify *HMMR* in breast, ovarian, and prostate cancers.

3.3.1 Novel cross-cancer DEGs in histological subtypes

From the cross-cancer analysis of histological breast, ovarian tissue, and prostate Gleason grade group subtypes, *HMMR*, centromere-associated protein E (*CENPE*), and centriolar assembly protein (*STIL*) gene expression (RNA) levels were at their highest in ovarian tissue location subtypes, slightly more prominently in fallopian tissue samples and at their lowest in breast cancer non-invasive DCIS (Figure 26). This was similarly observed in prostate carcinomas where Gleason grade 1 were observed to show lowest expression of the three genes.

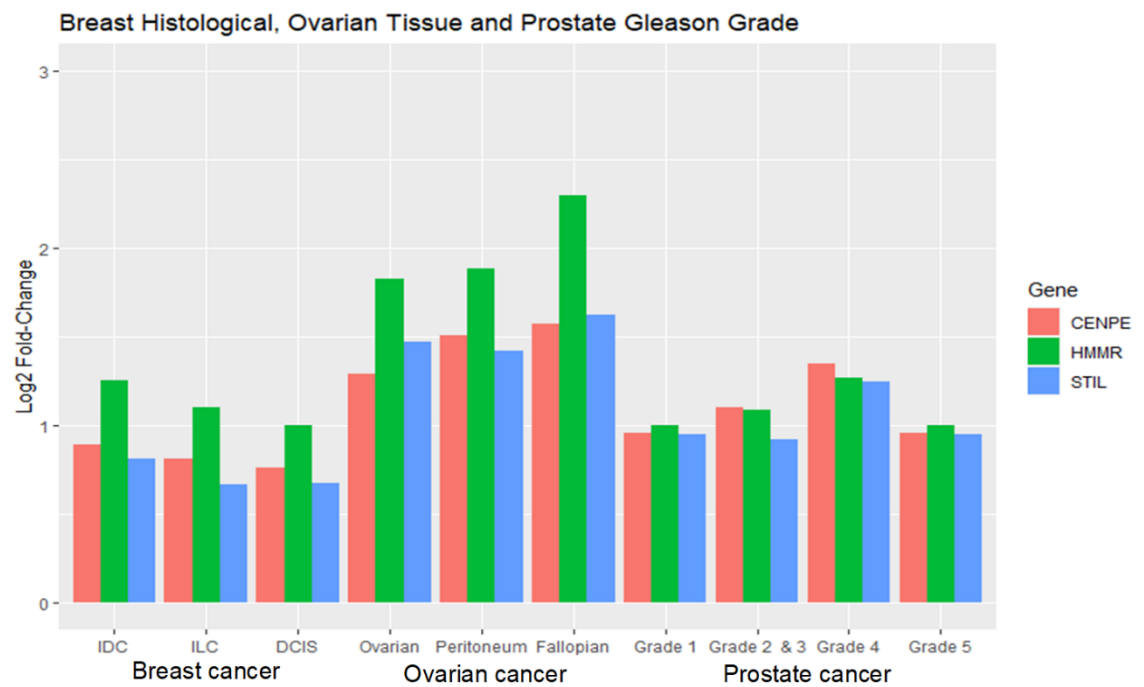


Figure 26. Log₂ Fold-change gene expression values for genes *HMMR*, *CENPE*, and *STIL* in breast histological, ovarian tissue, and prostate Gleason group subtypes.

The expression of the three cross-cancer DEGs was observed to be highest in ovarian cancer tissue location subtypes.

High levels of gene expression were identified in invasive subtypes (e.g. breast IDC and higher prostate Gleason grades) suggesting that these genes may contribute to tumour invasiveness. For example the highest levels in ovarian fallopian samples may reflect how fallopian tumours migrate to the ovaries (Lengyel, 2010) (section 1.3.2). This was tested further in the following survival and function/pathway analysis. The log₂ FC and adjusted P-values for each subtype for *HMMR*, *CENPE*, and *STIL* DEGs are shown in appendix Table 70.

3.3.2 Novel cross-cancer DEGs in molecular subtypes

The three DEGs (*HMMR*, *CENPE*, and *STIL*) were significantly upregulated in breast molecular, ovarian epithelial, and prostate Gleason score subtypes (Figure 27).

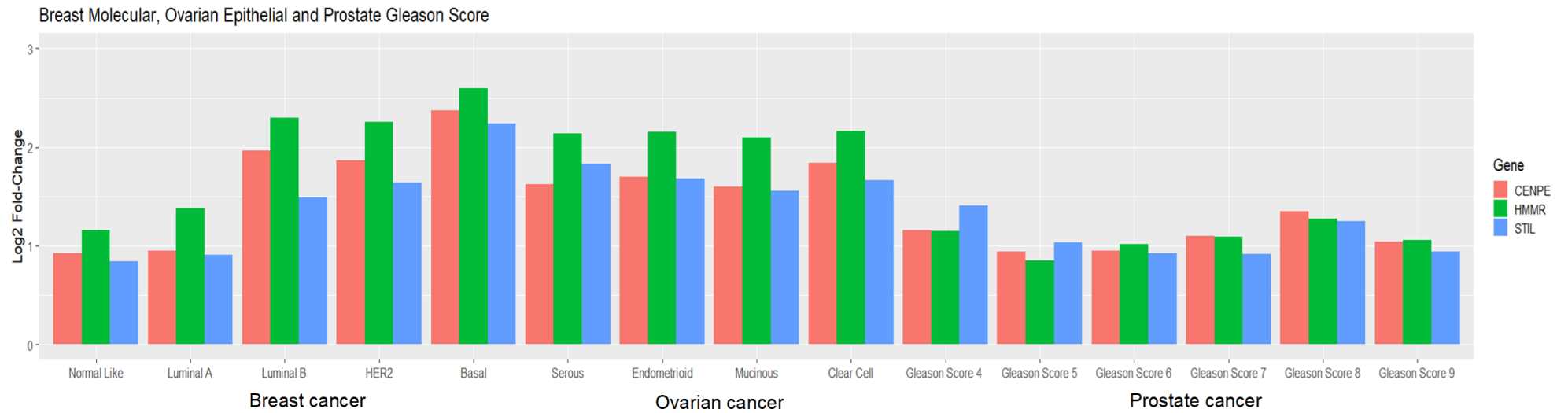


Figure 27. The Log₂ Fold-change gene expression values for genes HMMR, CENPE, and STIL in breast molecular, ovarian epithelial, and prostate Gleason score subtypes.

Breast cancer basal-like subtype showed the highest expression of the three cross-cancer DEGs with normal-like showing the lowest. Ovarian cancer epithelial subtypes HMMR expression was similar across the subtypes. CENPE expression was highest in clear cell whereas STIL was highest in serous. In prostate cancer Gleason scores the expression was similar in all but Gleason score 8.

Upregulation of the three DEGs was higher in more prognostically poor (more invasive) breast cancer subtypes; luminal B, HER2, and basal. These are considered subtypes with poorer survival (section 1.3.1, Figure 8) due to increased histologic grade and reduced receptor expression requiring chemotherapy and less effective hormone replacement therapy (HRT).

Gene expression levels in ovarian epithelial cancer subtypes did not follow the same pattern as in breast cancer molecular subtypes. *HMMR* expression remained relatively stable and constant throughout the ovarian subtypes. *CENPE* was highest in clear cell compared to the other ovarian subtypes. *STIL* expression was highest in the serous subtype.

In prostate cancer Gleason scores, Gleason score 8 had the highest changes in gene expression levels for the three cross-cancer DEGs; the reason for this is unclear. An increasing Gleason score is associated with poorer survival (section 1.3.3, Figure 11). If the expression of the three genes was associated with poorer survival subtypes (like observed in breast cancer molecular and histological subtypes) then the expression would be expected to increase from Gleason score 4-9. This is not quite observed, and it does not fully explain the peak expression change that is observed in Gleason score 8 samples. It could be that at Gleason score 8, the DEGs may be required more at this stage of tumour development for invasive capability. However, it is more plausible that the low sample size of Gleason score 8 (n=13) may explain this. Gleason score 4 had similar gene expression levels and this is likely due to low sample size (n=8) as well. The true gene expression levels would more likely be lower if the sample groups were larger. Increasing sample size for the prostate subtypes would potentially correct this and provide more accurate measurements. The Log2FC and adjusted P-value for each subtype are also given in appendix Table 71.

3.3.3 Novel cross-cancer gene identification summary

Three cross-cancer DEGs were identified as novel in breast, ovarian, and prostate cancers from reviewing literature for the three cancers to determine if they had previously been reported. The genes, *HMMR*, *CENPE*, and *STIL* were expressed at around double the level in the equivalent

normal tissue. The upregulation of the three genes was also increased in more invasive subtypes. This suggested a potential role of these genes in cancer progression/invasion, so additional survival analyses and functional analyses was performed to determine their potential prognostic value.

3.4 Survival analysis of the three novel cross-cancer DEGs

The three novel cross-cancer DEGs were tested to determine whether they could potentially affect survival times using primary patient survival data. Recurrence free survival (RFS) and overall survival (OS) for *HMMR*, *CENPE*, and *STIL* for each cancer were compared to determine whether their increased gene expression levels (RNA) associated to significantly shorter patient survival times. Drug treatment data, where available, were also compared within each cancer to identify any potential role of the three novel cross-cancer DEGs in drug resistance mechanisms within patient cohorts.

3.4.1 Recurrence free survival (RFS)

In breast cancer, high expression of *HMMR*, *CENPE*, and *STIL* individually significantly associated with reduced RFS (Figure 28 A-C). In ovarian and prostate cancer, high expression of *HMMR* and *CENPE* also significantly associated with reduced RFS (Figure 29-30 A-B), but not with *STIL* which did not reach significance ($P < 0.05$) (Figure 29-5 C).

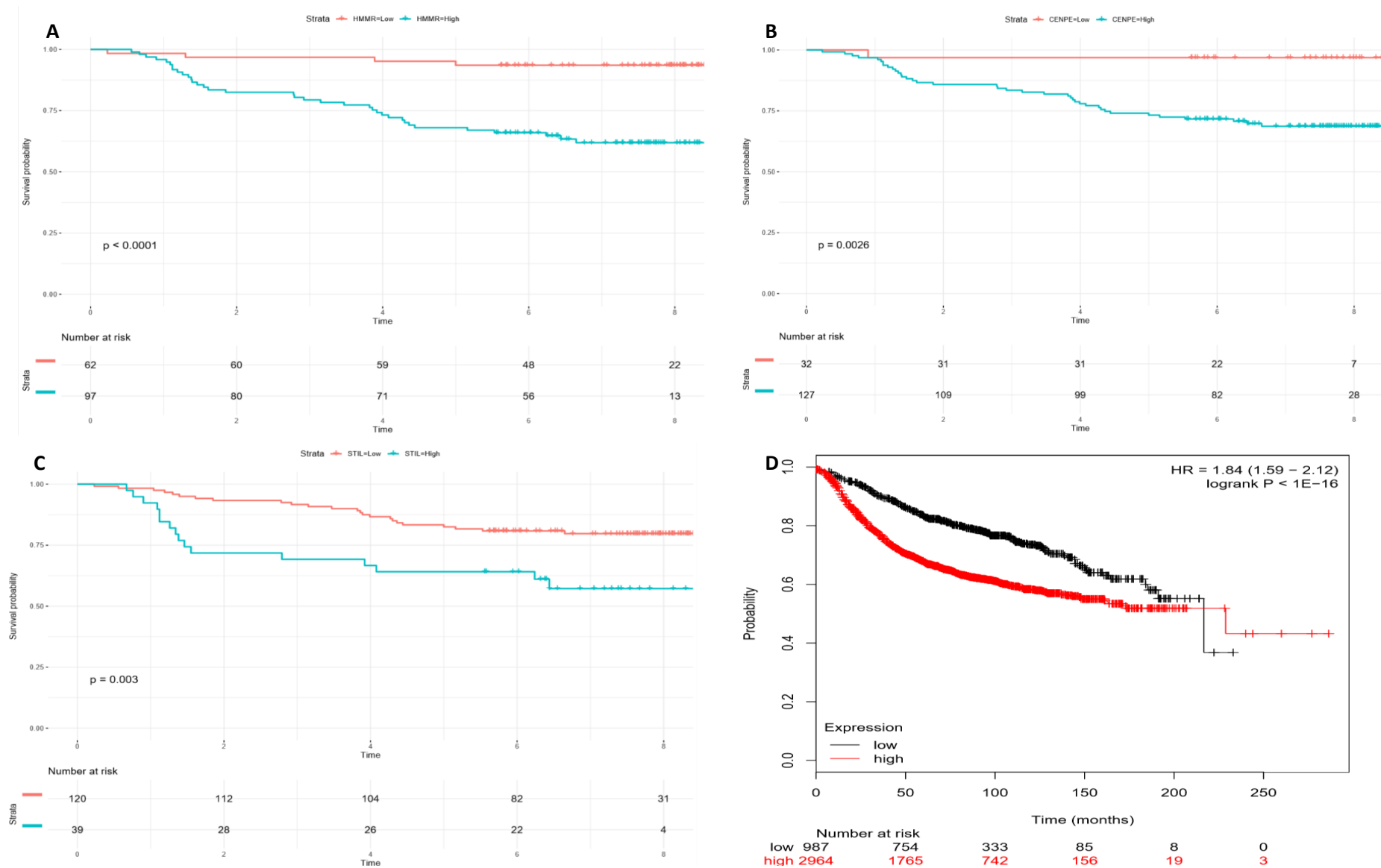


Figure 28. Kaplan Meier RFS survival curves in breast cancer primary tumour samples.

A) HMMR. B) CENPE. C) STIL. D) Combined (HMMR, CENPE, STIL). In plots A, B, and C blue shows high expression of genes and red shows low expression plotted using Survminer. In plot D red shows high expression and black, low expression using the K-M Plotter web tool for validation.

RFS: Recurrence Free Survival

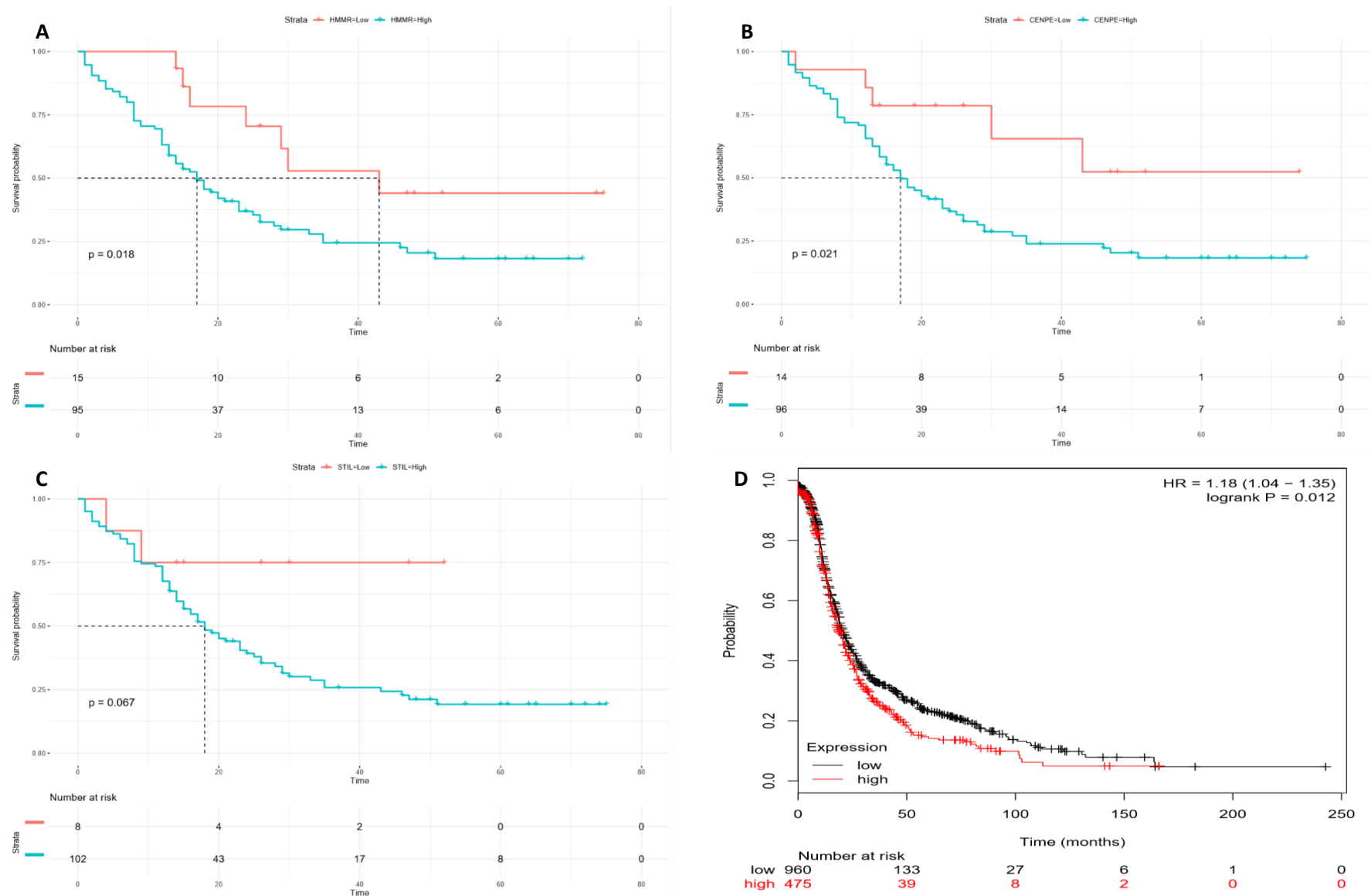


Figure 29. Kaplan Meier RFS survival curves in ovarian cancer primary tumour samples.

A) HMMR. B) CENPE. C) STIL. D) Combined (HMMR, CENPE, STIL). In plots A, B, and C blue shows high expression of genes and red shows low expression plotted using Survminer. In plot D red shows high expression and black, low expression using the K-M Plotter web tool for validation.

RFS: Recurrence Free Survival

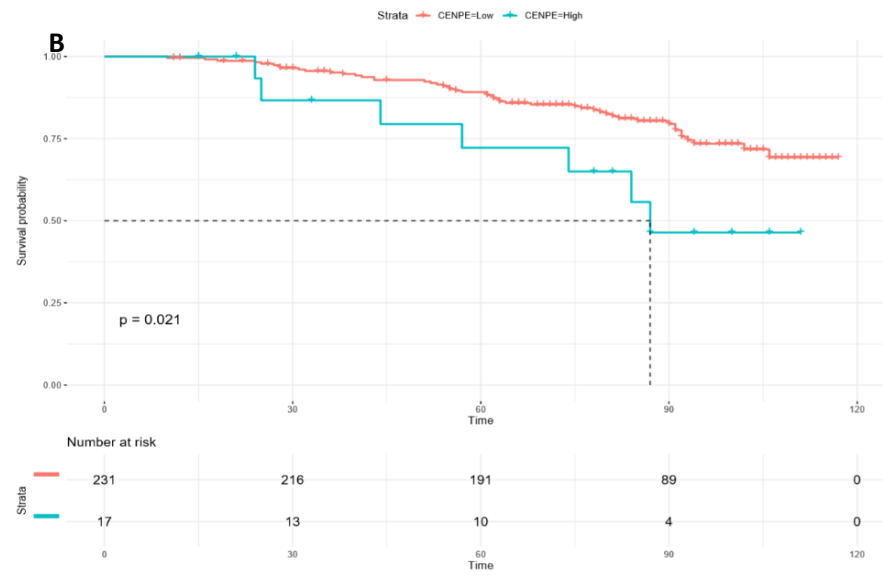
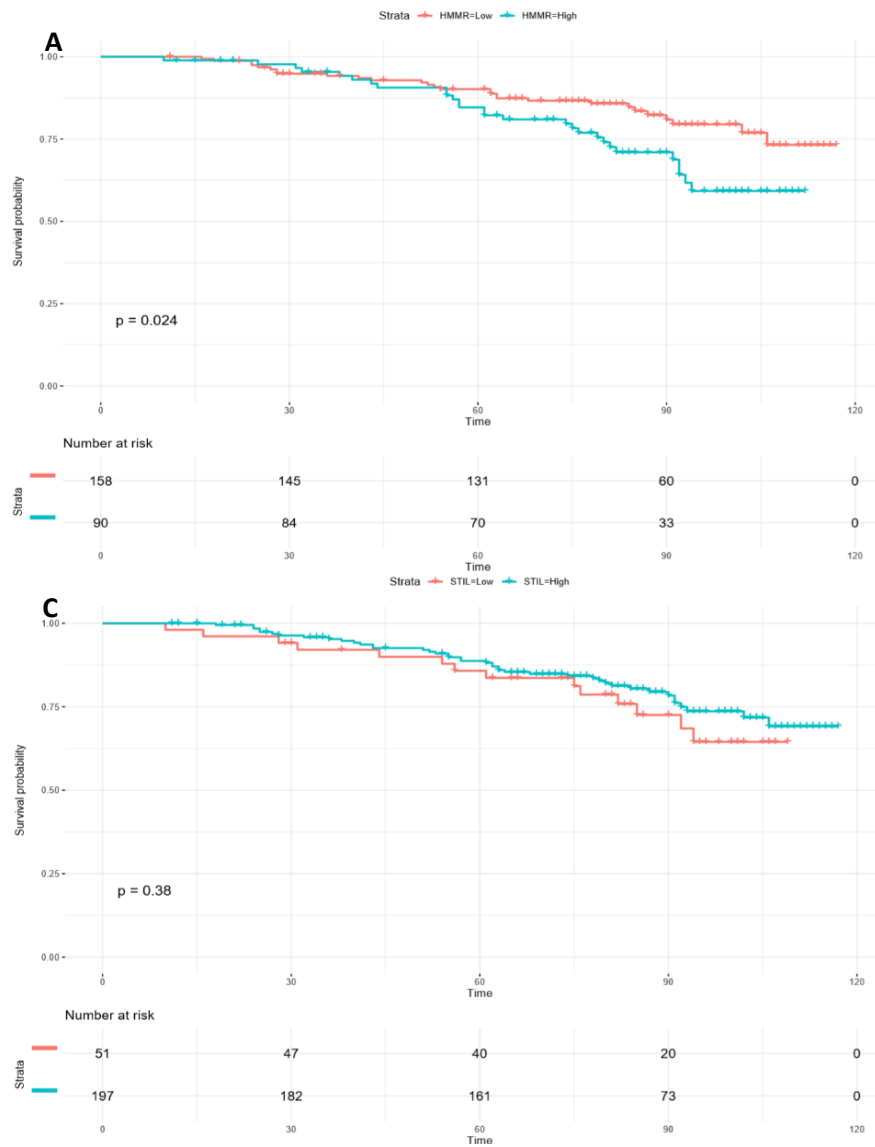


Figure 30. Kaplan Meier RFS survival curves in prostate cancer primary tumour samples.

A) HMMR. B) CENPE. C) STIL. In plots A, B, and C blue shows high expression of genes and red shows low expression plotted using Survminer. Data was not available for the combined expression of the three DEGs.

RFS: Recurrence Free Survival

Interestingly, in the combined expression of *HMMR*, *CENPE*, and *STIL* in breast and ovarian cancer, high expression significantly associated with reduced RFS (Figure 28 D and Figure 29 D); there is more of an effect observable in breast cancers than in ovarian cancers. This suggests a potentially oncogenic role of the three genes in progression - secondary recurrence and/or invasiveness. However as stated in section 2.10 of the methodology, there were no relevant RFS data for prostate cancer to be able to perform a combined gene expression survival analysis, which was a limitation. The hazard ratios (HR) and P-Values are summarised in (Table 40).

Table 40. RFS P-values and hazard ratios for each of the three novel cross-cancer DEGs for breast, ovarian, and prostate cancers.

	HMMR	CENPE	STIL	Combined
breast cancer				
Hazard ratio	1.87	3.01	2.29	1.84
P-Value	P<0.0001	P=0.0026	P=0.0030	P<0.0001
ovarian cancer				
Hazard ratio	1.21	1.22	1.11	1.18
P-Value	P=0.018	P=0.021	P=0.067*	P=0.012
prostate cancer				
	HMMR	CENPE	STIL	Combined
Hazard ratio	1.65	2.45	1.14	-
P-Value	P=0.024	P=0.021	P=0.38*	-

P-values highlighted with * identify those that are not significant (P<0.05). The 'combined' column is the expression of all three genes in relation to recurrence free survival (RFS).

An explanation for the *STIL* RFS results could be that the expression of *STIL* in ovarian carcinomas is not required for a tumour to progress from primary cancer treatment to secondary recurrence.

HMMR and *CENPE* may have more prominent roles in this, to potentially promote tumour progression. *STIL* is still relevant though, as it could be adding to the functions of *HMMR* and *CENPE* when combined. This also means that the results from each gene could be influenced by the combined effects of the three genes.

The potential mechanisms by which *STIL*, *HMMR*, and *CENPE* are acting are detailed in the following section 3.7. However, it maybe that *STIL* functions with either *HMMR*, *CENPE*, or both and thus contribute to reduced RFS.

3.4.2 Overall survival (OS)

In breast cancer, high gene expression of all three novel cross-cancer DEGs significantly associated with reduced OS (Figure 31 A-C). This was also seen in ovarian cancer (Figure 32 A-C) except for *CENPE*, where low *CENPE* expression associated to significantly reduced OS. Also *STIL* had more of an effect in breast than ovarian, as in ovarian, there were fewer samples for the high gene expression group, that seemed to skew the results. In prostate cancer, high *HMMR* and *CENPE* expression also associated with reduced OS survival time, but *STIL* was not significant ($P < 0.05$, Figure 33 A-C).

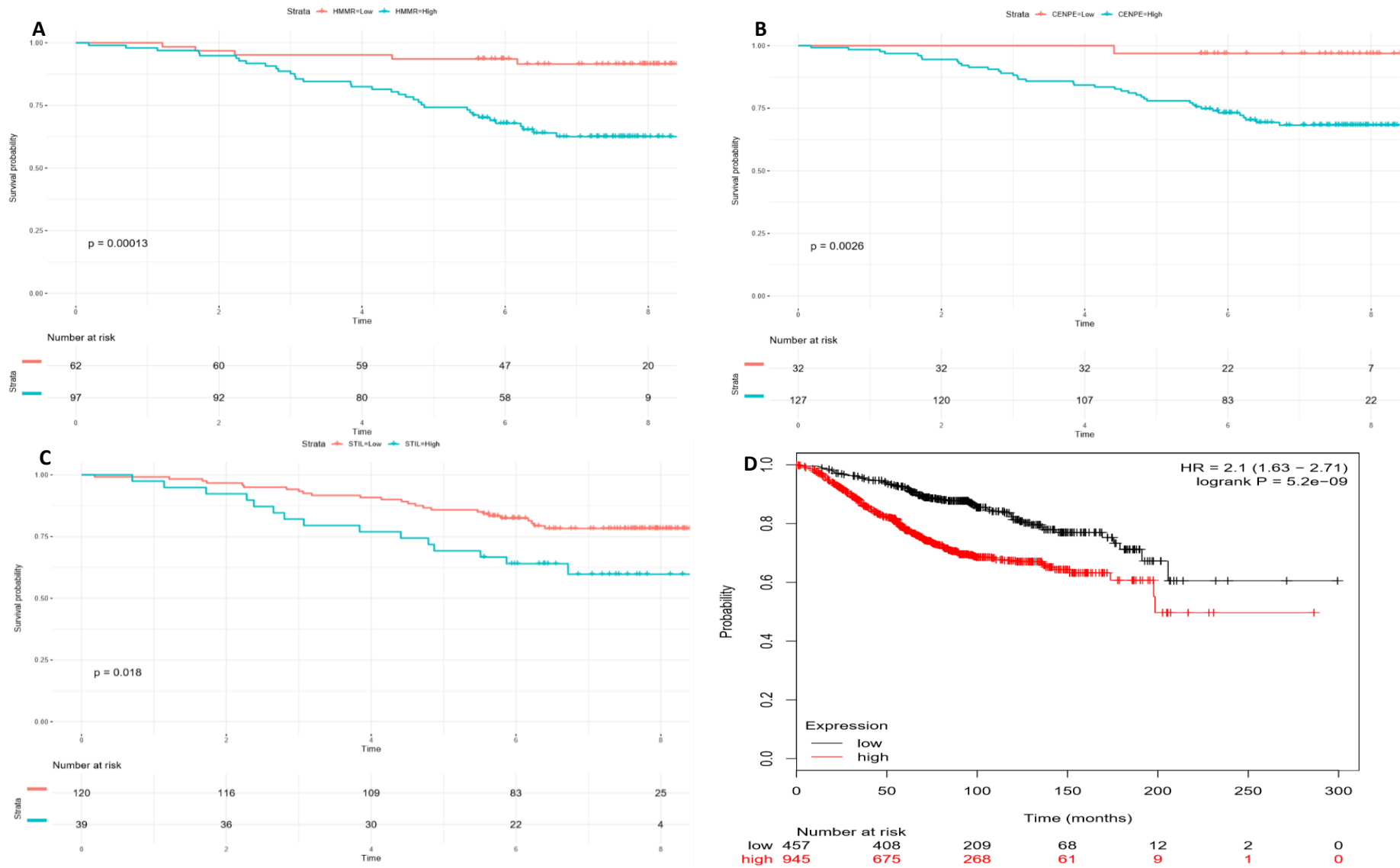


Figure 31. Kaplan-Meier OS survival curves in breast cancer primary tumour samples.

A) HMMR. B) CENPE. C) STIL. D) Combined (HMMR, CENPE, STIL). In plots A, B, and C blue shows high expression of genes and red shows low expression plotted using *Survminer*. In plot D red shows high expression and black, low expression using the *K-M Plotter* web tool for validation.

OS: Overall Survival

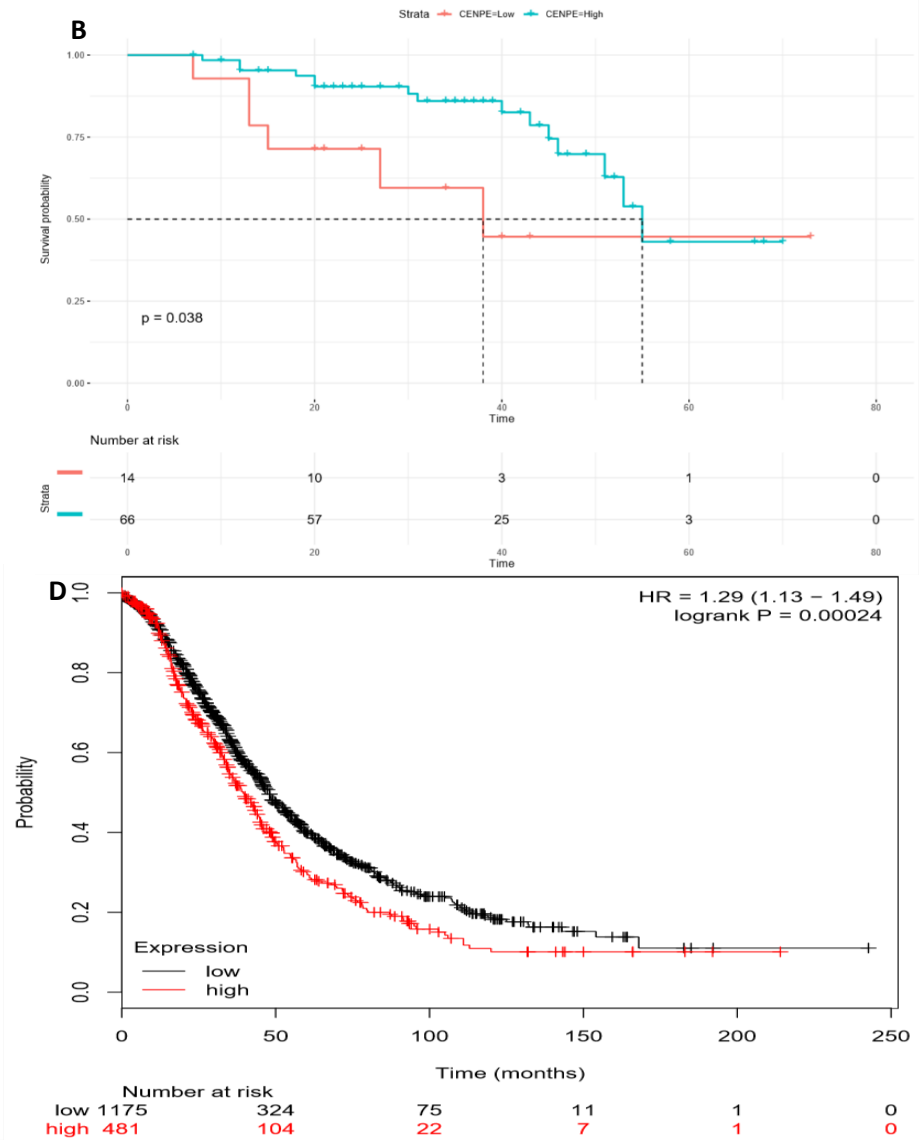
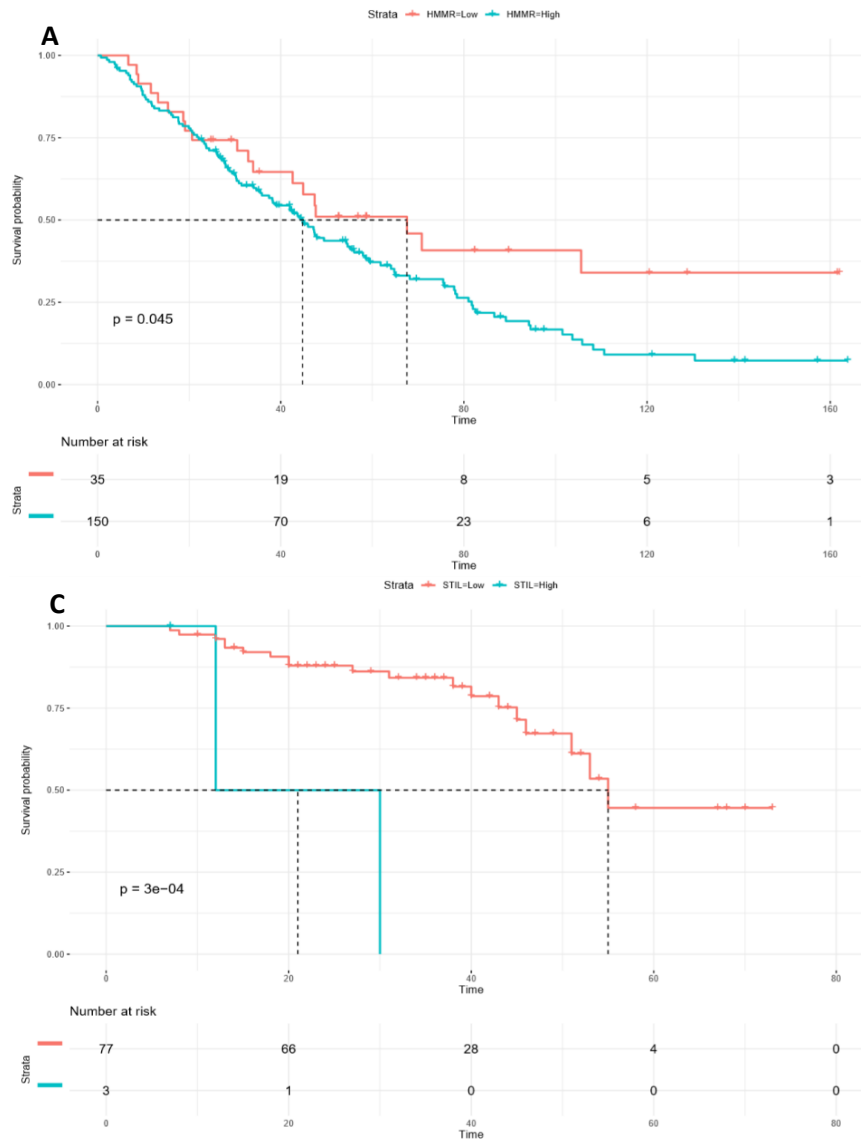
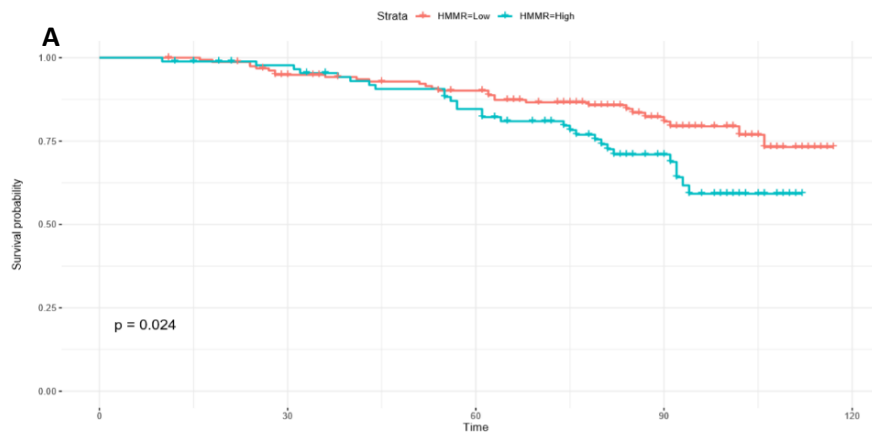


Figure 32. Kaplan Meier OS survival curves in ovarian cancer primary tumour samples.

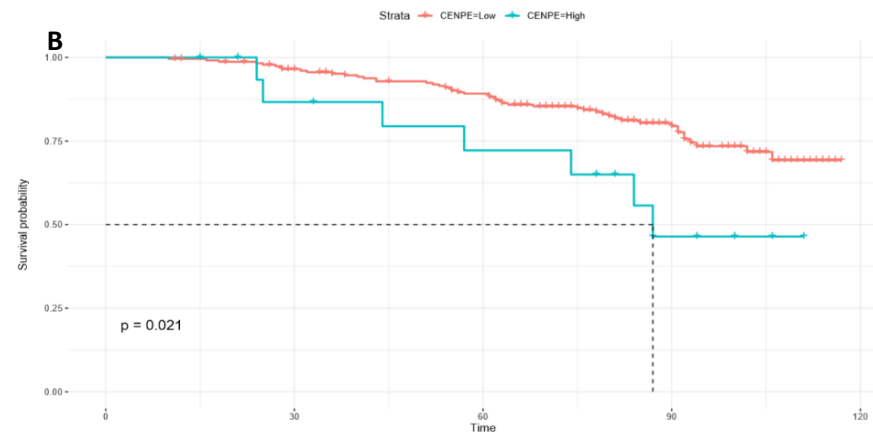
A) HMMR. B) CENPE. C) STIL. D) Combined (HMMR, CENPE, STIL). In plots A, B, and C blue shows high expression of genes and red shows low expression plotted using Survminer. In plot D red shows high expression and black, low expression using the K-M Plotter web tool for validation.

OS: Overall Survival



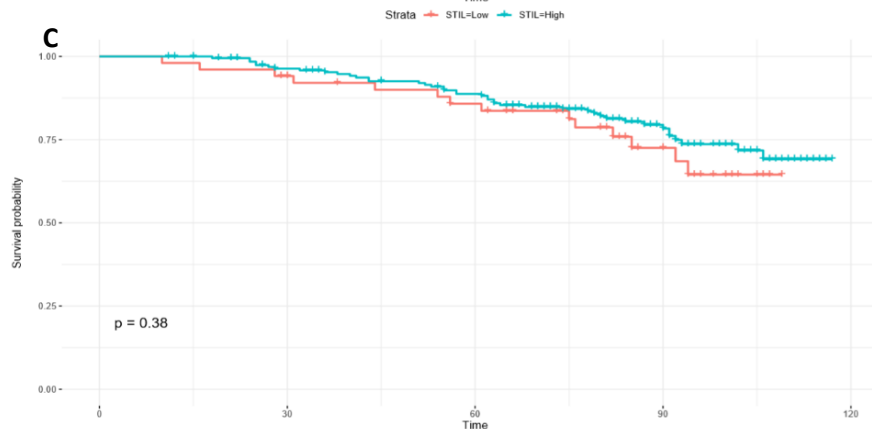
Number at risk

Strata	0	30	60	90	120
HMMR-Low	158	145	131	60	0
HMMR-High	90	84	70	33	0



Number at risk

Strata	0	30	60	90	120
CENPE-Low	231	216	191	89	0
CENPE-High	17	13	10	4	0



Number at risk

Strata	0	30	60	90	120
STIL-Low	51	47	40	20	0
STIL-High	197	182	161	73	0

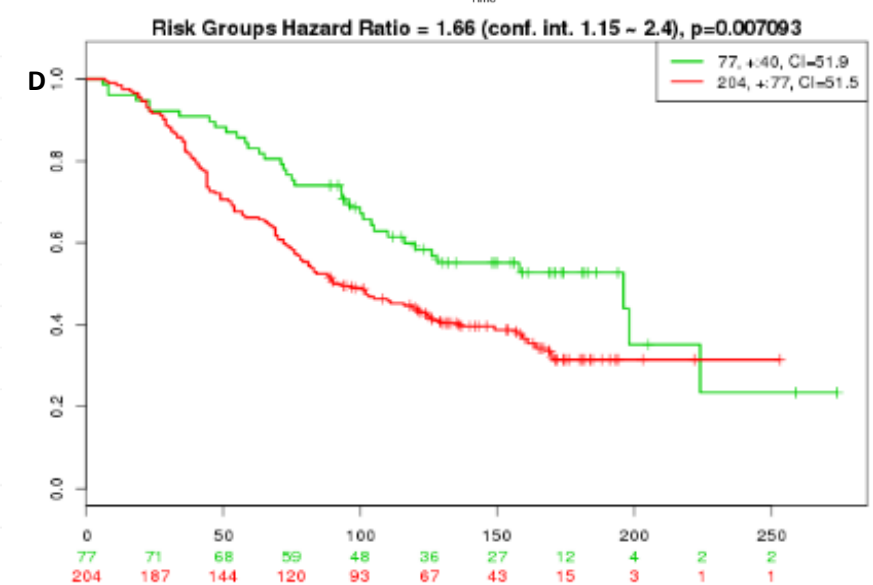


Figure 33. Kaplan Meier OS survival curves in prostate cancer primary tumour samples.

A) HMMR. B) CENPE. C) STIL. D) Combined (HMMR, CENPE, STIL). In plots A, B, and C blue shows high expression of genes and red shows low expression plotted using Surminer. In plot D red shows high expression and green, low expression using the SurvExpress Plotter web tool for validation.

OS: Overall Survival

The low *CENPE* expression associated with reduced OS in ovarian cancer was contrary to high expression in OS and RFS for breast and prostate cancer. This is most likely due to the samples which were identified as receiving taxane treatment. This could be that there is a selected group of tumours where taxane treatment is effective and the tumours that remain are those where microtubule dynamics are destabilised (by the taxane) and there are other mechanisms bypassing the effect of taxane and promoting proliferation leading to resistance.

In the combined gene expression analysis of all three genes, high gene expression significantly associated with reduced OS (Figure 31-33 D). This suggested that the three genes function in a potentially common mechanism/s of action that reduced OS and RFS. This may occur through a more aggressive primary/secondary (metastasis) cancer or acquiring a mutation allowing for treatment resistance. The hazard ratios (HR) and P-Values are summarised in (Table 41).

Table 41. OS P-values and hazard ratios for each of the three novel cross-cancer DEGs for breast, ovarian, and prostate cancers.

	HMMR	CENPE	STIL	Combined
breast cancer				
Hazard ratio	1.69	2.68	1.99	2.10
P-Value	P=0.00013	P=0.0026	P=0.018	P<0.0001
ovarian cancer				
Hazard ratio	1.19	1.27	1.53	1.29
P-Value	P=0.045	P=0.038	P<0.0001	P<0.0001
prostate cancer				
Hazard ratio	1.59	1.73	0.86	1.66
P-Value	P=0.024	P=0.021	P=0.38*	P=0.0071

P-values highlighted in * identify those that are not significant ($P<0.05$). The 'combined' column is the expression of all three genes in relation to overall survival (OS).

3.4.3 Survival analysis summary

High expression of *HMMR* and *CENPE* was significantly associated with reduced RFS and OS in breast, ovarian, and prostate cancers. This suggested that *HMMR* is involved in the progression of these cancers and potentially their recurrence and metastasis.

However, an exception was identified in ovarian cancers where lower expression of *CENPE* was significantly associated with reduced OS (rather than high expression seen in breast and prostate cancers in OS) and high *CENPE* expression associated to significantly lower RFS. This may be due to taxane treatment being prescribed. In ovarian cancers with high *CENPE* expression, recurrence (RFS) may occur much sooner and in response to taxane treatment, *CENPE* expression may be reduced as part of a resistance mechanism, thus low expression significantly reducing OS. A higher expression of *CENPE* is therefore more likely to lead to better prognosis due to its function

in microtubule dynamics. Low expression of *CENPE* is likely to lead to worse prognosis because other microtubule dynamics processes are also dysfunctional. The functions of these novel cross-cancer DEGs will be described in detail in section 3.7.

High *STIL* expression was associated with significantly reduced RFS and OS in breast cancer but not in ovarian and prostate cancer. This suggested that *STIL* may not be important in ovarian and prostate cancer prognosis on its own, but when combined with the other genes it enhances the impact on survival (RFS and OS).

Combined the three novel cross-cancer DEGs with high expression were significantly associated with reduced OS in breast, ovarian, and prostate cancer. However, high expression in RFS could only be confirmed in breast and ovarian cancer. High expression is likely to be the same in prostate cancer but requires future research on more samples to confirm. These results suggested that the three genes in combination may function in a similar related underlying pathway/s or action/s to promote tumorigenesis, aggressiveness and invasiveness (metastasis) that reduces survival time.

3.5 Summary of DEG, cross-cancer, and novel/survival results

The results from the DEG analyses helps in confirming aim 1 as known PAM50 genes were identified in breast cancer subtypes. Genes that were also known in cancers/subtypes previously were also identified which also helped in validating the integration method. Interestingly it was also found that some of these PAM50 genes were also present in ovarian and prostate cancers, which was also a novel finding suggesting their use as potential classification biomarkers in these cancers. Three cross-cancer DEGs have been identified that have met the criteria for the following network analyses. These three cross-cancer DEGs were *HMMR*, *CENPE*, and *STIL*. All three were upregulated and significantly differentially expressed in the individual cancer subtype analyses and in the histological and molecular (and corresponding ovarian and prostate subtypes) comparisons. These cross-cancer DEGs were novel in regard to the literature in these cancers and

subtypes. All three novel cross-cancer DEGs were associated with decreased RFS and OS survival either individually (*HMMR* and *CENPE*) or in combination (*HMMR*, *CENPE*, and *STIL*).

This sequential methodology approach has reduced the initial significant DEGs identified from thousands down to hundreds, with the survival analyses reducing this to below 100 for both RFS and OS, ending with three novel cross-cancer DEGs. As such, these three novel cross-cancer DEGs may be potential hub genes following the proceeding network analysis.

3.6 Network and hub gene analysis

Due to their consistently upregulated expression across the three cancers and subtypes, as well as their association with reduced survival (OS and RFS), the three novel cross-cancer DEGs of interest (*HMMR*, *CENPE*, and *STIL*) were taken forward for hub gene analysis; *STIL* was included as it may have functioned with the other two genes in a common mechanism or cellular pathway. All three novel cross-cancer DEGs had a module modality >0.8 indicating that they were highly linked and important to the co-expressed gene network modules. This was also observable in the networks when visualised. As such they are termed hub genes from here onwards.

Co-expressed gene networks were created for *HMMR*, *CENPE* and *STIL* in each of their subtypes and then compared across the cancers. Co-expression modules were chosen as they allow the visualisation of other co-expressed genes in each cancer subtype. The genes that were statistically predicted to have an interaction with the three hub genes (*HMMR*, *CENPE*, and *STIL*) could then be analysed in pathway analysis to determine their potential functions in the cancers.

For many of the cancers and subtypes, the three hub genes were found to be co-expressed within the same network module (co-expression module) despite them each being extracted and analysed separately. For example, in luminal A breast cancers, the three hub genes were identified as being co-expressed in the same module, but in a separate network analysis of Serous ovarian cancers, the three hub genes were also found to be co-expressed together. In these cases, the main co-expression network module was used to construct an individual subnetwork for *HMMR*, *CENPE*, and *STIL* from the K-nearest neighbours. The K-nearest neighbours' algorithm was selected as it extracts the genes that are predicted to have a direct interaction. This helped to reduce the co-expression modules to more interpretable sizes. This also aids in providing a more consistent list of genes for Gene ontology (GO) analysis. This step also removes unnecessary analogous terms due to the size of the original unfiltered network. The network images that will be shown in the following sections were chosen because they most clearly identified the hub genes and interactions with the hub genes.

Gene ontology (GO) terms were identified for each of the three hub gene's co-expression networks. For visualisation of the co-expression modules a further filtering step was added. This was because many of the modules still contained many directly interacting genes. After the GO analysis was completed, a heat diffusion algorithm was applied to reduce the modules. From these final reduced modules, co-expression images for each of the hub genes were produced. A heat diffusion of a maximum of 25 genes was found to provide interpretable co-expression module images. These are referred to as subnetworks hereafter. The analysis is presented in two parts for each of the cancer comparisons subtypes. The first part is a summarised co-expression network. One representative network from each cancer is shown here for simplicity with all the networks for each of the subtypes in the appendices (Figure 70-93). The second part is a table summarising the GO terms that were identified.

Many of the network analysis and GO analysis results were highly homogenous across subtypes and cancers. The results for each subtype comparison (i.e., breast cancer histological, ovarian tissue location, and prostate Gleason grade) are described at the beginning of each of the corresponding subsections.

3.6.1 Breast histological, ovarian tissue, and prostate Gleason grade networks

The biological processes identified for GO analysis for the three hub genes in these subtypes were related to cell cycle transition leading to proliferation. Each of the three hub genes was found to overlap in some manner in relation to GO function. Many of the GO terms identified from the subnetworks for each of the three hub genes were highly similar. This suggests similar networks and network modules of co-expressed genes. However, for some subtypes the same GO terms were found but the terms identified were not significant. This was observed in the fallopian subtype for ovarian tissue location analysis. This is most likely due to small sample size making the network underpowered. The *HMMR*, *CENPE*, and *STIL* co-expression networks were constructed for breast histological, ovarian tissue, and prostate Gleason grade classified tumours (appendices Figure 70-79), and the GO terms identified (Table 42). The co-expression subnetworks for IDC

breast cancer were chosen for *HMMR*, ovarian tissue for *CENPE*, and Gleason grade group 1 for *STIL* as representative examples (Figure 34 A-C).

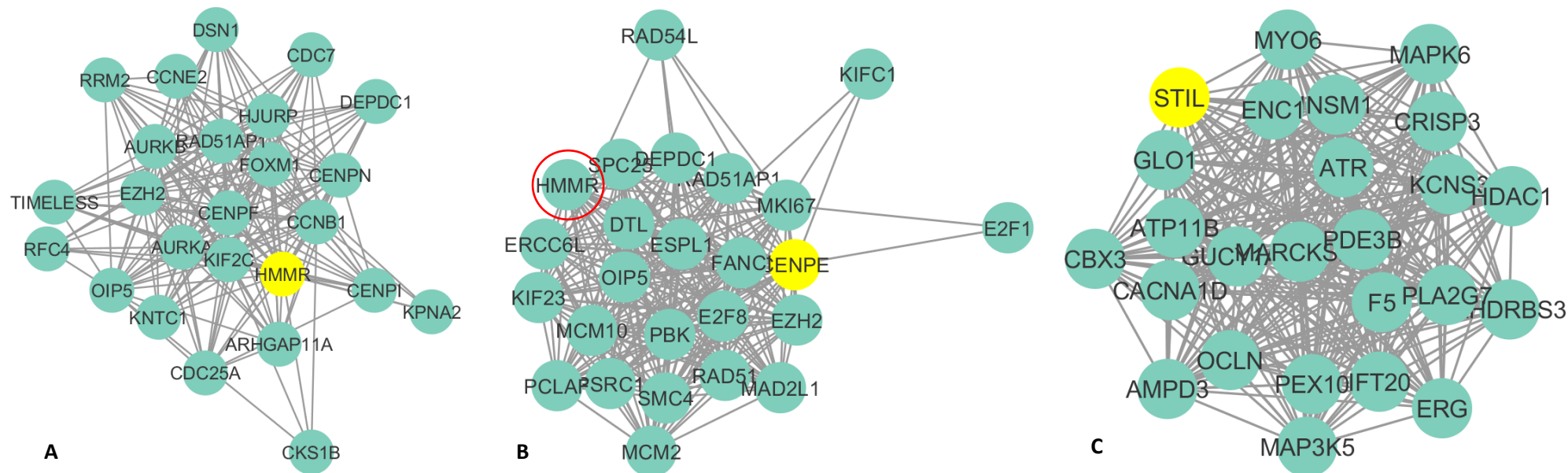


Figure 34. Co-expression networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for breast cancer IDC tumour samples. B) CENPE subnetwork from ovarian tissue samples. C) STIL subnetwork from Gleason grade 1 samples. Only the 25 most highly correlated genes based on heat diffusion are shown. HMMR was found to be co-expressed with CENPE and has been highlighted in red. Both CENPE and STIL were identified to be co-expressed within the same modules also however these are not shown due to restriction of the heat diffusion to the 25 most highly connected nodes.

Table 42. Gene Ontology (GO) terms identified in breast histological, ovarian tissue location, and prostate Gleason grade for HMMR, CENPE, and STIL.

Hub gene	GO term (cellular pathway or function)	P-value IDC	P-value ILC	P-value DCIS	P-value Ovarian	P-value Fallopian	P-value Peritoneum	P-value Gleason Grade 1	P-value Gleason Grade 2&3	P-value Gleason Grade 4	P-value Gleason Grade 5
HMMR	regulation of G2/M transition of mitotic cell cycle	5.15E-06	1.35E-05	5.64E-06	3.28E-06	P>0.05	1.59E-04	3.09E-04	2.59E-05	-	4.16E-05
	hyaluronan catabolic process	P>0.05	P>0.05	P>0.05	P>0.05	P>0.05	P>0.05	P>0.05	P>0.05	P>0.05	0.0455
CENPE	cell division	5.34E-14	3.21E-14	4.61E-14	2.26E-14	-	5.52E-14	9.96E-12	9.78E-14	-	-
	mitotic spindle organization	4.00E-13	2.75E-12	7.69E-13	9.35E-13	P>0.05	6.33E-13	2.35E-04	2.91E-07	-	-
	chromosome segregation	6.04E-13	1.49E-12	9.26E-10	7.75E-11	-	3.98E-10	1.01E-04	4.60E-06	-	-
	mitotic cell cycle	9.19E-13	3.21E-14	1.20E-12	2.26E-14	-	1.71E-12	7.32E-06	5.59E-09	-	-
	mitotic metaphase plate congression	5.08E-12	1.12E-08	-	-	-	-	-	-	-	-
	microtubule-based movement	-	-	7.01E-09	5.33E-11	-	1.06E-11	-	3.78E-03	-	-
	lateral attachment of mitotic spindle microtubules to kinetochore	-	-	-	-	P>0.05	-	-	-	0.0423	0.0295

Hub gene	GO term (cellular pathway or function)	P-value IDC	P-value ILC	P-value DCIS	P-value Ovarian	P-value Fallopian	P-value Peritoneum	P-value Gleason Grade 1	P-value Gleason Grade 2&3	P-value Gleason Grade 4	P-value Gleason Grade 5
STIL	mitotic spindle organization	1.81E-14	1.20E-12	1.95E-11	2.79E-12	P>0.05	6.02E-11	-	2.06E-07	-	P>0.05
	cell proliferation	4.46E-05	6.09E-07	8.75E-05	1.24E-07	-	2.96E-07	-	7.47E-04	-	-
	protein localization to centrosome	2.30E-03	9.48E-03	0.0132	5.96E-03	-	P>0.05	-	9.05E-03	-	-
	floor plate development	0.0441	P>0.05	P>0.05	P>0.05	P>0.05	P>0.05	P>0.05	-	3.40E-03	P>0.05
	embryonic axis specification	0.0441	-	-	-	P>0.05	-	P>0.05	-	4.76E-03	-
	centrosome duplication	-	-	-	0.0348	-	-	-	0.0462	9.50E-03	-

GO terms where they were not identified in are shown as –.

GO: Gene Ontology

For breast cancer histological subtypes, it was observed that the three hub genes occurred within the same module. The GO terms identified for *HMMR* were found to be primarily related to G2/M phase transition. This was observed in the histological subtypes (IDC, ILC, and DCIS), ovarian tissue subtypes (ovarian and peritoneum), and Gleason grade groups (1, 2 & 3, and 5). Notably, only in the Gleason grade 5 the hyaluronan catabolic process was significantly identified. In Gleason grade 4, the observation that *HMMR* was not being significantly assigned to any GO term is likely due to the network analysis parameters for network construction not being sensitive enough, although this will need future work to confirm. It is also possible that the minimum network modules were too large for this subtype and, therefore, no specific GO term/terms could be determined.

The GO terms for *CENPE* were found to be primarily related to processes involved in chromosome functions and microtubule dynamics. For example, the most common GO terms significantly identified were cell division, mitotic spindle organisation, microtubule-based movement, and chromosome segregation. Cell division and mitotic spindle organisation, and chromosome segregation were identified as significant in all of the breast histological, ovarian tissue location, and prostate Gleason groups except for the fallopian tissue subtype and prostate Gleason Grade groups 4 and 5. It was also observed that microtubule-based movement was significantly identified in DCIS, ovarian tissue, peritoneum tissue, and Gleason grade 2 and 3. However, this GO term notably overlaps in role with chromosome segregation and mitotic spindle organisation. Further to this, *CENPE* in the Gleason grade groups 4 and 5 were found to be more different from the other subtypes but highly similar to one another in the GO terms identified. Many of these GO terms were still found to be related to chromosome and microtubule organisation. For example, one of these significant GO terms in Gleason grade 4 and grade 5 was 'lateral attachment of mitotic spindle microtubules to kinetochore'. Despite the apparent dissimilarity in assigned GO terms, those that were assigned to the different co-expression modules were still interlinked in their overall functions. For example, those assigned to *CENPE* despite not being assigned directly to mitotic spindle organisation as observed in other Gleason grade group 1 and Gleason grade

group 2 & 3 they are all subsequent functions observed in microtubule spindle organisation of the chromosomes.

For *STIL*, the GO terms were found to be related primarily to roles in spindle organisation and protein localisation. The GO terms for mitotic spindle organisation which was also identified with *CENPE* was found to be significant in all the breast histological, ovarian tissue location with the exception of fallopian, and Gleason grade 2 and 3. The identification of both *CENPE* and *STIL* being assigned directly to the same GO term suggests there is some level of interlinking function between the two. Notably, for Gleason grade 4 the *STIL* subnetwork was significantly assigned to floor plate development. Genes involved in floor plate development via the sonic hedgehog (SHH) pathway have been observed in cancers prior but this is the first study to identify the function in prostate cancers. Interestingly in Gleason grade group 5, *STIL* was not found significantly assigned to a GO term.

3.6.2 Breast molecular, ovarian epithelial, and prostate Gleason score networks

Network analysis of the breast molecular, ovarian epithelial, and prostate Gleason score subtypes identified similar GO terms to the previous network analysis (section 3.6.1). Gleason score 4 were removed from network analysis due to insufficient sample size for accurate network construction. This could be addressed in future work with further samples being added when suitably identified as well as the integration of more sequencing platform data. The *HMMR*, *CENPE*, and *STIL* co-expression networks were constructed for breast molecular, ovarian epithelial, and prostate Gleason score classified tumours (appendices Figure 80-93). The GO terms identified in breast molecular and ovarian epithelial subtypes are shown in Table 43, and for prostate cancer Table 44. The co-expression subnetworks for luminal A breast cancer were chosen for *HMMR*, ovarian endometrioid for *CENPE*, and Gleason score 5 for *STIL* (Figure 35 A-C).

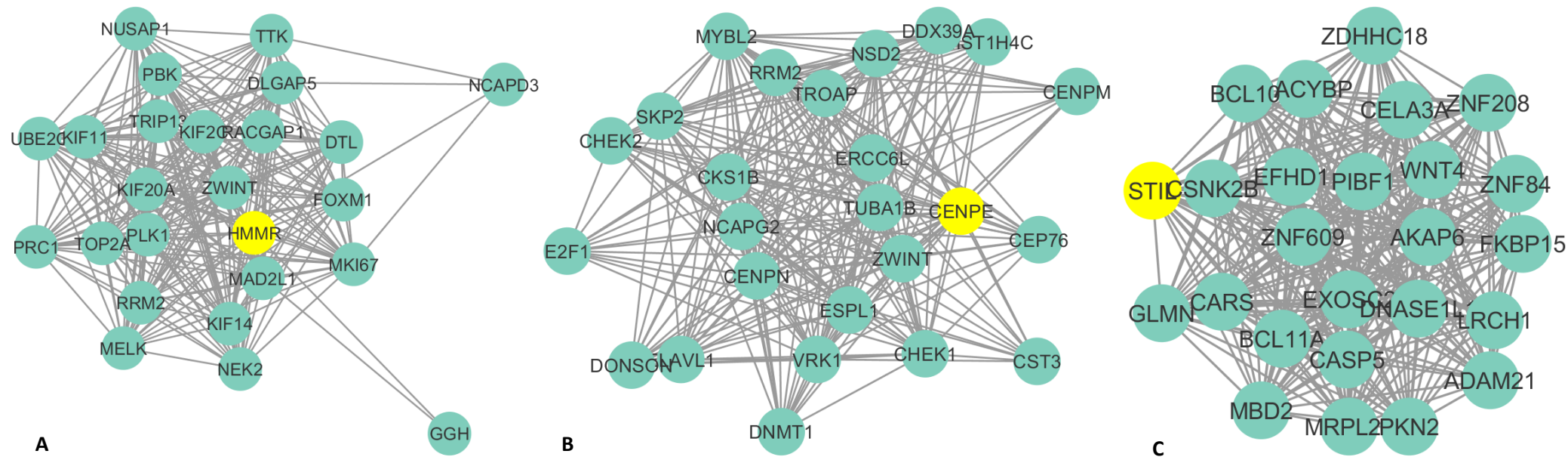


Figure 35. Co-expression networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for breast cancer luminal A tumour samples. B) CENPE subnetwork from ovarian Endometrioid samples. C) STIL subnetwork from Gleason score 5. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same modules also however these are not shown due to restriction of the heat diffusion to the 25 most highly connected nodes.

Table 43. Gene Ontology (GO) terms identified in breast molecular, and ovarian epithelial, for HMMR, CENPE, and STIL.

Hub gene	GO term (cellular pathway or function)	P-value Normal like	P-value Luminal A	P-value Luminal B	P-value HER2	P-value Basal	P-value Serous	P-value Endometrioid	P-value Mucinous	P-value Clear cell
HMMR	regulation of G2/M transition of mitotic cell cycle	3.99E-06	2.07E-05	3.69E-07	8.35E-04	8.70E-06	2.51E-06	1.87E-07	P>0.05	8.09E-07
CENPE	mitotic cell cycle	3.86E-14	1.95E-14	2.44E-14	1.48E-13	2.91E-14	1.97E-14	2.13E-14	0.0124	2.81E-14
	cell division	3.86E-14	1.95E-14	2.44E-14	5.56E-14	2.91E-14	1.97E-14	2.13E-14	4.39E-05	2.81E-14
	chromosome segregation	2.59E-12	3.72E-12	1.89E-12	2.28E-07	4.24E-11	1.54E-13	2.13E-10	1.55E-04	-
	microtubule-based movement	1.78E-07	4.08E-11	2.07E-10	1.63E-07	6.69E-09	2.93E-10	1.46E-10	-	2.71E-12
	mitotic spindle organization	1.94E-07	3.13E-14	3.38E-12	8.19E-08	9.15E-12	4.15E-12	-	P>0.05	3.75E-11
	mitotic metaphase plate congression	-	-	-	-	-	-	2.18E-10	P>0.05	8.14E-10
STIL	mitotic spindle organization	2.10E-14	3.75E-13	2.93E-11	1.04E-07	3.12E-11	1.26E-11	1.45E-10	0.0443	9.14E-08
	cell proliferation	4.91E-07	2.12E-08	3.53E-08	7.21E-05	4.36E-07	3.19E-09	7.92E-09	-	1.41E-08

Hub gene	GO term (cellular pathway or function)	P-value Normal like	P-value Luminal A	P-value Luminal B	P-value HER2	P-value Basal	P-value Serous	P-value Endometrioid	P-value Mucinous	P-value Clear cell
<i>STIL</i>	protein localization to centrosome	2.55E-03	2.07E-03	0.0126	P>0.05	0.018	0.017	0.016	P>0.05	-
	negative regulation of apoptotic process	0.0313	0.0426	-	P>0.05	-	-	-	-	0.0311
	floor plate development	0.0456	0.0426	P>0.05	-	P>0.05	P>0.05	-	P>0.05	0.041
	centrosome duplication	-	-	6.94E-03	P>0.05	0.0118	7.08E-05	0.0104	P>0.05	-

GO terms where they were not identified in are shown as -.

GO: Gene Ontology

Table 44. Gene Ontology (GO) terms identified in prostate Gleason scores for HMMR, CENPE, and STIL.

Hub gene	GO term (cellular pathway or function)	P-value Gleason score 5	P-value Gleason score 6	P-value Gleason score 7	P-value Gleason score 8	P-value Gleason score 9
HMMR	regulation of G2/M transition of mitotic cell cycle	0.0138	1.24E-04	2.59E-05	P>0.05	2.48E-04
CENPE	lateral attachment of mitotic spindle microtubules to kinetochore	2.19E-03	-	-	P>0.05	0.0275
	microtubule plus-end directed mitotic chromosome migration	2.19E-03	-	-	P>0.05	0.0275
	mitotic chromosome movement towards spindle pole	8.75E-03	-	-	P>0.05	0.0398
	regulation of mitotic metaphase/anaphase transition	0.0174	-	-	-	0.0445
	attachment of mitotic spindle microtubules to kinetochore	0.0196	-	-	-	0.0445
	mitotic cell cycle	-	1.17E-08	5.59E-09	P>0.05	-
	mitotic spindle organization	-	1.05E-04	2.91E-07	-	-
	microtubule-based movement	-	3.65E-04	3.78E-03	-	-
	chromosome segregation	-	4.22E-04	4.60E-06	-	-
STIL	protein localization to centrosome	0.0371	P>0.05	9.05E-03	-	-
	floor plate development	P>0.05	P>0.05	-	0.0302	0.0116

Hub gene	GO term (cellular pathway or function)	P-value Gleason score 5	P-value Gleason score 6	P-value Gleason score 7	P-value Gleason score 8	P-value Gleason score 9
STIL	embryonic axis specification	P>0.05	P>0.05	-	0.0172	0.0116
	regulation of centriole replication	P>0.05	P>0.05	-	0.0302	0.0132
	mitotic spindle organization	-	P>0.05	2.06E-07	-	9.16E-03

GO terms where they were not identified in are shown as –.

GO: Gene Ontology

HMMR subnetworks were observed again to be significantly identified with G2/M phase transition in all the breast cancer molecular, ovarian epithelial, and prostate Gleason score with the exception of ovarian endometrioid and prostate Gleason score 8. This was the only significant GO term identified for *HMMR* in the subtypes.

The GO terms for *CENPE* were found to be primarily related to chromosome and mitotic spindle functions. Mitotic cell cycle and cell division were found to be significant in the majority of subtypes with the exception of Gleason score 5, 8, and 9. This was also seen for mitotic spindle organisation and chromosome segregation. The GO terms for *CENPE* identified for Gleason score networks were more different than the other Gleason score subtypes. However, again these significant GO terms were related to the functions of microtubules and chromosomes.

For *STIL*, GO terms were significantly related to mitotic spindle organisation like that of *CENPE* as well as protein localisation to centrosomes. In the breast molecular and ovarian epithelial subtypes mitotic spindle organisation was significant in all subtypes. However, in prostate Gleason scores this was significant in the higher Gleason scores (Gleason scores 7-9) only. In prostate Gleason scores the two GO terms; floor plate development and regulation of centriole replication were found to be significant in higher Gleason scores (7 - 9) and not significant in lower Gleason scores. Notably it was also found that the *STIL* subnetwork was significantly associated with the 'negative regulation of apoptotic process'. This was only identified in the normal like, luminal A, and clear cell and not in the prostate Gleason scores.

3.6.3 Summary of gene network and functional analyses

The functions of the three hub genes and associated gene networks may be relevant for prognosis and survival. *HMMR*, *CENPE*, and *STIL* were analysed to identify the potential roles in these cancers and subtypes. These were assessed using network analysis to identify the co-expressed genes associated with *HMMR*, *CENPE*, and *STIL* independently in each of the cancer subtypes and then these gene lists analysed via GO.

Interestingly, the three hub genes were frequently identified to be co-expressed in the same network modules in nearly all subtypes and cancers. This suggests that they function in the same or similar pathways/functions. This, in combination with the results of survival analysis for these hub genes (particularly *HMMR* and *CENPE*), associated with reduced survival and are more likely to be pathways that promote tumorigenesis, invasiveness, and increased aggressiveness and poor prognosis.

The results of GO term analysis did find an underlying similarity in function of the co-expression modules further in terms that are involved in driving cell cycle progression (G2/M phase). The process of driving cell cycle proliferation is a keystone in cancers. This is a novel finding of this study and suggests that the level of similarity in function provides a unique treatment option to target highly interconnected hub genes. *HMMR* was predominantly significantly assigned to the regulation of G2/M transition of mitotic cell cycle, *CENPE* to mitotic cell cycle, and *STIL* to mitotic spindle organisation. This suggests that they pose a potential combined mechanism of action that together promotes carcinogenesis/progression in breast, ovarian, and prostate cancers.

In breast, ovarian, and prostate cancers, the three hub genes being co-expressed frequently in the same co-expression modules would emphasise their importance in these networks which was identified with high level of module membership (>0.8) due to the level of predicted interactions that they were found to have. This could also be seen when their co-expressed gene network modules were visualised. The number of interactions meant that a further heat diffusion step selecting the first-degree neighbours was used to help with visualisation and GO analysis. In fact, their interconnectivity to just first-degree neighbours (direct interactions) was commonly found to be greater than 50 genes and hence the reduced 25 presented in each network figure. This made both *HMMR*, *CENPE*, and to a lesser extent *STIL* (depending on the subtype) integral components of their network modules and frequently observed to be highly interconnected with other genes.

To determine if they are interlinked in their function directly, further investigation (wet laboratory experiments) would be required. These results from the network and pathway analyses were then taken further by combining with current literature.

3.7 Functions and mechanisms of HMMR, CENPE, and STIL

From the earlier analyses, *HMMR*, *CENPE*, and *STIL* were identified as putative hub genes in breast, ovarian, and prostate cancers, making them interesting biomarkers and potential treatment targets. This was supported by several pieces of evidence. First, they were all identified as significantly upregulated in breast, ovarian, and prostate cancer subtype comparisons. Secondly, all three hub genes were found to have large interconnectivity (high module modality) with other genes with interrelated functions identified in GO suggesting a common pathways/pathway of action; these primarily being G2/M phase, Mitotic spindle organisation, and cell division.

These GO terms were used to help understand the potential mechanism/mechanisms of action in of these hub genes in these cancers. A literature review was conducted using the GO terms identified from the previous chapter (section 3.6) in order to determine how the hub genes maybe interrelated in function and in order to propose a potential novel mechanism of action. Confirmation of this mechanism of action could be the aim of future research.

In the following sections the cellular roles of the HMMR, CENPE, and STIL proteins are given.

3.7.1 HMMR – A HA receptor

The hyaluronan mediated motility receptor (HMMR also referred to as RHAMM) is a hyaluronic acid (HA) binding receptor. It is one of four known HA binding receptors, with the main HA receptor identified as being CD44 (Marhaba and Zöller, 2004). Whilst CD44 has been identified as the main HA receptor, it was only significantly upregulated in breast cancer invasive subtypes and downregulated in ovarian and prostate cancer subtypes (see expression of *CD44* in Table 72). As such, it is unlikely that, at least in the subtypes analysed here, current inhibitors of CD44 (such as Verbascoside) will be cancer specific and are likely to show only a potential benefit in breast cancers. *HMMR* however, shows consistent significant upregulation across all the cancers and subtypes and is potentially the primary HA receptor. This makes it attractive as a potential treatment target across all three cancers.

The only structural difference between the CD44 and HMMR proteins is that HMMR does not contain a transmembrane domain, but is instead directly bound to the plasma membrane via GPI-anchoring (Glycosylphosphatidylinositol) (Misra *et al.*, 2015). GPI-anchor bound proteins have been suggested to also be attractive biomarkers for cancers (Dolezal *et al.*, 2014), due to their ability to be detected in plasma.

The role of HMMR in breast, ovarian, and prostate cancers

HMMR has been found to have multiple biological roles across many cellular functions and pathways. These range from roles in cell motility, proliferation, inflammation, wound healing, and angiogenesis. Importantly, these are all commonly part of the G2/M phase (identified in *HMMR* modules). It is likely that HMMR will influence some, but not all, these functions to varying degrees in these cancers and potentially in a subtype dependent manner. As such, further investigation of HMMR in these cancers and subtypes will be required to fully quantify the exact potential of this target gene in treatment.

Here the roles of HMMR identified from literature are described with context to previous findings in cancers if possible. Similar roles and functions have been combined where logical and appropriate.

HMMR and microtubule dynamics

The HMMR protein is a primary component controlling the microtubule dynamics that lead to alteration in cell polarity due to loss of apicobasal polarisation. Apicobasal polarity helps with mitotic spindle orientation and tissue formation in progenitor cells such as basal and luminal cells. Disruption to this has been observed during tumour progression in breast, ovarian and prostate cancer (Shafer *et al.*, 2017; Eder *et al.*, 2005; Chatterjee and McCaffrey, 2014). Loss of the regulation of apicobasal polarisation leads to incorrect tissue growth and early tumour invasion. It is believed that the BRCA1, HMMR and AURKA proteins together alter apicobasal polarity in epithelial cells (Maxwell *et al.*, 2011). Both *HMMR* and *BRCA1* were commonly found to be co-expressed within the same module in breast, ovarian, and prostate cancer subtypes in this study. These modules are shown in Appendix (Figure 94-96). In summary, AURKA has been identified as

regulating BRCA1 by preventing BRCA1 binding and ubiquitination of HMMR. This is by preventing phosphorylation by BRCA1 to HMMR preventing downregulation of microtubule nucleation at the centrosome (Maxwell *et al.*, 2011). This in turn leads to loss of BRCA1 regulation of cell polarisation due to microtubule nucleation downstream and loss of cell polarity. Over time, the loss of tissue polarity will facilitate tumour invasion into surrounding tissue such as stroma. *AURKA* expression is significantly upregulated in the cross-cancer analyses; histological breast cancer, ovarian tissue location, and prostate Gleason grade group ($\log_2FC = 1.00$, $P < 0.01$), and Molecular breast cancer, ovarian epithelial, and prostate Gleason score ($\log_2FC = 1.31$, $P < 0.01$). *AURKA* was also identified to be co-expressed with *HMMR* in all subtypes suggesting that HMMR bound *AURKA* is likely present in these cancers. *BRCA1* was also significantly upregulated in nearly all subtypes, but as it was not significant in all subtypes, it was not identified in the cross-cancer analyses. The expression changes of *BRCA1* in breast, ovarian, and prostate cancer subtypes are shown in appendix Table 73.

HMMR aids microtubule nucleation via *AURKA* and *TPX2* localisation

In addition to apicobasal polarity deregulation, the movement of HMMR-*AURKA* to the microtubule organisation centre (MTOC) promotes microtubule nucleation by recruitment of *TPX2* to the HMMR-*AURKA* complex at the centrosome. HMMR-*AURKA* at MTOCs has been found to promote *TPX2* localisation (Bayliss *et al.*, 2003). The movement of *TPX2* to the centrosome occurs via importin α across the nuclear membrane and the binding of activated RAN (RAN-gtp) (Fu and Zhang, 2010). The binding of RAN(gtp) to *TPX2*-importin(α) removes importin- α (Gruss *et al.*, 2001) and releases *TPX2* to HMMR-*AURKA* at the centrosome. Once *TPX2* localisation occurs, it activates *AURKA* via phosphorylation beginning microtubule formation (Chen *et al.*, 2014). *TPX2* was observed to frequently be co-expressed with *HMMR* in all breast, ovarian, and prostate cancer co-expression networks. Further to this, *AURKA* depletion has been observed to reduce intracellular HMMR which, as discussed previously, is believed to be controlled by the BRCA1 ubiquitination of HMMR (Maxwell *et al.*, 2011; Joukov *et al.*, 2006). The proposed mechanism for

the HMMR protein in these cancers is shown in Figure 36. The increased expression of HMMR in the various cancers would be predicted to increase capacity of microtubule spindle formation.

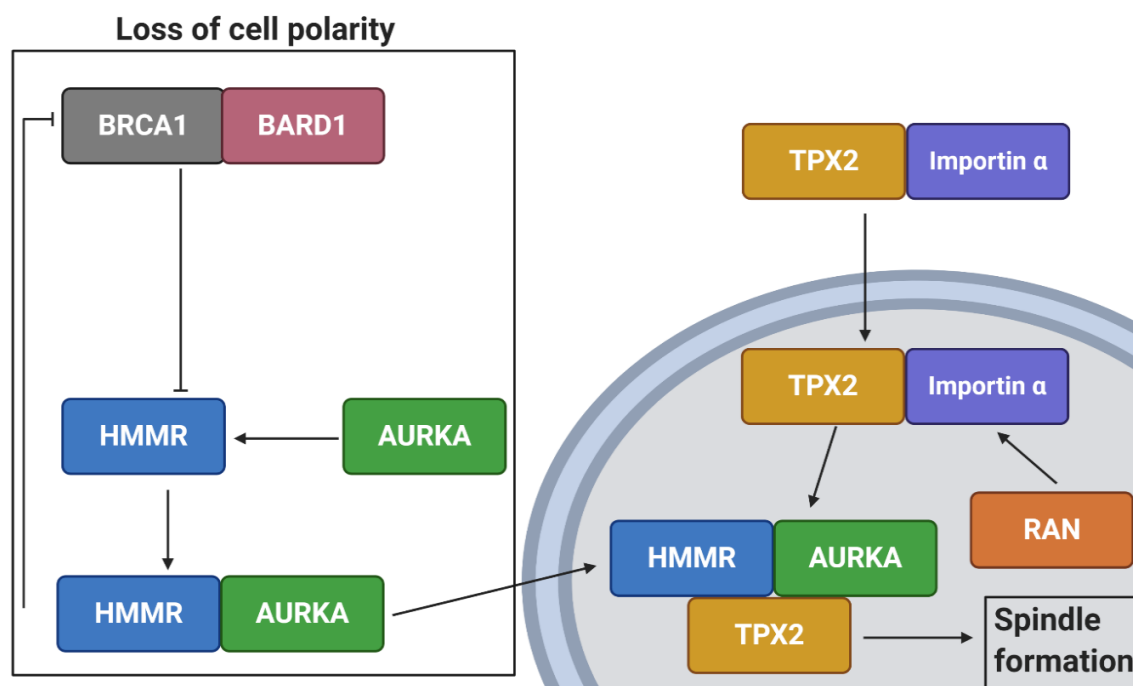


Figure 36. Proposed mechanism of HMMR regulated apicobasal polarity and microtubule formation.

BRCA1 regulation of microtubule spindle formation is lost due to AURKA binding of HMMR preventing ubiquitination of HMMR. AURKA-HMMR complex can then move to the MTOC at the centromere recruiting TPX2 to the complex to begin spindle formation. This causes the loss of cell polarity and invasion to surrounding tissue.

HMMR induces proliferation and anti-apoptosis pathways via Ras activation

HMMR has been found to be important for activation proliferation pathways and anti-apoptosis related pathways. These include the MAPK/ERK and PI3K/Akt signalling pathways.

HMMR is believed to regulate RAS signalling induction (Wang *et al.*, 1998) with initial activation of this pathway occurring via HMMR-bound HA binding to RAS associated RTKs (Misra *et al.*, 2015). This begins the RAS signal induction which, in turn, promotes downstream cell proliferation via PI3K/AKT and ERK/MAPK pathways (Vara *et al.*, 2004). This is a feature commonly observed in multiple carcinomas such as breast (Muga *et al.*, 2008), ovarian (Shayesteh *et al.*, 1999), and prostate (Butler *et al.*, 2017). In basal breast cancer cell lines (MDS-MB-231), increased HA still contributed to ERK signalling despite *CD44* knockdown, promoting invasiveness which was believed to be due to *HMMR* expression (Hamilton *et al.*, 2007; Nedvetzki *et al.*, 2004). This has also been observed in head and neck carcinomas where a stem cell renewal phenotypic trait is

maintained by both CD44 and HMMR (Shigeishi *et al.*, 2013). *HMMR* may therefore prove to be a novel CSC biomarker for breast, ovarian, and prostate cancers, potentially, in part, explaining the finding of high *HMMR* expression being significantly associated with reduced OS and RFS. This would be especially relevant in more aggressive/proliferative cancer subtypes such as basal breast cancers, high grade serous and endometrioid ovarian cancers, and Gleason grade 8 - 10 for prostate cancers where *HMMR* expression was observed to be highest. It is also important to note that RAS signal cascades can also activate anti-apoptotic pathway PI3K/AKT. Activation of the PI3K pathway can lead to both proliferation and anti-apoptosis via BAD (Mendoza, Er and Blenis, 2011). The role of HMMR in activation of cell proliferation and prevent apoptosis via RTKs and inducing RAS cascade signalling makes it an interesting target for treatment. A summary of these mechanisms is shown in Figure 37.

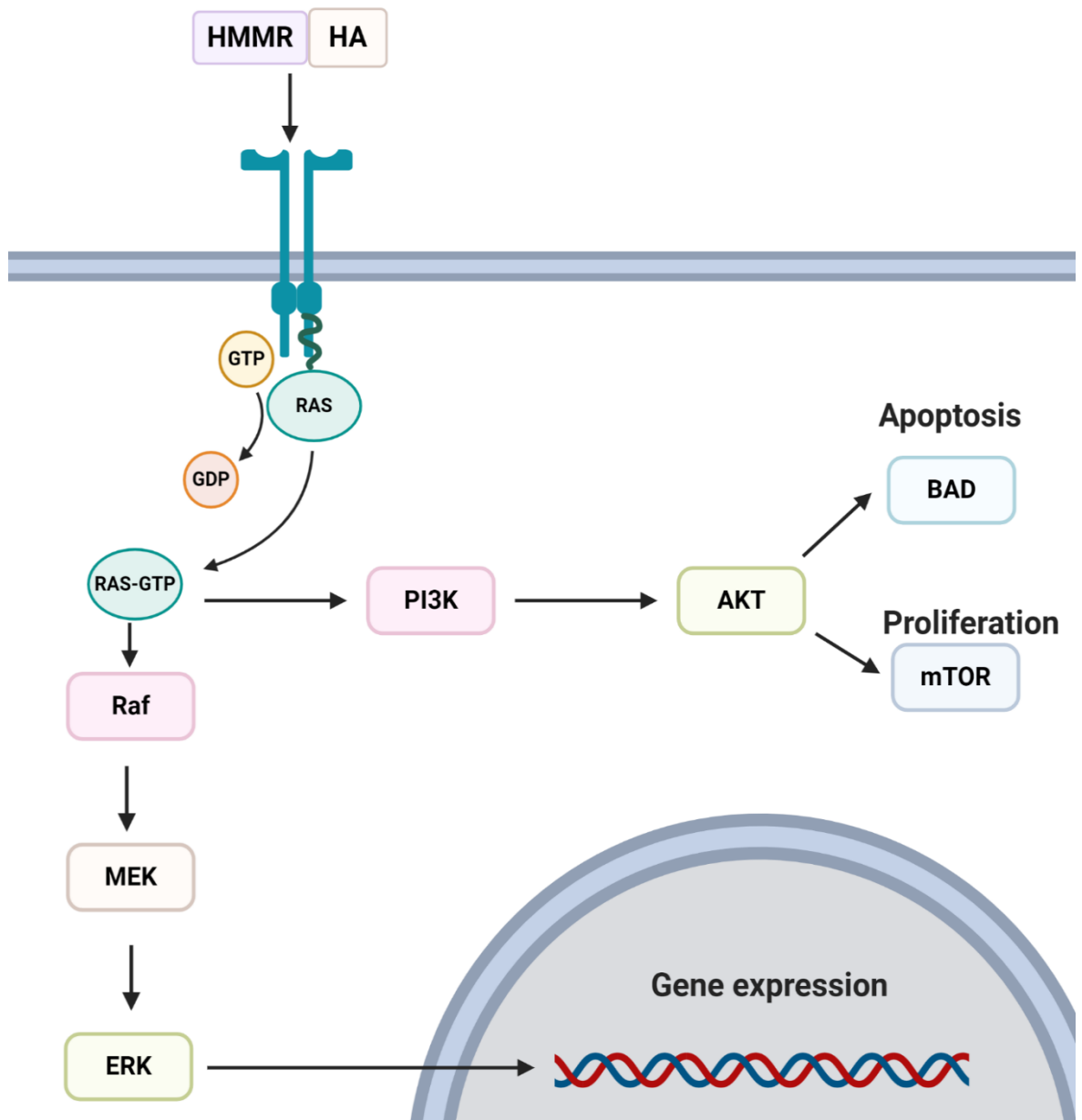


Figure 37. Proposed mechanisms of action for HMMR in cell proliferation.

HMMR binding to HA and RTK activation leads to RAS pathway activation. From RAS activation both PI3K and MEK pathways can be induced to lead to increased cell proliferation and/or apoptosis inhibition.

RTK: Receptor Tyrosine Kinase

HMMR and Inflammation – altering the tumour microenvironment

HMMR has been found to be functionally involved in promoting inflammatory response (Schwertfeger *et al.*, 2015). Inflammation has been observed in many cancers, including breast (Agnoli *et al.*, 2017), ovarian (Jia *et al.*, 2018), and prostate cancers (Stark, Livas and Kyprianou, 2015). Inflammation has a central role in controlling the tumour microenvironment with cytokines

and chemokines. Many carcinoma cells produce pro-inflammatory signals (chemokines and cytokines) to alter their microenvironment by controlling the recruitment of leukocytes and fibroblasts to the tumour microenvironment (Gatti-Mays *et al.*, 2019). This leads to a positive feedback loop in which further pro-inflammatory signals are produced that can benefit the carcinomas growth further. Importantly, leukocytes lead to the promotion of ROS (Liou and Storz, 2010; Storz, 2005) which has been found to be important for *HMMR* expression. Fibroblasts have also been found to produce HA which can be utilised by extracellular *HMMR* to activate RTKs and, again, promote proliferation and anti-apoptosis pathways.

The chemokine receptors *CXCR1* and *CXCR4* have been observed to be associated with leukocyte recruitment during inflammation. These have been reported as being important in breast (Muller *et al.*, 2001), ovarian (Muralidhar and Barbolina, 2013), and prostate (Salazar *et al.*, 2013) cancers previously. Within the cancer subtypes analysed here, the *CXCR1* gene was not found to be upregulated significantly. However, *CXCR4* which is also a marker of metastasis, was observed to be significantly upregulated in breast and ovarian cancer subtypes, with highest expression observed in more invasive subtypes (Table 74). For example, one potential mechanism is that with *HER2* being an RTK and *CD44* being reduced or absent, the extracellular *HMMR* can bind HA. This in turn promotes *CXCR4* expression leading to leukocyte (and macrophage) recruitment that can increase ROS production, which could also lead to further *HMMR* expression. Notably, this mechanism of leukocyte recruitment is commonly observed in response to inflammation. Importantly this can activate innate immune response because of inflammation signals (Newton and Dixit, 2012).

Toll-like receptor signalling is an important inducer for the innate immune system in response to the detection of inflammation (Kawai and Akira, 2006). Toll-like receptor pathways (TLRs) have been found to be regulated by *HMMR* (Foley *et al.*, 2012). The main role of the TLR pathway is to activate the production of TGF- β which results in innate immune response activation causing the recruitment of tumour associated macrophages (TAMs) and leukocytes. In later stage tumorigenesis, higher levels of TGF- β can be beneficial to cancers and act in an anti-inflammatory

capacity as a metastatic promoter (Jakowlew, 2006; Foley *et al.*, 2012). Importantly this has been found to require HMMR–HA but not CD44. In *CD44* knockdown mice, TLRs were still active for surfactant protein A (SPA) but those with *HMMR* knockdown showed decreased macrophage chemotaxis in response to SPA, HA, and TGF- β (Foley *et al.*, 2012). In this study, expression of TGF- β was significantly upregulated in breast cancer subtypes. It was also observed to decrease in molecular breast cancers from luminal B to Basal (Table 75). At higher grade and later stage Luminal B, HER2, and Basal breast cancers have been identified to show far lower inflammatory mediated immune response and TGF- β induced macrophage recruitment (Gatti-Mays *et al.*, 2019). The expression of TGF- β was not significantly upregulated in ovarian subtypes with the exception of clear cell. In prostate cancer TGF- β expression was not significant, rather, it was observed to decrease in higher Gleason scores (Table 75). Overall, this increases ROS production from the recruited macrophages/leukocytes. Increased ROS production itself is important for increasing *HMMR* expression. And as such works as a potential feedback mechanism. This is because in response to increased ROS, HA production increases by *TNF- α* expression. *TNF- α* was found to be significantly upregulated in the more invasive IDC and basal breast cancers, ovarian tissue and all ovarian epithelial subtypes, and Gleason Grade group 2 & 3 (appendix Table 78).

It is unlikely that increased *TNF- α* expression in these subtypes increased HA synthesis as its target genes the hyaluronic acid synthases (*HAS1*, 2, and 3) were not significantly upregulated and in fact, were actually found to be significantly downregulated in most subtypes (appendices Table 76-77). Therefore, the source of HA synthesis is unlikely from the tumours themselves and likely from another source. A potential answer is via fibroblast recruitment by tumours to the microenvironment, where they then provide HA. It is likely, therefore, that in these subtypes, ROS production is from leukocytes and HA production from fibroblasts with both cell types being potentially recruited by CXCR4 in response to MAPK pathway activation by HMMR. This is also important as ROS can lead to fragmentation of HA which is oncogenic in function (Bourguignon *et al.*, 2011). This fragmentation is from non-oncogenic high molecular weight HA (HMW-HA) to oncogenic pro-inflammatory low molecular weight HA (LMW-HA) (Basakran, 2015; Lokeshwar,

Mirza and Jordan, 2014). The role of HMMR the recruitment of leukocytes and fibroblasts altering the tumour microenvironment is shown in Figure 38.

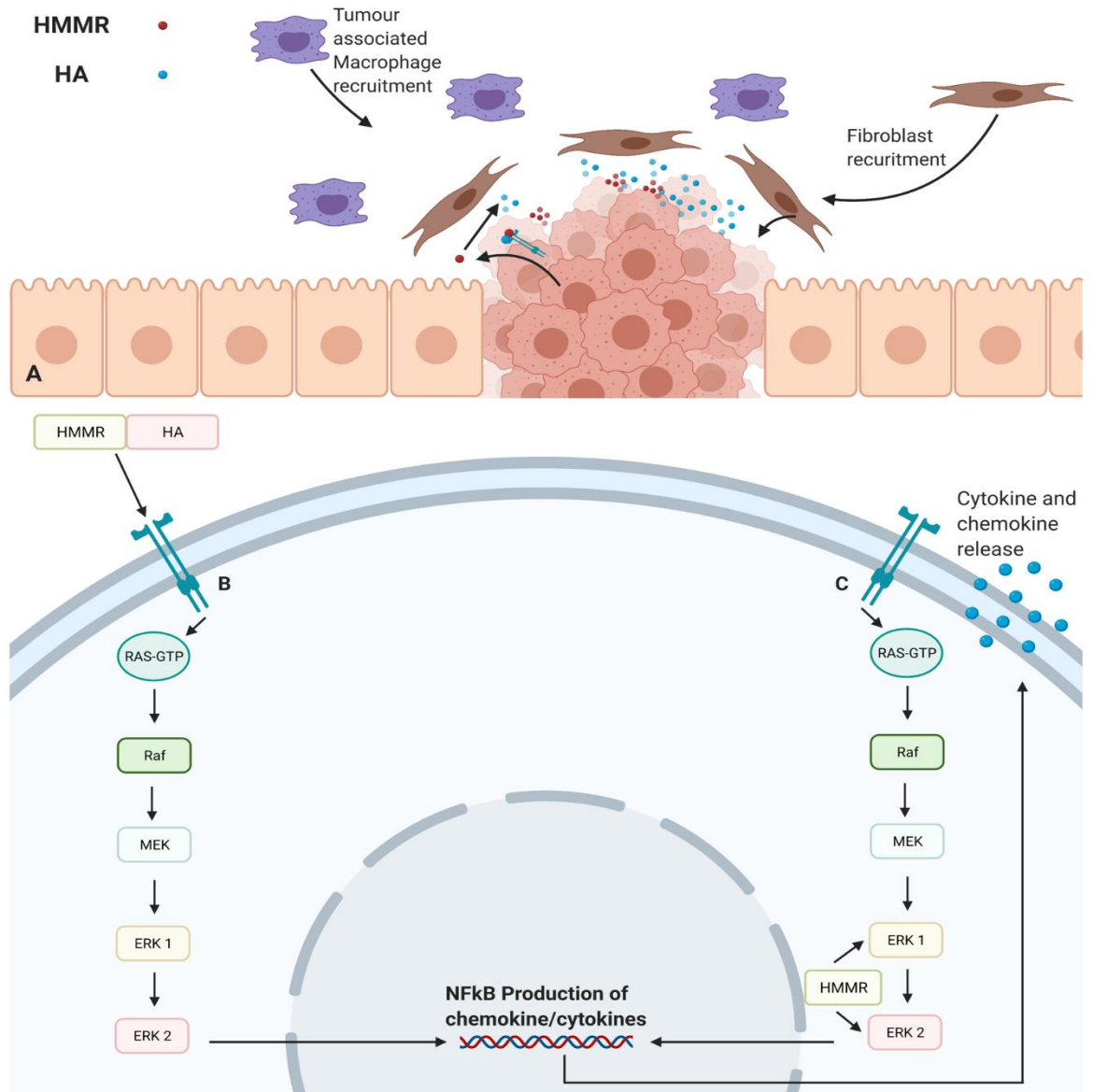


Figure 38. HMMR alterations to the tumour microenvironment.

A) HA can be produced by carcinomas themselves however restructuring of the tumour microenvironment leads to alternate sources of HA for utilisation. Extracellular HMMR can bind with this source of HA to initiate HMMR mediated pathways. B) Extracellular HMMR can bind with HA to induce RAS signalling to promote downstream cytokine and chemokine production. B) Intracellular HMMR can bind in MAPK/ERK signalling to form a complex with ERK1 and ERK2 to promote Cytokine and Chemokine signalling. C) Changes to the tumour microenvironment by chemokine and cytokine release leads to alternate sources of HA for utilisation from the recruitment of fibroblasts. Extracellular HMMR can bind with this source of HA to initiate HMMR mediated pathways further.

HMMR aids cell motility and metastasis

Increased levels of intracellular HMMR can lead to increased expression of matrix metalloproteinases (*MMPs*) which has been found to promote cell motility (Misra *et al.*, 2015). In relation to angiogenesis *MMPs* activate VEGFs to induce angiogenesis (Mills, 2017). Angiogenesis is a process exploited by tumours that facilitates continued tumour growth through new blood supplies, thus maintaining and increasing the requirements of the carcinoma. This greatly increases the metastatic potential of carcinomas by providing a biological 'highway' to other distal areas to which carcinoma cells can migrate.

HMMR has a role in initiating angiogenesis via TGF- β and *MMPs*. The expression of *MMPs* is regulated by the MAPK pathway, activation of which leads to increased expression of *MMPs* (Reddy *et al.*, 1999). As discussed earlier, HMMR is able to activate RTKs associated with MAPK signalling whilst also able to activate MAPK independent of these receptors intracellularly.

MMPs are active in degrading the ECM, thus allowing carcinoma cells to metastasise through blood vessels. In breast cancer, the *MMP9* is commonly observed to be increased via MAPK pathway activation (Reddy *et al.*, 1999). *MMP9* is a downstream target of MAPK/ERK pathway activation with HMMR-HA binding to RTKs to initiate this pathway. Intracellular HMMR has also been observed to promote *MMP9* expression (Tolg *et al.*, 2014). In these cancers, *MMP9* expression has been found to be significantly upregulated in nearly all cancers and subtypes except for ovarian peritoneum and mucinous subtypes. These are shown in appendices Table 79.

The expression of *MMP9* has also been found to promote epithelial to mesenchymal transition (EMT) in breast (Radisky and Radisky, 2010), ovarian (Zhang and Chen, 2017), and prostate cancers (Wang *et al.*, 2013). This, linked with HMMR downstream *MMP9* expression, provides evidence of an indirect HMMR-promoted EMT induction (and angiogenesis) through ECM remodelling via *MMP9*.

TGF- β signalling and *MMP9* are important initiators of EMT (Maier, Wirth and Beug, 2010; Huber *et al.*, 2004). The downregulation of TGF- β as discussed earlier is most likely in the benefit of

increasing ROS and subsequently aiding in LMW-HA production from HMW-HA in the surrounding microenvironment (a potential feedback loop). In those subtypes where loss of *TGF-β* was found EMT could be additionally driven by another means in these cancers such as HMMR mediated expression of *MMP9*. The loss of *TGF-β* has also been found to not be required for EMT in pancreatic carcinomas (Maier, Wirth and Beug, 2010). This suggests that previous evidence from ovarian and prostate cancers, where *TGF-β* expression was not significantly expressed, may not be a necessary requirement for EMT. It is also important to note that changes in apical-basal cell polarity is a key feature of EMT (Xu, Lamouille and Derynck, 2009), another function of HMMR described previously. The role of HMMR in EMT is shown in Figure 39.

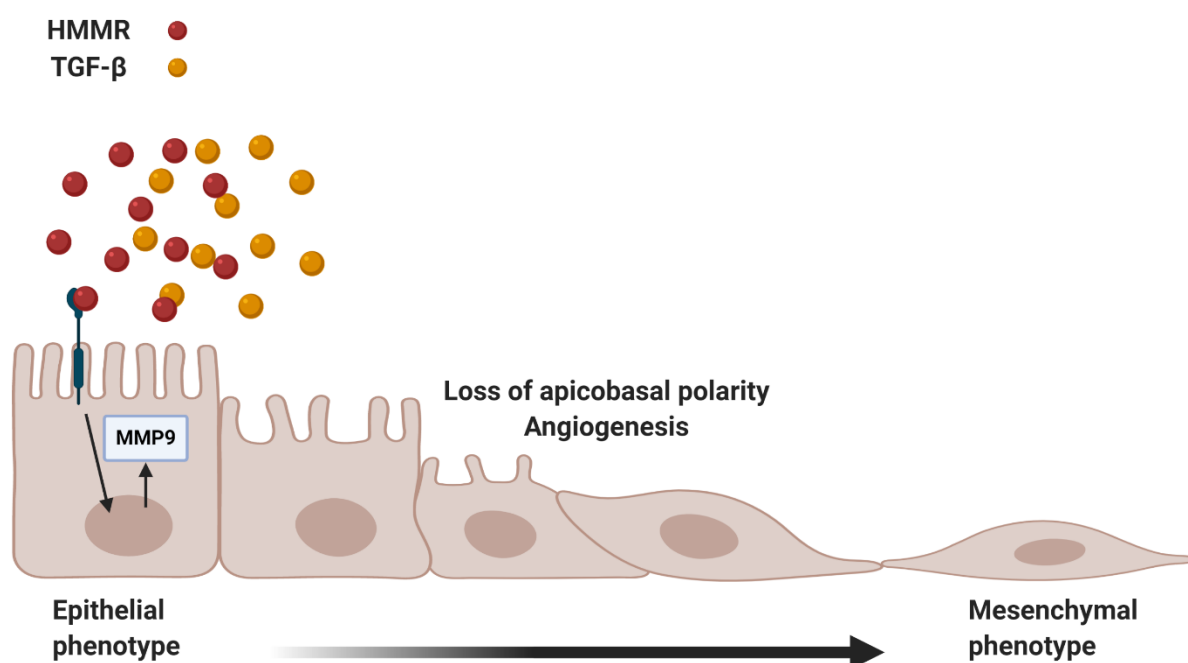


Figure 39. HMMR promotes EMT phenotype via utilizing *TGF-β* receptor and RTKs.

MAPK leads to increased *MMP9* expression promoting EMT phenotype. HMMR also leads to loss of apical-basal polarity via microtubule restructuring. Lastly HMMR also promotes angiogenesis which aids in metastasis.

HMMR as a hub gene target in breast, ovarian, and prostate cancers

There are multiple cellular roles that *HMMR* plays that make it's identification as a commonly disrupted hub gene in breast, ovarian, and prostate cancer subtypes of interest. The first is that extracellular HMMR-HA can facilitate RAS signal induction through RTKs (such as EGFR and FGFRs and HER2) as it has been observed that CD44 is not required for HMMR-HA mediated signal

induction. As such, current CD44 silencing would not inhibit cell proliferation and is possibly why HMMR was identified in G2/M phase transition here. Because of this, the role of HMMR in inducing cell proliferation makes it an interesting target for treatment.

HMMR effect on microtubule dynamics promotes recruitment of important mitotic factors such as AURKA and TPX2 to the centrosomes. This is an important factor in the loss of apicobasal polarity leading to contribution in both metastasis and cell cycle progression. Increased HMMR also leads to potential induction of metastasis via expression of angiogenesis related genes. This also explains why *HMMR* was observed to significantly associated with reduce OS and RFS.

The increased expression of *HMMR* may also promote changes in the tumour microenvironment by recruiting leukocytes and fibroblasts. Both of which may provide additional benefits to the tumour in the form of ROS from leukocytes and HA from fibroblasts. This makes it a promising target for treatment across breast, ovarian, and prostate cancers.

3.7.2 CENPE – A chromosome separator

The centromere-associated protein E (CENPE) is part of a family of kinesin motor like proteins. CENPE has been found to accumulate in the G2 phase of mitotic cell cycle (Yen *et al.*, 1992). Notably, both CENPE and HMMR accumulate at the G2/M phase providing a common functional association between the two proteins at this stage of cell cycle. This is most likely due to their roles in microtubule dynamics. *CENPE* and *HMMR* were frequently observed to be co-expressed within the same network modules (section 3.6.1 & 3.6.2). In breast cancer, *CENPE* has only been identified as being over-expressed in triple negative samples (Kung *et al.*, 2014). Its increased expression has not been previously identified in ovarian or prostate cancers or in any of the other breast cancer subtypes analysed here.

Role of CENPE in breast, ovarian, and prostate cancers

Unlike HMMR, which has a diverse set of roles in multiple cell cycle related pathways, CENPE appears to have a more restricted functionality, being primarily observed in organisation of microtubule spindles, promotion of cell cycle progression through checkpoint regulatory

mechanisms and chromosome separation. During metaphase, CENPE is primarily involved in the microtubule orientation that allows chromosomes to align correctly at the centromere; whilst HMMR is an important for initial spindle microtubule assembly through its effect on the localisation of AURKA to the MTOC (section 3.7.1).

During anaphase, CENPE is observed to be bound to BUBR1 allowing recruitment of AURKB to the kinetochores. At this point CENPE dissociates from this complex. This is an important part of the anaphase promoting complex (APC). The APC (APC2 and APC11) is key for the separation of chromosomes during anaphase and it regulates the 'speed' of sister chromatid separation. If chromosomes are separated into sister chromatids too quickly, genomic instability can occur such as aneuploidy (Weaver *et al.*, 2003).

Currently, one potential direct CENPE inhibitor, GSK923295, has been identified. This has been found to prevent the CENPE motor function of binding microtubules to kinetochores. This drug prevented chromosome alignment, forcing cell cycle arrest and apoptosis in cell lines and xenographs (Wood *et al.*, 2010). Other than this, taxanes are the current utilised treatment that alter microtubule dynamics. Taxanes and their mechanisms are discussed in more detail in the following sub chapters.

[CENPE promotes genomic instability via changes in microtubules](#)

CENPE has been reported to be important for kinetochore fibres (Yao, Anderson and Cleveland, 1997). Recently, it has been found that CENPE is part of plus end microtubule elongation of kinetochore fibres (K-fibres). The plus end of K-fibres is attached to the kinetochores via CENPE directly. CENPE has been found to move along the microtubules and bind to its plus end to the K-fibre (Sikirzhytski *et al.*, 2018; Cai *et al.*, 2009). It has also been reported that CENPE promotes microtubule elongation via α and β tubulin with ATP coupling (Sardar *et al.*, 2010). Extending the microtubules from the kinetochore to centrosomes (Walczak and Heald, 2008). These microtubules are vital for correct chromosome segregation (Bickenson, 2012).

Aberrant microtubule formation (such as K-fibres) in cancers promotes cell proliferation and potentially leads to genomic instability (Mukhtar, Adhami and Mukhtar, 2014; Stumpff *et al.*, 2014). This role of CENPEs in K-fibre microtubule/kinetochores function suggests it to be a potential target for the current taxane-based treatment in CENPE positive cancers. If incorrect formation of these microtubules occurs via *CENPE* overexpression, this would drive genomic instability via chromosome instability. This has been observed to occur prior in K-fibres not linked directly to CENPE (Milunović-Jevtić *et al.*, 2016). The mechanisms of which K-fibre nucleation occurs are still being explored, but CENPE is a possible cause from findings of previous studies.

Taxane treatment is the primary method to either promote stable microtubule formation or works to destabilise microtubules. In some cancer subtypes taxanes have been observed to be ineffective due to the development of resistance (Harrison, Holen and Liu, 2009). Many of the mechanisms underlying resistance to taxanes are yet to be identified, but with the identification of CENPE's role in microtubule elongation, it is possible that CENPE has a potential role in resistance. If CENPE was susceptible to taxane treatment, leading to microtubule stabilisation/destabilisation, it might be predicted to impact upon OS and/or RFS for patients treated with taxanes. This would be seen as an increased survival probability in patients with upregulated CENPE. This was examined, but due to limitations in the availability in public data for specific treatments, taxane-based treatment datasets for OS and RFS were only identified for ovarian cancers. This is primarily because taxanes have been the main treatment method for ovarian cancers and have shown to be an effective treatment (Schwab *et al.*, 2014). It was observed here that high *CENPE* expression in ovarian cancer was associated with significantly improved OS times (Figure 32 B). However, this did identify a potential taxane resistant group within the patients, where lower expression of *CENPE* conforms to a significantly lower survival probability in OS.

It is important to note that the expression of *CENPE* is higher in the survival data patients compared to normal samples in either group (high and low). This is because *CENPE* has been found to be significantly upregulated in ovarian, breast, and prostate cancer patients previously in

this thesis (sections 3.4.1 and 3.4.2). As such the classifier for ‘higher or lower’ expression groups in survival analysis that is applied will still be higher than that in normal tissue samples. Therefore, even the ‘low’ *CENPE* expression group observed here it will still have a higher expression of *CENPE* than that of normal tissue and the ‘high’ *CENPE* expression group even more so. However, one potential mechanism of resistance is that *CENPE* microtubule elongation is effectively targeted by taxanes and promotes stable microtubules. This would be expected to show better prognosis as the resistance of taxane treatment mechanism would come from *CENPE* secondary functions in chromosome segregation dysregulation rather than microtubule assembly. It is important to note that even if taxanes did affect *CENPE* associated microtubule formation this would not necessarily affect its secondary role in spindle checkpoint (discussed in more detail in following section). Therefore, taxane treatment would be an ineffective long-term treatment mechanism in patients with high *CENPE* expression. It would be interesting to determine whether, for breast and prostate cancers, there is a similar effect upon survival after taxane treatment as that observed for ovarian cancers. This is because they have previously been identified to be similar in the phenotypes as discussed in the introduction (section 1.5).

Loss SAC/APC regulation via *CENPE*

Genomic instability is a key hallmark of nearly all cancers with increased genomic stability being associated with a poorer prognosis (Negrini, Gorgoulis and Halazonetis, 2010). The anaphase promoting complex (APC) is a large ubiquitin ligase complex involved in the maintenance of genomic stability. *CENPE* has been shown to play a central role in regulating this complex (Rao *et al.*, 2009). *CENPE* is a component of the SAC complex (spindle assembly checkpoint) which acts to inhibit the APC if incorrect chromosome alignment is identified, therefore preventing separation. The SAC complex consists of the MAD1, MAD2, BUB1, BUBR1, and MPS1 proteins. In the SAC, *CENPE* has been found to recruit BUBR1 to form a BUBR1-*CENPE* complex, which is located at the kinetochores (Mao, Desai and Cleveland, 2005). This complex aides in the regulation of Securin (PTTG1) ubiquitination via APC and chromosome separation via Separase release (Huang *et al.*, 2005). BUBR1 is a checkpoint kinase involved in assessing the functioning of *CENPE*. *CENPE* has

been found to silence BUBR1, inhibiting regulation of checkpoint signalling of the APC (Mao, Desai and Cleveland, 2005). Loss of BUBR1 regulation has been found to prevent cell cycle arrest (Chan *et al.*, 1999). Increased expression of CENPE leads to premature movement into the anaphase and separation of chromatids promoting genomic instability. This would be through inhibition of BUBR1 removing/reducing BUBR1 regulation of APC activation driving progression into the anaphase (Figure 40 A & B). *BUBR1* was also found to be significantly upregulated in all breast, ovarian and prostate cancer subtypes. Histological breast, ovarian tissue location, and prostate Gleason grade (Log2 FC = 1.76, P<0.01), Molecular breast, ovarian epithelial, and prostate Gleason score (Log2 FC = 2.04, P<0.01). *BUBR1* (*BUB1B*) was found to be co-expressed with *CENPE* in the same networks.

BUBR1 binding to CENPE at the kinetochores releases BUBR1-bound CDC20 (Cleveland, Mao and Sullivan, 2003). APC requires association with activators CDC20 (Harkness, 2018). *CDC20* was found to be significantly upregulated in the cross-cancer analysis in all subtypes: Histological breast, ovarian tissue, and prostate Gleason grade (log2FC = 1.30, P<0.01), Molecular breast, ovarian epithelial, and prostate Gleason score (log2FC = 1.61, P<0.01). *CDC20* was also observed to be co-expressed with *CENPE*.

The inhibition of unattached kinetochores has been observed to be limited in the silencing anaphase promotion via BUBR1. In cases where relatively few unattached kinetochores to spindles are observed, the inhibitory signal by BUBR1 does not prevent APC/CDC20 ubiquitination of Securin and subsequent chromosome separation (Cleveland, Mao and Sullivan, 2003).

Therefore, it only takes a few unattached kinetochores to promote cell cycle progression. As such increased *CENPE* expression can induce a false positive kinetochore binding signals to promote incorrect separation in these cancers due to limitations in BUBR1 signal inhibition. This would be uniquely advantageous to these cancers as it will sequentially promote genomic instability at a potential non-lethal rate with incorrect separation being driven by only 1 or 2 incorrectly bound spindle assemblies due to CENPE.

In normal conditions, APC activation via CENPE targets downstream PTTG1 (Securin) for ubiquitination leading to chromosome separation if correct alignment is identified via BUBR1 (Cleveland, Mao and Sullivan, 2003). Securin ubiquitination promotes Separase activation and Cohesin degradation. Here *PTTG1* was observed to be significantly upregulated also in the cross-cancer analysis across subtypes; Histological breast, ovarian tissue, prostate Gleason grade ($\log_2FC = 1.43$, $P < 0.01$), Molecular breast, ovarian epithelial, and prostate Gleason score ($\log_2FC = 1.72$, $P < 0.01$). *PTTG1* was also frequently found to be co-expressed with *CENPE* also. Upregulation of Separase (encoded by the *ESPL1* gene) has also been observed to be significantly upregulated with *CENPE* in all cancer subtypes; Histological breast, ovarian tissue, and prostate Gleason grade ($\log_2FC = 0.75$, $P < 0.01$), Molecular breast, ovarian epithelial, and prostate Gleason score ($\log_2FC = 0.84$, $P < 0.01$). As with other CENPE target genes, *ESPL1* (Separase) was also identified in the network analysis to be co-expressed with *CENPE*. Aberrant activation of Securin via CENPE premature anaphase induction would be expected to cause premature Separase release and cohesin breakdown. This would promote premature chromosome separation possibly before correct alignment to the centrosome. It has been observed that increased Separase expression does result in premature chromosome separation. This has been observed to promote chromosomal instability and aneuploidy (Zhang and Pati, 2017). CENPE orientated genomic instability has been observed in in-vivo mice where aneuploidy was frequently observed (Weaver *et al.*, 2006). A summary of CENPE in microtubule dynamics and chromosome separation is shown in Figure 40.

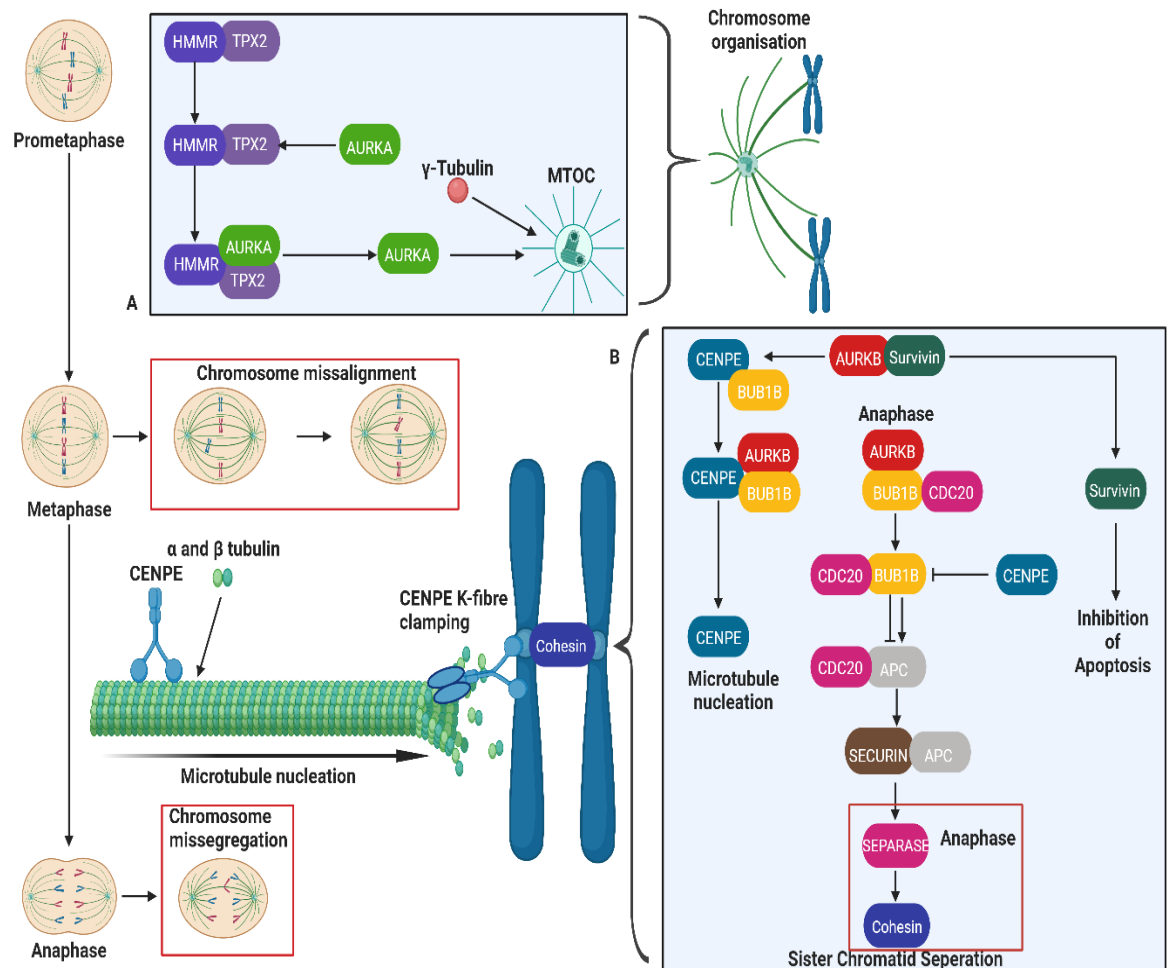


Figure 40. Summary of CENPE function in microtubule dynamics.

A) AURKA movement to the MTOC causes the activation of CENPE in microtubule nucleation specifically K-fibre elongation by adding α and β tubulin. CENPE moves along the K-fibre binding the plus end of the K-fibre to the centrosome. B) CENPE promotes incorrect or premature chromosome separation in the anaphase by inhibiting BUBR1 regulation of CDC20. This is through incorrectly 'overriding' the BUBR1 signal of incorrect chromosome alignment/attachment. This allows for APC binding to SECURIN leading to Cohesin ubiquitination by SEPARASE and premature/incorrect chromosome separation.

CENPE poses a potential resistance to taxane based treatments

Inhibition of CENPE may prevent both its kinetochore binding and microtubule organisation activities. This should prevent cell cycle progression through reactivation of SAC control. This may offer CENPE as a potential treatment target within these cancers as the BUBR1-CDC20 regulatory mechanism appears to be retained. The expression of BUBR1 and CDC20 was observed to be similarly identical in cancer subtypes (appendices Figure 97 & 98).

This would be particularly important in basal breast cancers where the levels of *BUBR1* and *CDC20* expression was found to be highest. By inhibiting CENPE function, sufficient *BUBR1* regulation should stall cell cycle progression to induce apoptosis.

It has been observed that taxane-based treatments in basal breast cancers are insensitive to CENPE expression (Kung *et al.*, 2014). Therefore, more direct treatments are required for CENPE. In cases where CENPE has been efficiently silenced this promotes cell cycle arrest (Chan *et al.*, 1999; Ohashi, Ohori and Iwai, 2016). This makes it an interesting target for treatment.

CENPE as a hub gene target in breast, ovarian, and prostate cancers

The role of *CENPE* as a hub gene in these cancers appears to act in both a protective and promoter manner. It acts to protect against chronic genomic instability that would most likely lead to apoptosis by stabilising kinetochore and spindle binding promoting cell cycle progression. This is evident in that there is consistent upregulation of both *BUBR1* and *CDC20* and observed to be co-expressed with *CENPE*. However, it has been identified previously that there is no perfect spindle assembly binding by CENPE with some kinetochores not being bound to spindles.

Relatively few of these unbound kinetochores have been found to be insufficient to blocking of CENPE induction of APC-CDC20 binding by 'overriding' BUBR1 regulation. APC-CDC20 induction and chromosome separation is likely occurring in these cancers due to the identification also of *CDC20* upregulation and co-expression across subtypes. This in part explains potentially how, in the *CENPE*-upregulated cancers (such as ovarian cancer), despite receiving a taxane based treatment, these showed significantly reduced RFS and OS in patients with higher or lower *CENPE* expression.

Currently, taxanes only target the role of CENPE microtubule formation of microtubule plus end elongation by stabilising the combination of α and β tubulin. However, this does not affect its role within the dysregulation of SAC and chromosome separation (Kung *et al.*, 2014). Further to this increased *CENPE* has also been observed to increase resistance to paclitaxel in ovarian cancer (Chong *et al.*, 2018).

3.7.3 STIL – A centriole assembler

The centriolar assembly protein, STIL, is a key regulator of the mitotic spindle checkpoint monitoring chromosome segregation. This indirectly links the predicted GO functions of STIL to CENPE. It has also been shown that direct binding of APC/CDC20 mediates STIL ubiquitination (Arquint and Nigg, 2014). Previously, it has been observed that mutations in *STIL* promote microcephaly due to incorrect centriole formation (Patwardhan *et al.*, 2018).

STIL deregulation leads to both centrosome instability and the appearance of supernumerary centrosomes, leading to mis-orientated spindles and chromosome instability. This has been predominantly observed in poorer-prognosis cancers such as lung, colon, and prostate cancers with an increase metastatic potential also being observed (Patwardhan *et al.*, 2018). Because of this, *STIL* poses a novel treatment target in these cancer subtypes.

Role of STIL in breast, ovarian, and prostate cancers

In normal ductal epithelium, *STIL* has been found to be expressed at low levels (Kasai *et al.*, 2008). *STIL* has also been previously identified as being dysregulated in gastric (Wang *et al.*, 2019a) and pancreatic cancers (Kasai *et al.*, 2008), but data presented in this thesis is the first reporting increased expression in breast, ovarian, and prostate cancer subtypes.

STIL is regulated by the spindle checkpoint control APC/C and CDC20 (Arquint and Nigg, 2014). Both of these are important targets for CENPE-mediated checkpoint progression and chromosome separation. Despite STIL having functional links to CENPE, its overexpression is unlikely to be a driver across all cancers. *STIL* increased expression is not significantly associated with an alteration individually in RFS or OS in most of the cancer subtypes studied here, with a significant survival reduction only being observed in breast cancer survival data (sections 3.4.1 and 3.4.2). This could simply be due to there being smaller survival data sample sizes for ovarian and prostate cancers compared to that of breast cancers but this is unlikely. *STIL* expression has been associated with increased cancer metastasis in lung cancer, ovarian cancer, and prostate cancer previously (Patwardhan *et al.*, 2018). *STIL* is likely to be more effective in combination with

CENPE treatment due to the interlinked roles that they pose in microtubule dynamics rather than a single treatment targeting *STIL* alone.

STIL promotes inconsistent centrosome development and loss of spindle organisation

STIL co-expression modules were identified as being significantly associated with centrosome formation and spindle organisation. Despite being identified as being involved in centrosome formation mechanisms, *STIL*'s direct role is yet to be fully determined, especially in cancers. Previously, *STIL* has been reported to promote procentriole formation which leads to increased centrosome maturation (Ohta *et al.*, 2014). The procentriole is an important microtubule nucleation site as well as providing a site for microtubule orientation activities (such as those of *CENPE*). The development of these structures is vital for normal cell proliferation and for the maintenance of chromosome stability and alignment. As discussed earlier for *CENPE*, the loss of correct microtubule spindles can promote genomic instability through errors in chromosome separation due to misalignment of the chromosomes at the centromere. This can cause copy number alterations and aneuploidy. *STIL* co-expression networks were shown to be significantly associated with spindle organisation in breast, ovarian, and prostate cancer subtypes in this study. Increased centrosome formation has been identified to promote genomic instability and aneuploidy also (Hollander and Fornace, 2002; Nigg and Raff, 2009). The presence of multiple, supernumerary, centrioles can lead to a number of centrosome-related effects, such as tetraploid cells due to uneven chromosome segregation (Nigg and Raff, 2009). Spindle organisation is also dependent on correct centrosome formation, the number of centrioles and their duplication (D'Assoro, Lingle and Salisbury, 2002). By the G2-M phase, centrosome duplication is complete in a normal cell cycle and there are only two centrosomes present in the cell, providing correct cell polarity and chromosome separation. *STIL* overexpression has been observed to increase the number of centrosomes leading to genomic instability via increased centriole formation Figure 41 (Castiel *et al.*, 2011; Leda, Holland and Goryachev, 2018).

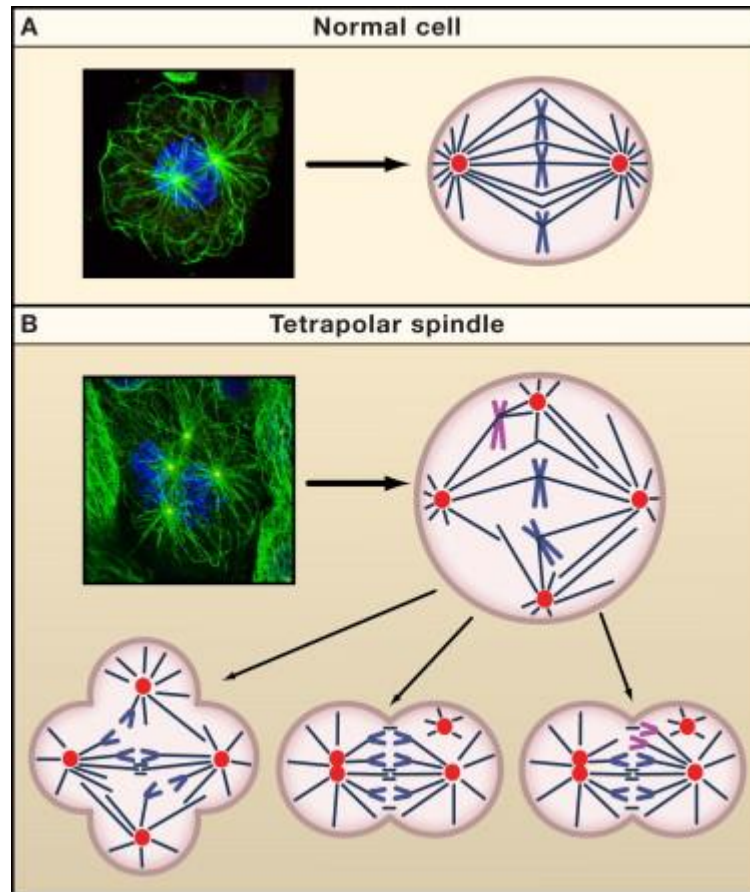


Figure 41. Centrosome duplication and genomic instability.

A) normal centrosome formation promotes even separation of chromosomes and genomic stability. B) four centrosomes (tetrapolar) are formed which leads to uneven separation of chromatids. From Nigg & Raff., 2009.

Importantly for centrosome formation, STIL has been found to localise to the protein matrix of the pericentriolar material (PCM) surrounding the centrioles/centrosome. The PCM has two major roles in ensuring normal genomic stability; the first is to nucleate microtubules, the second; is to organise microtubules as MTOCs (Varadarajan and Rusan, 2018). Overexpression of *STIL* has been found to promote increased centriole formation and maturation by formation of a multiple procentrioles for centrosome duplication (Vulprecht *et al.*, 2012; Tang *et al.*, 2011; Ohta *et al.*, 2014). In this action *STIL* interacts with *CENPJ* (CPAP) directly to induce procentriole formation around the parental centrioles (Tang *et al.*, 2011). This is initiated via *PLK4* and *CEP135* in G1-S phase moving to the parental centriole (Figure 42) (Kleylein-Sohn *et al.*, 2007). Once this has begun *SAS6*, *CEP135*, *CENPJ* in S-G2 phase promote procentriole formation with *STIL* to promote centriole elongation and maturation (Holland, Lan and Cleveland, 2010).

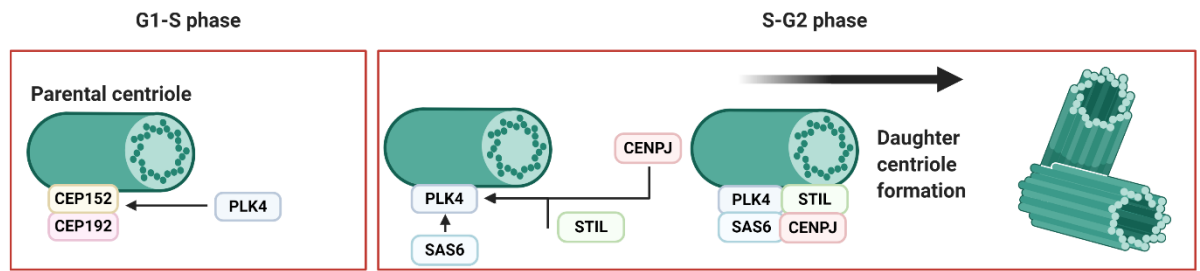


Figure 42. *STIL* promotes centriole formation in the S-G2 phase prior to G2-M phase transition.

STIL localises to the parental centriole where it aids in the formation of the daughter centriole in the transition of S phase to G2 phase of cell cycle.

In summary, *STIL* has been found to recruit *SAS6* to the centrosomes. *PLK4* is a key centriole duplication regulator that has been shown to promote tumorigenesis by phosphorylating *STIL*, thereby promoting *CENPK* binding and procentriole formation (Moyer and Holland, 2019). Despite their roles, *PLK4* and *CEP135* were not found to be significant in the cross-cancer analyses. *CEP135* was found to be significantly upregulated in breast cancers only and significantly downregulated in ovarian and prostate cancer subtypes (appendices Table 80). This was also observed for *CENPJ*, which was also found to be significantly upregulated and co-expressed in *STIL* modules. *CENPJ* was also identified as being significantly upregulated in invasive histological breast cancers and highest expression observed in the more invasive molecular subtypes. For ovarian cancer subtypes, the highest expression of *CENPJ* was observed in invasive subtypes also (serous and endometrioid) (appendix Table 81). Finally, for prostate cancer subtypes, *CENPJ* expression data was not present due to its absence from the microarray platforms used. This is one limitation of the array used to analyse prostate cancer tumours and *CENPJ* analysis of prostate tumours therefore requires future work.

STIL directly interacts and recruits *CENPJ* to the centrioles for centrosome replication.

Importantly, *CENPJ*, like *STIL*, is highly expressed in the G2-M phase and their proteins regulate *PLK1* activity (Takeshita *et al.*, 2019; Garcez *et al.*, 2015). *PLK1* was upregulated in the cross-cancer analysis across subtypes; Histological ($\log_2FC = 0.54$, $P < 0.01$), Molecular ($\log_2FC = 0.67$, $P < 0.01$). *PLK1* expression was identified across not only breast cancer subtypes but also ovarian and prostate cancers. Increased expression of *PLK1* is associated with *TP53* inactivation in G2-M

phase (Takeshita *et al.*, 2019). *PLK1* was identified to be co-expressed in networks with *STIL* and *CENPJ*.

The expression of *STIL* and other related genes involved in centriole formation (*PLK1*, *CENPJ*, *BUBR1*, and *CDK1*) identified, are predominantly seen in that of the G2-M phase transition. *HMMR* is also highly expressed and functional in this stage with *CENPE* also. This corresponds to findings that MTOC activity is peaked in late G2 phase and matured centrosomes are formed (Varadarajan and Rusan, 2018). Therefore, the centrioles would likely be formed into mature centrosomes by *STIL* facilitation and microtubule organisation (also part of *CENPE* function) at polar ends of the cell disrupted (also part of *HMMR* function). Whether the number of centrosomes in cancers here are abnormal requires further investigation but proposes a study to clarify *STIL* further.

Overall, both the loss of spindle organisation and centriole supernumerary promotes unregulated and non-uniform tissue development (Royer and Lu, 2011; Pease and Tirnauer, 2011).

Interestingly, this has also been observed to occur with *HMMR* expression (Tolg *et al.*, 2010). This is because there is decreased organisation from the centrosomes (MTOCs) promoting incorrect cell polarity, and as such proliferating cells to grow uncontrolled masses (tumours) (Basto *et al.*, 2008).

[STIL aids G2-M phase transition via SHH pathway](#)

STIL has been reported to act on the sonic hedgehog (SHH) pathway, with *STIL* overexpression leading to an increased SHH pathway signal transduction and *STIL* inhibition reducing cell proliferation via SHH (Sun *et al.*, 2014). This pathway has been observed to be concentrated within the cilia of cells present in epithelial luminal cells and basal cells (Higgins, Obaidi and McMorrow, 2019). It has been reported that basal cells have a higher numbers of cilia than luminal cells in breast cancer (Menzl *et al.*, 2014). Disruption to cilia SHH pathway has been observed in tumorigenesis (Nigg and Raff, 2009).

In the SHH pathway (Figure 43), STIL interacts with SUFU which acts as a signal suppressor of the SHH pathway on GLI. Binding of STIL to SUFU releases GLI1 suppression (Sun *et al.*, 2014). This allows non-active GLI1 (GLI) to move to the primary cilium and releases GLI-A (GLI1 active) from KIF7 which causes the accumulation of non-active GLI1. It is currently unknown what processes causes the change of the GLI1 protein from its non-active to its active form at the tip of the primary cilium. This facilitates downstream gene expression of GLI1s transcription factor roles which promotes cell proliferation (Yao *et al.*, 2019) and also metastatic potential (Zhang *et al.*, 2016a).

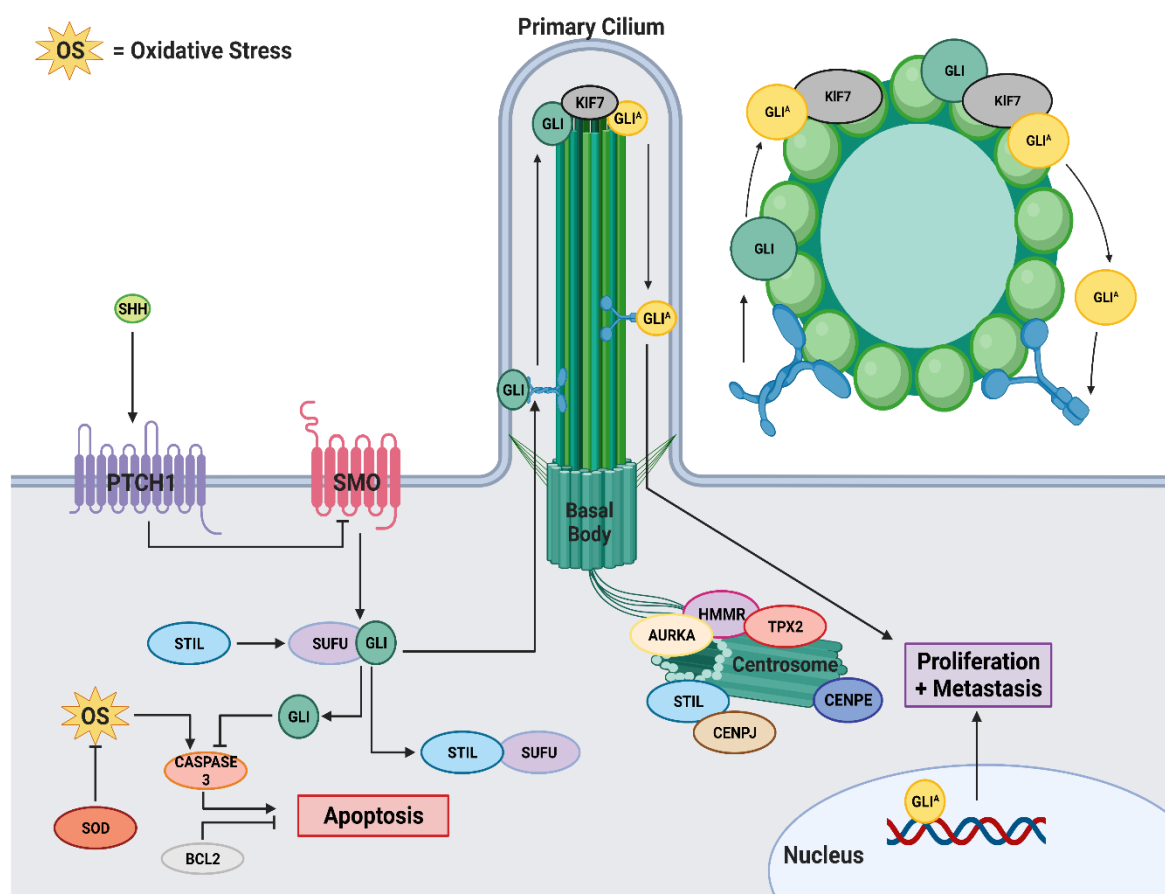


Figure 43. Proposed mechanism of STIL.

STIL acts on the SHH pathway to induce via SUFU release of GLI. This causes GLI movement to the tip of the primary cilium to release KIF7 bound GLI-active (GLI^A). This promotes cell proliferation prior to G2-M phase by increasing the expression of proliferation related genes. Additionally STIL also acts of promoting centriole duplication which can cause the abnormal formation of multiple centrioles. This promotes genomic instability via supernumerary.

In addition to SHH pathway activation, K-RAS (encoded by the *KRAS* gene) has been observed to enhance STIL association with SUFU (Kasai *et al.*, 2008). K-RAS has been found to increase *GLI1*

transcription by increasing STIL binding to SUFU thus releasing further *GLI1*. *KRAS* gene expression was not identified as significantly altered in cross-cancer analysis. It was, however, identified as being significantly upregulated in breast cancer subtypes, with the highest expression observed in invasive subtypes (IDC, ILC, HER2, and Basal). This was also seen in ovarian cancer subtypes. Prostate cancer did not show significantly altered *K-RAS* expression in the subtypes analysed. *GLI1* expression was not identified as being significantly upregulated in cross-cancer analysis, contrary to previous findings in other cancers. For example, *GLI1* has been reported to be upregulated in prostate cancers (Yang *et al.*, 2017). In this study, prostate cancer subtypes showed *GLI1* expression to not be significantly differentiated in cross-cancer analysis or in the individual subtype analysis. This was also observed in breast and ovarian cancer subtypes. However, *GLI1* overexpression in the G2-M phase of the cell cycle has been found to inhibit cell cycle progression (Galvin *et al.*, 2008).

As other genes discussed in this study are linked functionally to the G2/M phase of the cell cycle, it would in part explain the observed down regulation of *GLI1* in these cancers. *GLI1* release from SUFU has been observed to promote Cyclin B (*CCNB1*) and Cyclin dependent kinase 1 (*CDK1*) expression (Cai *et al.*, 2016). This has been observed to occur at the end of S phase prior to G2 phase with highest expression of *CDK1* and *CCNB1* observed (Hochegger, Takeda and Hunt, 2008). This suggests a potential 'switch' in these cancers where *GLI1* mediated transcription occurs via STIL binding to SUFU. Once sufficient *CDK1* and *CCNB1* expression has been achieved *GLI1* can be ubiquitinated and STIL can move to its centrosome functions in the G2/M phase. This explains why, in these cancers, there is an observed downregulation of *GLI1* in breast, ovarian, and prostate cancer subtypes contrary to previous finding in related studies. Both *CCNB1* and *CDK1* were observed to be significantly upregulated in cross-cancer analysis; histological breast, ovarian tissue, and prostate Gleason grade $\log_2FC = 1.48$ ($P < 0.01$) & $\log_2FC = 1.17$, ($P < 0.01$), Molecular breast, ovarian epithelial, and prostate Gleason score $\log_2FC = 1.78$ ($P < 0.01$) and $\log_2FC = 1.58$ ($P < 0.01$) for *CCNB1* and *CDK1* respectively.

In addition to *CDK1* and *CCNB1* expression, the G2-M phase progression also requires *E2F* expression. *E2F* was also observed to be significantly upregulated in cross-cancer analysis in subtypes. Histological breast, ovarian tissue, and prostate Gleason grade group ($\log_2FC = 0.65$, $P < 0.01$), Molecular breast, ovarian epithelial, and prostate Gleason score ($\log_2FC = 0.82$, $P < 0.01$). This suggests that the cancers in this study are associated with alterations in the G2-M phase transition signal explaining the loss of *GLI1* expression despite increased *STIL* expression.

Despite this finding, *STIL* is potentially an important marker and target for inhibiting G2-M phase progression. Further analysis is required to determine whether the SHH pathway plays any role in these cancers and the level of their progression.

STIL as a hub gene target in breast, ovarian, and prostate cancers

The role of *STIL* as a hub gene, compared to *HMMR* and *CENPE*, is more difficult to fully determine. This is primarily due to the paucity of research into its role in cancers. The *STIL* protein does function in centrosome formation and its disruption promotes genomic instability via supernumerary centrosomes and alterations in cell polarity. *STIL* has also been observed to function in releasing SHH pathway restriction by *SUFU* allowing *GLI* related gene transcription and cell cycle progression from G2-M phase Figure 43. This has identified that in this study a potential G2-M phase transition signal based on the co-expression and identification of further G2-M phase marker genes (*CDK1*, *CCNB1*, and *E2F*). This was identified with downregulation *GLI* genes which have been found to be a target of *STIL* prior to G2/M phase and downregulated in transition to G2/M. Together with *HMMR*, and *CENPE* treatment of *STIL* would provide a potentially effective target. This is due to its overlapping roles with *HMMR* and *CENPE*. Independent treatment of *STIL* would likely not be beneficial as its high expression with the exception to breast cancers shows no improved survival.

3.7.4 Combined mechanism of action

All three of the hub genes show a potential overlap of function in that they contribute to the promotion of cell-cycle transition specifically the G2/M phase. In summary extracellular and/or

intracellular HMMR can activate proliferation related pathways (PI3K/AKT or MAPK/ERK). This could also be in combination with STIL release of SUFU regulation promoting *CCNB1* and *CDK1* upregulation helping to promote mitosis. HMMR also recruits AURKA and TPX2 to the MTOC's which initiates microtubule nucleation again promoting cell cycle progression to mitosis. STIL promotes multiple centriole formation (centriole supernumerary) and subsequently multiple MTOCs where CENPE at the MTOCs potentially aides in microtubule nucleation. CENPE binds the K-fibres to the kinetochores which causes incorrect chromosome alignment due to loss of cell polarity (initiated both by HMMR and STIL). Aberrant and premature chromosome separation due to misalignment of chromosomes and loss of regulation of SAC/APC leads to premature separation of chromosomes (Figure 44). This causes genomic instability and aneuploidy which can promote more aggressive recurrence (treatment resistance) (Zhang *et al.*, 2017) and shorter RFS (Correa *et al.*, 2020; Gemoll *et al.*, 2015) and OS (Gemoll *et al.*, 2015; Zhang *et al.*, 2021). This means that these three hub genes could potentially be useful biomarkers representing this cell cycle transition. A combined targeted treatment of these hub genes may also be affective across these cancers and subtypes. A summary of the functions of HMMR, CENPE, and STIL is shown in Figure 44.

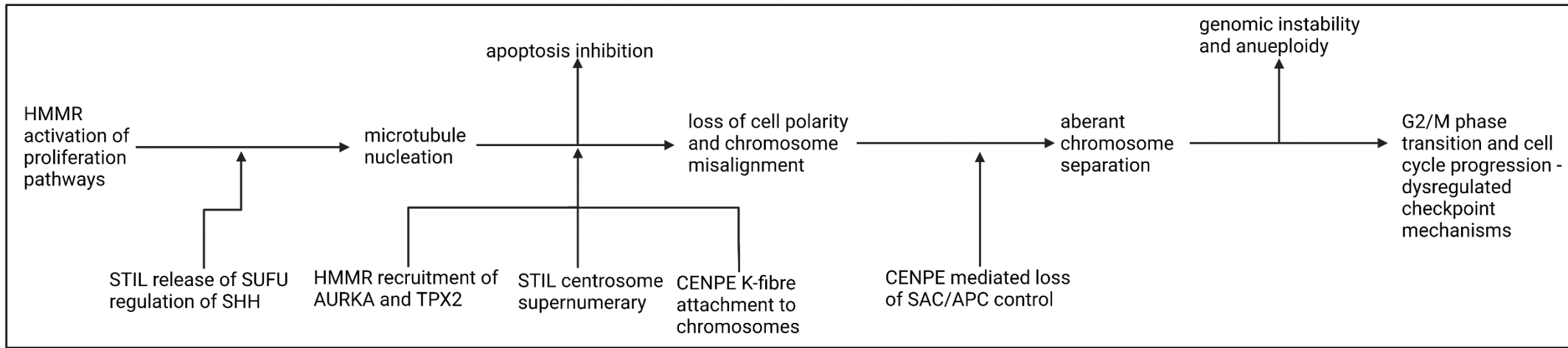


Figure 44. Summary of the proposed effect of hub genes overexpression as encoded proteins HMMR, CENPE, and STIL in the development and progression of breast, ovarian, and prostate cancer.

Following the activation of proliferation related pathways which can occur either by extracellular HMMR or by intracellular HMMR. HMMR recruits AURKA and TPX2 to the MTOCs of centrosomes where CENPE and STIL are also present. STIL causes the formation of multiple centrosomes (supernumerary) leading to multiple sites of microtubule nucleation and multiple K-fibre binding to centrosomes by CENPE leading to loss of cell polarity and incorrect chromosome alignment. Insufficient binding of centrosomes by CENPE can cause a loss of regulation of the SAC/APC complex leading to aberrant chromosome separation causing genomic instability and aneuploidy. This process promotes overall G2/M phase transition through loss and dysregulation of checkpoint mechanisms.

SAC: spindle assembly checkpoint; APC: anaphase promoting complex; SHH: Sonic Hedgehog pathway

Chapter 4. Discussion

The results presented in this thesis represent a novel approach to the identification of new genes of interest for three of the most common cancers. From a literature review of the commonly used classification systems for breast, ovarian, and prostate cancer it was possible to create a strategy to analyse and compare cancer tumour subtypes. In Begg *et al.*, 2017, it was highlighted that basal breast cancer subtypes are similar in aetiology to serous ovarian cancer.

In this chapter the results of each aim will be discussed.

[Aim 1: Develop a method and a workflow to integrate publicly available microarray gene expression data from breast, ovarian, and prostate cancer subtypes.](#)

A novel cross-platform microarray data integration method was developed that identified DEGs and hub genes that are commonly altered in breast, ovarian and prostate cancers. As a consequence of this approach, it identified cross-platform batch effects that arise from the different separate platform designs. The ability to integrate multiple microarray datasets allowed for a significantly increased power over individually analysed cancer datasets. As has been shown, this approach provides a novel method for the identification of cross-cancer hub genes that would otherwise be missed with lower samples sizes, due to the ability to identify more subtle changes in expression.

The integration of microarray data from multiple sources has been demonstrated previously but did not identify a cross-platform batch effect. These prior studies have either combined the datasets and normalised the data en masse (as one larger dataset), or normalised the datasets individually and then combined afterwards, without checking for cross platform effects which would affect the accuracy of any direct comparisons made (i.e. cancer compared to normal). This approach would likely introduce potential bias with the potential for misleading results due to the underlying batch effect. In this study, a cross-platform batch effect was identified within combined microarray datasets and an empirical bayes method (ComBat) used which was

originally designed to remove batch effects within individual studies (Johnson, Li and Rabinovic, 2007; Leek *et al.*, 2012). It was determined that this could be applied to reduce/remove batch effects between studies that had already been normalised but which also allowed data to retain biological differences (variance). Indeed, as a validation of the approach used, the integrated data sets identified changes in genes that were expected to be upregulated or downregulated in each of the cancers.

One of the main benefits of integrating multiple microarray datasets is the increase in sample size which increases the ability to identify novel genes and gene signatures that would otherwise not be found in individual studies with smaller sample sizes (which are often <100 samples). This effect has been seen in other microarray studies where increased sample size allowed the identification of gene signatures that would have otherwise been missed (Xu *et al.*, 2008; Xu, Geman and Winslow, 2007; Stretch *et al.*, 2013; Creighton, 2007; Kim, 2009). This approach also beneficially reduces the effects of study-specific bias, giving more robust results (Kim, 2009) and, as identified here, identifies observable effects in other cancers. For example, PAM50 genes were identified in the breast cancer subtypes analysed here and many of these changes in gene expression were found to be similarly altered in ovarian and prostate cancers. This suggests that these genes may be useful for classification of these cancers. Currently, prostate tumours are classified into luminal and basal subtypes, similar to the breast cancer molecular subtypes (Zhao *et al.*, 2017; Grist *et al.*, 2020), but this is not the case for ovarian cancers. The gene changes identified here may therefore provide a 'truer' method of molecular classification of ovarian cancers than the current epithelial subtypes.

As an alternative approach to data integration, a meta-analysis approach could have been used. However, integration beneficially allows the use of studies which contain only cancer samples, with no control tissue data, or vice-versa. For example, using integration, a study consisting of only normal breast tissue can be compared to a study consisting of only breast cancer samples making data relevant for use. This is beneficial over meta-analysis approach would require studies to have both normal and cancer samples. The majority of cancer studies that are publicly available

are asymmetric in their sample content, for example, as has been reported with datasets from hormonally driven cancers such as breast, ovarian, and prostate (Dawany and Tozeren, 2010). This study also found that there were many cancer samples from public datasets where data from normal tissue samples were not present (or vice versa). The use of data integration allowed these cancer or normal samples to be included – each would otherwise have been excluded using a comparison-based analysis.

It has also been identified in studies previously that integrating datasets rather than performing a meta-analysis, yields results that are more similar to those in the cancer literature where microarray analyses had been excluded (Dawany and Tozeren, 2010; Dawany, Dampier and Tozeren, 2011) such as results from RNA-Seq analyses. i.e. the results of a microarray study that has data integrated produces results that are similar to RNA-Seq studies despite being from different sequencing technologies. This means that results from this analysis are more likely to be reproducible in regard to other sequencing technologies. This could be analysed in future work by applying the integration methodology to other sequencing data (RNA-Seq data) to determine the limitations of the robustness of the integration. However, it is unlikely that the integration approach used here could correct for different types of sequencing technologies i.e. microarrays with RNA-Seq. This is because of the finding that microarrays from different manufacturers could not be combined accurately (i.e. Affymetrix and Illumina). But it is promising that the integration correction could be applied to RNA-Seq data from different studies.

The approach to the integration of microarray in this study has been found to have some caveats. First is the issue that not all platforms, even those from the same manufacturer, do not measure the same number of genes. This was adjusted for, in part, with the inclusion of a missing-value imputation which allowed for the retention of data from genes that would have otherwise been removed. However, this did lead to a reduction in the number of genes studied here, as not all 'missing' data could be accurately imputed. The inclusion of more samples will likely help to correct for these missing data points as the use of additional samples and subtypes also improves the accuracy of imputed data points, therefore allowing the retention of more genes in the final

dataset. The approach taken in this study to impute data and therefore decrease the number of genes lost from the final dataset is novel. In earlier studies, for example, in a study by Warnat *et al.*, 2005, in which raw microarray data sets were integrated by a simple averaging of array probe data, the majority of genes were lost as they were not present in all datasets (Warnat, Eils and Brors, 2005).

The second caveat is that not all platforms yield data that can be integrated despite each measuring global gene expression (mRNA) in a similar manner. Agilent arrays could not be combined due to being unable to accurately annotate the probes used. The Illumina bead array platforms could not be combined due to issues with corrections for the cross-platform batch effect in expression measurement. This is likely due to Affymetrix and Agilent array data being based on their measuring expression using sets of 25-nucleotide (Affymetrix) and 60-nucleotide (Agilent) probes per gene, whereas Illumina utilises multiple copies of a single 50-nucleotide probe attach to silica beads. These differences in the core technology may be why it was not possible to resolve the technology-based batch effects in the Illumina arrays with other array technologies such as Affymetrix.

It was found that it was possible to correct for batch effects and to combine the Illumina bead array data sets similarly to the Affymetrix arrays used in this study i.e. the Illumina array data sets could be sufficiently combined with other Illumina arrays, but not with other manufacturers. This means the expression measured on some array technology is not comparable despite an attempted correction. This is also in concordance with other studies that have looked to compare different microarray technologies where they also found incompatibility with expression levels (Kuo *et al.*, 2002; Mah *et al.*, 2004). These studies did not look to correct for cross platform correction but instead looked to compare the resulting gene expression levels following analysis (identified a cross-platform batch effect). This finding shows that there is a limitation to the ability to integrate platforms. However, it is possible that with future work these platforms could be integrated with modifications to the method developed here with emphasis on the Agilent arrays which showed promising integration with Affymetrix despite issues with gene annotation.

Aim 2: Use this approach to identify differentially expressed genes, and of those, further identify novel candidate hub genes common across the cancer subtypes.

A large number of genes were identified as DEGs common across breast, ovarian and prostate cancers. Some of the genes identified were already known to be DEGs in at least one of the three cancers prior to this study, providing a validation for the approach used. Many of the DEGs were similarly expressed in all three cancers and subtypes, suggesting a potential common (conserved) gene/DEG signature. These 'cross-cancer' DEGs were also found to be associated with the same functional mechanisms in each cancer. The gene/DEG signatures were identifiable because of the large sample sizes obtained from data integration as discussed earlier. This is important and useful when identifying genes for the classification and use in prognosis as it provides additional reproducibility between cancers and subtypes.

In the analysis of breast histological, ovarian tissue location, and prostate Gleason grade groups 395 genes were identified as being significantly upregulated DEG's. In the analysis of the molecular breast cancer, ovarian epithelial, and prostate Gleason scores a total of 75 genes were identified as being significantly upregulated across the subtypes. This suggests that these genes may play a conserved role in a particular pathway or gene network. From both of these analyses, three novel cross-cancer DEGs were identified with a common mechanism of action within the G2/M phase of the cell cycle. This was observed in all the cancers and subtypes from GO analysis of the gene co-expression networks. Future work could be performed to silence/inhibit the upregulated genes in cancer cell lines that have elevated expression to determine their mechanism of action that is predicted from GO analysis of the co-expression network. The downregulated DEGs identified from both analyses were not analysed further due to the large number of DEGs identified. These downregulated DEGs are still relevant targets for further research as potential biomarkers or treatment targets. For example, reactivation (upregulation) of these genes might be expected to have adverse effects on tumour survival and these DEGs could therefore be novel tumour suppressors in these cancers. Studies have been successful in

reactivating downregulated tumour suppressor genes which has been identified as a promising approach for cancer treatment (Karpf and Jones, 2002; Beltran *et al.*, 2019; Kazanets *et al.*, 2016).

Aim 3: Determine whether the novel candidate hub genes identified, may be useful as potential biomarkers and/or treatment targets for these cancers.

The three novel cross-cancer DEGs identified from aim 2 were also found to be candidate hub genes in network analysis. The three DEGs identified were:

1. Hyaluronan Mediated Motility Receptor (*HMMR*), a HA receptor protein active in the G2 phase of the cell cycle.
2. Centromere Protein E (*CENPE*), a centrosome motor protein active in the G2 phase of the cell cycle.
3. Centriolar Assembly Protein (*STIL*), a regulator of the mitotic spindle checkpoint in the M phase of the cell cycle.

None of these genes had been identified in these three cancers/subtypes together prior to this study. The identification of these genes provides a novel treatment potential that is not limited to only certain cancer subtypes, but is common to all breast, ovarian, and prostate cancer classifications. There was also no evidence in the literature for the three hub genes being housekeeping genes. The targeting of these hub genes (in particular *HMMR*) should not affect normal cell functions/processes because if upregulation of these genes is targeted, the expression can be reduced to approximately normal concentrations reducing genomic instability and inducing apoptosis. This reduces the risk of inducing cell necrosis over apoptosis. This adds value for the three genes as novel targets for treatment and/or the targeting of the G2/M phase transition that they are involved in.

HMMR gene

HMMR was identified as a novel hub gene in all three cancers studied and in all subtype classifications. *HMMR* has previously been observed as being upregulated in triple negative breast cancer, where *HMMR* protein can be ubiquitinated by BRCA1 (Podo *et al.*, 2010). This is similar to

the findings here, where *HMMR* was found to be co-expressed with *BRCA1* in all breast subtypes, but this study has extended this to now include all ovarian and prostate cancer subtypes.

In the literature, CD44 is considered to be the primary HA receptor. Increased *CD44* expression has previously been found to positively correlate with that of basal cytokeratin and negatively correlate with expression of the luminal marker, *FOXA1* (Xu *et al.*, 2016). This is similar to findings in the breast cancers studied here, where the expression of *CD44* was significantly upregulated in invasive subtypes (IDC and ILC), but not in non-invasive subtypes (DCIS). However, contrary to previous findings in ovarian cancer, where increased *CD44* expression has been found to promote increased metastasis and reduced survival (Zhou *et al.*, 2019). In the ovarian cancers/subtypes studied here, *CD44* was found to be significantly downregulated in the serous subtype. *CD44* was not significantly differentially expressed in the other ovarian cancer subtypes analysed.

Interestingly, the serous subtype in fallopian tube tissue carcinomas have been observed to migrate to the ovaries (Labidi-Galy *et al.*, 2017). But *CD44* expression has been found to be low in serous subtype located to the fallopian tubes, suggesting a pre-invasive cancer stem cell phenotype (CSC) (Chene *et al.*, 2015). This may be what is being observed in this study in ovarian cancers, a pre-invasive signature. In prostate cancer, *CD44* expression was significantly downregulated across all Gleason scores and grade groups, whereas other studies have found that its expression was significantly upregulated (Iczkowski, 2010). This is the first study to identify *CD44* expression being significantly downregulated in ovarian and prostate cancers with *HMMR* expression significantly upregulated. However, the consistent upregulated expression of *HMMR* across the cancers may suggest that it is a potentially novel CSC marker due to its similar role in functions to CD44 and may 'replace' *CD44* in low expressing/downregulated cancers. This is important as the CSC phenotype is known to be related to chemotherapy resistance (Phi *et al.*, 2018), metastasis (Shiozawa *et al.*, 2013), and shorter recurrence time (Ayob and Ramasamy, 2018). This is potentially evident in the survival data presented in this study where reduced RFS and OS is observed across these cancers. This does suggest that further analysis is required to

clarify the role of CD44 and HMMR in breast, ovarian and prostate cancers to fully determine whether HMMR is the primary active HA receptor.

The identification of *HMMR* as the primary HA receptor in these cancers is important because it offers a potential resistance mechanism in these cancers to *CD44* inhibitors. This is because, even when *CD44* target knockdown/silencing is achieved, the inhibition of proliferation might not occur as the HMMR protein can activate many of the same mechanisms or pathways as CD44.

Extracellular HMMR can promote cell proliferation pathways in cancers by activation of RTKs to induce a RAS signalling cascade. Intracellular HMMR promotes MAPK pathway activation also promoting cell proliferation and acts as novel resistance mechanism. In addition, RAS-signalling directed treatments are also likely ineffective in these cancers. This is because intracellular HMMR can activate MAPK regardless of external RAS signal induction. This is evident from the observation that an increased expression of *HMMR* was found to be significantly associated with reduced survival in breast, ovarian, and prostate cancer patients. This study provides the first reported instance of *HMMR* being identified with reduced OS and RFS in breast, ovarian, and prostate cancers.

Taken together, both of the observations detailed above suggest that *HMMR* plays an important role in breast, ovarian and prostate cancer progression and metastasis and has been observed with lung cancers, where it is associated with increased proliferation and metastasis (Cai *et al.*, 2019b). This is consistent with findings from the network analysis where it was found that *HMMR* is associated with the G2/M phase transition (as are the other two genes, *CENPE* and *STIL*), something that would promote the cell cycle and hence cancer progression and invasion. The other genes (including *CENPE* and *STIL*) identified in the *HMMR* network co-expression module could also be used as potential biomarkers for G2/M phase disruption in these cancers and subtypes. For example, the genes that were first degree neighbours (i.e. predicted direct interactors) of *HMMR* could serve as potential biomarkers in these cancers to help prognostic diagnosis as many of these genes are also associated with the loss of regulation in the G2/M phase transition. Targeted treatment of *HMMR* alone could be beneficial in reducing cell

proliferation at the G2/M phase due to HMMR's role in the initiation of microtubule formation and alterations of the extracellular matrix (which aids in the promotion of metastasis).

An increased expression of *AURKA*, which was also identified in this study, would lead to altered ubiquitination of HMMR by BRCA1, most likely leading to increased microtubule nucleation.

Additionally, it is important to note that *AURKA* expression and activity peaks in the G2/M phase with *HMMR* (Marumoto *et al.*, 2002; Maxwell *et al.*, 2005). Targeted inhibition of *HMMR* would likely remove this microtubule initiation event as the HMMR protein is required for correct *AURKA* location to the microtubule organising centre (MTOCs).

From these observations, it is clear that HMMR acts as an important hub gene in all three cancers, with a mechanism that appears to be conserved and highly interlinked with CENPE and STIL functions. Increased *HMMR* expression is associated with reduced survival in all three cancers and subtypes likely due to its role in promoting G2/M phase transition through microtubule functions.

CENPE gene

A second novel hub gene identified in this thesis is *CENPE*, and this is the first study to identify it in multiple cancers and subtypes. *CENPE* was found to be co-expressed with other genes whose encoded proteins are involved with microtubule organisation and chromosome separation. It was also identified as being commonly co-expressed within the same gene networks as *HMMR*.

Importantly this highlights a shared function between these two genes with a predicted interaction. This interaction may not be directly *via* the binding of the two proteins, but instead through their common functional roles in microtubule dynamics. HMMR's role is in the initiation and facilitation of *AURKA* movement to the MTOCs, where CENPE helps coordinate spindle organisation following microtubule nucleation that has been facilitated by HMMR recruitment of *AURKA*. Furthermore, the expression of *CENPE* has been identified to peak in the G2 phase, similar to *HMMR* (and *AURKA*) with the protein accumulating around centrosomes (Zhu *et al.*, 2019).

CENPE has not been identified as a gene of interest in prostate cancer prior to this study but has been identified as being upregulated in triple negative breast cancer (Kung *et al.*, 2014) and taxane-resistant ovarian cancer (Ju *et al.*, 2009). However, *CENPE* expression has also been found to be decreased in taxane-resistant ovarian cancer cells (Chong *et al.*, 2018). Interestingly, the conflicting results from survival analysis here suggests that both these earlier findings may be correct and further research is required to fully clarify this. In RFS, of the patients that had received taxane treatment, those with high expression of *CENPE* had a significantly reduced survival (Figure 29). However, for OS, in patients that also received taxane treatment, low expression of *CENPE* had a better prognosis (Figure 32). Similar to the RFS and OS findings here, it has been observed that for ovarian cancers treated with platinum or taxanes, RFS is ~18 months compared to a longer OS of around ~44 months (du Bois *et al.*, 2009). This provides an interesting and novel treatment target for localised recurrence in high *CENPE* expression cancers. Prior to this study, in ovarian cancer, progression free survival (PFS) (similarly to RFS identified here) is reduced despite receiving taxane and platinum treatment (Yoshihara *et al.*, 2010) suggesting in RFS taxane and platinum treatment is less effective. The main difference between PFS and RFS being that RFS takes local resurgence and metastatic spread into consideration where PFS looks for regrowth of the cancer (locally).

Even with an effective taxane treatment, the role of *CENPE* in kinetochore binding to chromosomes will still promote genomic instability and cancer progression via an aberrant chromosome separation. The BUBR1 regulatory mechanism is insufficient to prevent cell cycle progression if *CENPE* is upregulated (Cleveland, Mao and Sullivan, 2003). As such, taxane treatment will not be effective as these primarily affect the microtubule nucleation mechanisms of *CENPE*, something that may provide an evolving resistance over time.

Despite *CENPE* posing a potential route to resistance (specifically ovarian cancers), treatment in breast and prostate cancers that have increased expression of *CENPE* still makes it a useful target for future treatment options. Targeting of *CENPE* could destabilise and prevent microtubule orientation and nucleation potentially stalling cell cycle. This would be more effective with

targeting with HMMR as the role of both genes in microtubule dynamics should cause a cascade effect at both, the initiation of this mechanism by preventing AURKA recruitment, and later orientation and kinetochore binding by CENPE.

STIL gene

The third novel gene identified here is *STIL*. *STIL* was found to be associated with genes involved in centriole duplication and as being important in monitoring chromosome segregation. Unlike the other two hub genes identified, *STIL* was found to be more of an outlier than *HMMR* and *CENPE*. Firstly, it has been found to primarily function in the M phase following G2 phase (where *HMMR* and *CENPE* function). Secondly, whilst *STIL* is an important hub gene in the majority of network analyses, as a treatment target, *STIL* would most likely be ineffective in prostate cancers. This is due to it not being associated with any significant reduction in survival (OS or RFS). As such, its role in these cancers is unlikely to be driving progression. It is still conserved across these cancers and is therefore likely functioning in mechanisms involving *HMMR* and *CENPE*. The common function is likely to be by promoting centriole supernumerary for microtubule nucleation to occur by *HMMR* and *CENPE* orientation functions. This would also provide multiple centrioles for *CENPE* binding for chromosome separation, increasing genomic instability and potentially leading to further mutations and potential resistance.

HMMR, CENPE and *STIL* combined gene targeting

The three genes were found to be highly significant in survival analysis emphasising their joint importance, and all three were also identified as novel hub genes in all the cancers and subtypes studied here. The identification of all three genes encoding proteins that are functional in the G2/M phase suggests that the remaining co-expressed genes for each of the *HMMR*, *CENPE*, and *STIL* modules could also be potential novel biomarkers linked to disruption in G2/M in these and other cancers. They could potentially be used to identify disruption of this cell cycle transition in further cancers as a novel classification and treatment target. With their observed common functionality and with all three being predominantly expressed in the G2/M phase, a combined target linked to all three genes would be the most promising novel target for treatment.

A targeted treatment directed at the hub gene *HMMR* alone would be beneficial in reducing cell proliferation at the G2/M phase. This would be expected to reduce downstream CENPE and STIL function due to destabilisation of initial microtubule formation. A combined treatment would additionally be effective in regard to the downstream mechanisms that occur following HMMR's role in microtubule initiation. As CENPE can promote resistance *via* other microtubule activities, a treatment that could target CENPE's kinetochore binding role would likely be more effective and pose more difficulty for resistance to occur. This would prevent chromosome separation and induce apoptosis.

Potential impact

The research presented in this thesis allows for the identification of hub genes not limited to these cancers. This methodology can be applied to further carcinomas of interest. The identification and use of hub genes as both biomarkers (diagnosis and prognosis) and treatment targets will provide far reaching treatment options particularly for resistant tumours.

Three genes/proteins have been identified as having consistently increased expression in breast, ovarian, and prostate cancers and overlapping functions in the promotion of cell cycle progression during the G2/M phase. The G2/M phase is particularly important in many cancers as it is a point at which the integrity of the cell's genetic material (DNA damage etc.) is determined. As such, genes/proteins that are involved in its progression, such as those identified in this study (*HMMR*, CENPE, and STIL) provide novel and attractive targets for treatment.

The expression of two of the three genes (*HMMR* and *CENPE*) could potentially be used as a novel 'gene expression signature' that could help determine treatment options as their expression is strongly associated with reduced survival. Their first-degree neighbours identified in the networks could be expanded upon and added to this gene signature to provide a novel signature across breast, ovarian, and prostate cancers and subtypes.

The identification of two novel genes that may contribute to the development of resistance to current treatments is relevant for clinicians in order to guide treatment choice. The use of taxanes

is widespread in the treatment of ovarian carcinomas. In this study, sub-groups of ovarian cancer individuals were identified. Some were found to have significantly lower OS despite receiving taxane and platinum treatments in combination. In individuals with lower levels of *CENPE* expression, taxane treatment was not as beneficial as those with higher expression as OS was significantly reduced, thus reducing the efficiency of the treatment overall. This is most likely because of the other roles of *CENPE* (chromosome separation) that are likely unaffected by taxanes. The use of *CENPE* as a biomarker to determine whether treatment resistance may occur to taxanes could be performed on biopsies from a breast, ovarian or prostate cancer regardless of subtype as these could be profiled for *CENPE* expression. Therefore, those with high expression of *CENPE* would be less likely to benefit from taxane treatment in the long term other than potentially extending recurrence free time. However, overall survival would not improve in high *CENPE* expressing cancers due to potential resistance occurring throughout the course of treatment. This is likely through the selection of low *CENPE* expression cancer cells surviving taxane treatment allowing for a more aggressive recurring cancer which reduces overall survival. An alternative treatment could be prescribed, that may otherwise not have been considered.

Instead of *CENPE*, HMMR would likely be a more appropriate treatment target for clinicians as it is a much clearer factor in these cancers. This is emphasised more so in high *CENPE* tumours in which taxane resistance is likely to occur. Additionally, the use of HA or CD44 inhibitors again would likely only slow the cancer progression as intracellular HMMR can likely still drive G2/M phase transition.

As a guide for clinicians the subtype specific genes identified here in addition to the three cross-cancer related hub genes can be used as a novel gene signature also. This could benefit in classification of difficult subtypes. One such example highlighted in this thesis was normal-like and luminal A breast which are frequently found to be highly phenotypically similar. This makes them difficult to differentiate and diagnose. In prostate cancers this is more evident with Gleason scoring which encouraged the adoption of the newer Gleason grade groupings. These subtype specific genes would provide an unbiased classification that would be independent of the

clinicians or technicians experience. For example, a biopsy that is collected from a tumour of unknown subtype, the subtype specific genes and expression signature found here could be used to initially classify the tumour in an unbiased manner. The three genes for *HMMR*, *CENPE* and *STIL* could guide timing of treatment as to when is best in order to maximise treatment efficiency and reduce the possibility of resistance. In summary *HMMR* is a potentially useful biomarker and treatment target in these cancers. *CENPE* is a potential biomarker and treatment target in high *CENPE* expression cancers. *STIL* is a biomarker but would likely not be a beneficial treatment target.

As a final point the *HMMR*, *CENPE*, *STIL*, which can be named the 'HCS three-gene signature', observed in this study provides both a potential set of treatment targets and a potential diagnostic tool or set of biomarkers that could also be explored in other cancers. As an example, lung cancer which is the most predominant cancer in males but also highly prevalent in females the HCS signature is significantly associated with reduced OS (appendices Figure 99). This is also found in gastric cancers associated with reduced OS (appendices Figure 100). This suggests the HCS signature may be useful in other cancers.

Chapter 5. Conclusions

A novel integrated data approach using R programming was developed that combines microarray tissue datasets from multiple breast, ovarian, and prostate cancers. These datasets encompass the majority and most frequently used classifications. Three novel hub genes (*HMMR*, *CENPE*, and *STIL*) were identified as being significantly upregulated across breast, ovarian, and prostate cancer subtypes. Two of the three hub genes identified (*HMMR* and *CENPE*) were important across these cancers with their increased expression being associated with significantly reduced survival times and the *STIL* gene possibly contributing in the overall functions of *HMMR* and *CENPE*. *HMMR* and *CENPE* were also co-related in their functions within the G2/M phase transition of cell cycle. These genes were therefore proposed as potential novel biomarkers and treatment targets that also indicate cell phase dysregulation. Based upon the current understanding of these genes functions, a mechanism of action for each of the genes was proposed. Genes were also identified that were subtype and cancer specific and these might prove useful for the development of novel classifications for tumours. These three novel hub genes, 'HSC three-gene signature', may also be potentially important in guiding treatment decisions and/or provide new treatment targets.

Chapter 6. Future work

The primary aim of this study was to identify novel genes for further research for treatment using an integrated data approach. One main aim for future work would be to determine the limits of the integration method when used with data such as RNA-Seq-derived datasets in order to determine whether a similar approach can be applied. This takes the application developed here beyond microarray studies. This would also be a further confirmation of the methodology. A second aim would be to determine whether the approach used here can be applied to data from other cancers, to identify additional novel hub genes.

The *HMMR*, *CENPE*, *STIL* (HCS three-gene) signature observed in this study provides a potential set of treatment targets that could also be explored in these or other cancers. One approach would be to assess the impact of knockdown experiments by targeting mRNA both individually and in combination in cell lines. A second approach would be to screen for small molecule inhibitors that inhibit the encoded proteins. Both these approaches could be initially performed in-vitro on cell lines with promising candidates taken forward to animal models.

It would be interesting to explore whether these three hub genes are mutated in tumours (for example whether mutations may be involved in the increased expression observed here), or whether underlying gene variants may be associated with elevated expression potentially using the TCGA database. Similarly, other epigenetic mechanisms that lead to increased expression could be explored.

Appendices



Figure 45. Volcano plot of breast cancer histological subtype IDC tumours.

Adjusted *P*-value (*Q*-value) and gene expression (*Log*₂ Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a *log*FC of 2 threshold.

IDC: Invasive Ductal Carcinoma



Figure 46. Volcano plot of breast cancer histological subtype ILC tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

ILC: Invasive Lobular Carcinoma

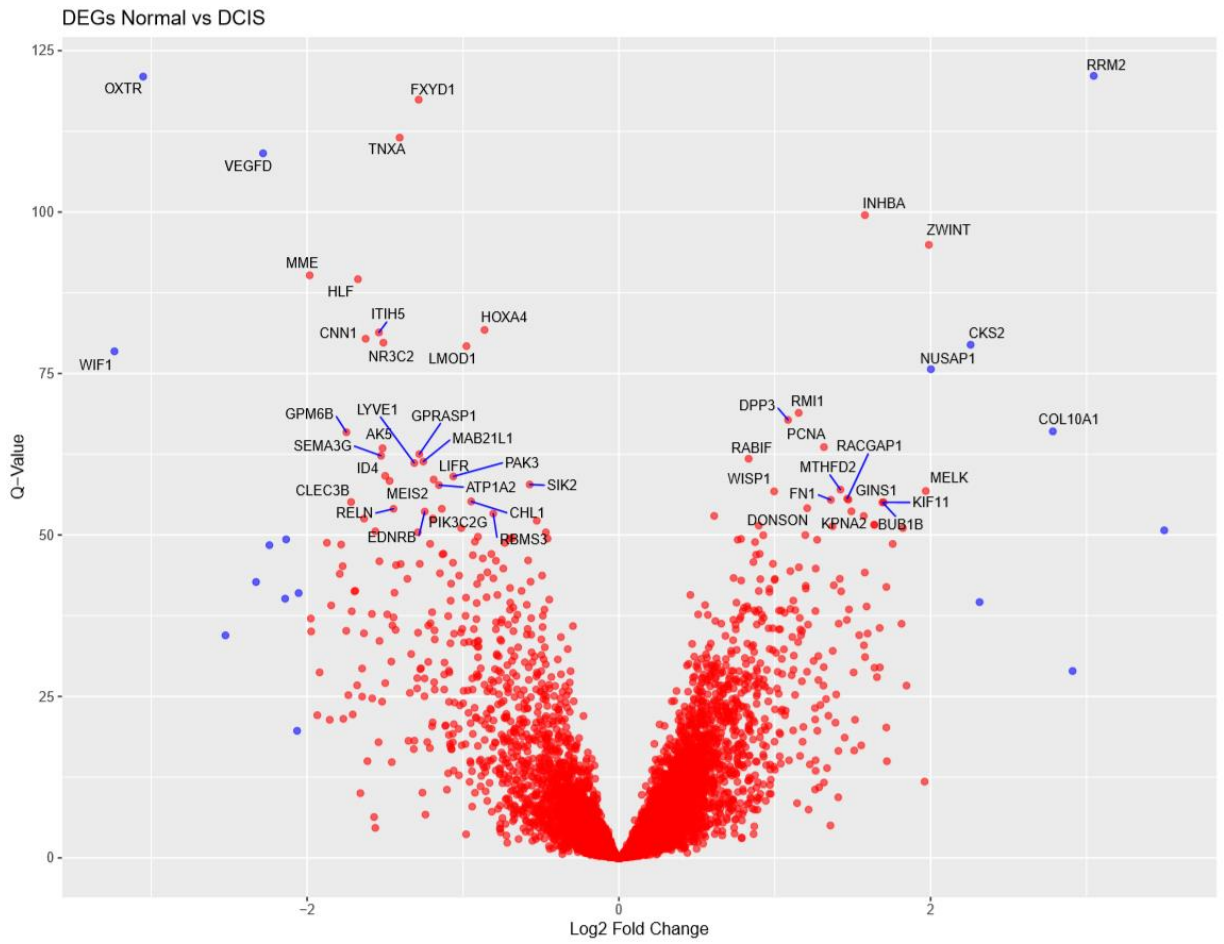


Figure 47. Volcano plot of breast cancer histological subtype DCIS tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

DCIS: Ductal Carcinoma in-situ

DEGs Normal vs Ovarian

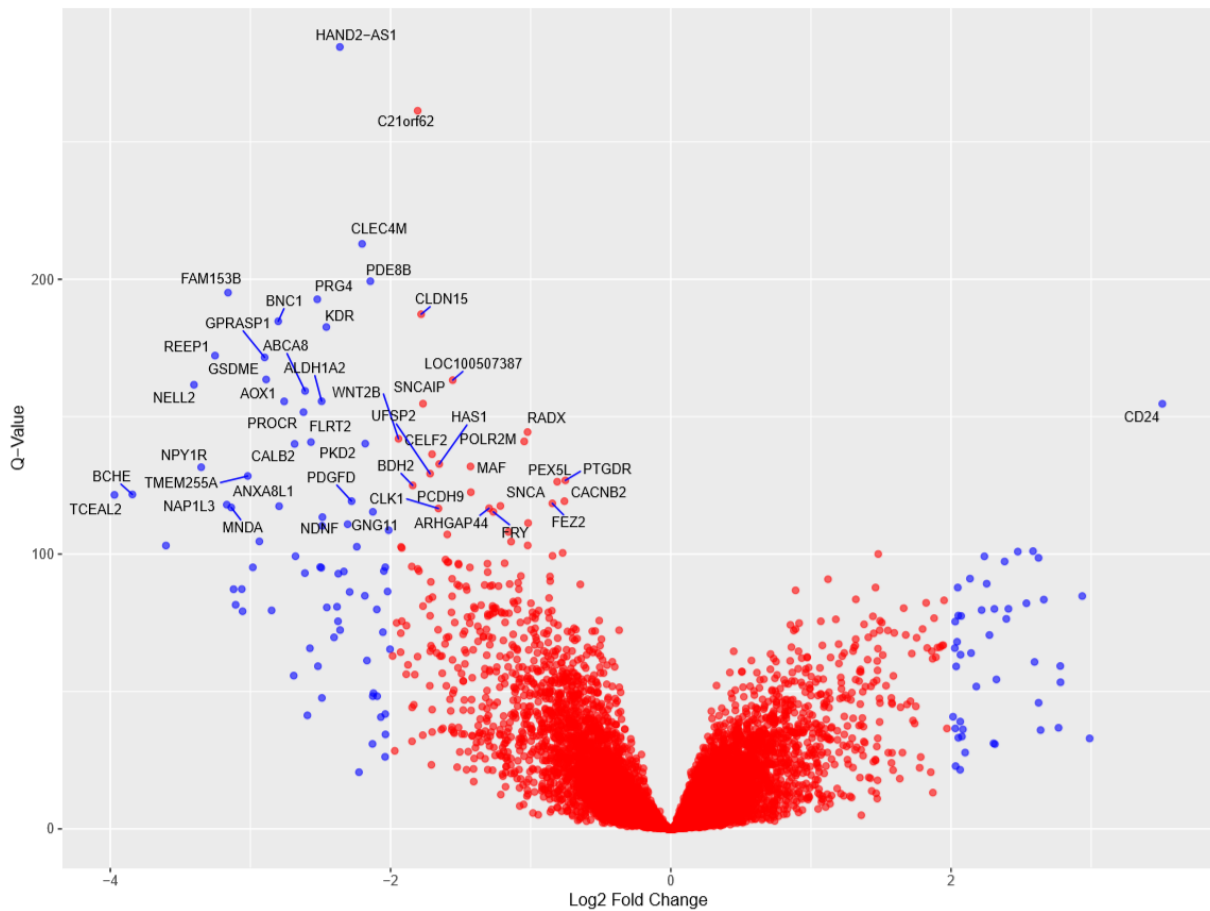


Figure 48. Volcano plot of ovarian cancer tissue location ovarian tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

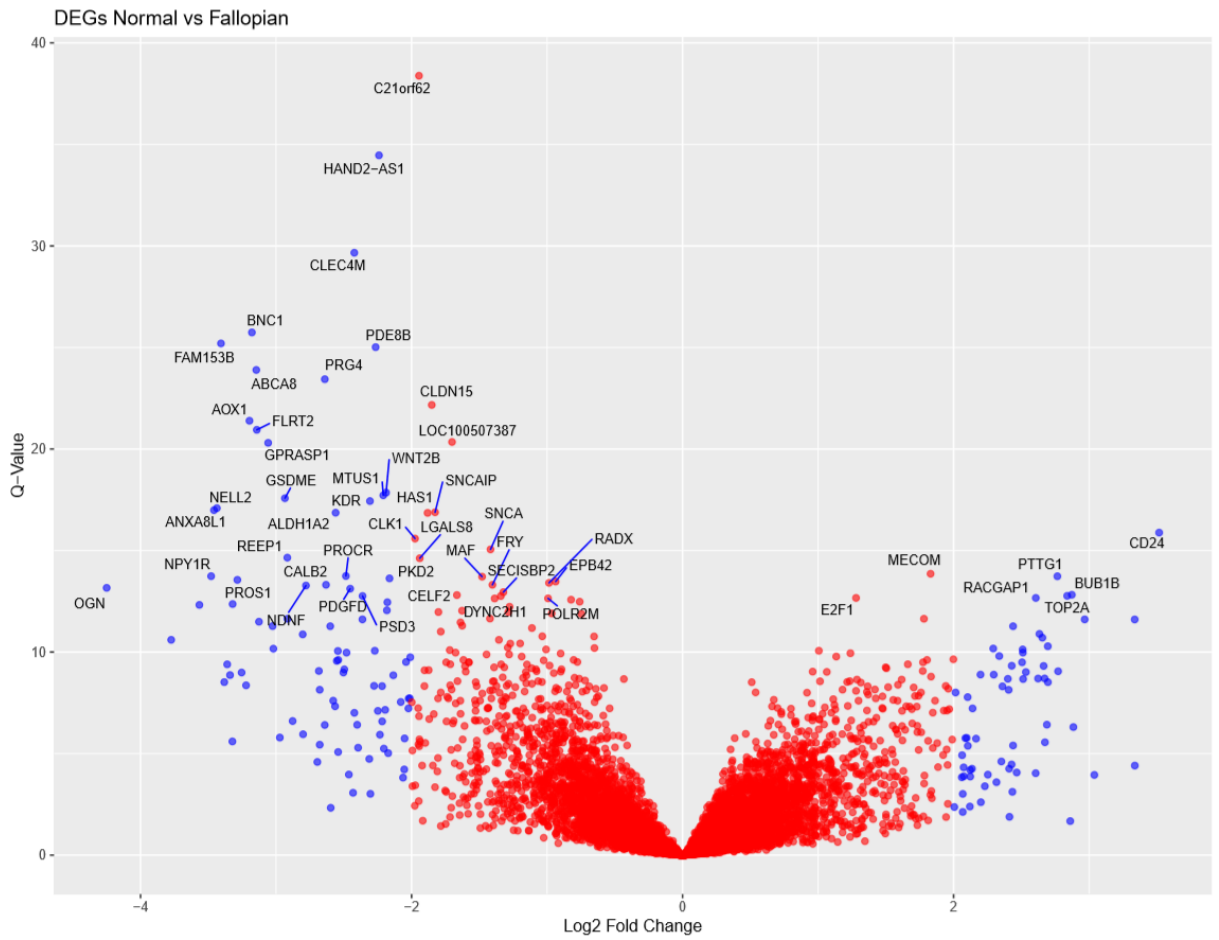


Figure 49. Volcano plot of ovarian cancer tissue location fallopian tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

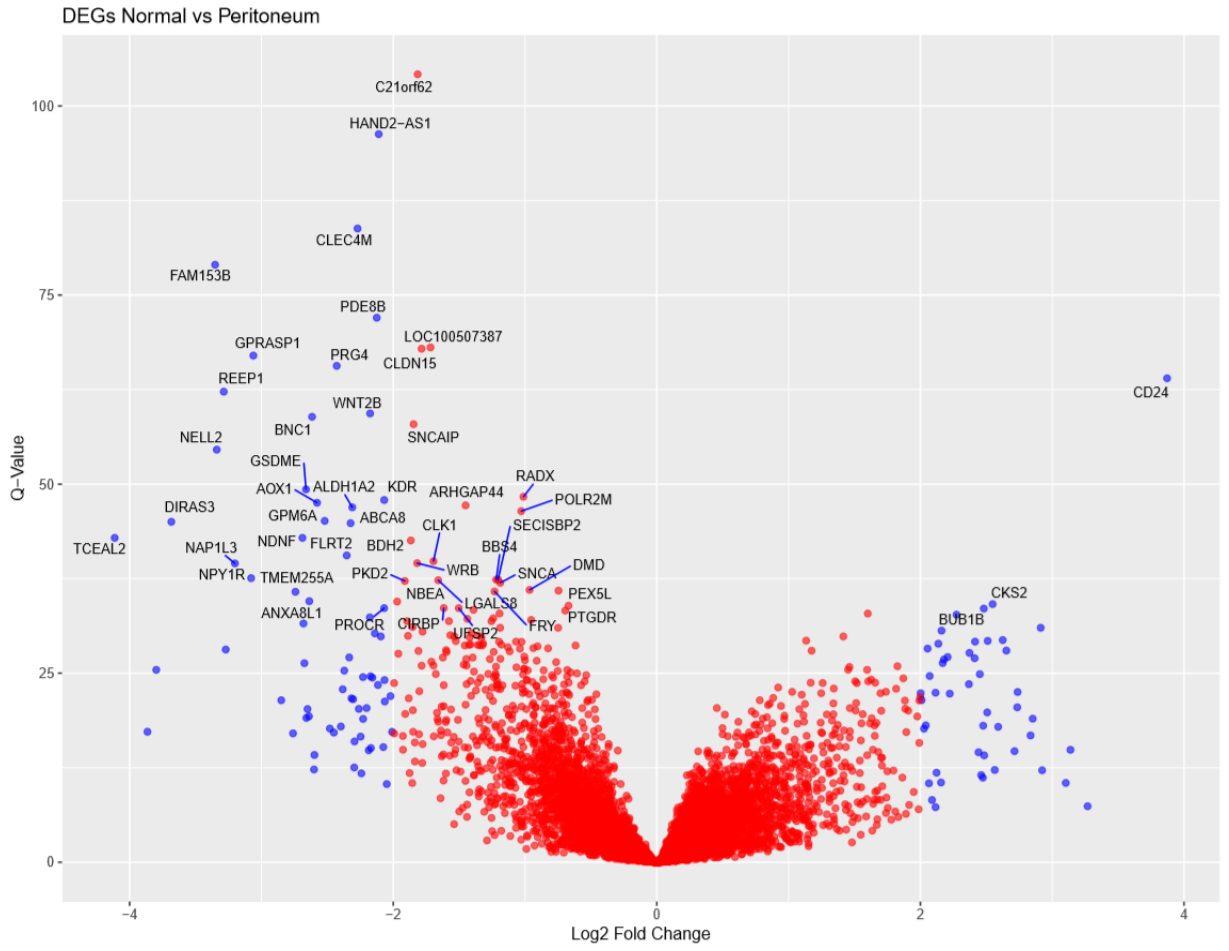


Figure 50. Volcano plot of ovarian cancer tissue location peritoneum tumours.

Adjusted *P*-value (*Q*-value) and gene expression (*Log*₂ Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a *log*₂FC of 2 threshold.

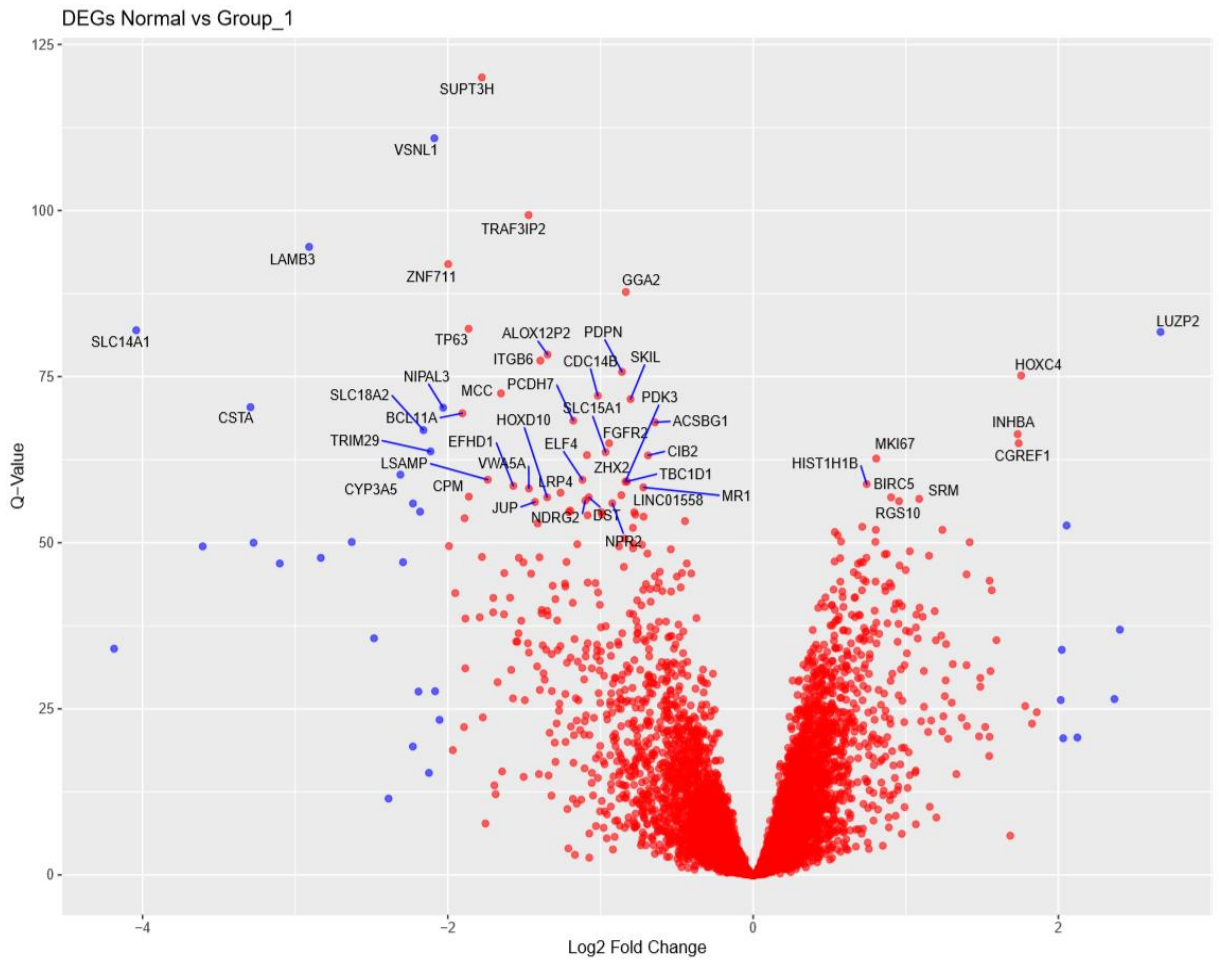


Figure 51. Volcano plot of prostate cancer Gleason grade 1 tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

DEGs Normal vs Group_2_3

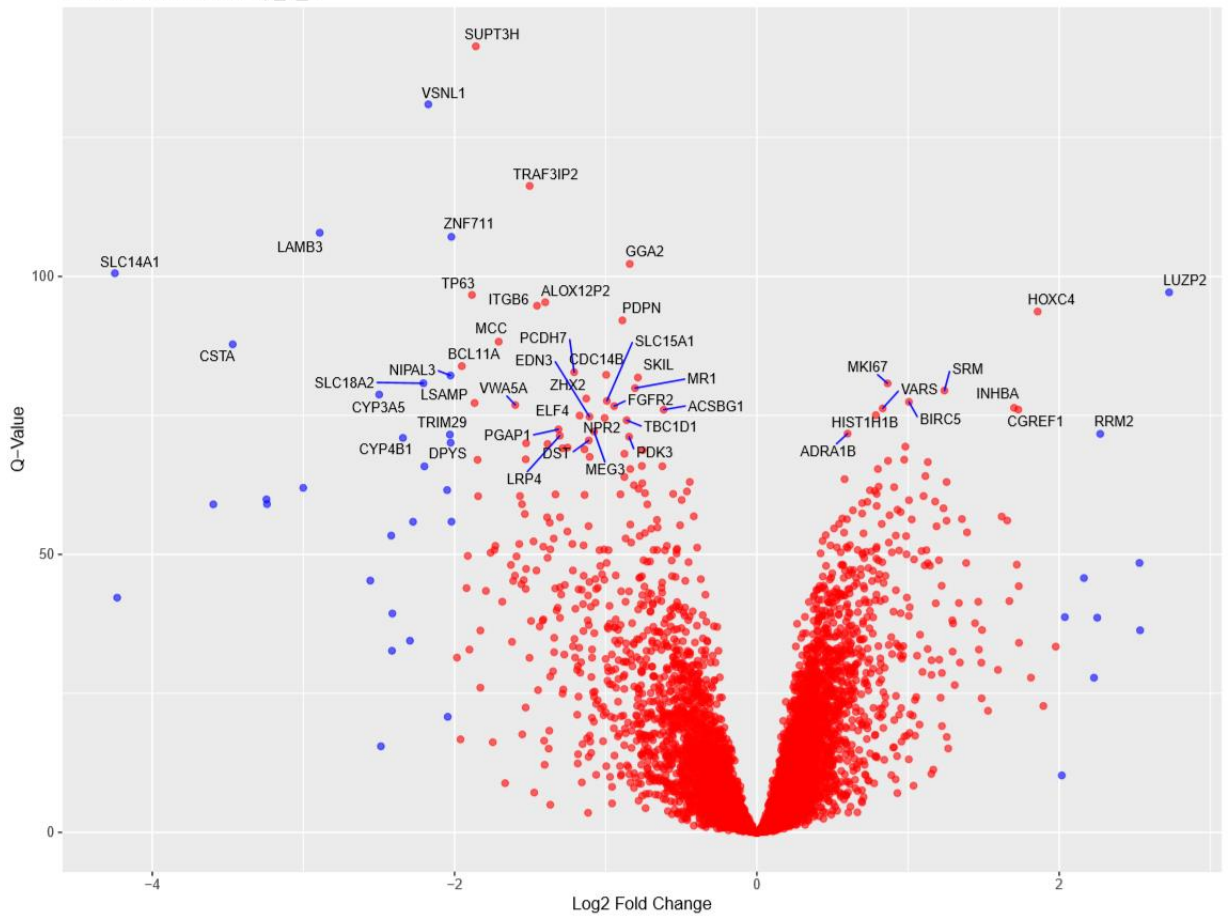


Figure 52. Volcano plot of prostate cancer Gleason grade 2 & 3 tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.



Figure 53. Volcano plot of prostate cancer Gleason grade 4 tumours.

Adjusted *P*-value (*Q*-value) and gene expression (Log₂ Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

DEGs Normal vs Group_5

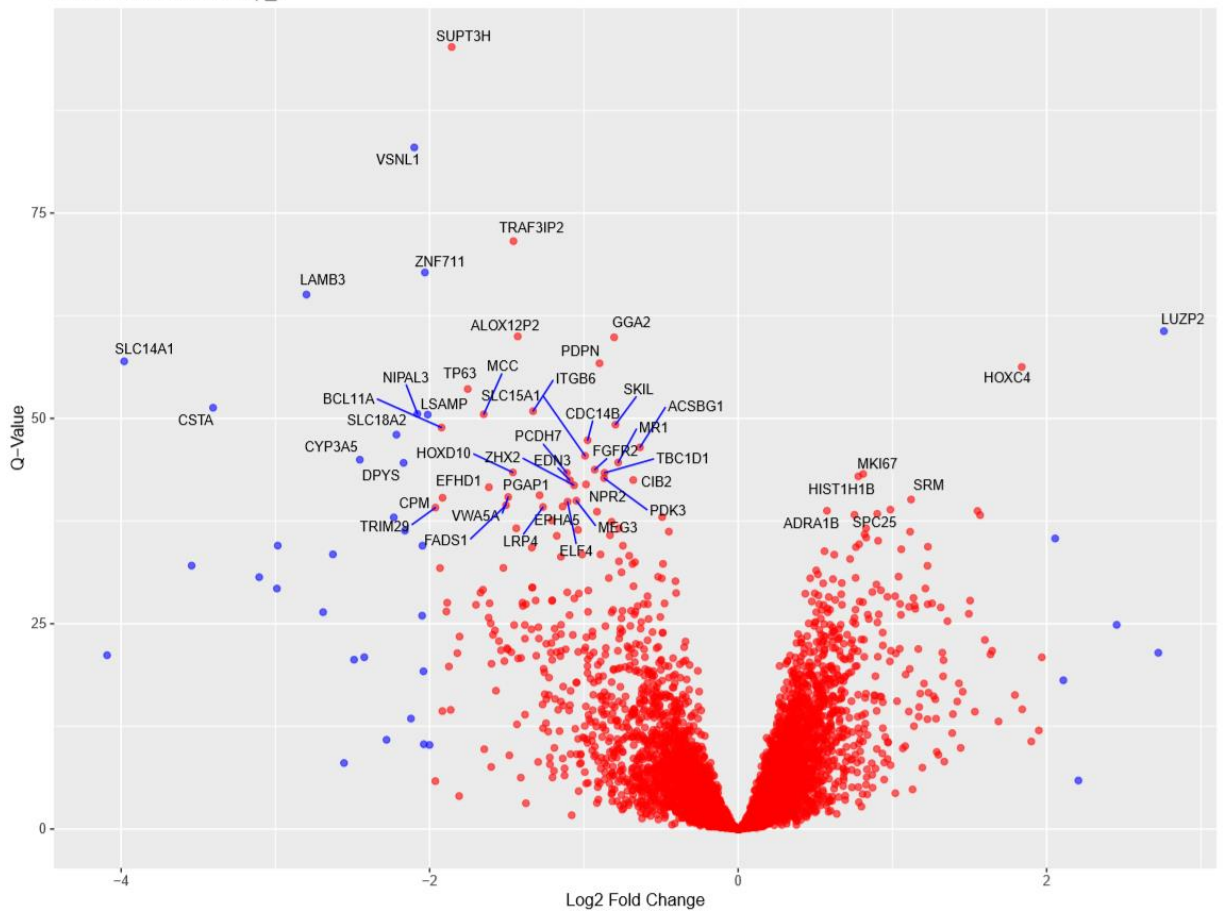


Figure 54. Volcano plot of prostate cancer Gleason grade 5 tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

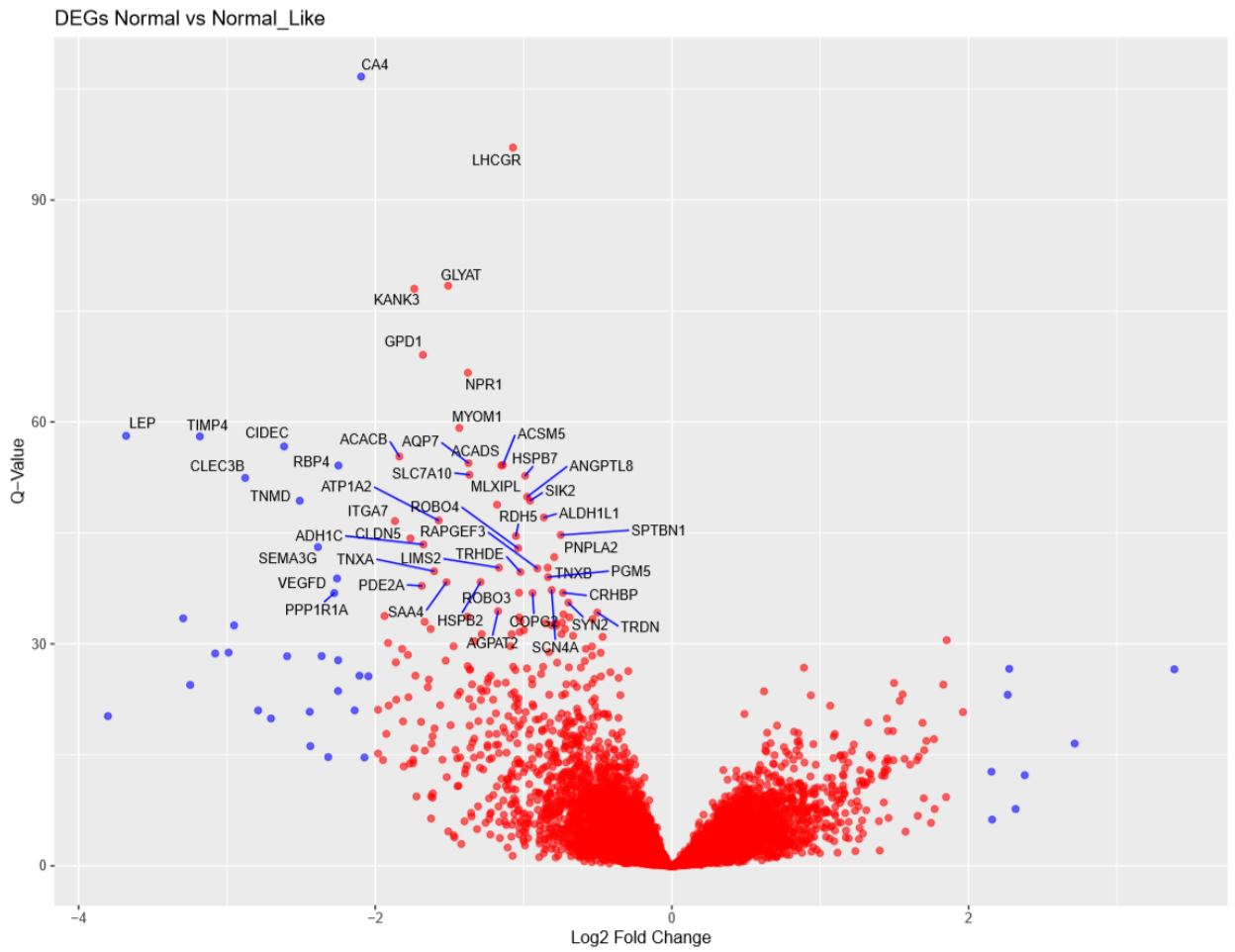


Figure 55. Volcano plot of breast cancer molecular subtype normal like tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

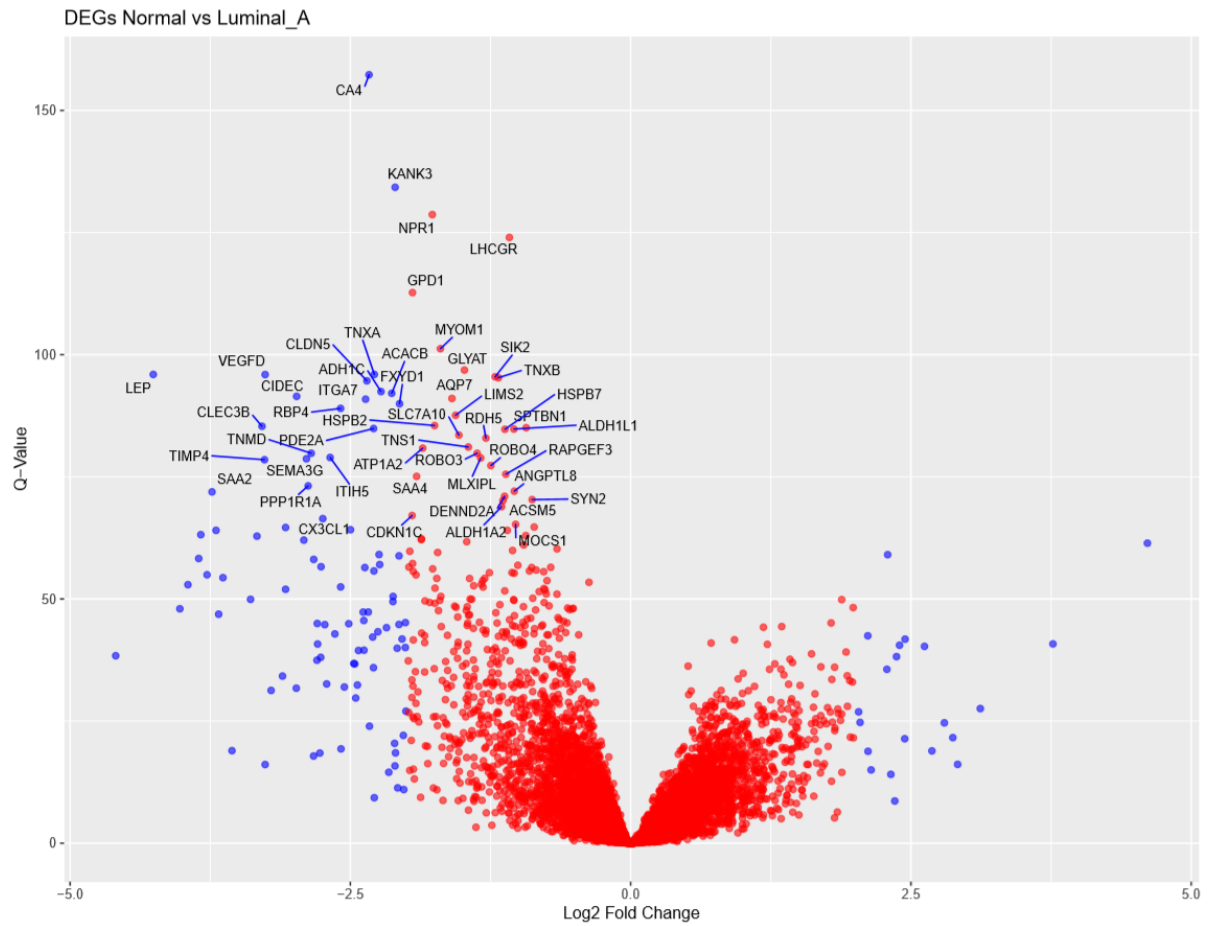


Figure 56. Volcano plot of breast cancer molecular subtype luminal A tumours.

Adjusted *P*-value (*Q*-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

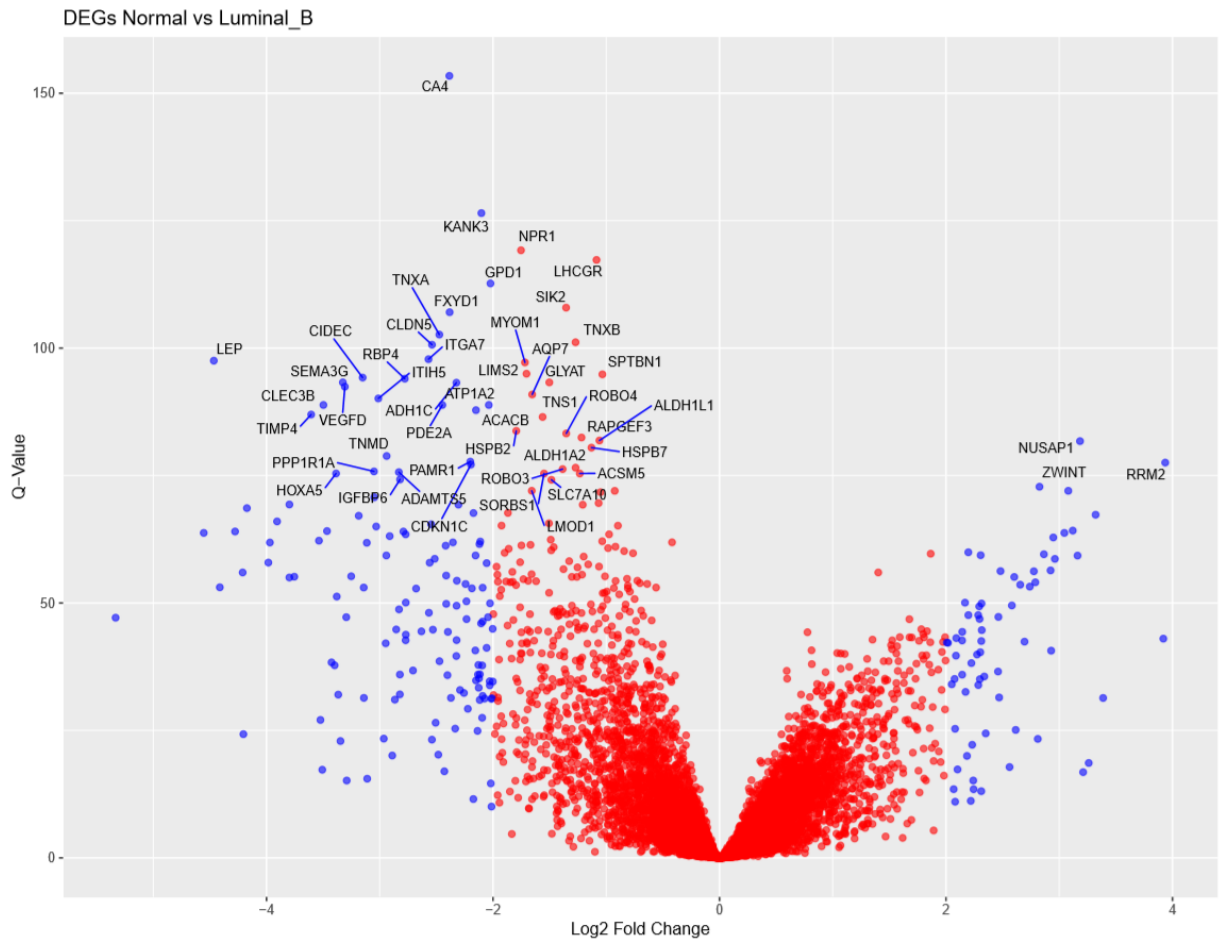


Figure 57. Volcano plot of breast cancer molecular subtype luminal B tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

DEGs Normal vs HER2



Figure 58. Volcano plot of breast cancer molecular subtype HER2 tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

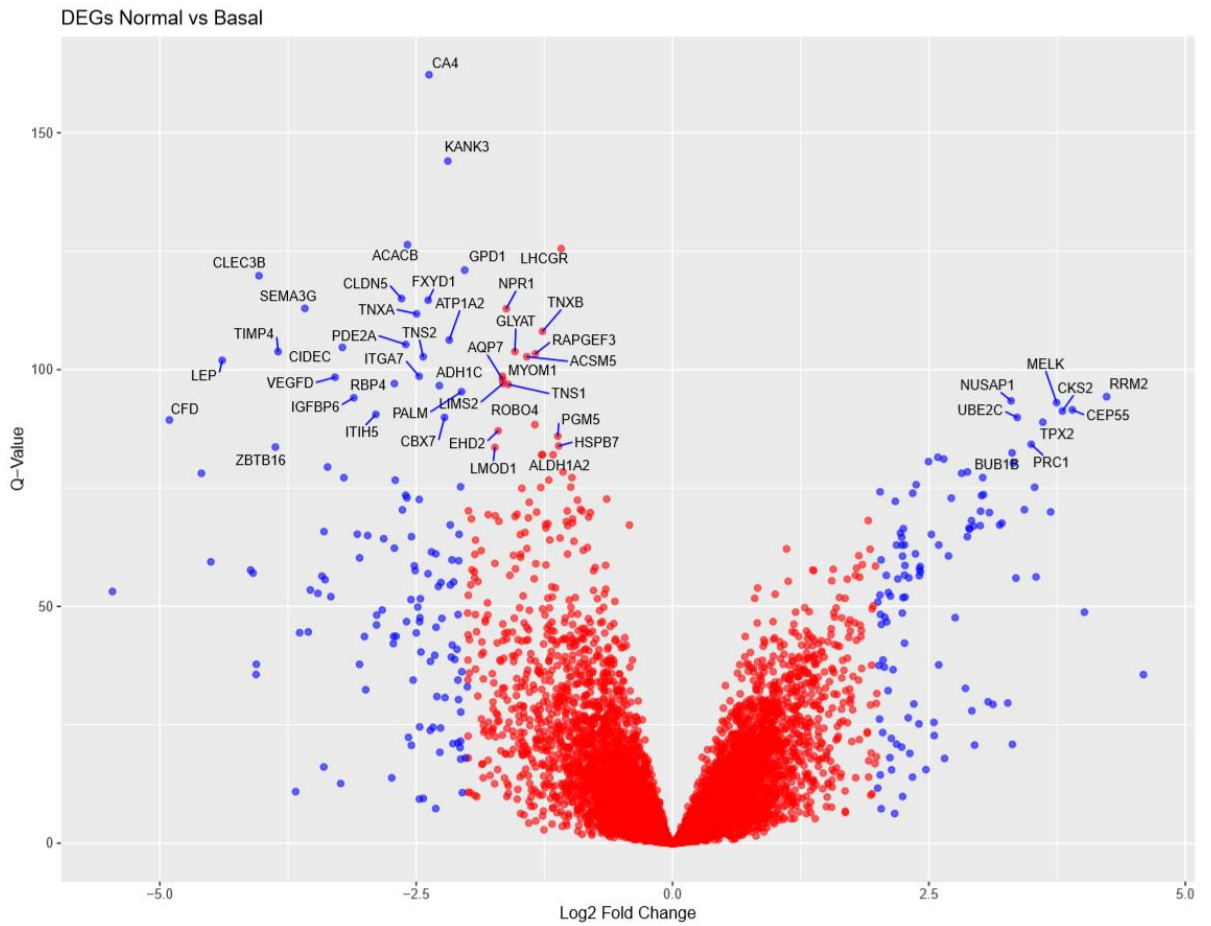


Figure 59. Volcano plot of breast cancer molecular subtype basal tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

DEGs Normal vs Serous



Figure 60. Volcano plot of ovarian cancer epithelial subtype serous tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

DEGs Normal vs Endometrioid

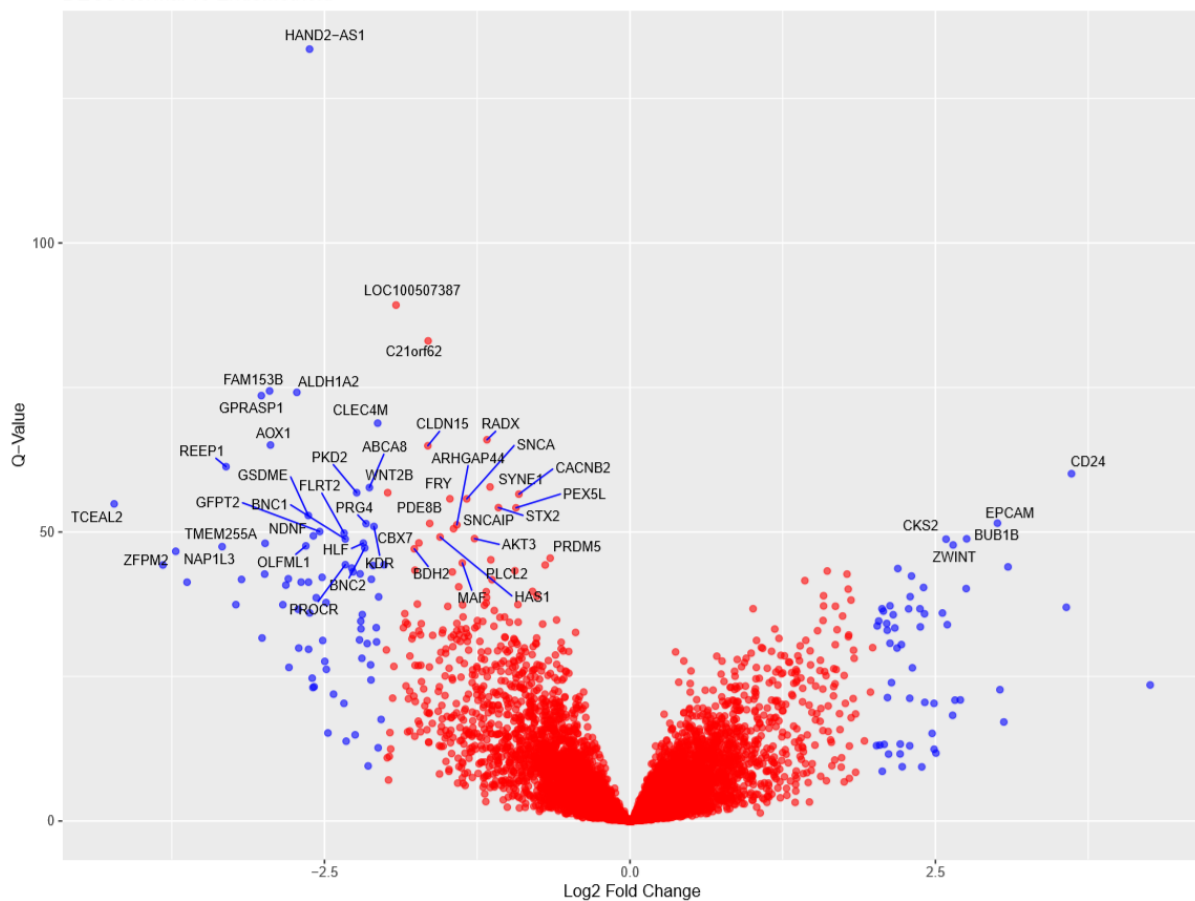


Figure 61. Volcano plot of ovarian cancer epithelial subtype endometrioid tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

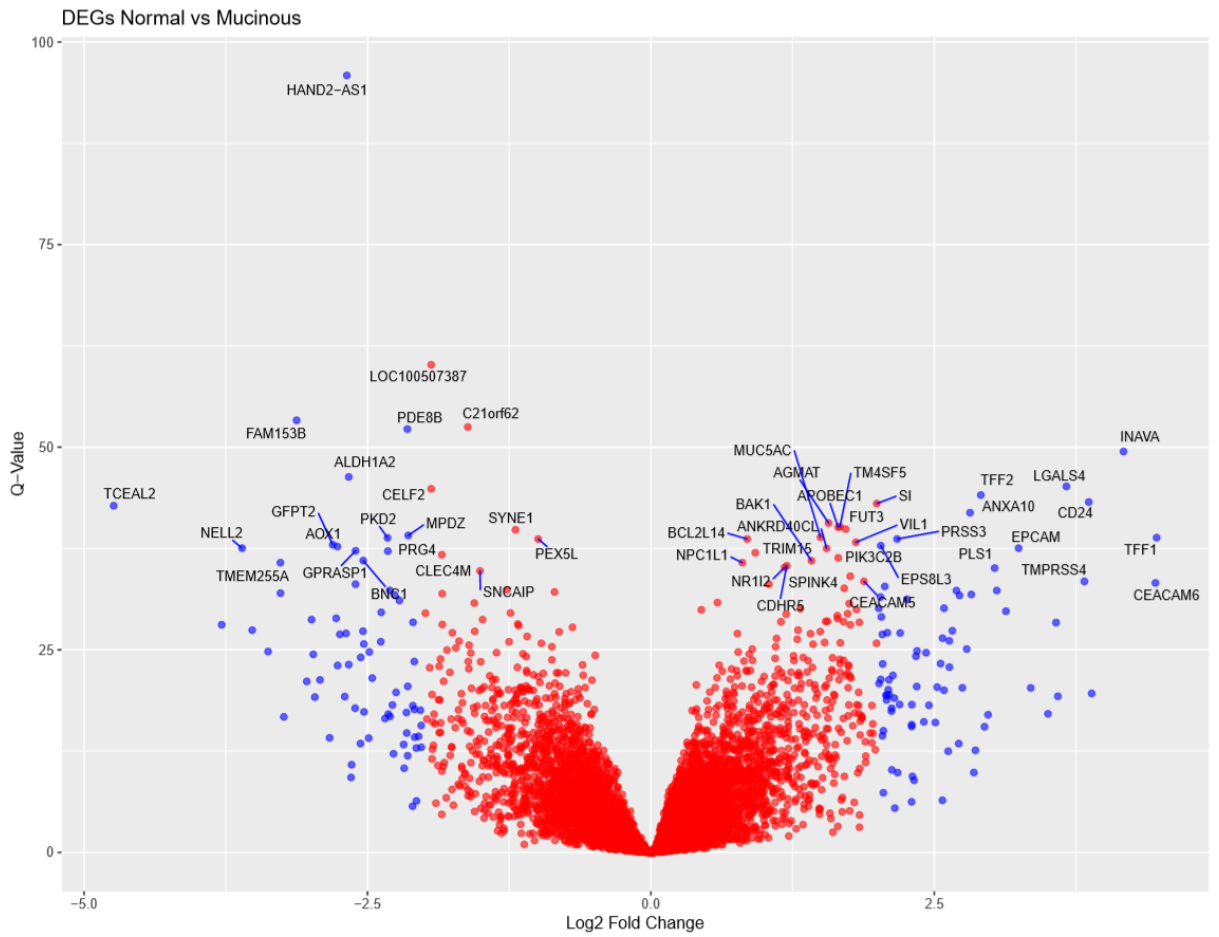


Figure 62. Volcano plot of ovarian cancer epithelial subtype mucinous tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

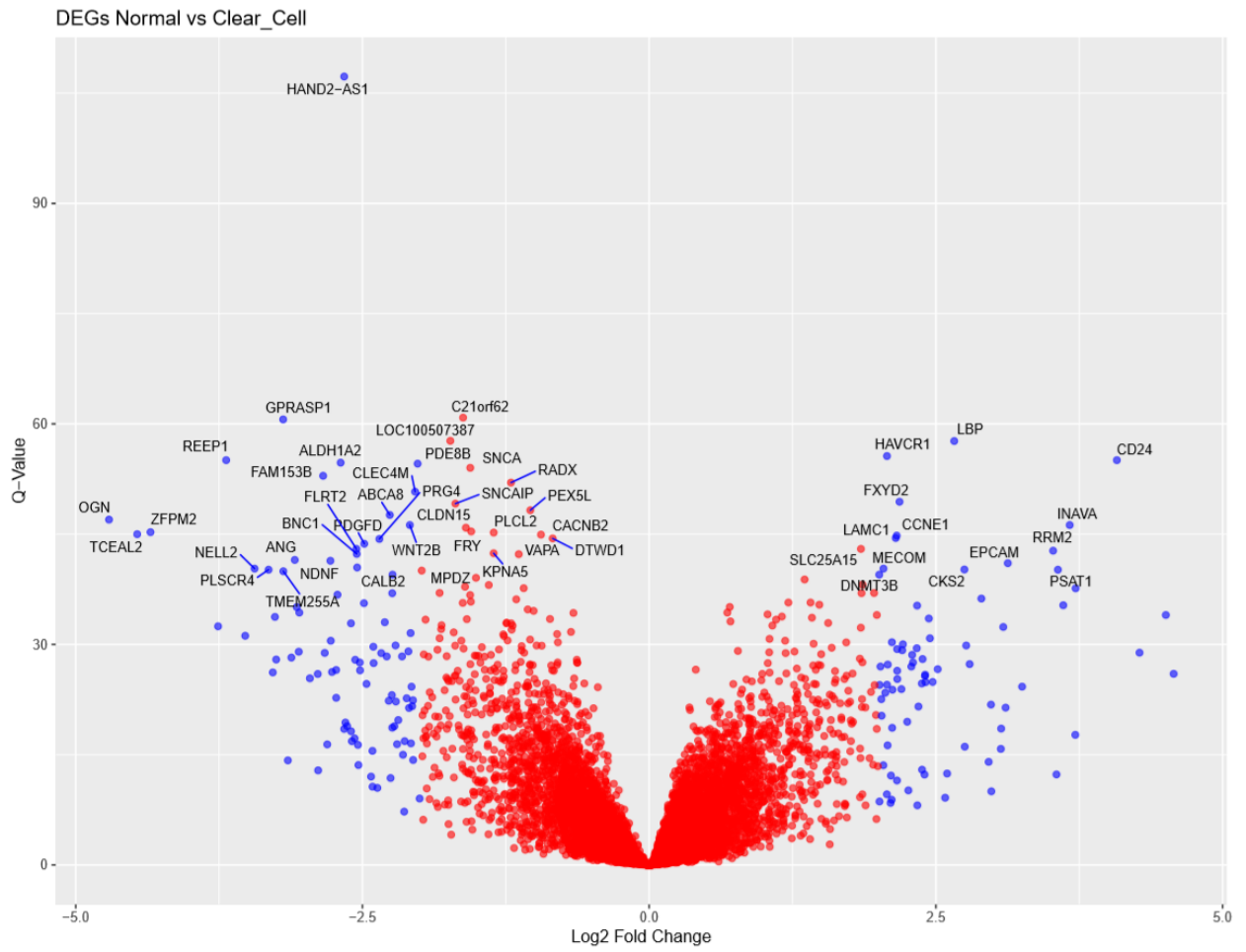


Figure 63. Volcano plot of ovarian cancer epithelial subtype clear cell tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

DEGs Normal vs Score_4

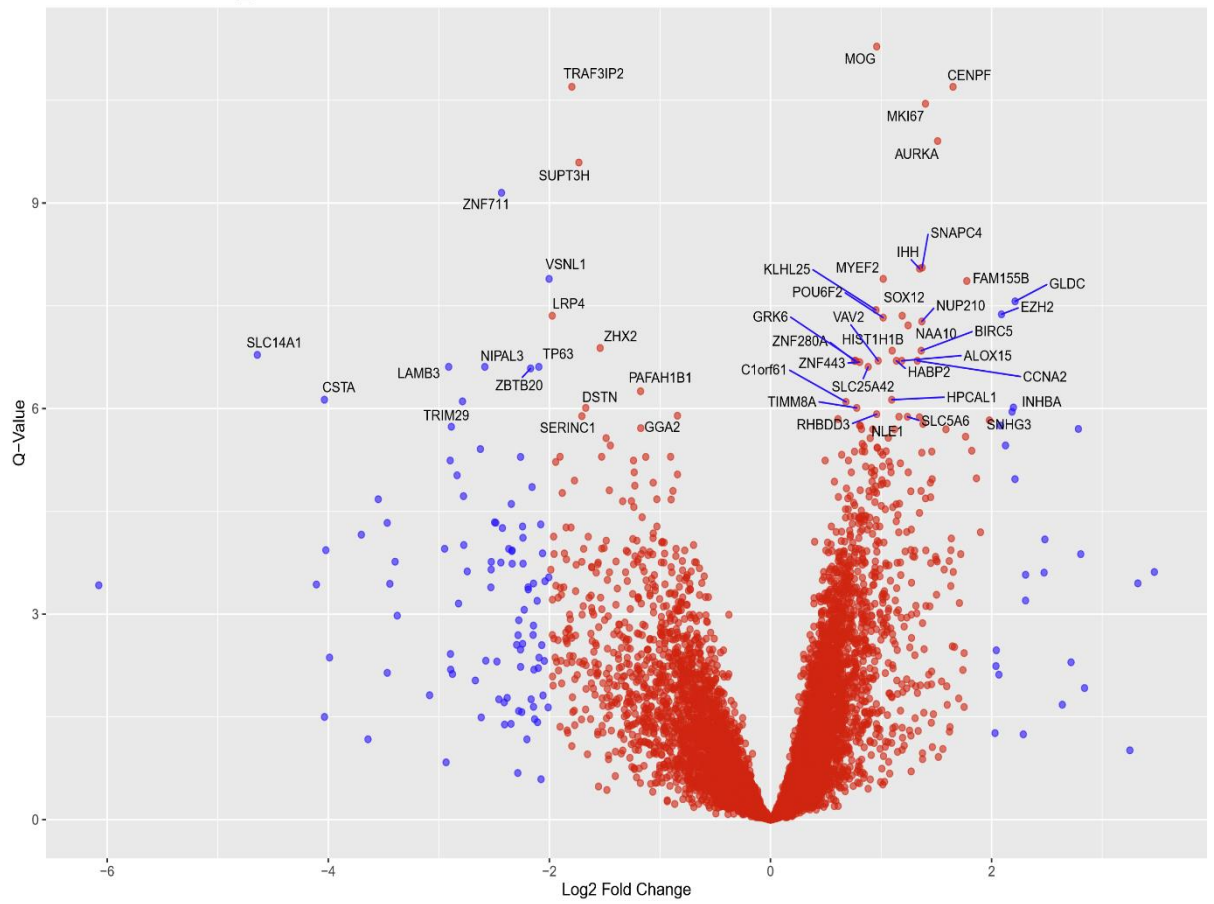


Figure 64. Volcano plot of prostate cancer Gleason score 4 tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

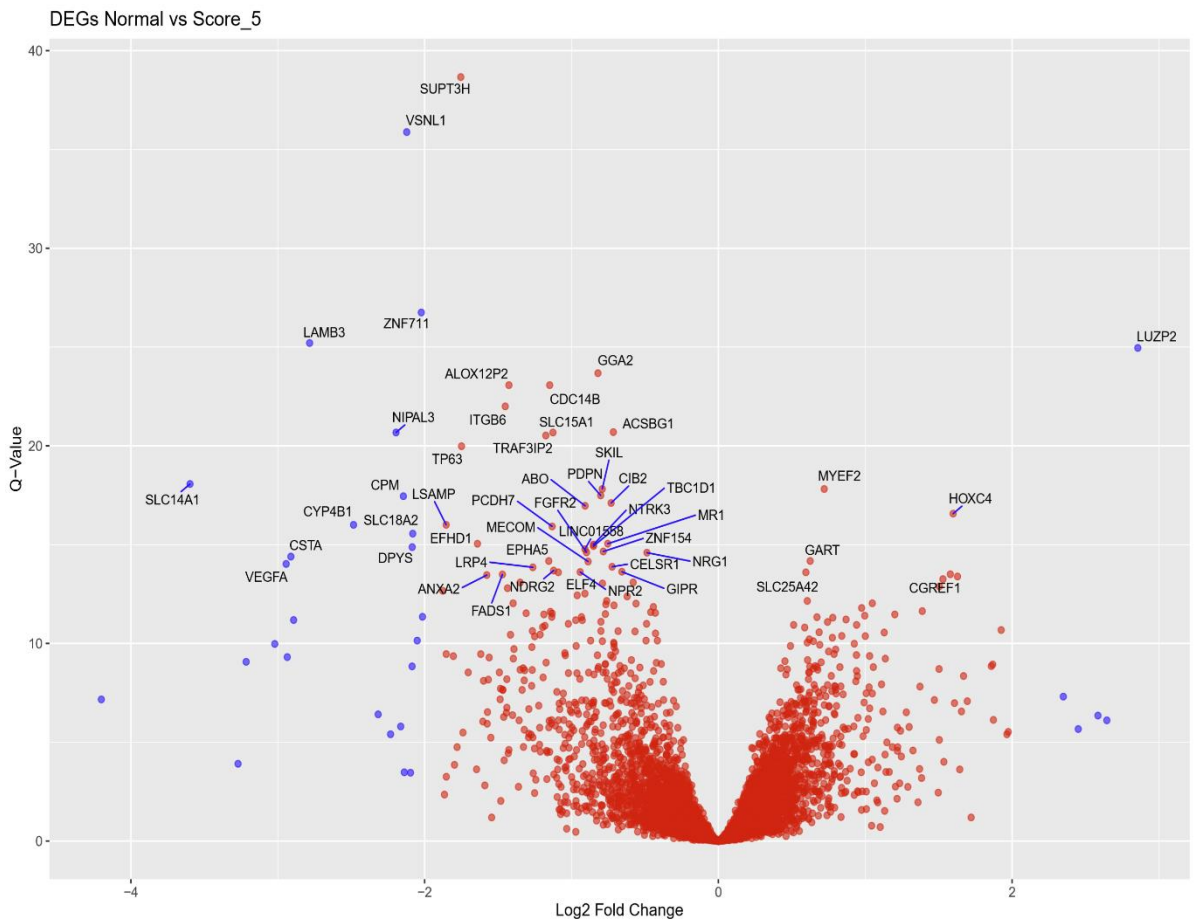


Figure 65. Volcano plot of prostate cancer Gleason score 5 tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

DEGs Normal vs Score_6

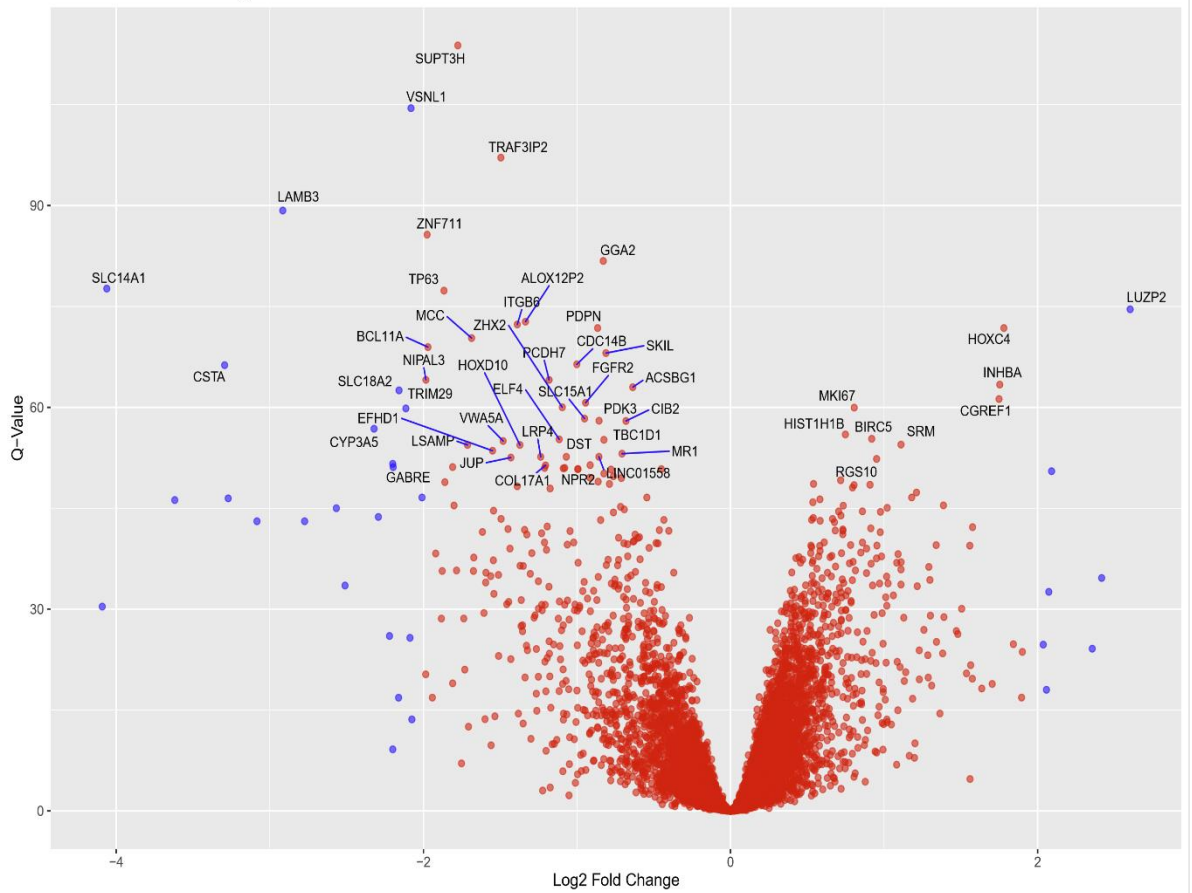


Figure 66. Volcano plot of prostate cancer Gleason score 6 tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

DEGs Normal vs Score_7

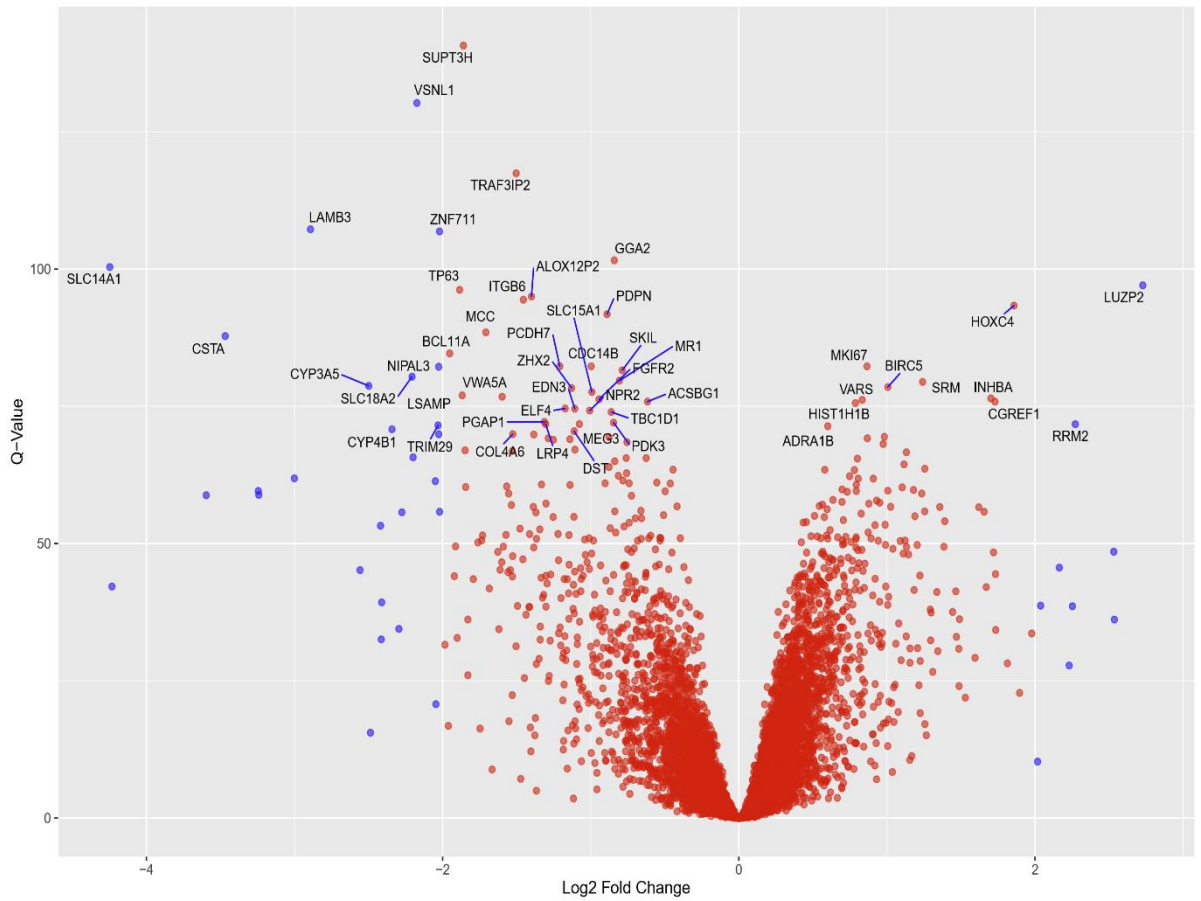


Figure 67. Volcano plot of prostate cancer Gleason score 7 tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

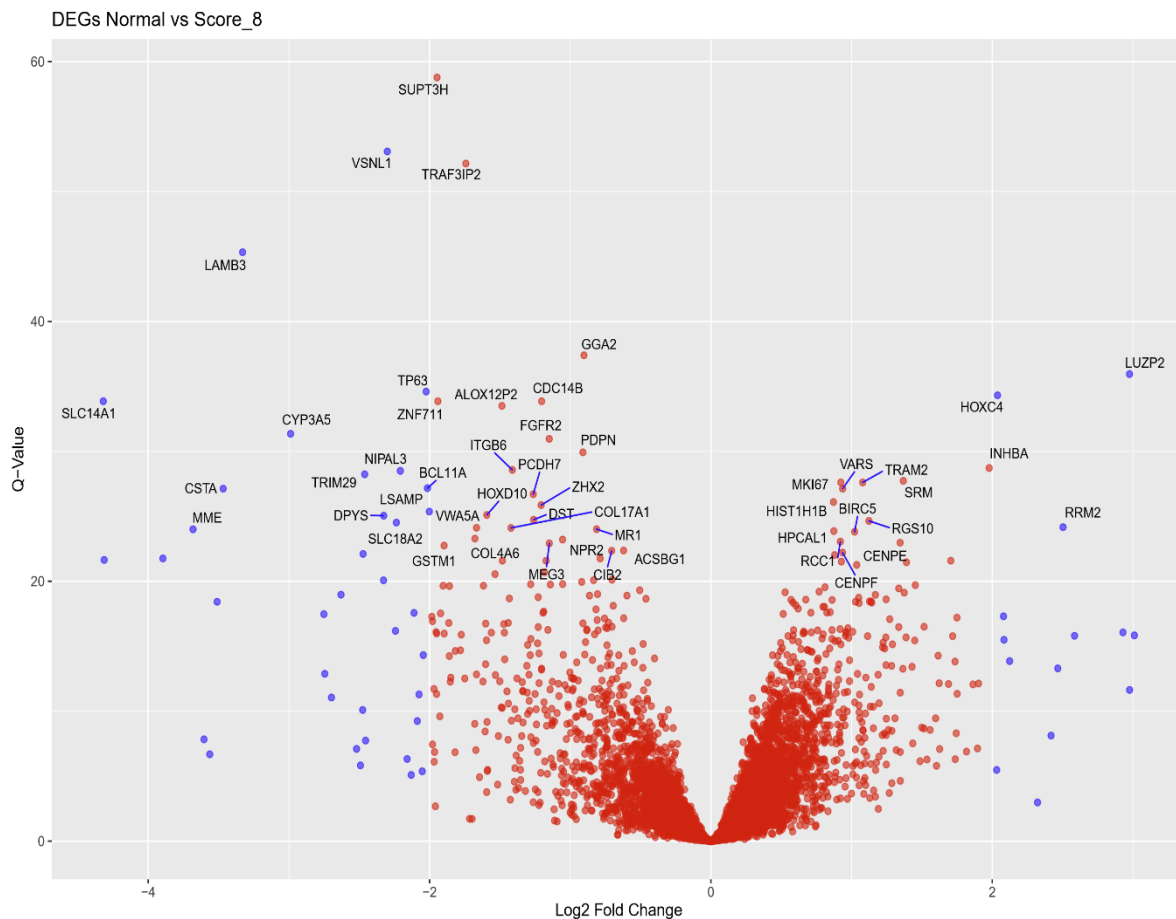


Figure 68. Volcano plot of prostate cancer Gleason score 8 tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

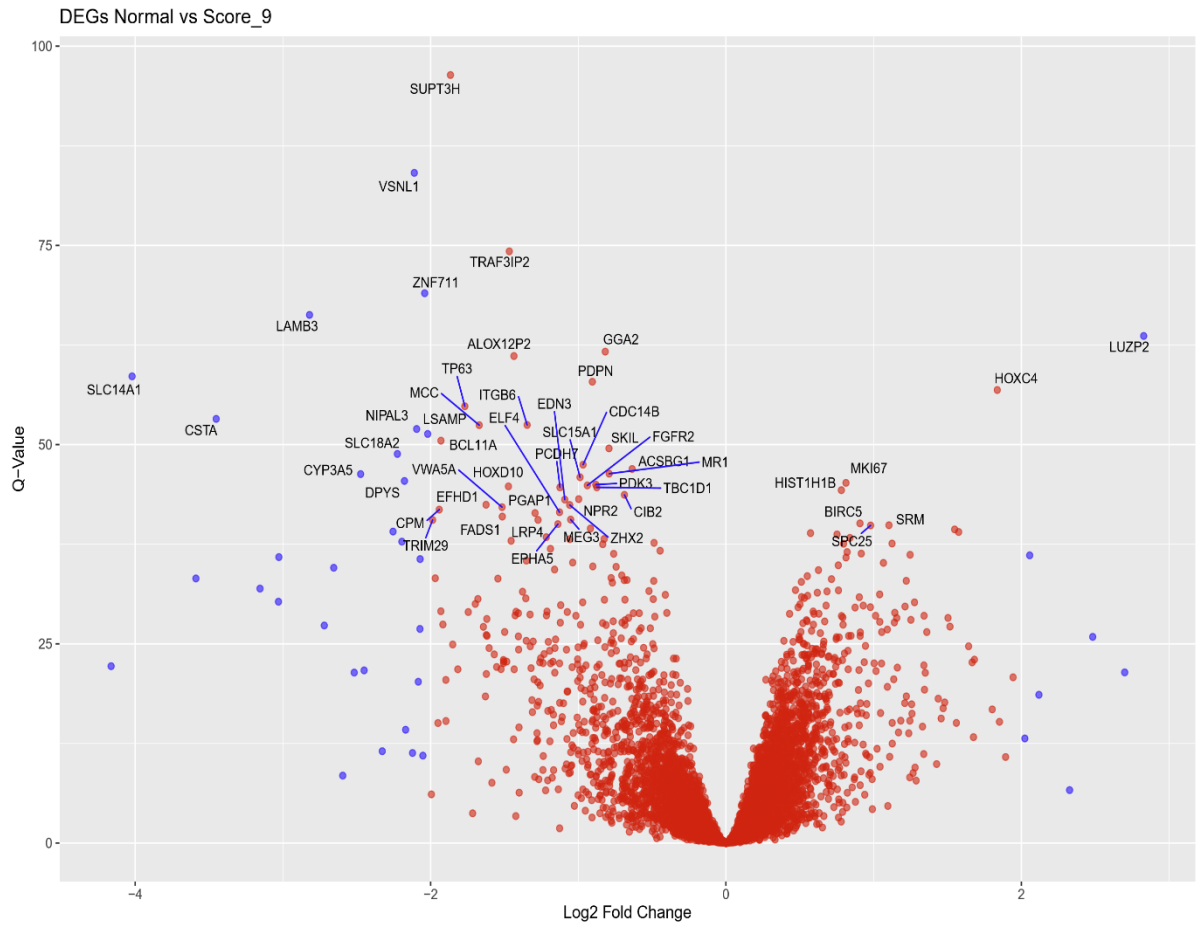


Figure 69. Volcano plot of prostate cancer Gleason score 9 tumours.

Adjusted P-value (Q-value) and gene expression (Log2 Fold change) plotted for all genes. The 50 most significant genes are labelled. Genes highlighted in blue are significant and above a logFC of 2 threshold.

Table 45. Top 50 most significant DEGs in breast cancer IDC.

Gene	Log2 FC	adj.P.Val
<i>OXTR</i>	-3.38099	1.59E-291
<i>FXYD1</i>	-1.42595	2.59E-285
<i>TNXA</i>	-1.56599	1.12E-274
<i>INHBA</i>	1.764755	1.69E-251
<i>VEGFD</i>	-2.41766	1.42E-250
<i>RRM2 (PAM50)</i>	3.022999	1.32E-249
<i>CNN1</i>	-1.97124	4.77E-238
<i>ZWINT</i>	2.134749	2.00E-227
<i>MME</i>	-2.15898	2.56E-222
<i>NR3C2</i>	-1.73165	7.04E-217
<i>COL10A1</i>	3.522949	8.80E-216
<i>WIF1</i>	-3.67476	1.39E-210
<i>LMOD1</i>	-1.10275	3.90E-210
<i>ITIH5</i>	-1.7048	2.88E-209
<i>HLF</i>	-1.75092	5.74E-208
<i>CKS2</i>	2.439067	2.77E-197
<i>NUSAP1</i>	2.198924	5.64E-194
<i>HOXA4</i>	-0.89351	2.15E-189
<i>AK5</i>	-1.72667	2.23E-176
<i>SYNM</i>	-1.58059	3.27E-174
<i>GPRASP1</i>	-1.45398	1.90E-173
<i>FN1</i>	1.641929	1.74E-172
<i>ID4</i>	-1.73951	3.57E-171
<i>LYVE1</i>	-1.48447	6.00E-169
<i>GPM6B</i>	-1.89613	1.19E-168
<i>EDNRB</i>	-1.48958	9.28E-165
<i>LIFR</i>	-1.3515	3.09E-164
<i>CLEC3B</i>	-2.01903	4.20E-164
<i>SEMA3G</i>	-1.66379	8.54E-162
<i>DPP3</i>	1.125729	8.28E-161
<i>RACGAP1</i>	1.672122	8.68E-158

Gene	Log2 FC	adj.P.Val
<i>RMI1</i>	1.163426	8.80E-156
<i>SIK2</i>	-0.63188	1.40E-154
<i>LAMB3</i>	-1.75667	3.30E-154
<i>WISP1</i>	1.106318	1.25E-153
<i>PAMR1</i>	-1.30554	9.05E-153
<i>IGFBP6</i>	-1.88218	4.79E-152
<i>SPRY2</i>	-1.72518	1.30E-150
<i>MEIS2</i>	-1.58719	4.13E-150
<i>ATP1A2</i>	-1.25121	9.95E-150
<i>DST</i>	-0.8234	1.73E-149
<i>MAB21L1</i>	-1.31219	4.88E-149
<i>CCNB1 (PAM50)</i>	1.878344	9.47E-149
<i>RABIF</i>	0.86529	1.30E-148
<i>CDK1</i>	1.566463	3.10E-147
<i>MELK (PAM50)</i>	2.125556	1.08E-146
<i>KIF11</i>	1.861772	2.79E-146
<i>CBX7</i>	-1.34131	3.54E-146
<i>PCNA</i>	1.323598	1.65E-144
<i>NR3C1</i>	-1.07071	2.50E-144

Known PAM50 Genes are shown in brackets.

Table 46. Top 50 most significant DEGs in breast cancer DCIS.

Gene	Log2 FC	adj.P.Val
<i>OXTR</i>	-3.25021	2.17E-166
<i>TNXA</i>	-1.51109	2.56E-156
<i>INHBA</i>	1.734853	6.05E-146
<i>FXYP1</i>	-1.28393	8.41E-146
<i>VEGFD</i>	-2.29032	1.30E-136
<i>CNN1</i>	-1.89171	9.78E-132
<i>RRM2 (PAM50)</i>	2.727986	3.54E-125
<i>COL10A1</i>	3.486471	2.14E-124
<i>WIF1</i>	-3.634	6.30E-121
<i>ZWINT</i>	1.998021	5.93E-120
<i>MME</i>	-1.9998	8.46E-115
<i>NR3C2</i>	-1.61447	6.50E-113
<i>HLF</i>	-1.66001	7.58E-111
<i>LMOD1</i>	-1.0049	1.25E-104
<i>HOXA4</i>	-0.85404	2.69E-101
<i>ITIH5</i>	-1.51591	1.00E-99
<i>SYNM</i>	-1.53317	7.89E-95
<i>GPM6B</i>	-1.86987	3.13E-94
<i>WISP1</i>	1.147686	3.42E-93
<i>NUSAP1</i>	1.968323	1.23E-92
<i>CKS2</i>	2.152713	4.28E-92
<i>ID4</i>	-1.66957	2.52E-91
<i>FN1</i>	1.555668	5.27E-90
<i>GPRASP1</i>	-1.36473	4.56E-89
<i>LAMB3</i>	-1.76981	6.96E-89
<i>AK5</i>	-1.60158	7.12E-89
<i>LIFR</i>	-1.2707	4.26E-84
<i>ATP1A2</i>	-1.21966	4.50E-81
<i>RMI1</i>	1.098729	5.65E-80
<i>EDNRB</i>	-1.35605	1.28E-79
<i>MAB21L1</i>	-1.25745	3.25E-78

Gene	Log2 FC	adj.P.Val
<i>SEMA3G</i>	-1.50802	2.08E-77
<i>LYVE1</i>	-1.30118	1.89E-76
<i>SPRY2</i>	-1.62041	2.51E-76
<i>CHL1</i>	-0.98976	4.53E-76
<i>SIK2</i>	-0.58167	9.55E-76
<i>RAB1F</i>	0.812223	3.95E-75
<i>MEIS2</i>	-1.47849	6.11E-75
<i>PAK3</i>	-1.05909	9.53E-75
<i>MATN2</i>	-1.39918	1.64E-74
<i>DMD</i>	-0.76826	1.80E-74
<i>CLEC3B</i>	-1.76561	9.42E-74
<i>DPP3</i>	0.995435	1.37E-73
<i>PAMR1</i>	-1.18773	3.13E-73
<i>PDGFA</i>	-0.9369	4.91E-73
<i>IL33</i>	-1.72573	1.56E-72
<i>KLHL21</i>	-1.06499	2.21E-72
<i>IGFBP6</i>	-1.70249	4.13E-72
<i>DST</i>	-0.75205	4.89E-72
<i>FAM189A2</i>	-1.59968	3.87E-71

Known PAM50 Genes are shown in brackets.

Table 47. Top 50 most significant DEGs in breast cancer ILC.

Gene	Log2 FC	adj.P.Val
<i>RRM2 (PAM50)</i>	3.045954	8.10E-122
<i>OXTR</i>	-3.05208	1.04E-121
<i>FXYP1</i>	-1.28446	3.95E-118
<i>TNXA</i>	-1.40649	2.96E-112
<i>VEGFD</i>	-2.28264	7.88E-110
<i>INHBA</i>	1.578129	2.92E-100
<i>ZWINT</i>	1.988331	1.17E-95
<i>MME</i>	-1.98404	6.20E-91
<i>HLF</i>	-1.67405	2.45E-90
<i>HOXA4</i>	-0.86206	1.77E-82
<i>ITIH5</i>	-1.53939	4.38E-82
<i>CNN1</i>	-1.62477	4.11E-81
<i>NR3C2</i>	-1.51068	1.68E-80
<i>CKS2</i>	2.25618	3.47E-80
<i>LMOD1</i>	-0.97815	5.63E-80
<i>WIF1</i>	-3.23613	3.63E-79
<i>NUSAP1</i>	2.001342	2.15E-76
<i>RMI1</i>	1.153228	1.23E-69
<i>DPP3</i>	1.08513	1.52E-68
<i>COL10A1</i>	2.783655	9.05E-67
<i>GPM6B</i>	-1.7479	1.22E-66
<i>PCNA</i>	1.315299	2.43E-64
<i>AK5</i>	-1.51635	3.73E-64
<i>GPRASP1</i>	-1.2803	3.01E-63
<i>SEMA3G</i>	-1.52568	5.43E-63
<i>RABIF</i>	0.832248	1.55E-62
<i>MAB21L1</i>	-1.25494	4.25E-62
<i>LYVE1</i>	-1.31239	7.11E-62
<i>ID4</i>	-1.49898	6.98E-60
<i>PAK3</i>	-1.06216	8.52E-60
<i>LIFR</i>	-1.18825	2.62E-59

Gene	Log2 FC	adj.P.Val
<i>MEIS2</i>	-1.47105	4.22E-59
<i>SIK2</i>	-0.57234	1.52E-58
<i>ATP1A2</i>	-1.15437	1.89E-58
<i>MTHFD2</i>	1.421074	9.61E-58
<i>MELK (PAM50)</i>	1.968695	1.54E-57
<i>WISP1</i>	0.995719	1.74E-57
<i>RACGAP1</i>	1.464846	2.43E-56
<i>GIN51</i>	1.472427	3.38E-56
<i>FN1</i>	1.359457	3.59E-56
<i>CHL1</i>	-0.94735	6.40E-56
<i>CLEC3B</i>	-1.71801	7.84E-56
<i>KIF11</i>	1.698038	8.68E-56
<i>BUB1B</i>	1.690546	8.76E-56
<i>DONSON</i>	1.207777	7.07E-55
<i>PIK3C2G</i>	-1.13623	8.77E-55
<i>RELN</i>	-1.44592	8.77E-55
<i>KPNA2</i>	1.491565	2.14E-54
<i>EDNRB</i>	-1.24566	2.35E-54
<i>RBMS3</i>	-0.80585	4.83E-54

Known PAM50 Genes are shown in brackets.

Table 48. Top 50 most significant DEGs in ovarian cancer ovarian tissue.

Gene	Log2 FC	adj.P.Val
<i>HAND2-AS1</i>	-2.36211	2.82E-285
<i>C21orf62</i>	-1.80682	4.52E-262
<i>CLEC4M</i>	-2.20341	1.27E-213
<i>PDE8B</i>	-2.14447	4.99E-200
<i>FAM153B</i>	-3.1602	7.00E-196
<i>PRG4</i>	-2.52328	1.88E-193
<i>CLDN15</i>	-1.78203	5.07E-188
<i>BNC1</i>	-2.80198	1.84E-185
<i>KDR</i>	-2.45864	2.38E-183
<i>REEP1</i>	-3.2529	5.75E-173
<i>GPRASP1</i>	-2.89899	2.59E-172
<i>GSDME</i>	-2.88818	2.79E-164
<i>LOC100507387</i>	-1.55684	4.93E-164
<i>NELL2</i>	-3.40325	2.36E-162
<i>ABCA8</i>	-2.61028	4.23E-160
<i>ALDH1A2</i>	-2.49227	2.38E-156
<i>AOX1</i>	-2.75955	2.63E-156
<i>SNCAIP</i>	-1.76877	1.79E-155
<i>CD24</i>	3.507658	1.98E-155
<i>PROCR</i>	-2.62163	2.21E-152
<i>RADX</i>	-1.02236	3.77E-145
<i>WNT2B</i>	-1.94358	1.05E-142
<i>POLR2M</i>	-1.04645	9.62E-142
<i>FLRT2</i>	-2.56849	1.72E-141
<i>PKD2</i>	-2.18095	6.13E-141
<i>CALB2</i>	-2.68515	7.95E-141
<i>CELF2</i>	-1.70421	4.32E-137
<i>HAS1</i>	-1.65285	1.64E-133
<i>MAF</i>	-1.42994	1.28E-132
<i>NPY1R</i>	-3.35183	2.41E-132
<i>UFSP2</i>	-1.71795	5.23E-130

Gene	Log2 FC	adj.P.Val
<i>TMEM255A</i>	-3.01937	3.76E-129
<i>PTGDR</i>	-0.75292	1.49E-127
<i>PEX5L</i>	-0.81105	4.50E-127
<i>BDH2</i>	-1.84213	1.01E-125
<i>PCDH9</i>	-1.42729	2.97E-123
<i>BCHE</i>	-3.84276	1.94E-122
<i>TCEAL2</i>	-3.97093	2.80E-122
<i>CACNB2</i>	-0.75948	5.39E-120
<i>PDGFD</i>	-2.27837	5.39E-120
<i>FEZ2</i>	-0.84581	3.48E-119
<i>NAP1L3</i>	-3.16912	1.01E-118
<i>SNCA</i>	-1.21654	2.95E-118
<i>ANXA8L1</i>	-2.79572	3.60E-118
<i>MNDA</i>	-3.13733	1.02E-117
<i>ARHGAP44</i>	-1.29683	2.00E-117
<i>CLK1</i>	-1.65839	2.46E-117
<i>GNG11</i>	-2.12668	4.03E-116
<i>FRY</i>	-1.26684	4.26E-116
<i>NDNF</i>	-2.48681	3.47E-114

Known PAM50 Genes are shown in brackets.

Table 49. Top 50 most significant DEGs in ovarian fallopian

Gene	Log2 FC	adj.P.Val
<i>C21orf62</i>	-1.94539	4.15E-39
<i>HAND2-AS1</i>	-2.24066	3.44E-35
<i>CLEC4M</i>	-2.42283	2.16E-30
<i>BNC1</i>	-3.17862	1.81E-26
<i>FAM153B</i>	-3.40663	6.29E-26
<i>PDE8B</i>	-2.26544	9.60E-26
<i>ABCA8</i>	-3.14625	1.27E-24
<i>PRG4</i>	-2.6406	3.64E-24
<i>CLDN15</i>	-1.85122	6.72E-23
<i>AOX1</i>	-3.19646	4.04E-22
<i>FLRT2</i>	-3.14252	1.13E-21
<i>LOC100507387</i>	-1.70159	4.49E-21
<i>GPRASP1</i>	-3.05858	4.97E-21
<i>WNT2B</i>	-2.18827	1.40E-18
<i>MTUS1</i>	-2.20907	1.90E-18
<i>GSDME</i>	-2.93516	2.64E-18
<i>KDR</i>	-2.30735	3.65E-18
<i>NELL2</i>	-3.43687	8.15E-18
<i>ANXA8L1</i>	-3.4584	1.02E-17
<i>SNCAIP</i>	-1.82683	1.30E-17
<i>ALDH1A2</i>	-2.56079	1.37E-17
<i>HAS1</i>	-1.88178	1.40E-17
<i>CD24</i>	3.516502	1.30E-16
<i>CLK1</i>	-1.97473	2.57E-16
<i>SNCA</i>	-1.41773	8.74E-16
<i>REEP1</i>	-2.91713	2.23E-15
<i>LGALS8</i>	-1.94009	2.39E-15
<i>MECOM</i>	1.829885	1.38E-14
<i>PROCR</i>	-2.48475	1.79E-14
<i>NPY1R</i>	-3.47946	1.82E-14
<i>PTTG1 (PAM50)</i>	2.765441	1.84E-14

Gene	Log2 FC	adj.P.Val
<i>MAF</i>	-1.47972	1.93E-14
<i>PKD2</i>	-2.16315	2.34E-14
<i>PROS1</i>	-3.28589	2.77E-14
<i>EPB42</i>	-0.93728	3.35E-14
<i>RADX</i>	-0.9859	3.91E-14
<i>CALB2</i>	-2.63177	4.86E-14
<i>FRY</i>	-1.40367	4.86E-14
<i>NDNF</i>	-2.7795	5.26E-14
<i>OGN</i>	-4.25043	6.84E-14
<i>PDGFD</i>	-2.45336	7.52E-14
<i>SECISBP2</i>	-1.32278	1.11E-13
<i>BUB1B</i>	2.872202	1.50E-13
<i>CELF2</i>	-1.66571	1.55E-13
<i>TOP2A</i>	2.838172	1.69E-13
<i>PSD3</i>	-2.36198	1.73E-13
<i>DYNC2H1</i>	-1.34188	1.73E-13
<i>RACGAP1</i>	2.60645	2.15E-13
<i>E2F1</i>	1.279763	2.15E-13
<i>POLR2M</i>	-0.99305	2.25E-13

Known PAM50 Genes are shown in brackets.

Table 50. Top 50 most significant DEGs in ovarian peritoneum

Gene	Log2 FC	adj.P.Val
<i>C21orf62</i>	-1.81485	6.26E-105
<i>HAND2-AS1</i>	-2.11021	5.18E-97
<i>CLEC4M</i>	-2.27082	1.57E-84
<i>FAM153B</i>	-3.35156	9.72E-80
<i>PDE8B</i>	-2.12494	9.80E-73
<i>LOC100507387</i>	-1.71703	8.20E-69
<i>CLDN15</i>	-1.78503	1.23E-68
<i>GPRASP1</i>	-3.06155	9.96E-68
<i>PRG4</i>	-2.42883	2.29E-66
<i>CD24</i>	3.871085	9.89E-65
<i>REEP1</i>	-3.28587	5.91E-63
<i>WNT2B</i>	-2.17547	4.47E-60
<i>BNC1</i>	-2.61589	1.24E-59
<i>SNCAIP</i>	-1.84586	1.17E-58
<i>NELL2</i>	-3.33868	2.75E-55
<i>GSDME</i>	-2.66183	4.88E-50
<i>RADX</i>	-1.01223	4.70E-49
<i>KDR</i>	-2.06849	1.26E-48
<i>AOX1</i>	-2.57796	2.89E-48
<i>ARHGAP44</i>	-1.45118	6.38E-48
<i>ALDH1A2</i>	-2.31066	1.17E-47
<i>POLR2M</i>	-1.02894	3.73E-47
<i>GPM6A</i>	-2.51996	7.18E-46
<i>DIRAS3</i>	-3.68378	9.47E-46
<i>ABCA8</i>	-2.32364	1.43E-45
<i>TCEAL2</i>	-4.11325	1.26E-43
<i>NDNF</i>	-2.68901	1.26E-43
<i>BDH2</i>	-1.86654	2.82E-43
<i>FLRT2</i>	-2.3529	2.65E-41
<i>CLK1</i>	-1.6948	1.55E-40
<i>WRB</i>	-1.81846	2.81E-40

Gene	Log2 FC	adj.P.Val
<i>NAP1L3</i>	-3.20108	3.02E-40
<i>NPY1R</i>	-3.07827	2.69E-38
<i>SECISBP2</i>	-1.20479	4.09E-38
<i>BBS4</i>	-1.21875	4.24E-38
<i>LGALS8</i>	-1.65868	5.09E-38
<i>PKD2</i>	-1.91208	6.37E-38
<i>SNCA</i>	-1.1885	1.13E-37
<i>DMD</i>	-0.96592	9.73E-37
<i>PEX5L</i>	-0.74605	1.20E-36
<i>FRY</i>	-1.23154	1.56E-36
<i>TMEM255A</i>	-2.74204	1.70E-36
<i>ANXA8L1</i>	-2.63686	2.99E-35
<i>NBEA</i>	-1.97041	3.38E-35
<i>CKS2</i>	2.547338	7.29E-35
<i>PTGDR</i>	-0.67035	1.18E-34
<i>CIRBP</i>	-1.61645	2.43E-34
<i>UFSP2</i>	-1.50315	2.48E-34
<i>PROCR</i>	-2.06901	2.48E-34
<i>BUB1B</i>	2.480381	2.84E-34

Known PAM50 Genes are shown in brackets.

Table 51. Top 50 most significant DEGs in Prostate Gleason grade group 1.

Gene	Log2 FC	adj.P.Val
<i>SUPT3H</i>	-1.77727	9.03E-121
<i>VSNL1</i>	-2.08861	1.27E-111
<i>TRAF3IP2</i>	-1.47139	4.65E-100
<i>LAMB3</i>	-2.9091	2.89E-95
<i>ZNF711</i>	-1.99773	1.14E-92
<i>GGA2</i>	-0.83394	1.78E-88
<i>TP63</i>	-1.86414	6.07E-83
<i>SLC14A1</i>	-4.04182	1.03E-82
<i>LUZP2</i>	2.668938	1.83E-82
<i>ALOX12P2</i>	-1.34675	4.74E-79
<i>ITGB6</i>	-1.3943	3.79E-78
<i>PDPN</i>	-0.85911	1.87E-76
<i>HOXC4</i>	1.756112	6.77E-76
<i>MCC</i>	-1.65202	3.26E-73
<i>CDC14B</i>	-1.01761	7.39E-73
<i>SKIL</i>	-0.80313	2.36E-72
<i>CSTA</i>	-3.29341	3.87E-71
<i>NIPAL3</i>	-2.03085	4.65E-71
<i>BCL11A</i>	-1.90538	3.20E-70
<i>PCDH7</i>	-1.17822	4.16E-69
<i>ACSBG1</i>	-0.64295	7.53E-69
<i>SLC18A2</i>	-2.15982	1.11E-67
<i>INHBA</i>	1.732399	4.43E-67
<i>CGREF1</i>	1.73944	1.05E-65
<i>FGFR2</i>	-0.94341	1.05E-65
<i>TRIM29</i>	-2.11342	1.76E-64
<i>SLC15A1</i>	-0.96659	2.24E-64
<i>ZHX2</i>	-1.08888	6.45E-64
<i>CIB2</i>	-0.68892	7.23E-64
<i>MKI67 (PAM50)</i>	0.806164	2.10E-63
<i>CYP3A5</i>	-2.31094	5.60E-61

Gene	Log2 FC	adj.P.Val
<i>LSAMP</i>	-1.73887	3.22E-60
<i>ELF4</i>	-1.11846	3.32E-60
<i>TBC1D1</i>	-0.82632	5.99E-60
<i>PDK3</i>	-0.83708	6.21E-60
<i>HIST1H1B</i>	0.743663	1.42E-59
<i>EFHD1</i>	-1.57107	2.73E-59
<i>MR1</i>	-0.71951	4.57E-59
<i>VWA5A</i>	-1.46906	6.69E-59
<i>LRP4</i>	-1.26175	2.90E-58
<i>LINC01558</i>	-0.86331	6.83E-58
<i>CPM</i>	-1.86273	1.14E-57
<i>BIRC5 (PAM50)</i>	0.903496	1.45E-57
<i>HOXD10</i>	-1.34951	1.45E-57
<i>DST</i>	-1.07641	1.45E-57
<i>SRM</i>	1.087773	2.51E-57
<i>NDRG2</i>	-1.09981	4.51E-57
<i>RGS10</i>	0.95542	5.41E-57
<i>JUP</i>	-1.42899	7.07E-57
<i>NPR2</i>	-0.92251	1.10E-56

Known PAM50 Genes are shown in brackets.

Table 52. Top 50 most significant DEGs in prostate cancer Gleason grade group 2 & 3.

Gene	Log2 FC	adj.P.Val
<i>SUPT3H</i>	-1.85968	4.03E-142
<i>VSNL1</i>	-2.17405	1.12E-131
<i>TRAF3IP2</i>	-1.50381	5.18E-117
<i>LAMB3</i>	-2.89227	1.38E-108
<i>ZNF711</i>	-2.02109	7.57E-108
<i>GGA2</i>	-0.84112	5.66E-103
<i>SLC14A1</i>	-4.24759	2.69E-101
<i>LUZP2</i>	2.727414	7.22E-98
<i>TP63</i>	-1.88532	2.12E-97
<i>ALOX12P2</i>	-1.4002	4.46E-96
<i>ITGB6</i>	-1.45482	1.83E-95
<i>HOXC4</i>	1.857511	2.10E-94
<i>PDPN</i>	-0.88996	8.02E-93
<i>MCC</i>	-1.70842	5.43E-89
<i>CSTA</i>	-3.46805	1.61E-88
<i>BCL11A</i>	-1.95227	1.35E-84
<i>PCDH7</i>	-1.20899	1.66E-83
<i>CDC14B</i>	-0.99697	4.99E-83
<i>NIPAL3</i>	-2.02676	6.56E-83
<i>SKIL</i>	-0.78721	1.50E-82
<i>SLC18A2</i>	-2.20645	1.64E-81
<i>MKI67 (PAM50)</i>	0.864996	1.74E-81
<i>MR1</i>	-0.80631	1.24E-80
<i>SRM</i>	1.24074	3.33E-80
<i>CYP3A5</i>	-2.49916	1.75E-79
<i>ZHX2</i>	-1.12923	9.20E-79
<i>SLC15A1</i>	-0.99252	2.48E-78
<i>BIRC5 (PAM50)</i>	1.006009	3.56E-78
<i>LSAMP</i>	-1.86788	5.71E-78
<i>VWA5A</i>	-1.59842	1.35E-77
<i>FGFR2</i>	-0.9444	2.13E-77

Gene	Log2 FC	adj.P.Val
<i>INHBA</i>	1.701586	4.33E-77
<i>VAR5</i>	0.832368	5.38E-77
<i>CGREF1</i>	1.729253	9.03E-77
<i>ACSBG1</i>	-0.61675	9.38E-77
<i>HIST1H1B</i>	0.786951	8.79E-76
<i>ELF4</i>	-1.17291	1.09E-75
<i>EDN3</i>	-1.10768	1.72E-75
<i>NPR2</i>	-1.00717	2.84E-75
<i>TBC1D1</i>	-0.86163	7.28E-75
<i>PGAP1</i>	-1.313	3.24E-73
<i>MEG3</i>	-1.07713	8.03E-73
<i>ADRA1B</i>	0.600239	1.90E-72
<i>RRM2 (PAM50)</i>	2.271539	2.09E-72
<i>TRIM29</i>	-2.03107	2.80E-72
<i>LRP4</i>	-1.3041	4.48E-72
<i>PDK3</i>	-0.84577	6.69E-72
<i>CYP4B1</i>	-2.34217	1.12E-71
<i>DST</i>	-1.11238	3.13E-71
<i>DPYS</i>	-2.02612	8.03E-71

Known PAM50 Genes are shown in brackets.

Table 53. Top 50 most significant DEGs in prostate cancer Gleason grade group 4.

Gene	Log2 FC	adj.P.Val
<i>SUPT3H</i>	-1.94701	8.91E-60
<i>VSNL1</i>	-2.30081	4.23E-54
<i>TRAF3IP2</i>	-1.74225	5.17E-52
<i>LAMB3</i>	-3.32927	2.52E-46
<i>GGA2</i>	-0.90252	2.21E-38
<i>LUZP2</i>	2.975654	1.19E-36
<i>TP63</i>	-2.02434	1.72E-35
<i>HOXC4</i>	2.037305	3.84E-35
<i>ZNF711</i>	-1.94103	1.47E-34
<i>SLC14A1</i>	-4.31974	1.47E-34
<i>CDC14B</i>	-1.20343	1.57E-34
<i>ALOX12P2</i>	-1.48582	2.49E-34
<i>CYP3A5</i>	-2.98807	4.81E-32
<i>FGFR2</i>	-1.14962	7.34E-32
<i>PDPN</i>	-0.91016	9.78E-31
<i>INHBA</i>	1.977945	2.28E-29
<i>ITGB6</i>	-1.41177	2.28E-29
<i>NIPAL3</i>	-2.20802	3.60E-29
<i>TRIM29</i>	-2.46139	6.26E-29
<i>SRM</i>	1.367442	2.03E-28
<i>TRAM2</i>	1.078532	4.29E-28
<i>VAR5</i>	0.936243	8.20E-28
<i>CSTA</i>	-3.46746	9.36E-28
<i>MKI67 (PAM50)</i>	0.923398	1.37E-27
<i>PCDH7</i>	-1.26314	1.42E-27
<i>BCL11A</i>	-2.01543	1.55E-27
<i>HIST1H1B</i>	0.870742	1.60E-26
<i>ZHX2</i>	-1.20713	2.05E-26
<i>LSAMP</i>	-2.00197	3.68E-26
<i>DPYS</i>	-2.3263	8.26E-26
<i>HOXD10</i>	-1.59334	8.34E-26

Gene	Log2 FC	adj.P.Val
<i>DST</i>	-1.26049	1.90E-25
<i>SLC18A2</i>	-2.23621	2.35E-25
<i>RGS10</i>	1.123211	2.54E-25
<i>COL17A1</i>	-1.42152	6.22E-25
<i>VWA5A</i>	-1.66663	7.50E-25
<i>RRM2</i> (PAM50)	2.503003	7.50E-25
<i>MR1</i>	-0.81288	8.75E-25
<i>MME</i>	-3.68156	9.58E-25
<i>HPCAL1</i>	0.873514	3.12E-24
<i>BIRC5</i> (PAM50)	1.02122	5.05E-24
<i>NPR2</i>	-1.05453	5.05E-24
<i>COL4A6</i>	-1.67798	5.30E-24
<i>CENPE</i>	1.345071	8.69E-24
<i>MEG3</i>	-1.14897	9.59E-24
<i>RCC1</i>	0.919953	1.04E-23
<i>GSTM1</i>	-1.89705	1.73E-23
<i>CIB2</i>	-0.70558	3.59E-23
<i>ACSBG1</i>	-0.62119	3.96E-23
<i>CYP4B1</i>	-2.47297	7.62E-23

Known PAM50 Genes are shown in brackets.

Table 54. Top 50 most significant DEGs in prostate cancer Gleason grade group 5.

Gene	Log2 FC	adj.P.Val
<i>SUPT3H</i>	-1.85742	5.95E-96
<i>VSNL1</i>	-2.10009	1.02E-83
<i>TRAF3IP2</i>	-1.4568	2.62E-72
<i>ZNF711</i>	-2.03106	1.70E-68
<i>LAMB3</i>	-2.79905	8.27E-66
<i>LUZP2</i>	2.759514	2.38E-61
<i>ALOX12P2</i>	-1.42915	1.03E-60
<i>GGA2</i>	-0.80435	1.34E-60
<i>SLC14A1</i>	-3.98079	1.12E-57
<i>PDPN</i>	-0.89909	1.95E-57
<i>HOXC4</i>	1.838087	5.42E-57
<i>TP63</i>	-1.75332	2.64E-54
<i>CSTA</i>	-3.40411	5.01E-52
<i>ITGB6</i>	-1.32985	1.30E-51
<i>NIPAL3</i>	-2.07942	2.77E-51
<i>MCC</i>	-1.64977	3.14E-51
<i>LSAMP</i>	-2.01261	3.44E-51
<i>SKIL</i>	-0.79579	5.78E-50
<i>BCL11A</i>	-1.92313	1.30E-49
<i>SLC18A2</i>	-2.21571	9.53E-49
<i>CDC14B</i>	-0.97627	4.58E-48
<i>ACSBG1</i>	-0.63693	3.39E-47
<i>SLC15A1</i>	-0.99285	3.58E-46
<i>CYP3A5</i>	-2.4531	1.07E-45
<i>MR1</i>	-0.77817	2.34E-45
<i>DPYS</i>	-2.16919	2.57E-45
<i>FGFR2</i>	-0.93049	1.74E-44
<i>HOXD10</i>	-1.46069	3.77E-44
<i>TBC1D1</i>	-0.86673	4.30E-44
<i>PCDH7</i>	-1.11095	4.30E-44
<i>MKI67 (PAM50)</i>	0.809906	5.80E-44

Gene	Log2 FC	adj.P.Val
<i>HIST1H1B</i>	0.778639	1.12E-43
<i>PKD3</i>	-0.86954	1.86E-43
<i>CIB2</i>	-0.68038	3.22E-43
<i>EDN3</i>	-1.09101	3.43E-43
<i>NPR2</i>	-0.98667	1.10E-42
<i>ZHX2</i>	-1.06386	1.38E-42
<i>EFHD1</i>	-1.61638	2.40E-42
<i>PGAP1</i>	-1.28773	2.26E-41
<i>VWA5A</i>	-1.49095	3.74E-41
<i>CPM</i>	-1.91676	4.56E-41
<i>SRM</i>	1.120581	7.56E-41
<i>MEG3</i>	-1.0494	1.04E-40
<i>ELF4</i>	-1.1061	1.38E-40
<i>FADS1</i>	-1.50512	3.65E-40
<i>EPHA5</i>	-1.13827	5.32E-40
<i>LRP4</i>	-1.26499	6.07E-40
<i>TRIM29</i>	-1.96293	7.29E-40
<i>SPC25</i>	0.986493	1.28E-39
<i>ADRA1B</i>	0.575254	1.76E-39

Known PAM50 Genes are shown in brackets.

Table 55. Top 50 most significant DEGs in breast cancer molecular normal like.

Gene	Log2 FC	adj.P.Val
<i>CA4</i>	-2.09508	2.06E-107
<i>LHCGR</i>	-1.07178	8.08E-98
<i>GLYAT</i>	-1.50893	3.90E-79
<i>KANK3</i>	-1.73746	9.73E-79
<i>GPD1</i>	-1.67887	8.46E-70
<i>NPR1</i>	-1.37563	2.19E-67
<i>MYOM1</i>	-1.43402	6.48E-60
<i>LEP</i>	-3.67988	7.43E-59
<i>TIMP4</i>	-3.18297	8.90E-59
<i>CIDEA</i>	-2.61489	2.00E-57
<i>ACACB</i>	-1.83766	4.58E-56
<i>AQP7</i>	-1.37161	3.70E-55
<i>ACSM5</i>	-1.13955	6.42E-55
<i>ACADS</i>	-1.15054	7.62E-55
<i>RBP4</i>	-2.24848	7.62E-55
<i>SLC7A10</i>	-1.3649	1.38E-53
<i>HSPB7</i>	-0.98965	1.86E-53
<i>CLEC3B</i>	-2.87719	3.61E-53
<i>ANGPTL8</i>	-0.97636	1.30E-50
<i>SIK2</i>	-0.95648	4.31E-50
<i>TNMD</i>	-2.50932	4.42E-50
<i>MLXIPL</i>	-1.17958	1.49E-49
<i>ALDH1L1</i>	-0.86319	8.02E-48
<i>ATP1A2</i>	-1.57273	2.01E-47
<i>ITGA7</i>	-1.86776	2.47E-47
<i>SPTBN1</i>	-0.74953	1.88E-45
<i>RDH5</i>	-1.05233	2.56E-45
<i>CLDN5</i>	-1.76343	5.46E-45
<i>ADH1C</i>	-1.67563	3.60E-44
<i>SEMA3G</i>	-2.38635	8.11E-44
<i>ROBO4</i>	-1.03737	1.17E-43

Gene	Log2 FC	adj.P.Val
<i>PNPLA2</i>	-0.79363	1.82E-42
<i>TNXB</i>	-0.8374	4.89E-41
<i>LIMS2</i>	-1.16663	4.94E-41
<i>RAPGEF3</i>	-0.90685	6.33E-41
<i>TNXA</i>	-1.60356	1.49E-40
<i>TRHDE</i>	-1.02165	1.87E-40
<i>PGM5</i>	-0.83523	9.25E-40
<i>VEGFD</i>	-2.25798	1.40E-39
<i>HSPB2</i>	-1.29032	4.30E-39
<i>SAA4</i>	-1.52017	4.52E-39
<i>PDE2A</i>	-1.68676	1.42E-38
<i>SCN4A</i>	-0.8112	5.28E-38
<i>CRHBP</i>	-0.73501	1.18E-37
<i>ROBO3</i>	-1.03104	1.18E-37
<i>COPG2</i>	-0.93948	1.32E-37
<i>PPP1R1A</i>	-2.27661	1.32E-37
<i>SYN2</i>	-0.69823	2.64E-36
<i>AGPAT2</i>	-1.17258	3.73E-35
<i>TRDN</i>	-0.50355	5.71E-35

Known PAM50 Genes are shown in brackets.

Table 56. Top 50 most significant DEGs in breast cancer luminal A

Gene	Log2 FC	adj.P.Val
<i>CA4</i>	-2.33391	4.87E-158
<i>KANK3</i>	-2.1012	5.30E-135
<i>NPR1</i>	-1.76961	2.04E-129
<i>LHCGR</i>	-1.08202	9.84E-125
<i>GPD1</i>	-1.94631	1.92E-113
<i>MYOM1</i>	-1.69729	5.86E-102
<i>GLYAT</i>	-1.48284	1.38E-97
<i>TNXA</i>	-2.28657	1.08E-96
<i>LEP</i>	-4.25839	1.08E-96
<i>VEGFD</i>	-3.26074	1.13E-96
<i>SIK2</i>	-1.20932	3.24E-96
<i>TNXB</i>	-1.17978	5.06E-96
<i>CLDN5</i>	-2.35329	2.19E-95
<i>ADH1C</i>	-2.22698	3.50E-93
<i>ACACB</i>	-2.13157	7.73E-93
<i>CIDEA</i>	-2.97965	3.30E-92
<i>AQP7</i>	-1.59433	8.93E-92
<i>ITGA7</i>	-2.36448	1.25E-91
<i>FXD1</i>	-2.06186	1.07E-90
<i>RBP4</i>	-2.58773	9.86E-90
<i>LIMS2</i>	-1.56212	2.46E-88
<i>HSPB2</i>	-1.74929	3.09E-86
<i>CLEC3B</i>	-3.28862	4.43E-86
<i>SPTBN1</i>	-0.93225	8.72E-86
<i>PDE2A</i>	-2.29318	1.27E-85
<i>ALDH1L1</i>	-1.04176	1.66E-85
<i>HSPB7</i>	-1.12226	1.83E-85
<i>SLC7A10</i>	-1.53251	2.84E-84
<i>RDH5</i>	-1.29134	1.17E-83
<i>TNS1</i>	-1.44793	7.85E-82
<i>ATP1A2</i>	-1.85476	1.32E-81

Gene	Log2 FC	adj.P.Val
<i>ROBO3</i>	-1.36952	1.49E-80
<i>TNMD</i>	-2.84926	1.49E-80
<i>ITIH5</i>	-2.68099	1.09E-79
<i>MLXIPL</i>	-1.33794	1.32E-79
<i>SEMA3G</i>	-2.8925	2.01E-79
<i>TIMP4</i>	-3.26644	3.15E-79
<i>ROBO4</i>	-1.2471	4.84E-78
<i>RAPGEF3</i>	-1.11418	2.94E-76
<i>SAA4</i>	-1.90954	8.09E-76
<i>PPP1R1A</i>	-2.87701	6.95E-74
<i>ANGPTL8</i>	-1.03834	8.76E-73
<i>SAA2</i>	-3.73463	1.18E-72
<i>DENND2A</i>	-1.12439	9.40E-72
<i>SYN2</i>	-0.87911	4.56E-71
<i>ACSM5</i>	-1.13939	7.12E-71
<i>ALDH1A2</i>	-1.1517	1.26E-69
<i>CDKN1C</i>	-1.9497	8.23E-68
<i>CX3CL1</i>	-2.74652	3.53E-67
<i>MOCS1</i>	-1.02648	5.18E-66

Known PAM50 Genes are shown in brackets.

Table 57. Top 50 most significant DEGs in breast cancer luminal B.

Gene	Log2 FC	adj.P.Val
<i>CA4</i>	-2.38567	3.92E-154
<i>KANK3</i>	-2.10284	3.26E-127
<i>NPR1</i>	-1.75311	6.45E-120
<i>LHCGR</i>	-1.08635	5.08E-118
<i>GPD1</i>	-2.0227	2.04E-113
<i>SIK2</i>	-1.35496	1.15E-108
<i>FXVD1</i>	-2.38344	9.04E-108
<i>TNXA</i>	-2.47418	2.24E-103
<i>TNXB</i>	-1.27157	7.15E-102
<i>CLDN5</i>	-2.53902	2.17E-101
<i>ITGA7</i>	-2.57017	1.51E-98
<i>LEP</i>	-4.46563	2.91E-98
<i>MYOM1</i>	-1.71823	6.53E-98
<i>LIMS2</i>	-1.70466	1.05E-95
<i>SPTBN1</i>	-1.0344	1.40E-95
<i>CIDEA</i>	-3.15148	6.01E-95
<i>RBP4</i>	-2.77895	9.87E-95
<i>SEMA3G</i>	-3.32633	5.29E-94
<i>GLYAT</i>	-1.50311	5.32E-94
<i>ADH1C</i>	-2.32367	6.06E-94
<i>VEGFD</i>	-3.30855	3.67E-93
<i>AQP7</i>	-1.65441	1.24E-91
<i>ITIH5</i>	-3.01276	7.51E-91
<i>ATP1A2</i>	-2.03719	1.43E-89
<i>PDE2A</i>	-2.44761	1.43E-89
<i>CLEC3B</i>	-3.49802	1.43E-89
<i>ACACB</i>	-2.15084	1.48E-88
<i>TIMP4</i>	-3.60583	1.07E-87
<i>TNS1</i>	-1.56241	3.20E-87
<i>HSPB2</i>	-1.7949	1.71E-84
<i>ROBO4</i>	-1.35368	5.47E-84

Gene	Log2 FC	adj.P.Val
<i>RAPGEF3</i>	-1.21882	3.44E-83
<i>ALDH1L1</i>	-1.06041	1.31E-82
<i>NUSAP1</i>	3.182897	1.77E-82
<i>HSPB7</i>	-1.13123	3.34E-81
<i>TNMD</i>	-2.93947	1.46E-79
<i>PAMR1</i>	-2.20076	1.80E-78
<i>RRM2 (PAM50)</i>	3.935397	2.85E-78
<i>CDKN1C</i>	-2.19523	7.09E-78
<i>ALDH1A2</i>	-1.27146	2.87E-77
<i>ROBO3</i>	-1.38563	5.40E-77
<i>PPP1R1A</i>	-3.0514	1.55E-76
<i>ADAMTS5</i>	-2.8318	2.14E-76
<i>HOXA5</i>	-3.386	3.91E-76
<i>ACSM5</i>	-1.23391	3.99E-76
<i>SORBS1</i>	-1.54996	3.99E-76
<i>IGFBP6</i>	-2.82122	5.84E-75
<i>SLC7A10</i>	-1.48492	6.81E-75
<i>ZWINT</i>	2.824858	1.56E-73
<i>LMOD1</i>	-1.65843	9.56E-73

Known PAM50 Genes are shown in brackets.

Table 58. Top 50 most significant DEGs in breast molecular HER2.

Gene	Log2 FC	adj.P.Val
<i>CA4</i>	-2.41796	5.34E-167
<i>KANK3</i>	-2.20085	2.01E-145
<i>LHCGR</i>	-1.1142	6.75E-131
<i>NPR1</i>	-1.74786	2.03E-127
<i>GPD1</i>	-2.0613	9.39E-125
<i>ACACB</i>	-2.4564	3.53E-117
<i>CLDN5</i>	-2.59409	4.46E-112
<i>GLYAT</i>	-1.59892	6.52E-111
<i>MYOM1</i>	-1.77842	1.31E-110
<i>CLEC3B</i>	-3.81172	4.93E-110
<i>FXYD1</i>	-2.30403	3.83E-109
<i>ITGA7</i>	-2.62511	3.93E-109
<i>SEMA3G</i>	-3.49738	4.58E-109
<i>TIMP4</i>	-3.94615	1.12E-108
<i>LEP</i>	-4.5373	5.92E-108
<i>CIDEA</i>	-3.26324	1.46E-107
<i>TNXA</i>	-2.40164	1.48E-105
<i>TNXB</i>	-1.22408	1.37E-102
<i>PDE2A</i>	-2.54554	3.28E-102
<i>ATP1A2</i>	-2.09861	1.28E-100
<i>VEGFD</i>	-3.30492	1.68E-99
<i>AQP7</i>	-1.66788	2.15E-99
<i>RBP4</i>	-2.72026	6.00E-98
<i>SIK2</i>	-1.20361	4.26E-96
<i>SPTBN1</i>	-0.9811	7.74E-94
<i>ACSM5</i>	-1.33447	5.54E-93
<i>ITIH5</i>	-2.9167	3.59E-92
<i>HSPB7</i>	-1.16198	5.84E-91
<i>LIMS2</i>	-1.5804	1.61E-90
<i>TNS1</i>	-1.52095	2.84E-89
<i>ADH1C</i>	-2.15533	8.52E-89

Gene	Log2 FC	adj.P.Val
<i>RAPGEF3</i>	-1.21269	2.28E-88
<i>ALDH1L1</i>	-1.05359	1.51E-87
<i>TNMD</i>	-2.97365	5.99E-87
<i>RRM2</i> (PAM50)	4.016526	7.64E-87
<i>RDH5</i>	-1.30972	4.52E-86
<i>ROBO4</i>	-1.31151	2.40E-85
<i>TNS2</i>	-2.16086	6.26E-85
<i>ACADS</i>	-1.27525	1.08E-84
<i>HSPB2</i>	-1.70747	2.87E-83
<i>SLC7A10</i>	-1.50975	1.42E-82
<i>PGM5</i>	-1.08447	6.42E-82
<i>PALM</i>	-1.85737	1.03E-80
<i>CDKN1C</i>	-2.14825	1.66E-80
<i>IGFBP6</i>	-2.81853	3.02E-80
<i>CBX7</i>	-2.0667	6.98E-80
<i>NUSAP1</i>	2.979143	4.61E-79
<i>ALDH1A2</i>	-1.22407	1.09E-77
<i>EHD2</i>	-1.57705	4.58E-77
<i>MLXIPL</i>	-1.3076	6.38E-77

Known PAM50 Genes are shown in brackets.

Table 59. Top 50 most significant DEGs in breast molecular Basal.

Gene	Log2 FC	adj.P.Val
CA4	-2.37541	5.56E-163
KANK3	-2.19167	9.10E-145
ACACB	-2.58572	4.22E-127
LHCGR	-1.08705	2.84E-126
GPD1	-2.02722	9.99E-122
CLEC3B	-4.03348	1.51E-120
CLDN5	-2.64287	9.80E-116
FXYP1	-2.38348	2.13E-115
SEMA3G	-3.58604	1.18E-113
NPR1	-1.62132	1.34E-113
TNXA	-2.49704	1.66E-112
TNXB	-1.26851	8.33E-109
ATP1A2	-2.17795	5.55E-107
PDE2A	-2.60195	4.84E-106
CIDEA	-3.22008	2.00E-105
GLYAT	-1.53775	1.48E-104
TIMP4	-3.84684	1.58E-104
RAPGEF3	-1.33793	4.54E-104
ACSM5	-1.42215	1.88E-103
TNS2	-2.43227	1.88E-103
LEP	-4.39196	1.10E-102
MYOM1	-1.65971	2.53E-99
ITGA7	-2.47197	2.53E-99
VEGFD	-3.29198	4.10E-99
AQP7	-1.65792	1.15E-98
LIMS2	-1.65339	8.66E-98
RBP4	-2.71312	8.66E-98
TNS1	-1.6024	1.32E-97
ADH1C	-2.2743	2.41E-97
PALM	-2.05813	4.60E-96
RRM2 (PAM50)	4.232733	5.27E-95

Gene	Log2 FC	adj.P.Val
<i>IGFBP6</i>	-3.10858	8.79E-95
<i>NUSAP1</i>	3.300774	3.92E-94
<i>MELK</i> (PAM50)	3.744769	9.14E-94
<i>CEP55</i> (PAM50)	3.898225	3.16E-92
<i>CKS2</i>	3.801096	5.47E-92
<i>ITIH5</i>	-2.89374	2.81E-91
<i>UBE2C</i> (PAM50)	3.360469	1.23E-90
<i>CBX7</i>	-2.22423	1.23E-90
<i>CFD</i>	-4.90723	4.36E-90
<i>TPX2</i>	3.610411	1.25E-89
<i>ROBO4</i>	-1.34334	3.95E-89
<i>EHD2</i>	-1.70074	8.05E-88
<i>PGM5</i>	-1.11943	1.13E-86
<i>PRC1</i>	3.497016	5.97E-85
<i>HSPB7</i>	-1.11044	1.31E-84
<i>ZBTB16</i>	-3.87314	2.28E-84
<i>LMOD1</i>	-1.73284	2.63E-84
<i>BUB1B</i>	3.310458	3.69E-83
<i>ALDH1A2</i>	-1.26846	8.76E-83

Known PAM50 Genes are shown in brackets.

Table 60. Top 50 most significant DEGs in ovarian cancer serous.

Gene	Log2 FC	adj.P.Val
<i>HAND2-AS1</i>	-2.67506	9.12E-185
<i>C21orf62</i>	-1.71183	1.41E-124
<i>LOC100507387</i>	-1.86751	1.66E-122
<i>GPRASP1</i>	-3.23016	2.98E-117
<i>PDE8B</i>	-2.15653	2.59E-115
<i>ABCA8</i>	-2.36854	8.69E-100
<i>CLEC4M</i>	-2.04893	3.65E-99
<i>ALDH1A2</i>	-2.52996	2.66E-96
<i>RADX</i>	-1.16439	8.84E-96
<i>WNT2B</i>	-2.15618	1.48E-95
<i>PEX5L</i>	-1.04217	3.00E-95
<i>FAM153B</i>	-2.65337	1.70E-92
<i>AOX1</i>	-2.82546	7.27E-90
<i>CLDN15</i>	-1.59138	7.27E-90
<i>REEP1</i>	-3.22976	2.48E-87
<i>NDNF</i>	-2.88006	4.47E-87
<i>CD24</i>	3.539177	3.91E-86
<i>PRG4</i>	-2.3121	2.08E-85
<i>BDH2</i>	-1.95635	4.71E-83
<i>GSDME</i>	-2.72138	4.78E-83
<i>SYNE1</i>	-1.11585	1.93E-82
<i>FLRT2</i>	-2.49199	2.34E-82
<i>SNCAIP</i>	-1.51575	1.20E-81
<i>PKD2</i>	-2.16222	4.04E-80
<i>OGN</i>	-4.25528	2.93E-79
<i>KDR</i>	-2.13719	1.10E-78
<i>SNCA</i>	-1.28464	5.36E-78
<i>CACNB2</i>	-0.86585	6.37E-78
<i>BUB1B</i>	2.844556	3.92E-77
<i>GNG11</i>	-2.26222	2.00E-76
<i>NELL2</i>	-3.31966	2.16E-76

Gene	Log2 FC	adj.P.Val
<i>CELF2</i>	-1.57261	7.20E-74
<i>PTGDR</i>	-0.73847	8.03E-74
<i>PTTG1</i> (PAM50)	2.512072	8.62E-74
<i>PRDM5</i>	-0.67311	6.43E-72
<i>CKS2</i>	2.553683	8.22E-72
<i>CENPF</i> (PAM50)	1.700463	3.32E-71
<i>CEP55</i> (PAM50)	3.22718	3.53E-71
<i>PCDH17</i>	-2.39234	3.58E-71
<i>BNC1</i>	-2.27624	1.21E-70
<i>AURKA</i>	1.758035	2.19E-70
<i>TCEAL2</i>	-3.83052	2.70E-70
<i>FRY</i>	-1.31956	7.21E-70
<i>CDCA8</i>	1.932543	2.19E-69
<i>HAS1</i>	-1.48661	8.48E-69
<i>RACGAP1</i>	2.243754	1.03E-68
<i>SESN1</i>	-1.51331	1.71E-68
<i>FAM13C</i>	-1.22655	3.92E-68
<i>HLF</i>	-2.10146	3.92E-68
<i>STX2</i>	-0.96431	4.41E-68

Known PAM50 Genes are shown in brackets.

Table 61. Top 50 most significant DEGs in ovarian cancer endometrioid.

Gene	Log2 FC	adj.P.Val
<i>HAND2-AS1</i>	-2.6228	2.86E-134
<i>LOC100507387</i>	-1.91538	5.42E-90
<i>C21orf62</i>	-1.65235	8.53E-84
<i>FAM153B</i>	-2.9499	3.90E-75
<i>ALDH1A2</i>	-2.72699	6.86E-75
<i>GPRASP1</i>	-3.01729	2.52E-74
<i>CLEC4M</i>	-2.06528	1.48E-69
<i>RADX</i>	-1.1711	1.10E-66
<i>AOX1</i>	-2.94251	8.85E-66
<i>CLDN15</i>	-1.65469	1.18E-65
<i>REEP1</i>	-3.30616	4.90E-62
<i>CD24</i>	3.612618	8.09E-61
<i>SYNE1</i>	-1.14565	1.59E-58
<i>ABCA8</i>	-2.13289	2.01E-58
<i>WNT2B</i>	-1.98341	1.55E-57
<i>PKD2</i>	-2.23712	1.56E-57
<i>CACNB2</i>	-0.90977	2.78E-57
<i>FRY</i>	-1.47421	1.78E-56
<i>SNCA</i>	-1.33661	1.86E-56
<i>TCEAL2</i>	-4.22178	1.34E-55
<i>STX2</i>	-1.07709	5.97E-55
<i>PEX5L</i>	-0.93339	6.41E-55
<i>GSDME</i>	-2.63258	1.42E-53
<i>EPCAM</i>	3.006439	2.91E-52
<i>PDE8B</i>	-1.63962	3.24E-52
<i>PRG4</i>	-2.16067	3.59E-52
<i>ARHGAP44</i>	-1.41727	5.76E-52
<i>KDR</i>	-2.09587	1.08E-51
<i>SNCAIP</i>	-1.44441	2.68E-51
<i>GFPT2</i>	-2.53964	8.26E-51
<i>FLRT2</i>	-2.33996	1.55E-50

Gene	Log2 FC	adj.P.Val
<i>NDNF</i>	-2.5914	4.54E-50
<i>HAS1</i>	-1.55497	7.22E-50
<i>AKT3</i>	-1.27276	1.39E-49
<i>BUB1B</i>	2.754415	1.51E-49
<i>BNC1</i>	-2.33041	1.64E-49
<i>CKS2</i>	2.586096	1.69E-49
<i>CBX7</i>	-1.72834	7.86E-49
<i>HLF</i>	-2.18286	8.32E-49
<i>TMEM255A</i>	-2.98656	8.99E-49
<i>ZWINT</i>	2.643745	1.75E-48
<i>OLFML1</i>	-2.65299	2.50E-48
<i>NAP1L3</i>	-3.33838	3.39E-48
<i>BNC2</i>	-2.17049	5.71E-48
<i>BDH2</i>	-1.76539	8.10E-48
<i>ZFPM2</i>	-3.71787	2.15E-47
<i>PRDM5</i>	-0.65374	3.27E-46
<i>PLCL2</i>	-1.13936	6.40E-46
<i>MAF</i>	-1.3729	2.20E-45
<i>PROCR</i>	-2.32997	4.48E-45

Known PAM50 Genes are shown in brackets.

Table 62. Top 50 most significant DEGs in ovarian cancer mucinous.

Gene	Log2 FC	adj.P.Val
<i>HAND2-AS1</i>	-2.68236	1.33E-96
<i>LOC100507387</i>	-1.93822	6.78E-61
<i>FAM153B</i>	-3.1252	4.80E-54
<i>C21orf62</i>	-1.61572	3.16E-53
<i>PDE8B</i>	-2.14824	5.79E-53
<i>INAVA</i>	4.168723	3.38E-50
<i>ALDH1A2</i>	-2.66554	4.65E-47
<i>LGALS4</i>	3.664237	6.85E-46
<i>CELF2</i>	-1.9385	1.36E-45
<i>TFF2</i>	2.909556	7.92E-45
<i>CD24</i>	3.862293	5.79E-44
<i>SI</i>	1.991527	8.90E-44
<i>TCEAL2</i>	-4.73861	1.66E-43
<i>ANXA10</i>	2.814647	1.21E-42
<i>AGMAT</i>	1.566965	2.27E-41
<i>TM4SF5</i>	1.667538	6.18E-41
<i>APOBEC1</i>	1.65158	7.53E-41
<i>FUT3</i>	1.717657	1.29E-40
<i>SYNE1</i>	-1.19427	1.54E-40
<i>MPDZ</i>	-2.14088	7.74E-40
<i>MUC5AC</i>	1.495833	1.27E-39
<i>TFF1</i>	4.460571	1.45E-39
<i>PKD2</i>	-2.32285	1.53E-39
<i>PRSS3</i>	2.171331	2.12E-39
<i>BCL2L14</i>	0.849211	2.12E-39
<i>PEX5L</i>	-0.99479	2.12E-39
<i>VIL1</i>	1.807502	5.30E-39
<i>GFPT2</i>	-2.80703	1.08E-38
<i>EPS8L3</i>	2.027445	1.42E-38
<i>AOX1</i>	-2.76611	1.76E-38
<i>NELL2</i>	-3.60626	2.86E-38

Gene	Log2 FC	adj.P.Val
<i>EPCAM</i>	3.242815	2.86E-38
<i>ANKRD40CL</i>	1.549938	3.05E-38
<i>GPRASP1</i>	-2.60491	6.00E-38
<i>PRG4</i>	-2.32042	6.51E-38
<i>TRIM15</i>	0.920305	1.05E-37
<i>CLEC4M</i>	-1.84451	1.83E-37
<i>PIK3C2B</i>	1.651996	4.47E-37
<i>BNC1</i>	-2.53772	9.70E-37
<i>BAK1</i>	1.416054	1.02E-36
<i>TMEM255A</i>	-3.26867	1.75E-36
<i>NPC1L1</i>	0.806698	1.78E-36
<i>CDHR5</i>	1.198287	4.20E-36
<i>NR1I2</i>	1.178768	6.86E-36
<i>PLS1</i>	3.031519	8.22E-36
<i>SNCAIP</i>	-1.50774	1.86E-35
<i>SPINK4</i>	1.75803	8.26E-35
<i>TMPRSS4</i>	3.822974	3.57E-34
<i>CEACAM5</i>	1.87886	3.57E-34
<i>CEACAM6</i>	4.450116	5.55E-34

Known PAM50 Genes are shown in brackets.

Table 63. Top 50 most significant DEGs in ovarian cancer clear cell.

Gene	Log2 FC	adj.P.Val
<i>HAND2-AS1</i>	-2.65958	5.49E-108
<i>C21orf62</i>	-1.62282	1.39E-61
<i>GPRASP1</i>	-3.19146	2.43E-61
<i>LOC100507387</i>	-1.73388	1.99E-58
<i>LBP</i>	2.659436	2.11E-58
<i>HAVCR1</i>	2.073131	2.33E-56
<i>REEP1</i>	-3.68831	8.45E-56
<i>CD24</i>	4.077482	8.45E-56
<i>ALDH1A2</i>	-2.69035	1.86E-55
<i>PDE8B</i>	-2.01921	2.55E-55
<i>SNCA</i>	-1.56032	9.29E-55
<i>FAM153B</i>	-2.84319	1.14E-53
<i>RADX</i>	-1.20526	9.59E-53
<i>CLEC4M</i>	-2.0414	1.65E-51
<i>FXVD2</i>	2.182469	3.87E-50
<i>SNCAIP</i>	-1.68989	6.97E-50
<i>PEX5L</i>	-1.03674	5.53E-49
<i>ABCA8</i>	-2.26179	2.40E-48
<i>OGN</i>	-4.71018	1.04E-47
<i>WNT2B</i>	-2.08707	5.67E-47
<i>INAVA</i>	3.66643	5.67E-47
<i>CLDN15</i>	-1.5983	1.30E-46
<i>FRY</i>	-1.55184	4.28E-46
<i>ZFPM2</i>	-4.34827	5.28E-46
<i>PLCL2</i>	-1.35589	6.09E-46
<i>TCEAL2</i>	-4.46418	9.92E-46
<i>CACNB2</i>	-0.94356	1.15E-45
<i>CCNE1 (PAM50)</i>	2.157852	1.56E-45
<i>LAMC1</i>	2.149501	3.34E-45
<i>DTWD1</i>	-0.8418	3.72E-45
<i>PRG4</i>	-2.35281	4.45E-45

Gene	Log2 FC	adj.P.Val
<i>PDGFD</i>	-2.48421	2.00E-44
<i>FLRT2</i>	-2.55233	1.00E-43
<i>SLC25A15</i>	1.846205	1.00E-43
<i>RRM2</i> (PAM50)	3.52174	1.77E-43
<i>KPNA5</i>	-1.35624	3.90E-43
<i>BNC1</i>	-2.54982	4.84E-43
<i>VAPA</i>	-1.13739	5.23E-43
<i>ANG</i>	-3.09019	3.15E-42
<i>NDNF</i>	-2.7796	4.13E-42
<i>EPCAM</i>	3.126884	8.73E-42
<i>CALB2</i>	-2.54633	3.38E-41
<i>NELL2</i>	-3.44078	4.82E-41
<i>MECOM</i>	2.04267	4.82E-41
<i>CKS2</i>	2.748026	6.31E-41
<i>PSAT1</i>	3.563642	6.75E-41
<i>PLSCR4</i>	-3.3188	6.75E-41
<i>MPDZ</i>	-1.98369	8.88E-41
<i>TMEM255A</i>	-3.18976	1.10E-40
<i>DNMT3B</i>	2.005804	3.21E-40

Known PAM50 Genes are shown in brackets.

Table 64. Top 50 most significant DEGs in prostate Gleason score 4.

Gene	Log2 FC	adj.P.Val
<i>MOG</i>	0.960844	5.19E-12
<i>CENPF</i> (PAM50)	1.65075	2.01E-11
<i>TRAF3IP2</i>	-1.79677	2.01E-11
<i>MKI67</i> (PAM50)	1.401803	3.55E-11
<i>AURKA</i>	1.511434	1.24E-10
<i>SUPT3H</i>	-1.73317	2.56E-10
<i>ZNF711</i>	-2.43264	7.08E-10
<i>SNAPC4</i>	1.371348	8.75E-09
<i>IHH</i>	1.349173	9.07E-09
<i>MYEF2</i>	1.019572	1.28E-08
<i>VSNL1</i>	-2.00262	1.28E-08
<i>FAM155B</i>	1.77558	1.37E-08
<i>GLDC</i>	2.212495	2.72E-08
<i>KLHL25</i>	0.953727	3.64E-08
<i>EZH2</i>	2.089204	4.20E-08
<i>LRP4</i>	-1.97456	4.41E-08
<i>SOX12</i>	1.1912	4.41E-08
<i>POU6F2</i>	1.018926	4.70E-08
<i>NUP210</i>	1.369325	5.35E-08
<i>NAA10</i>	1.244621	6.10E-08
<i>ZHX2</i>	-1.54081	1.31E-07
<i>HIST1H1B</i>	1.100577	1.43E-07
<i>BIRC5</i> (PAM50)	1.363489	1.43E-07
<i>SLC14A1</i>	-4.64177	1.64E-07
<i>ZNF280A</i>	0.759135	2.01E-07
<i>VAV2</i>	0.974138	2.01E-07
<i>CCNA2</i>	1.328618	2.01E-07
<i>ALOX15</i>	1.187138	2.01E-07
<i>HABP2</i>	1.139064	2.01E-07
<i>GRK6</i>	0.771272	2.01E-07
<i>ZNF443</i>	0.806564	2.10E-07

Gene	Log2 FC	adj.P.Val
<i>NIPAL3</i>	-2.58397	2.46E-07
<i>SLC25A42</i>	0.883847	2.46E-07
<i>LAMB3</i>	-2.91108	2.46E-07
<i>TP63</i>	-2.09566	2.46E-07
<i>ZBTB20</i>	-2.16972	2.59E-07
<i>PAFAH1B1</i>	-1.17522	5.58E-07
<i>CSTA</i>	-4.03484	7.42E-07
<i>HPCAL1</i>	1.096508	7.42E-07
<i>TRIM29</i>	-2.78591	7.84E-07
<i>C1orf61</i>	0.682584	7.97E-07
<i>INHBA</i>	2.197402	9.62E-07
<i>TIMM8A</i>	0.779811	9.77E-07
<i>DSTN</i>	-1.67093	9.77E-07
<i>SNHG3</i>	2.184858	1.11E-06
<i>RHBDD3</i>	0.960424	1.21E-06
<i>GGA2</i>	-0.84148	1.27E-06
<i>SERINC1</i>	-1.7062	1.29E-06
<i>SLC5A6</i>	1.238867	1.31E-06
<i>NLE1</i>	1.161909	1.31E-06

Known PAM50 Genes are shown in brackets.

Table 65. Top 50 most significant DEGs in prostate Gleason score 5.

Gene	Log2 FC	adj.P.Val
<i>SUPT3H</i>	-1.75374	2.16E-39
<i>VSNL1</i>	-2.12208	1.31E-36
<i>ZNF711</i>	-2.02254	1.78E-27
<i>LAMB3</i>	-2.78306	6.25E-26
<i>LUZP2</i>	2.85598	1.11E-25
<i>GGA2</i>	-0.81983	2.09E-24
<i>CDC14B</i>	-1.1483	8.45E-24
<i>ALOX12P2</i>	-1.42618	8.45E-24
<i>ITGB6</i>	-1.451	1.01E-22
<i>ACSBG1</i>	-0.71492	2.01E-21
<i>NIPAL3</i>	-2.19486	2.11E-21
<i>SLC15A1</i>	-1.12622	2.11E-21
<i>TRAF3IP2</i>	-1.17494	2.99E-21
<i>TP63</i>	-1.74877	1.05E-20
<i>SLC14A1</i>	-3.59748	8.50E-19
<i>SKIL</i>	-0.79069	1.53E-18
<i>MYEF2</i>	0.720841	1.53E-18
<i>PDPN</i>	-0.80116	3.24E-18
<i>CPM</i>	-2.14479	3.55E-18
<i>CIB2</i>	-0.73025	7.84E-18
<i>ABO</i>	-0.9071	1.09E-17
<i>HOXC4</i>	1.598597	2.71E-17
<i>CYP4B1</i>	-2.48419	9.94E-17
<i>LSAMP</i>	-1.8535	9.94E-17
<i>PCDH7</i>	-1.13155	1.20E-16
<i>SLC18A2</i>	-2.08087	2.78E-16
<i>MR1</i>	-0.75302	8.95E-16
<i>EFHD1</i>	-1.64039	9.00E-16
<i>NTRK3</i>	-0.85025	9.93E-16
<i>TBC1D1</i>	-0.85029	1.18E-15
<i>DPYS</i>	-2.08462	1.33E-15

Gene	Log2 FC	adj.P.Val
<i>LINC01558</i>	-0.90823	1.68E-15
<i>ZNF154</i>	-0.78339	2.22E-15
<i>FGFR2</i>	-0.89762	2.46E-15
<i>NRG1</i>	-0.48628	2.55E-15
<i>CSTA</i>	-2.91069	4.02E-15
<i>EPHA5</i>	-1.1546	6.72E-15
<i>GART</i>	0.625267	6.72E-15
<i>MECOM</i>	-0.88677	7.15E-15
<i>VEGFA</i>	-2.94291	9.47E-15
<i>CELSR1</i>	-0.72341	1.32E-14
<i>LRP4</i>	-1.26321	1.40E-14
<i>NDRG2</i>	-1.12182	2.05E-14
<i>GIPR</i>	-0.6569	2.35E-14
<i>NPR2</i>	-0.94089	2.43E-14
<i>SLC25A42</i>	0.595964	2.51E-14
<i>ELF4</i>	-1.08978	2.51E-14
<i>CGREF1</i>	1.579313	3.15E-14
<i>FADS1</i>	-1.47028	3.23E-14
<i>ANXA2</i>	-1.5767	3.47E-14

Known PAM50 Genes are shown in brackets.

Table 66. Top 50 most significant DEGs in prostate Gleason score 6.

Gene	Log2 FC	adj.P.Val
<i>SUPT3H</i>	-1.77599	1.56E-114
<i>VSNL1</i>	-2.08077	3.30E-105
<i>TRAF3IP2</i>	-1.49559	7.46E-98
<i>LAMB3</i>	-2.91593	5.45E-90
<i>ZNF711</i>	-1.9761	2.11E-86
<i>GGA2</i>	-0.82884	1.70E-82
<i>SLC14A1</i>	-4.06167	2.20E-78
<i>TP63</i>	-1.86591	4.43E-78
<i>LUZP2</i>	2.601961	2.66E-75
<i>ALOX12P2</i>	-1.33533	1.83E-73
<i>ITGB6</i>	-1.38862	4.79E-73
<i>HOXC4</i>	1.779915	1.61E-72
<i>PDPN</i>	-0.86546	1.61E-72
<i>MCC</i>	-1.68624	4.98E-71
<i>BCL11A</i>	-1.97049	1.10E-69
<i>SKIL</i>	-0.81149	8.63E-69
<i>CDC14B</i>	-1.00093	3.97E-67
<i>CSTA</i>	-3.29447	5.30E-67
<i>PCDH7</i>	-1.18149	8.15E-65
<i>NIPAL3</i>	-1.98389	8.15E-65
<i>INHBA</i>	1.753077	4.17E-64
<i>ACSBG1</i>	-0.63667	1.00E-63
<i>SLC18A2</i>	-2.15889	2.99E-63
<i>CGREF1</i>	1.748158	5.79E-62
<i>FGFR2</i>	-0.94399	2.07E-61
<i>ZHX2</i>	-1.09585	9.69E-61
<i>MKI67 (PAM50)</i>	0.805177	1.10E-60
<i>TRIM29</i>	-2.11427	1.44E-60
<i>SLC15A1</i>	-0.95061	4.67E-59
<i>PKD3</i>	-0.85623	9.76E-59
<i>CIB2</i>	-0.68094	1.06E-58

Gene	Log2 FC	adj.P.Val
<i>CYP3A5</i>	-2.32133	1.52E-57
<i>HIST1H1B</i>	0.747819	1.07E-56
<i>BIRC5</i> (PAM50)	0.919168	4.76E-56
<i>ELF4</i>	-1.11577	5.63E-56
<i>TBC1D1</i>	-0.82454	6.67E-56
<i>VWA5A</i>	-1.481	9.99E-56
<i>SRM</i>	1.110015	3.41E-55
<i>LSAMP</i>	-1.71264	3.94E-55
<i>HOXD10</i>	-1.37143	3.94E-55
<i>EFHD1</i>	-1.54975	2.80E-54
<i>MR1</i>	-0.70691	7.70E-54
<i>LRP4</i>	-1.23631	2.25E-53
<i>LINC01558</i>	-0.85715	2.25E-53
<i>DST</i>	-1.06947	2.25E-53
<i>JUP</i>	-1.43034	2.88E-53
<i>RGS10</i>	0.951154	4.64E-53
<i>GABRE</i>	-2.19983	2.51E-52
<i>NPR2</i>	-0.9143	3.92E-52
<i>COL17A1</i>	-1.20453	4.20E-52

Known PAM50 Genes are shown in brackets.

Table 67. Top 50 most significant DEGs in prostate Gleason score 7.

Gene	Log2 FC	adj.P.Val
<i>SUPT3H</i>	-1.85968	1.96E-141
<i>VSNL1</i>	-2.17405	5.64E-131
<i>TRAF3IP2</i>	-1.50381	3.59E-118
<i>LAMB3</i>	-2.89227	5.75E-108
<i>ZNF711</i>	-2.02109	1.39E-107
<i>GGA2</i>	-0.84112	2.54E-102
<i>SLC14A1</i>	-4.24759	4.31E-101
<i>LUZP2</i>	2.727414	9.27E-98
<i>TP63</i>	-1.88532	6.19E-97
<i>ALOX12P2</i>	-1.4002	1.00E-95
<i>ITGB6</i>	-1.45482	3.83E-95
<i>HOXC4</i>	1.857511	4.56E-94
<i>PDPN</i>	-0.88996	1.71E-92
<i>MCC</i>	-1.70842	3.38E-89
<i>CSTA</i>	-3.46805	1.63E-88
<i>BCL11A</i>	-1.95227	2.36E-85
<i>PCDH7</i>	-1.20899	5.04E-83
<i>CDC14B</i>	-0.99697	5.17E-83
<i>MKI67 (PAM50)</i>	0.864996	5.20E-83
<i>NIPAL3</i>	-2.02676	6.34E-83
<i>SKIL</i>	-0.78721	2.81E-82
<i>SLC18A2</i>	-2.20645	3.98E-81
<i>MR1</i>	-0.80631	2.11E-80
<i>SRM</i>	1.24074	3.60E-80
<i>CYP3A5</i>	-2.49916	1.90E-79
<i>BIRC5 (PAM50)</i>	1.006009	3.12E-79
<i>ZHX2</i>	-1.12923	4.77E-79
<i>SLC15A1</i>	-0.99252	2.53E-78
<i>LSAMP</i>	-1.86788	1.04E-77
<i>VWA5A</i>	-1.59842	1.81E-77
<i>INHBA</i>	1.701586	3.67E-77

Gene	Log2 FC	adj.P.Val
<i>FGFR2</i>	-0.9444	5.30E-77
<i>VAR5</i>	0.832368	6.83E-77
<i>CGREF1</i>	1.729253	1.41E-76
<i>ACSBG1</i>	-0.61675	1.42E-76
<i>HIST1H1B</i>	0.786951	2.58E-76
<i>ELF4</i>	-1.17291	2.51E-75
<i>EDN3</i>	-1.10768	2.75E-75
<i>NPR2</i>	-1.00717	6.03E-75
<i>TBC1D1</i>	-0.86163	1.10E-74
<i>PGAP1</i>	-1.313	6.99E-73
<i>PDK3</i>	-0.84577	9.52E-73
<i>MEG3</i>	-1.07713	1.74E-72
<i>LRP4</i>	-1.3041	1.80E-72
RRM2 (PAM50)	2.271539	1.84E-72
<i>TRIM29</i>	-2.03107	2.94E-72
<i>ADRA1B</i>	0.600239	4.27E-72
<i>CYP4B1</i>	-2.34217	1.54E-71
<i>DST</i>	-1.11238	3.52E-71
<i>COL4A6</i>	-1.52673	1.19E-70

Known PAM50 Genes are shown in brackets.

Table 68. Top 50 most significant DEGs in prostate Gleason score 8.

Gene	Log2 FC	adj.P.Val
<i>SUPT3H</i>	-1.94701	1.65E-59
<i>VSNL1</i>	-2.30081	8.16E-54
<i>TRAF3IP2</i>	-1.74225	6.99E-53
<i>LAMB3</i>	-3.32927	4.68E-46
<i>GGA2</i>	-0.90252	4.02E-38
<i>LUZP2</i>	2.975654	1.12E-36
<i>TP63</i>	-2.02434	2.49E-35
<i>HOXC4</i>	2.037305	4.83E-35
<i>ZNF711</i>	-1.94103	1.39E-34
<i>SLC14A1</i>	-4.31974	1.39E-34
<i>CDC14B</i>	-1.20343	1.39E-34
<i>ALOX12P2</i>	-1.48582	3.13E-34
<i>CYP3A5</i>	-2.98807	4.41E-32
<i>FGFR2</i>	-1.14962	1.08E-31
<i>PDPN</i>	-0.91016	1.19E-30
<i>INHBA</i>	1.977945	1.91E-29
<i>ITGB6</i>	-1.41177	2.67E-29
<i>NIPAL3</i>	-2.20802	3.16E-29
<i>TRIM29</i>	-2.46139	5.83E-29
<i>SRM</i>	1.367442	1.85E-28
<i>MKI67 (PAM50)</i>	0.923398	2.42E-28
<i>TRAM2</i>	1.078532	2.46E-28
<i>BCL11A</i>	-2.01543	6.77E-28
<i>VAR5</i>	0.936243	7.43E-28
<i>CSTA</i>	-3.46746	7.43E-28
<i>PCDH7</i>	-1.26314	1.99E-27
<i>HIST1H1B</i>	0.870742	7.91E-27
<i>ZHX2</i>	-1.20713	1.36E-26
<i>LSAMP</i>	-2.00197	4.30E-26
<i>HOXD10</i>	-1.59334	7.89E-26
<i>DPYS</i>	-2.3263	8.96E-26

Gene	Log2 FC	adj.P.Val
<i>DST</i>	-1.26049	1.81E-25
<i>RGS10</i>	1.123211	2.24E-25
<i>SLC18A2</i>	-2.23621	3.02E-25
<i>RRM2</i> (PAM50)	2.503003	6.79E-25
<i>VWA5A</i>	-1.66663	7.61E-25
<i>COL17A1</i>	-1.42152	7.80E-25
<i>MR1</i>	-0.81288	9.72E-25
<i>MME</i>	-3.68156	9.96E-25
<i>HPCAL1</i>	0.873514	1.36E-24
<i>BIRC5</i> (PAM50)	1.02122	1.53E-24
<i>COL4A6</i>	-1.67798	5.20E-24
<i>NPR2</i>	-1.05453	6.15E-24
<i>RCC1</i>	0.919953	8.78E-24
<i>CENPE</i>	1.345071	1.07E-23
<i>MEG3</i>	-1.14897	1.20E-23
<i>GSTM1</i>	-1.89705	1.78E-23
<i>ACSBG1</i>	-0.62119	4.30E-23
<i>CIB2</i>	-0.70558	4.46E-23
<i>CENPF</i> (PAM50)	0.935464	6.18E-23

Known PAM50 Genes are shown in brackets.

Table 69. Top 50 most significant DEGs in prostate cancer Gleason score 9.

Gene	Log2 FC	adj.P.Val
<i>SUPT3H</i>	-1.86571	4.18E-97
<i>VSNL1</i>	-2.11069	7.96E-85
<i>TRAF3IP2</i>	-1.46726	5.60E-75
<i>ZNF711</i>	-2.04036	1.02E-69
<i>LAMB3</i>	-2.82002	5.35E-67
<i>LUZP2</i>	2.828417	2.34E-64
<i>GGA2</i>	-0.81755	2.16E-62
<i>ALOX12P2</i>	-1.43605	7.76E-62
<i>SLC14A1</i>	-4.02162	2.73E-59
<i>PDPN</i>	-0.90495	1.30E-58
<i>HOXC4</i>	1.83695	1.41E-57
<i>TP63</i>	-1.76861	1.55E-55
<i>CSTA</i>	-3.45218	5.96E-54
<i>ITGB6</i>	-1.34642	3.45E-53
<i>MCC</i>	-1.67103	3.64E-53
<i>NIPAL3</i>	-2.09461	1.11E-52
<i>LSAMP</i>	-2.01975	4.62E-52
<i>BCL11A</i>	-1.93015	3.23E-51
<i>SKIL</i>	-0.79235	3.01E-50
<i>SLC18A2</i>	-2.22515	1.49E-49
<i>CDC14B</i>	-0.96781	3.34E-48
<i>ACSBG1</i>	-0.63551	1.15E-47
<i>MR1</i>	-0.79103	4.53E-47
<i>CYP3A5</i>	-2.47393	5.05E-47
<i>SLC15A1</i>	-0.98843	1.25E-46
<i>DPYS</i>	-2.17719	3.67E-46
<i>MKI67 (PAM50)</i>	0.812231	6.38E-46
<i>PDK3</i>	-0.88153	1.06E-45
<i>FGFR2</i>	-0.93904	1.34E-45
<i>HOXD10</i>	-1.47312	1.77E-45
<i>PCDH7</i>	-1.12433	2.34E-45

Gene	Log2 FC	adj.P.Val
<i>TBC1D1</i>	-0.87476	2.34E-45
<i>HIST1H1B</i>	0.781072	5.27E-45
<i>CIB2</i>	-0.68777	2.03E-44
<i>NPR2</i>	-0.99736	6.97E-44
<i>EDN3</i>	-1.09175	7.98E-44
<i>EFHD1</i>	-1.62486	3.57E-43
<i>ZHX2</i>	-1.05836	3.79E-43
<i>VWA5A</i>	-1.51648	7.09E-43
<i>CPM</i>	-1.94182	1.45E-42
<i>ELF4</i>	-1.12732	3.08E-42
<i>PGAP1</i>	-1.29301	3.90E-42
<i>FADS1</i>	-1.5148	1.08E-41
<i>MEG3</i>	-1.05181	2.56E-41
<i>LRP4</i>	-1.27277	2.82E-41
<i>TRIM29</i>	-1.98667	2.86E-41
<i>BIRC5 (PAM50)</i>	0.907081	7.72E-41
<i>EPHA5</i>	-1.13829	9.63E-41
<i>SRM</i>	1.103495	1.33E-40
<i>SPC25</i>	0.979236	1.40E-40

Known PAM50 Genes are shown in brackets.

Table 70. HMMR, CENPE, and STIL expression in histological (and equivalent) subtype comparison.

Gene	IDC LogFC	IDC Adj P-Value	ILC LogFC	ILC Adj P-Value	DCIS LogFC	DCIS Adj P-Value	Ovarian LogFC	Ovarian Adj P-Value	Peritoneum LogFC	Peritoneum Adj P-Value	Fallopian LogFC	Fallopian Adj P-Value	Grade 1 LogFC	Grade 1 Adj P-Value	Grade 2 & 3 LogFC	Grade 2 & 3 Adj P-Value	Grade 4 LogFC	Grade 4 Adj P-Value	Grade 5 LogFC	Grade 5 Adj P-Value
HMMR	1.25	2.15E-82	1.10	6.84E-36	1.00	1.02E-23	1.83	1.40E-66	1.88	4.17E-22	2.30	1.29E-09	1.00	1.70E-37	1.09	2.49E-51	1.27	7.66E-18	1.00	1.70E-37
CENPE	0.90	3.23E-57	0.81	9.08E-26	0.76	2.43E-18	1.29	2.49E-50	1.51	8.62E-21	1.57	6.62E-07	0.96	1.16E-41	1.10	3.19E-61	1.35	8.69E-24	0.96	1.16E-41
STIL	0.82	2.04E-49	0.67	2.09E-18	0.67	6.12E-15	1.47	7.96E-62	1.42	3.76E-18	1.62	4.38E-07	0.95	1.59E-38	0.92	2.79E-44	1.25	2.63E-19	0.95	1.59E-38

Table 71. HMMR, CENPE, and STIL expression in molecular (and equivalent) subtype comparison.

Gene	Normal_Like LogFC	Normal_Like Adj P-Value	Luminal_A LogFC	Luminal_A Adj P-Value	Luminal_B LogFC	Luminal_B Adj P-Value	HER2 LogFC	HER2 Adj P-Value	Basal LogFC	Basal Adj P-Value	Serous LogFC	Serous Adj P-Value	Endometrioid LogFC	Endometrioid Adj P-Value	Mucinous LogFC	Mucinous Adj P-Value	Clear_Cell LogFC	Clear_Cell Adj P-Value	Gleason_Score_4 LogFC	Gleason_Score_4 Adj P-Value	Gleason_Score_5 LogFC	Gleason_Score_5 Adj P-Value	Gleason_Score_6 LogFC	Gleason_Score_6 Adj P-Value	Gleason_Score_7 LogFC	Gleason_Score_7 Adj P-Value	Gleason_Score_8 LogFC	Gleason_Score_8 Adj P-Value	Gleason_Score_9 LogFC	Gleason_Score_9 Adj P-Value
HMMR	1.16	4.96E-11	1.38	2.97E-20	2.30	1.33E-47	2.26	1.29E-49	2.60	1.07E-63	2.14	1.43E-54	2.15	2.00E-36	2.10	5.03E-22	2.16	3.77E-27	1.15	0.003344012	0.85	1.77E-06	1.02	1.43E-35	1.09	3.41E-51	1.27	8.22E-18	1.05	3.37E-27
CENPE	0.92	1.74E-08	0.95	7.58E-12	1.96	3.94E-41	1.87	2.09E-40	2.37	7.88E-62	1.62	9.21E-48	1.69	7.93E-34	1.60	2.48E-19	1.84	8.54E-29	1.16	0.000872666	0.94	2.46E-09	0.95	2.84E-38	1.10	5.88E-61	1.35	1.07E-23	1.04	6.34E-32
STIL	0.84	3.42E-08	0.91	1.88E-12	1.49	1.56E-28	1.64	1.34E-36	2.24	1.20E-63	1.83	6.47E-55	1.68	3.12E-31	1.56	2.57E-17	1.67	1.01E-22	1.41	0.000117388	1.03	4.31E-10	0.92	2.98E-34	0.92	2.61E-44	1.25	2.44E-19	0.94	1.82E-25

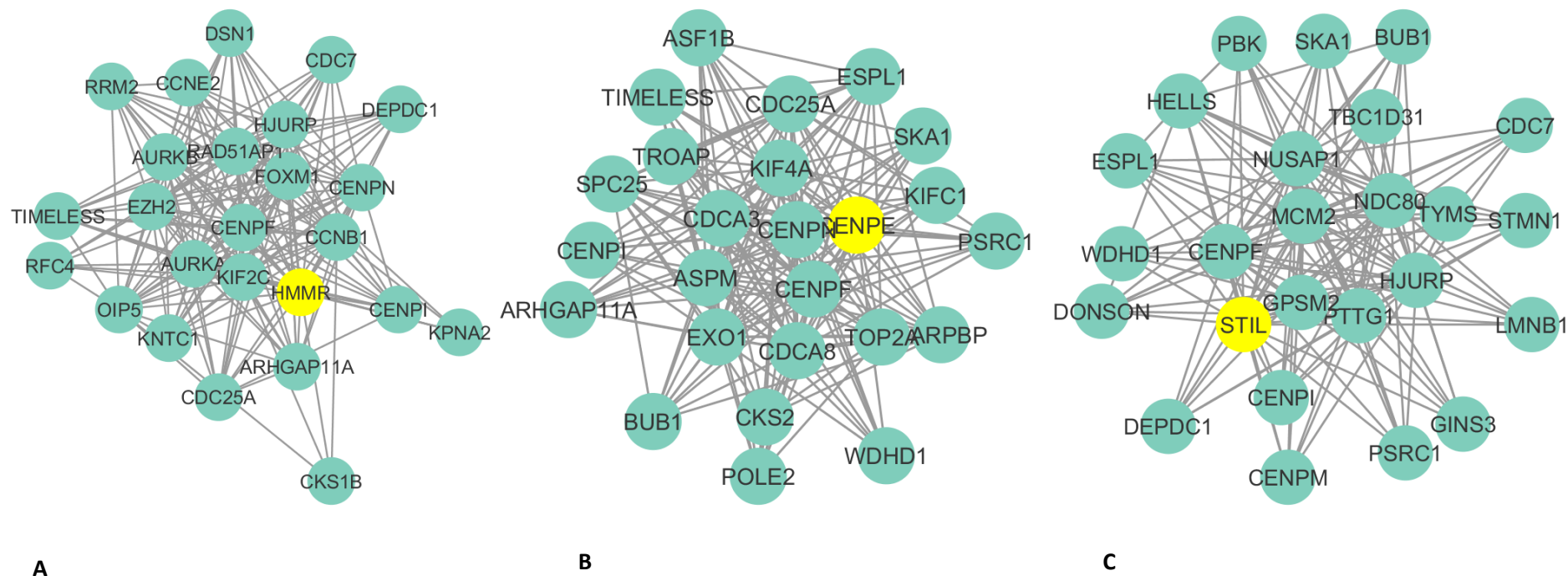


Figure 70. IDC breast tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for breast cancer IDC tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

IDC: Invasive Ductal Carcinoma

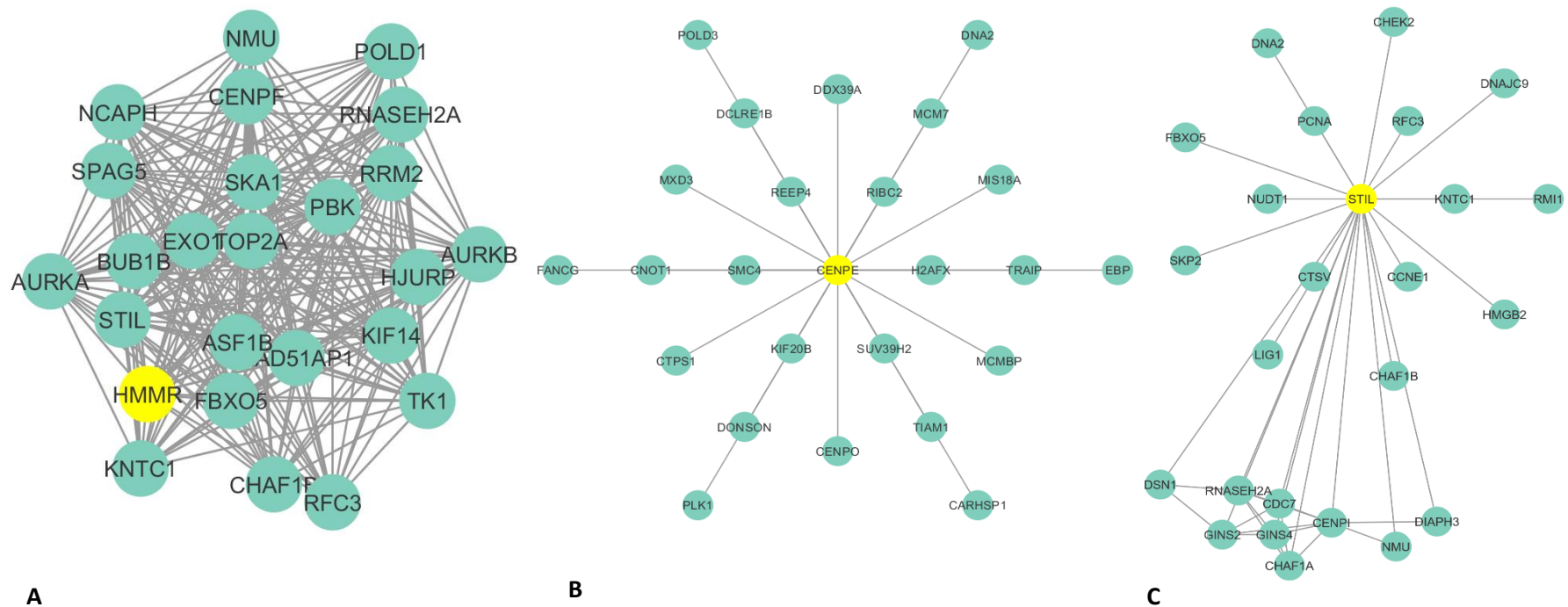
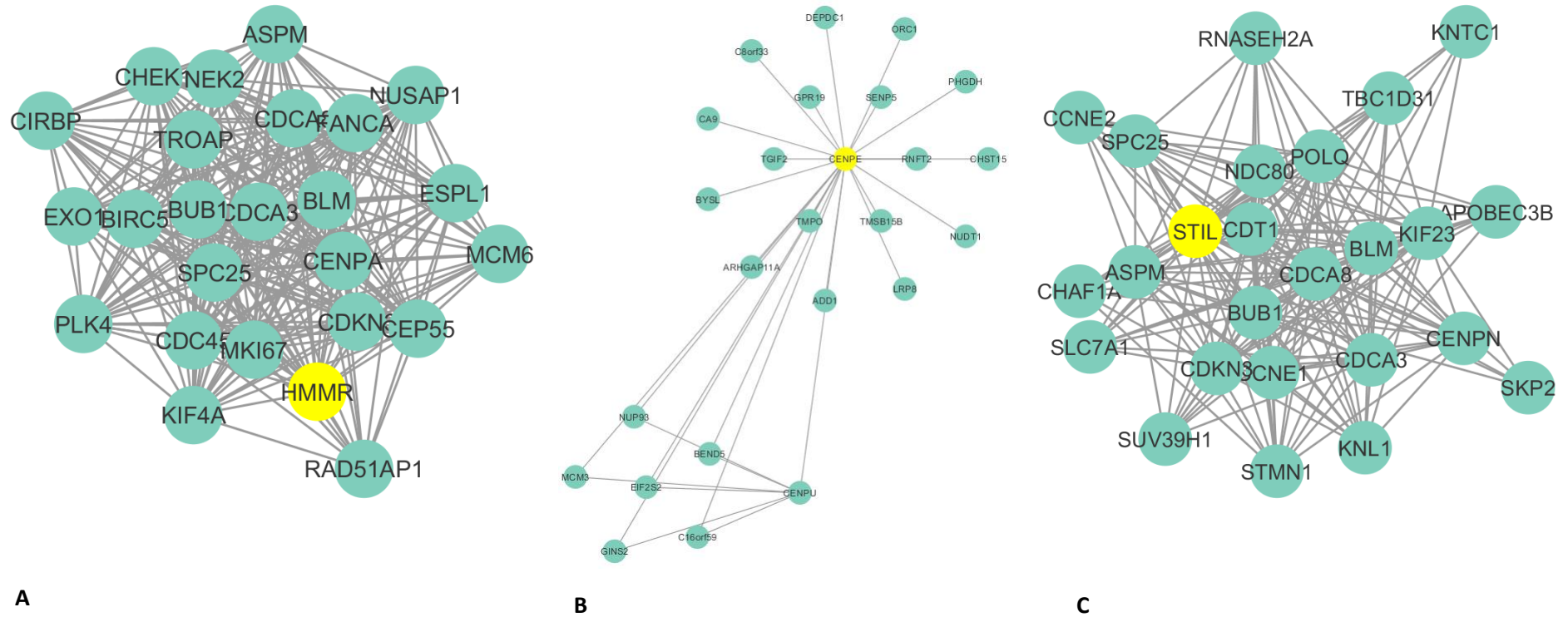


Figure 71. ILC breast tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for breast cancer ILC tumour samples. STIL was found to be co-expressed with HMMR and shown highlighted in red. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

LC: Invasive Lobular Carcinoma



A **B** **C**
 Figure 72. DCIS breast tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for breast cancer DCIS tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

DCIS: Ductal Carcinoma in-situ

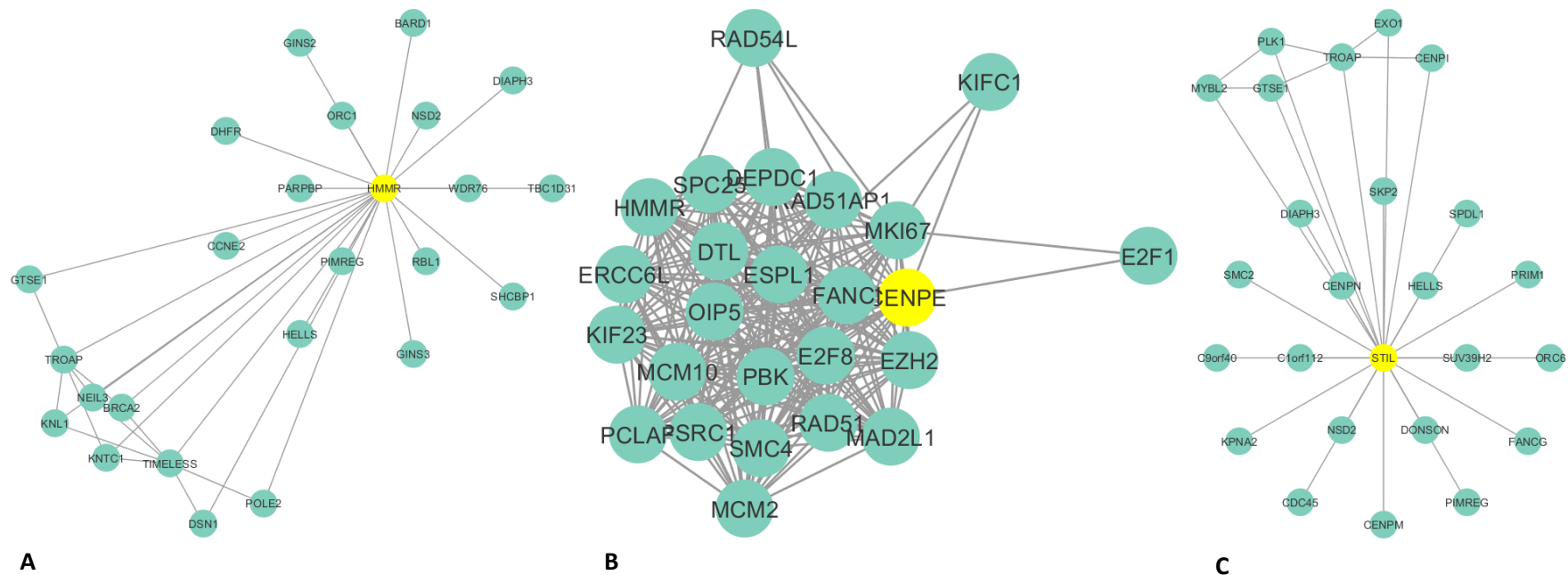


Figure 73. Ovarian tissue tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for ovarian cancer ovarian tissue tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

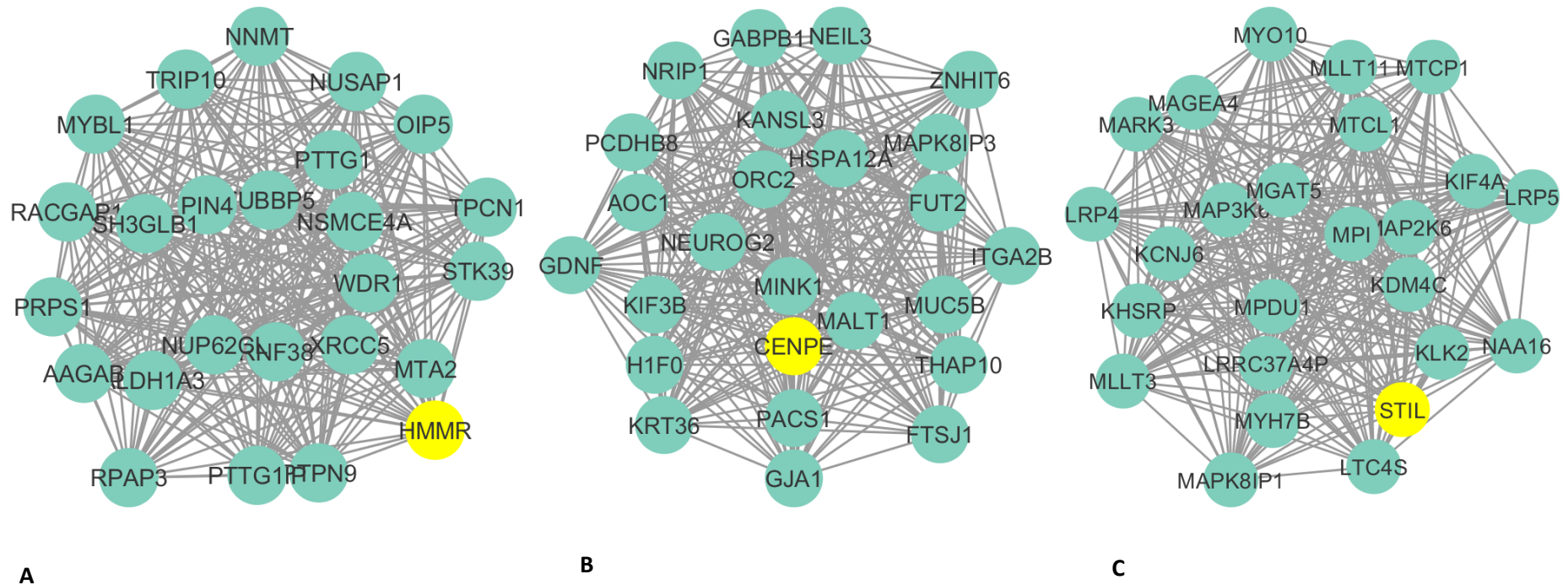
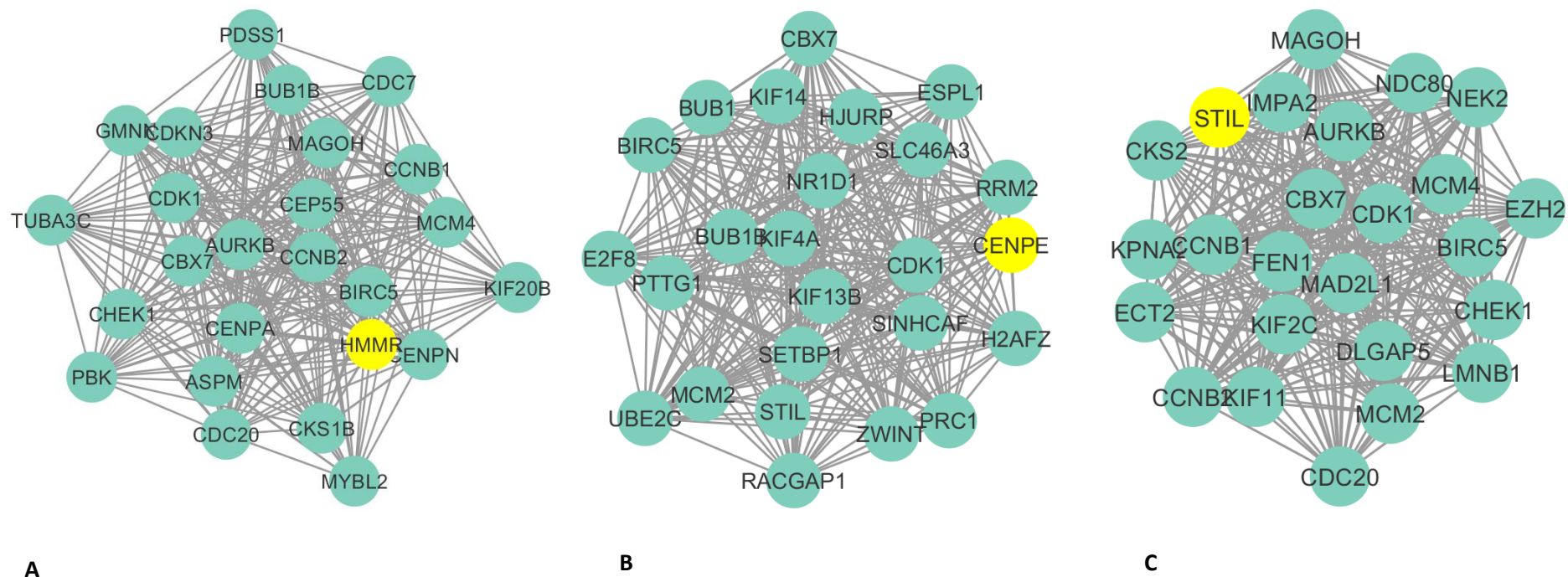


Figure 74. Fallopian tissue tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for ovarian cancer fallopian tissue tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).



A **B** **C**
 Figure 75. Peritoneum tissue tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for ovarian cancer peritoneum tissue tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

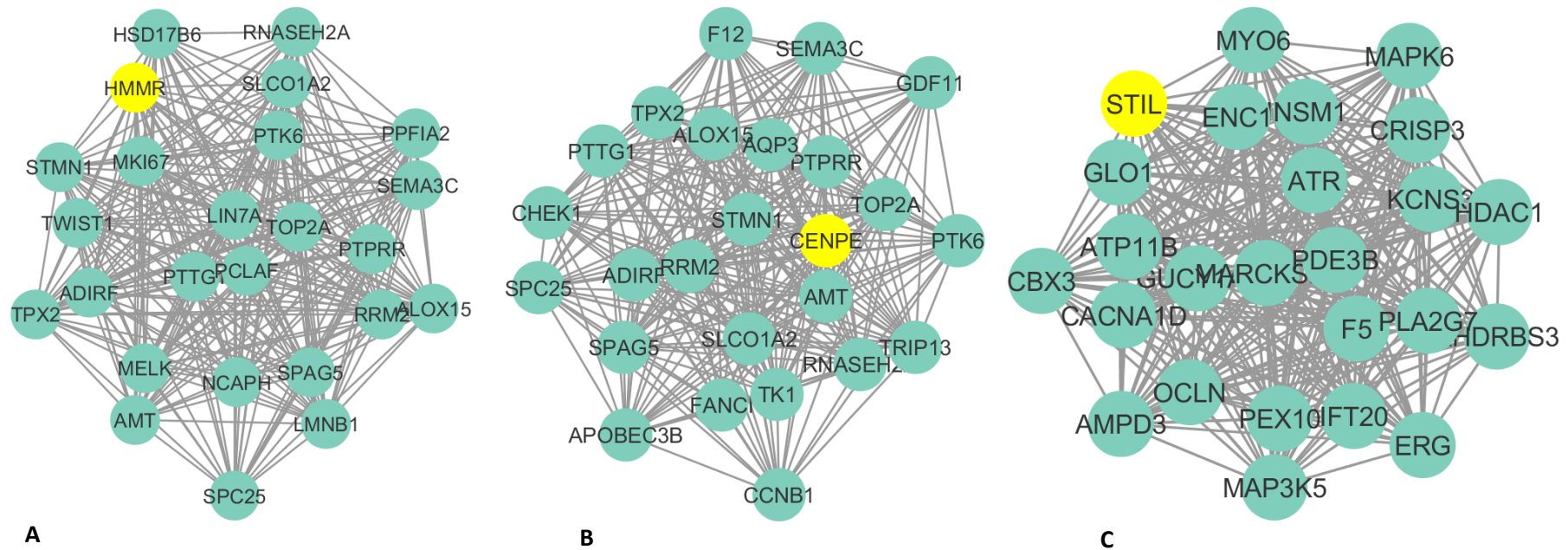


Figure 76. Gleason grade 1 tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for prostate cancer Gleason grade group 1 tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

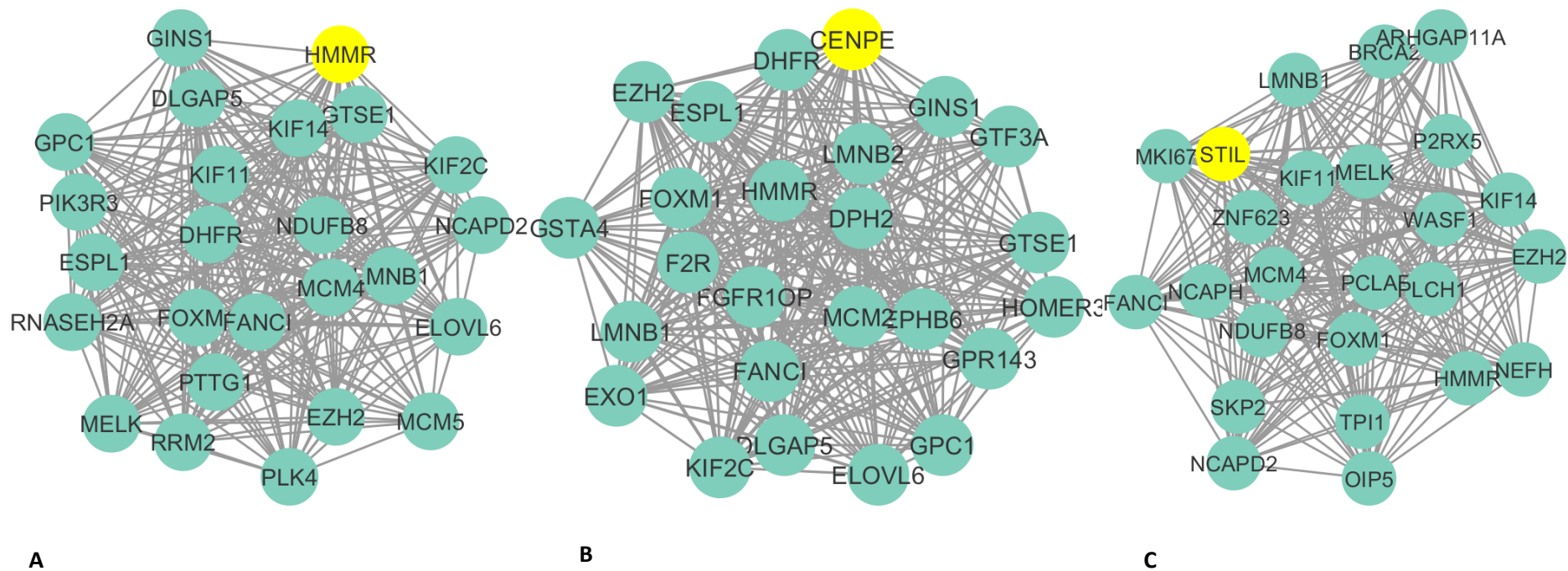


Figure 77. Gleason grade 2 & 3 tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for prostate cancer Gleason grade group 2 & 3 tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

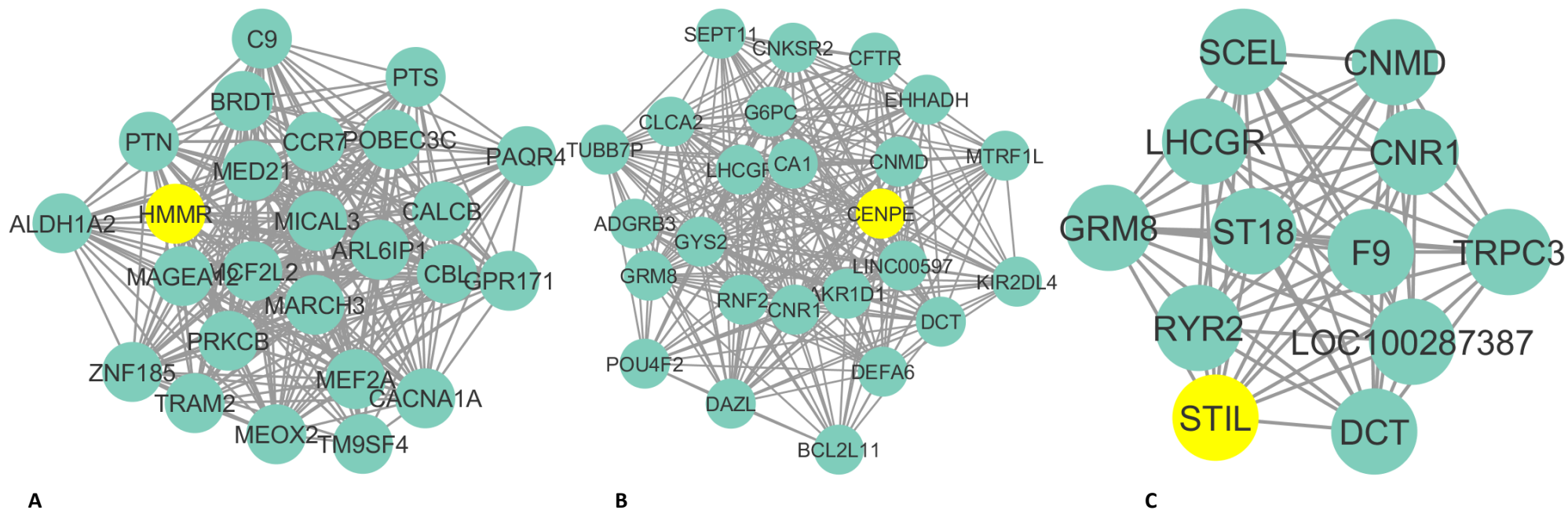


Figure 78. Gleason grade 4 tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for prostate cancer Gleason grade group 4 tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

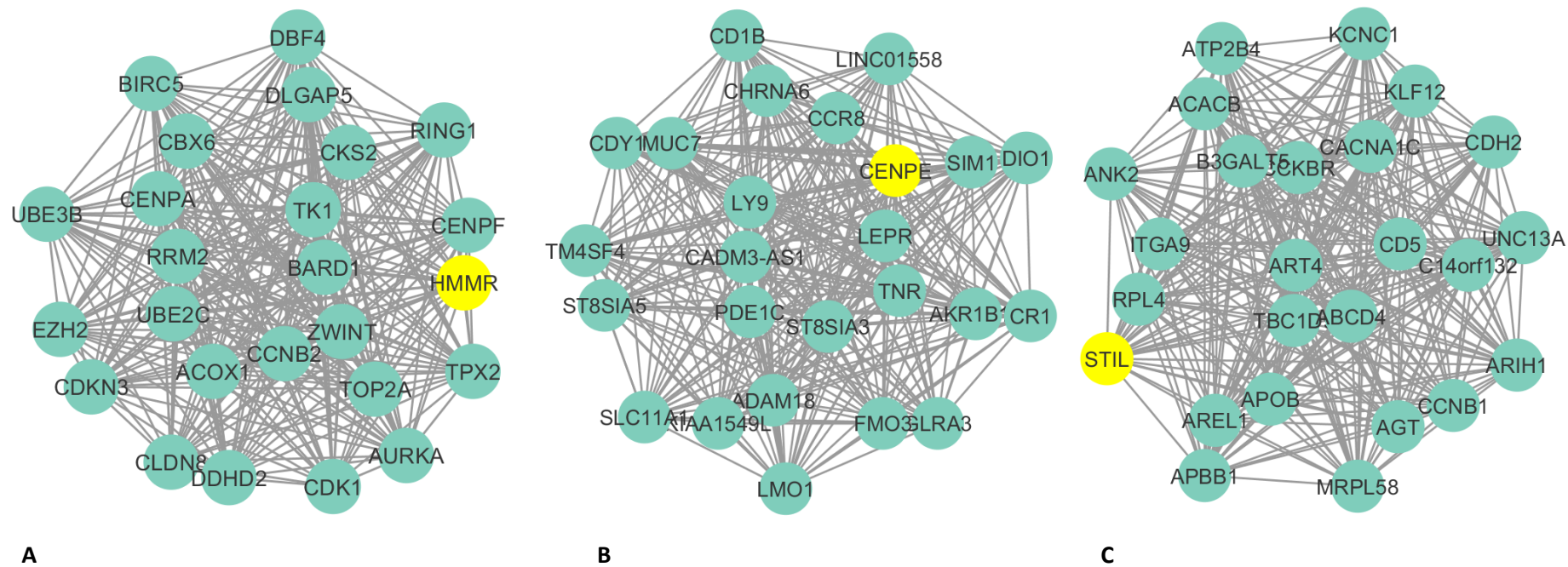


Figure 79. Gleason grade 5 tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for prostate cancer Gleason grade group 5 tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

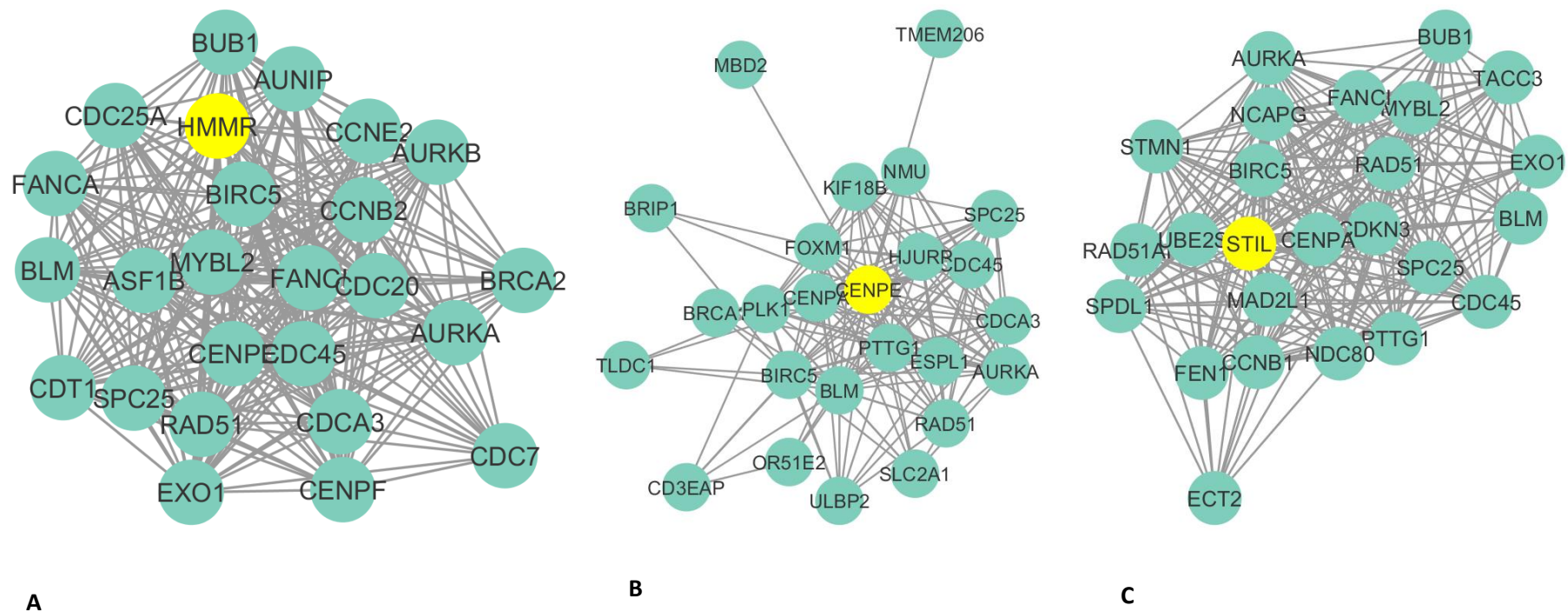


Figure 80. Normal-like breast tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for breast cancer normal like tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

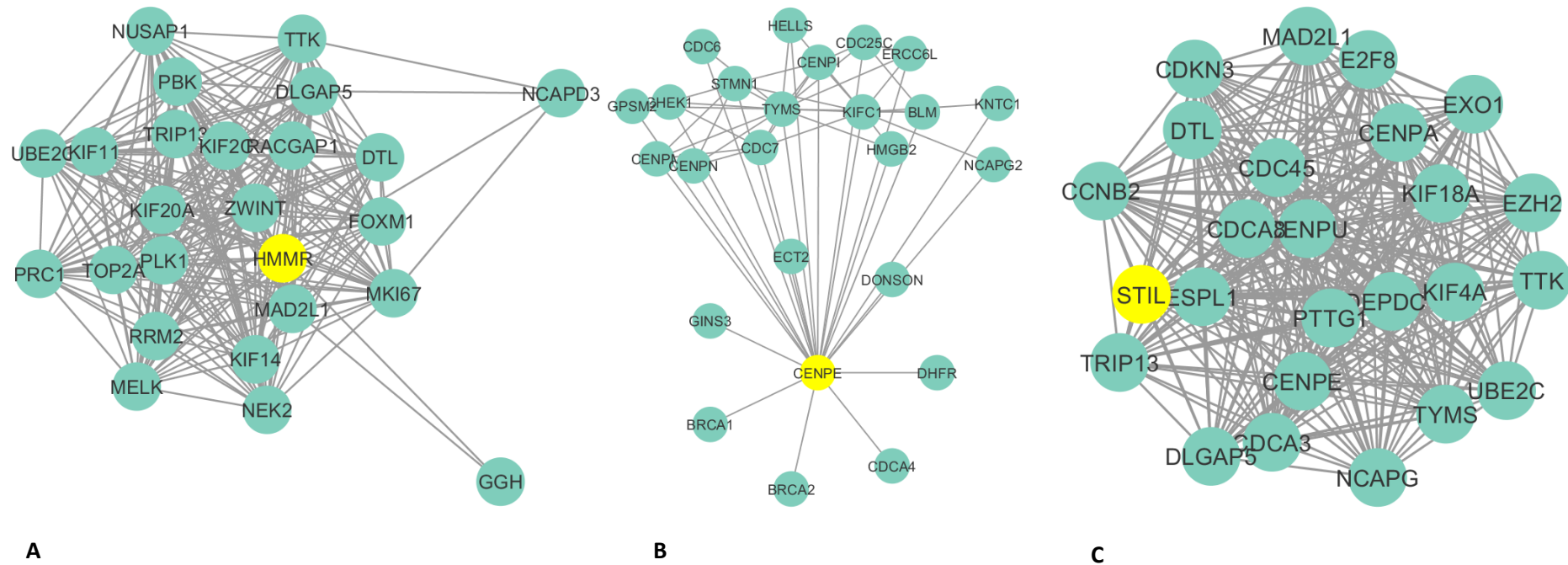


Figure 81. Luminal A breast tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for breast cancer Luminal A tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

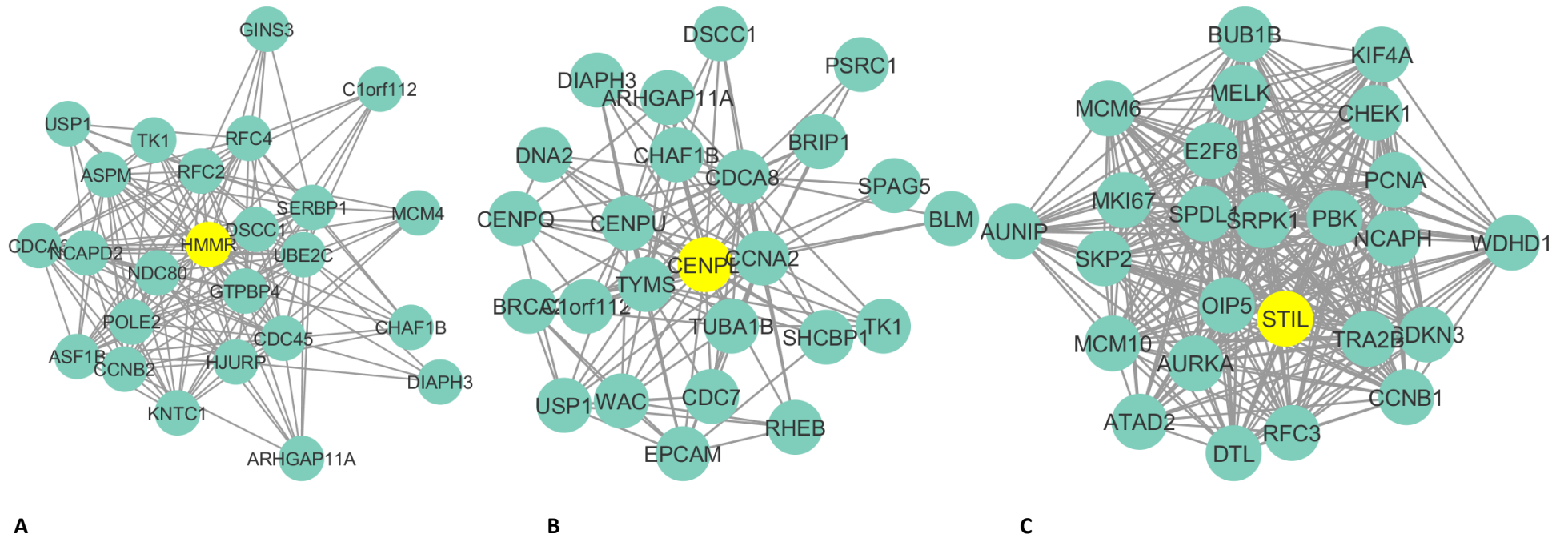


Figure 82. Luminal B breast tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for breast cancer Luminal B tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

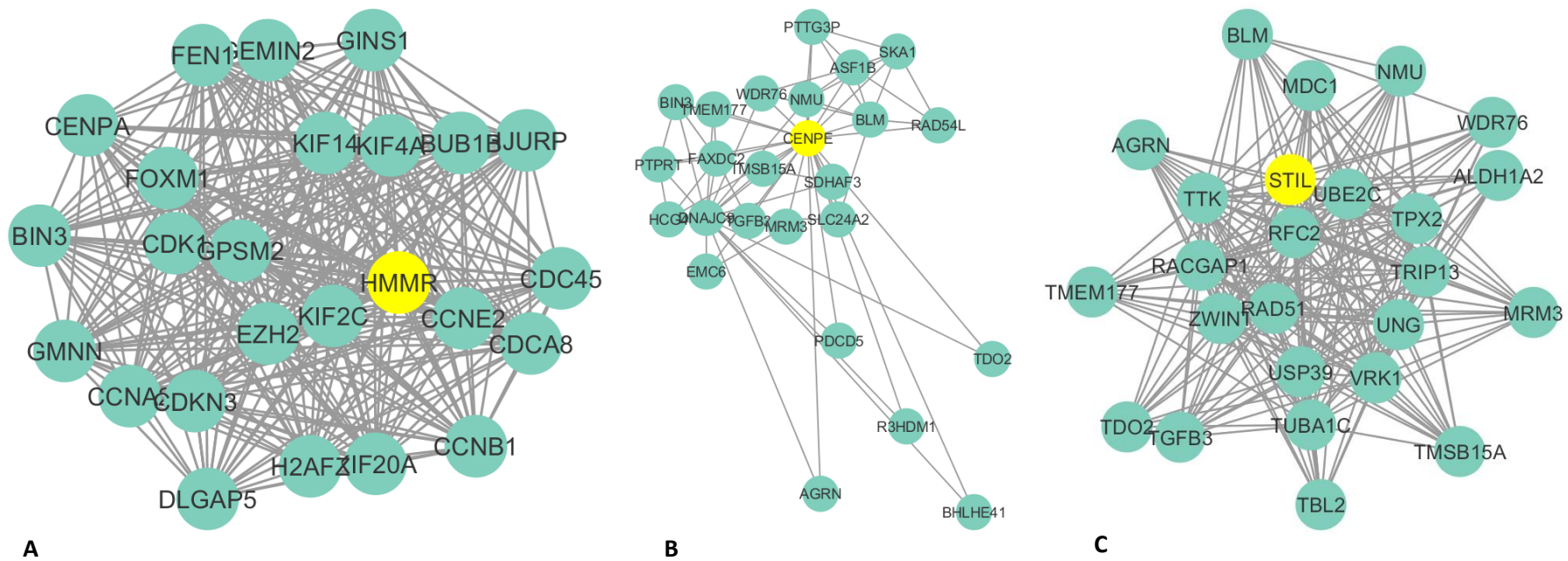


Figure 83. HER2 breast tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for breast cancer HER2 tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

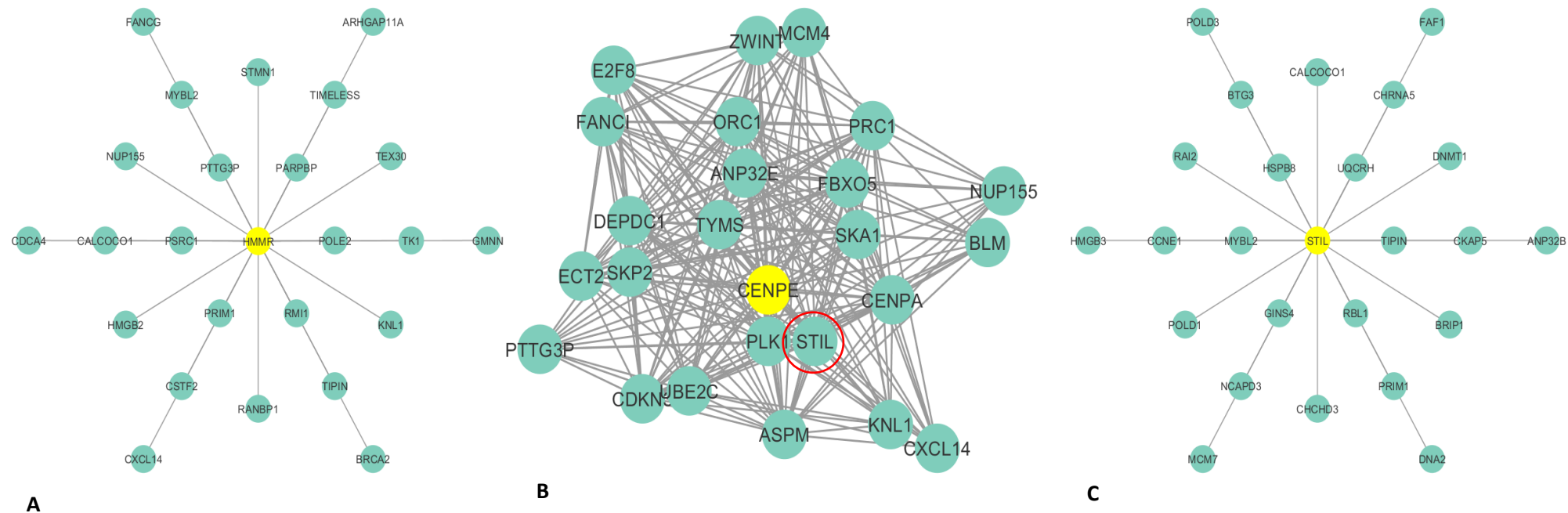


Figure 84. Basal breast tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for breast cancer basal tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

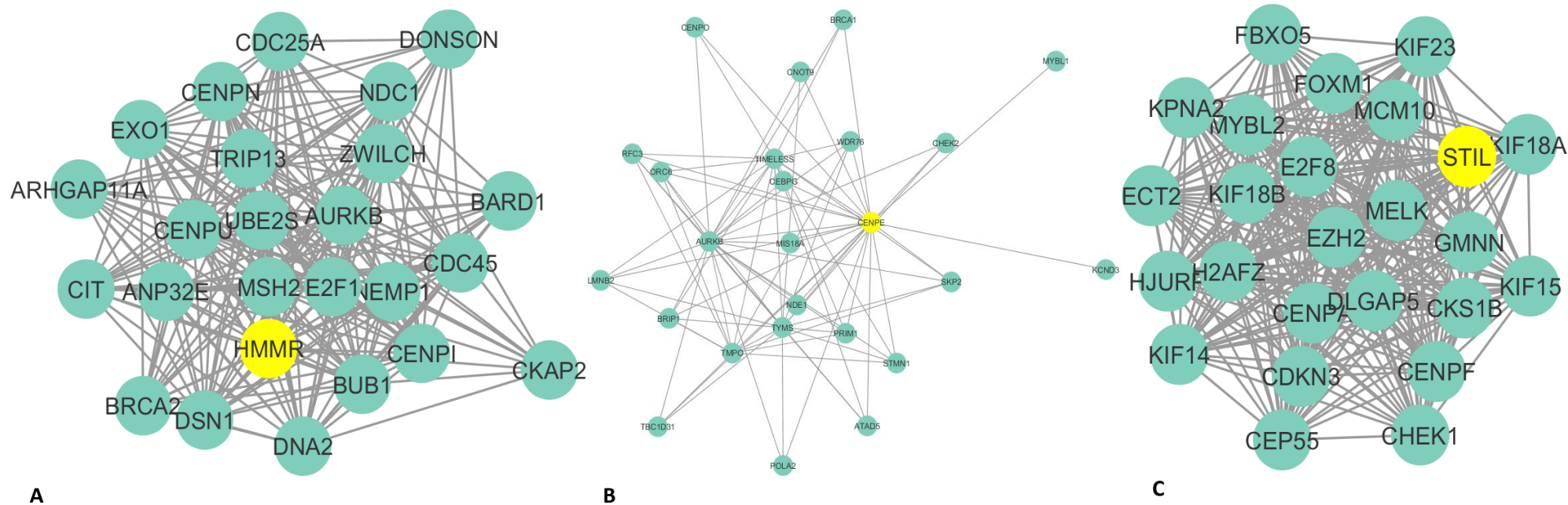


Figure 85. Serous tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for ovarian cancer serous tissue tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

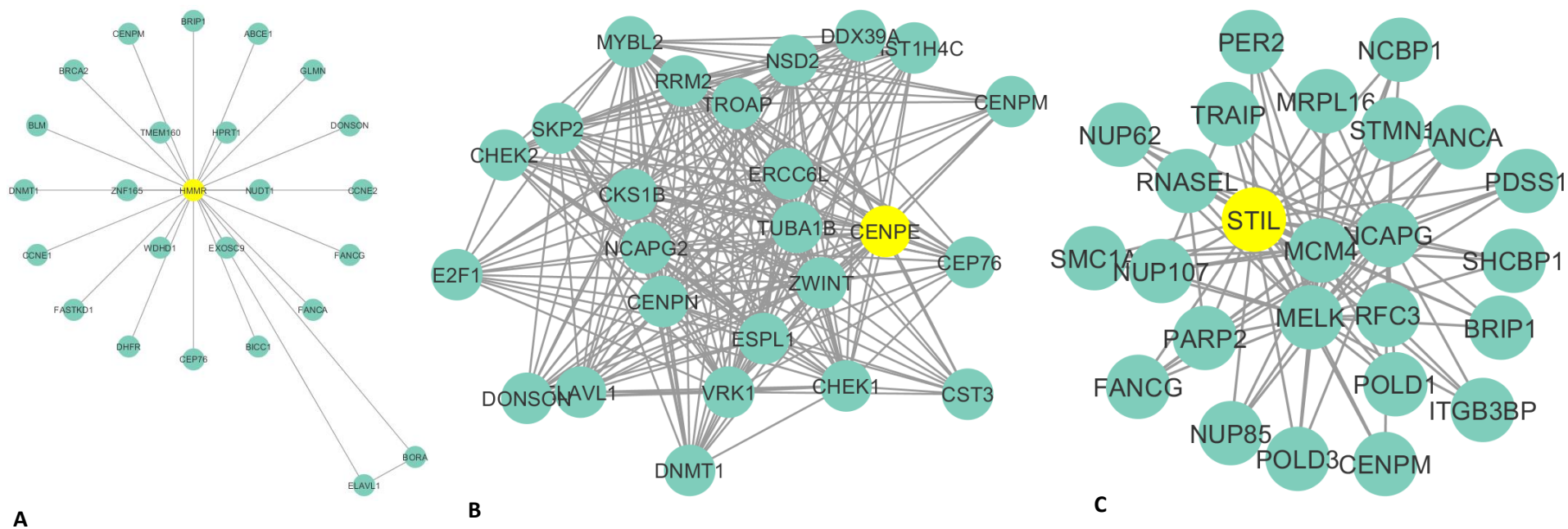


Figure 86. Endometrioid tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for ovarian cancer endometrioid tissue tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

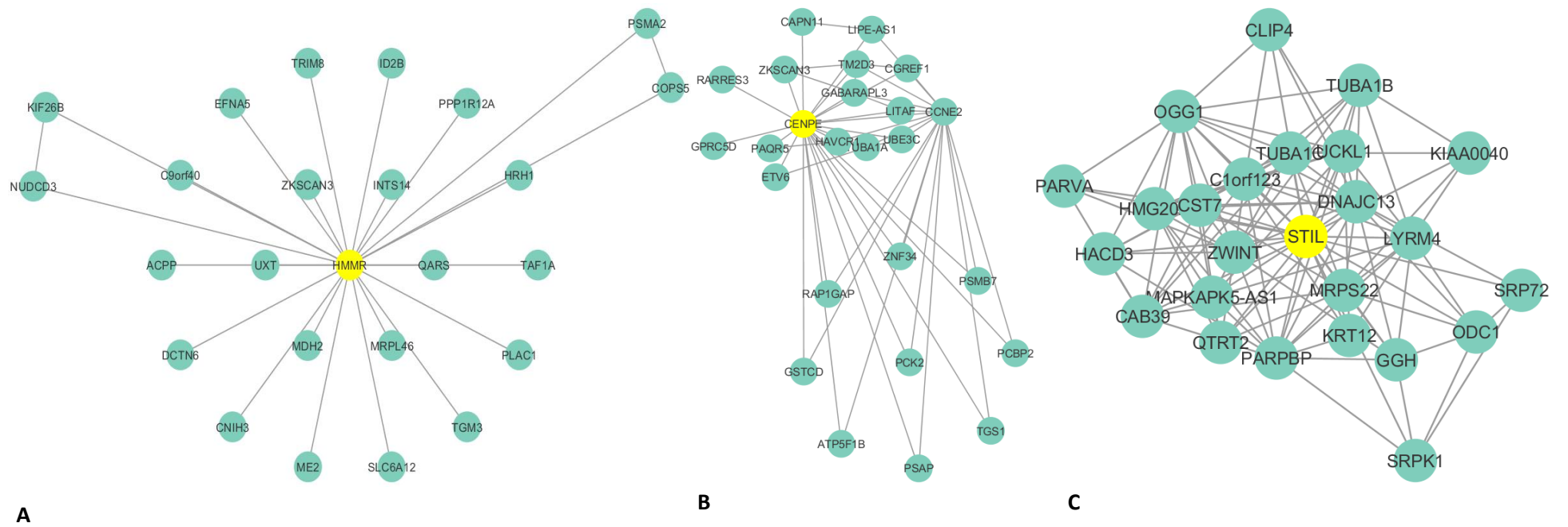


Figure 87. Mucinous tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for ovarian cancer mucinous tissue tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

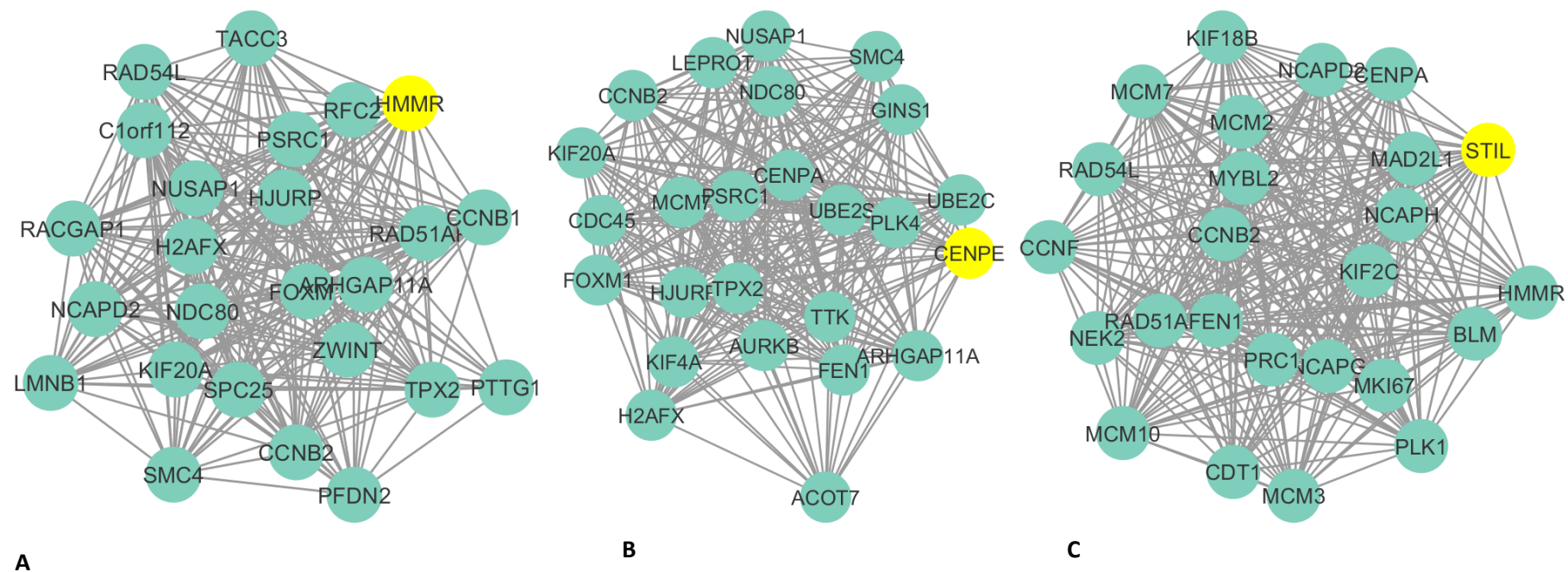


Figure 88. Clear cell tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for ovarian cancer clear cell tissue tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

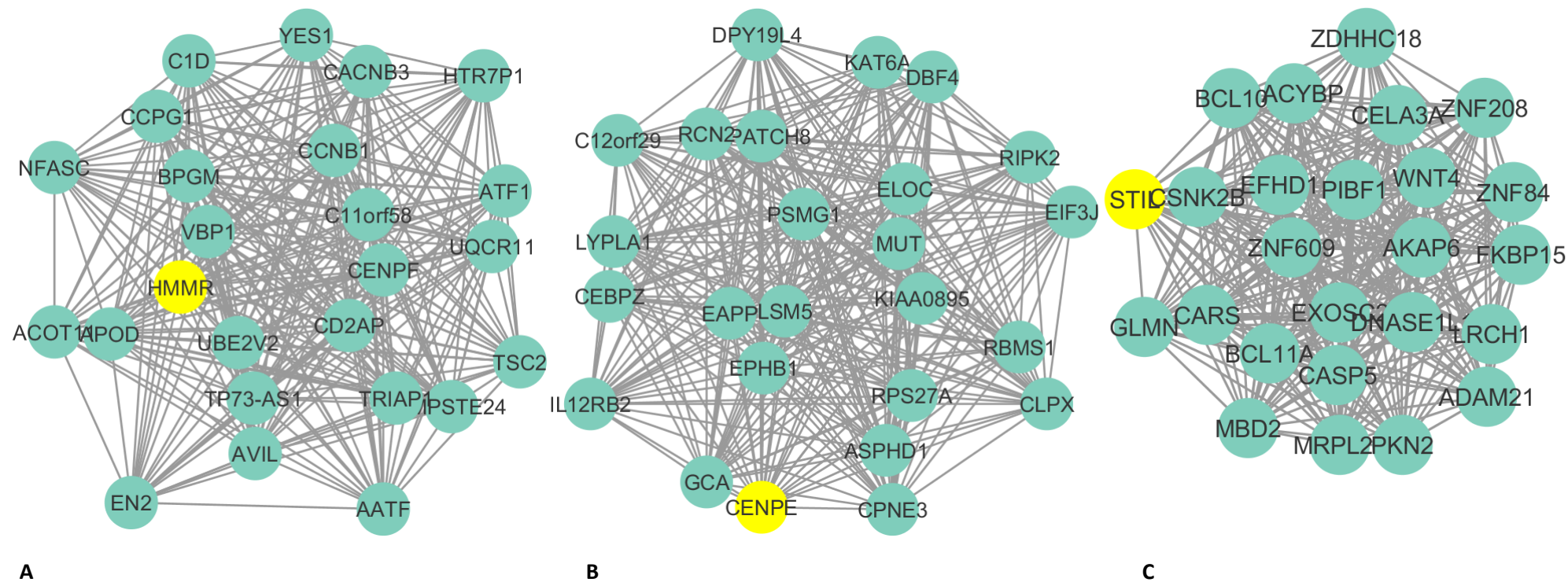


Figure 89. Gleason score 5 tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for prostate cancer Gleason score 5 tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

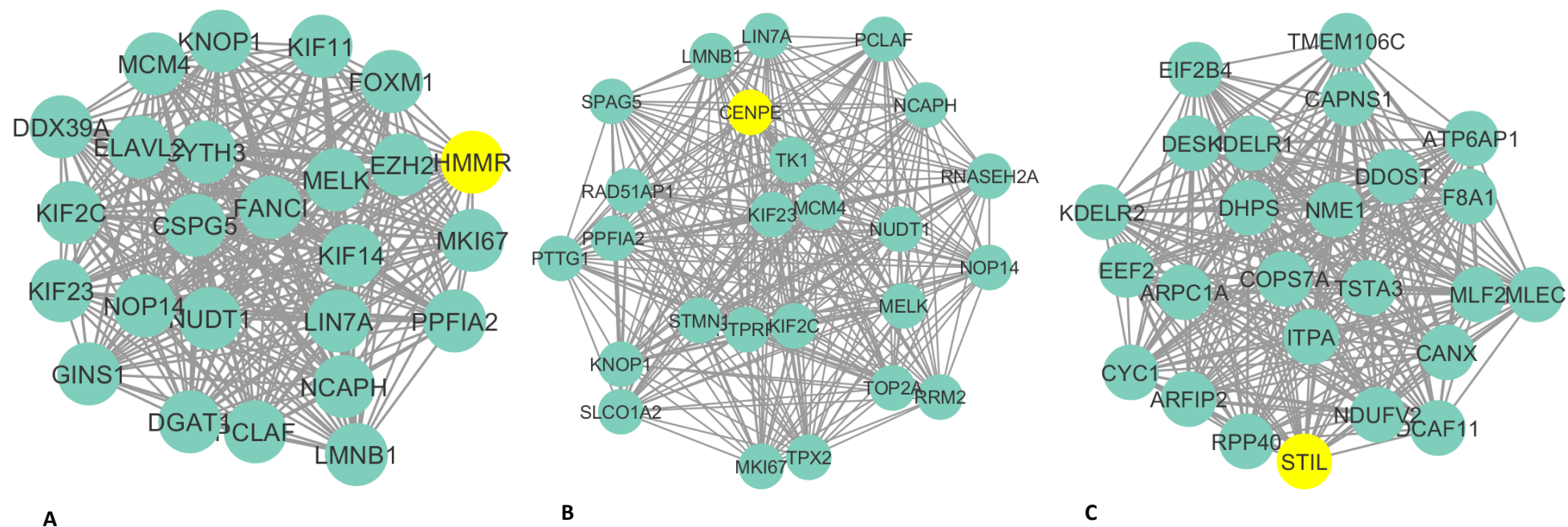


Figure 90. Gleason score 6 tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for prostate cancer Gleason score 6 tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

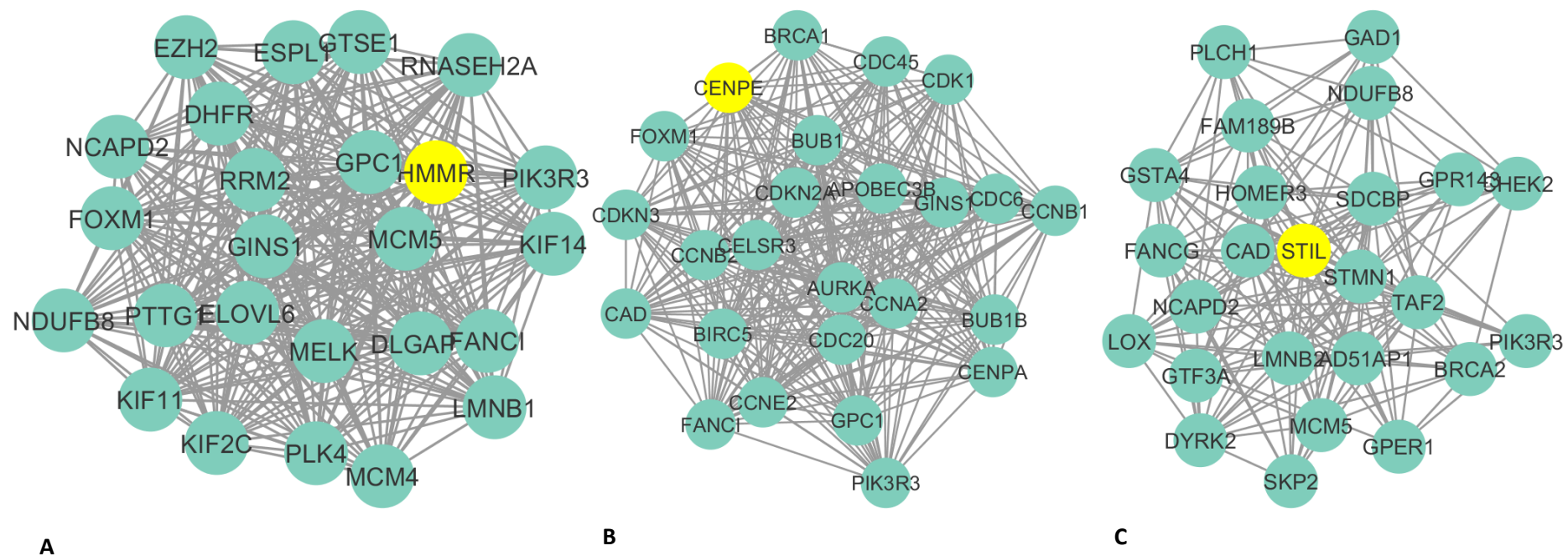


Figure 91. Gleason score 7 tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for prostate cancer Gleason score 7 tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

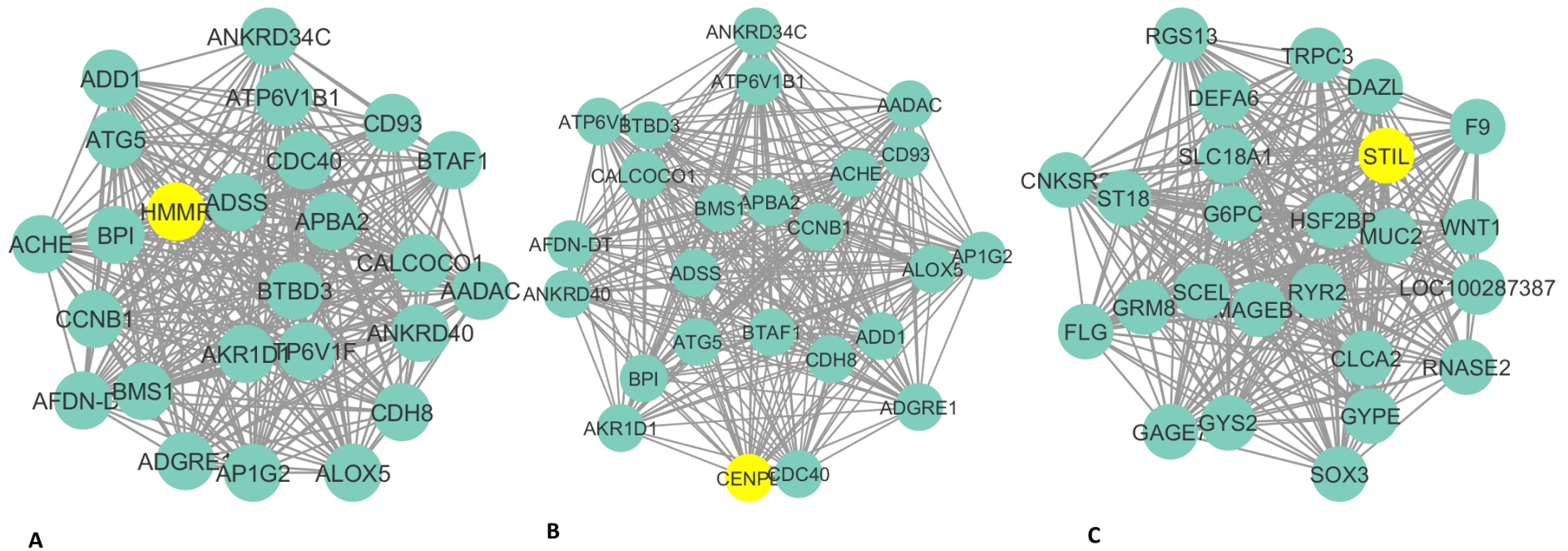


Figure 92. Gleason score 8 tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for prostate cancer Gleason score 8 tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

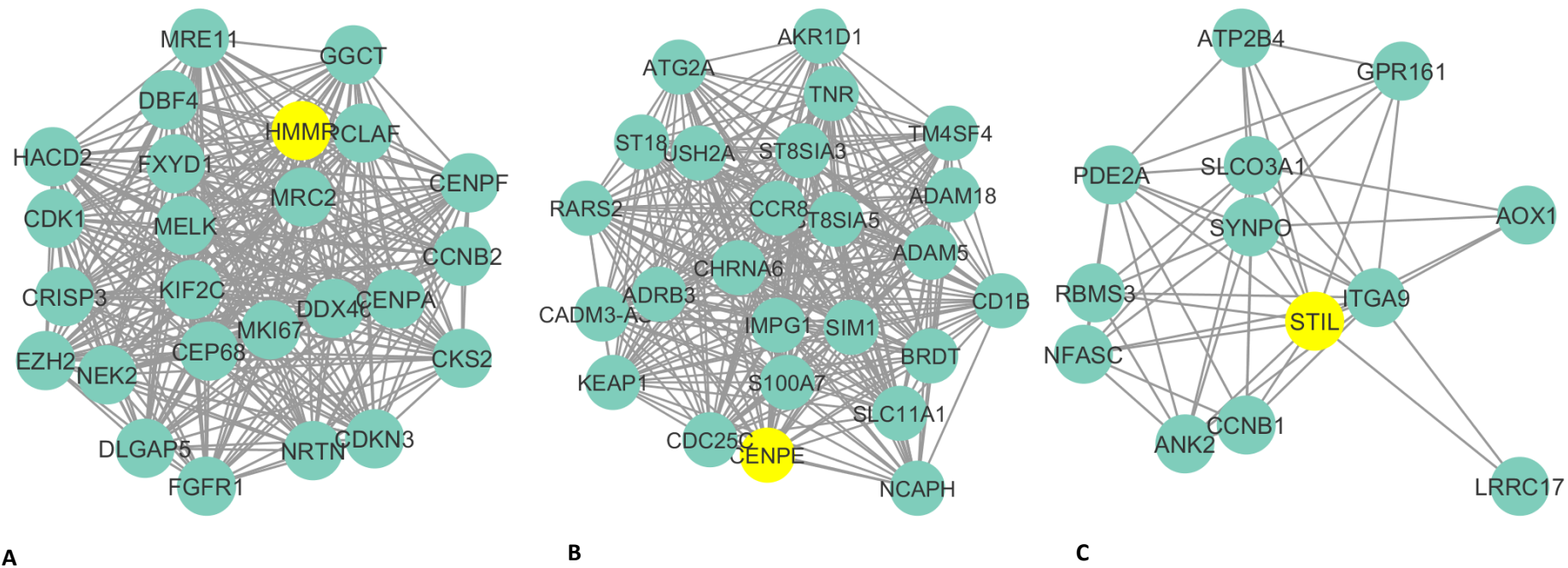


Figure 93. Gleason score 9 tumour networks for HMMR, CENPE, and STIL.

A) Subnetwork of HMMR for prostate cancer Gleason score 9 tumour samples. B) CENPE subnetwork. C) STIL subnetwork. Only the 25 most highly correlated genes based on heat diffusion are shown. Both CENPE and STIL were identified to be co-expressed within the same module (not shown).

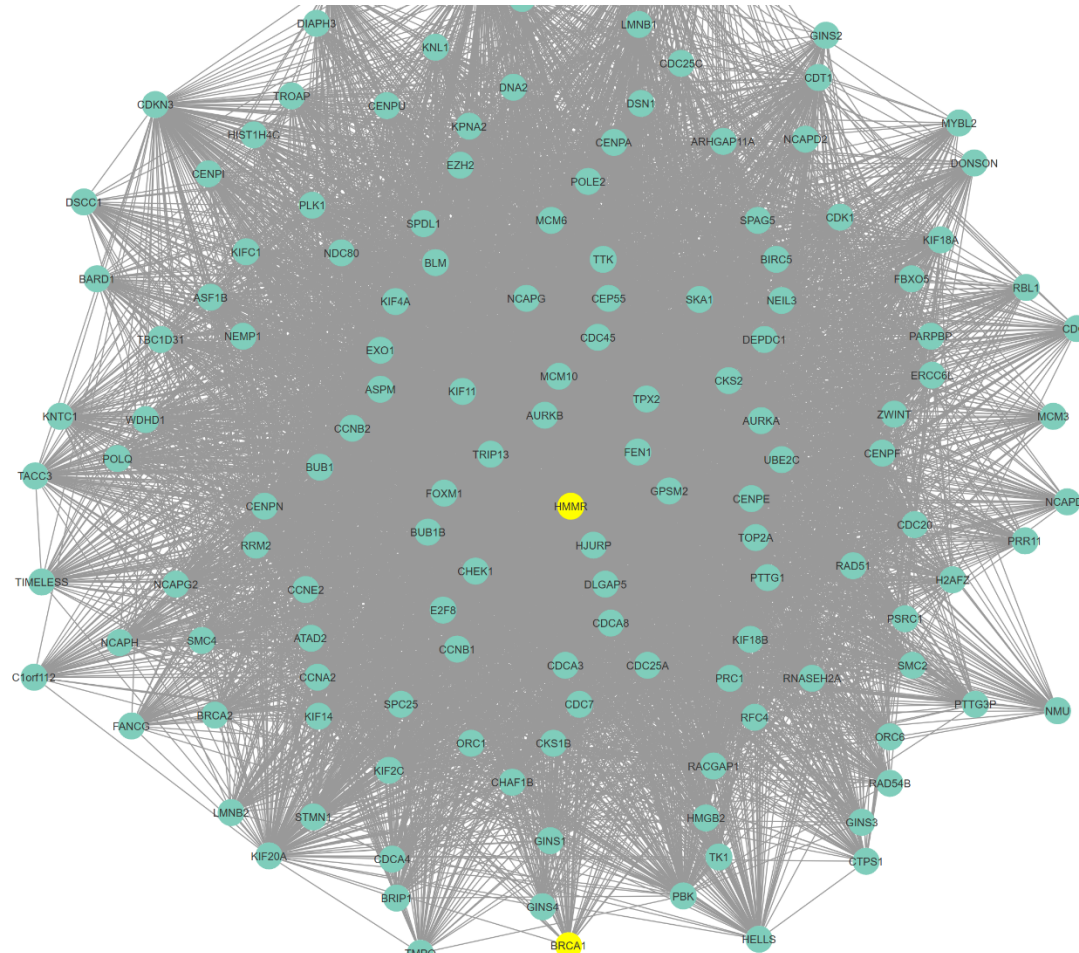


Figure 94. Expanded network of HMMR in breast tumour samples.

BRCA1 was found to be co-expressed with HMMR in this expanded view. Both HMMR and BRCA1 are highlighted in yellow.

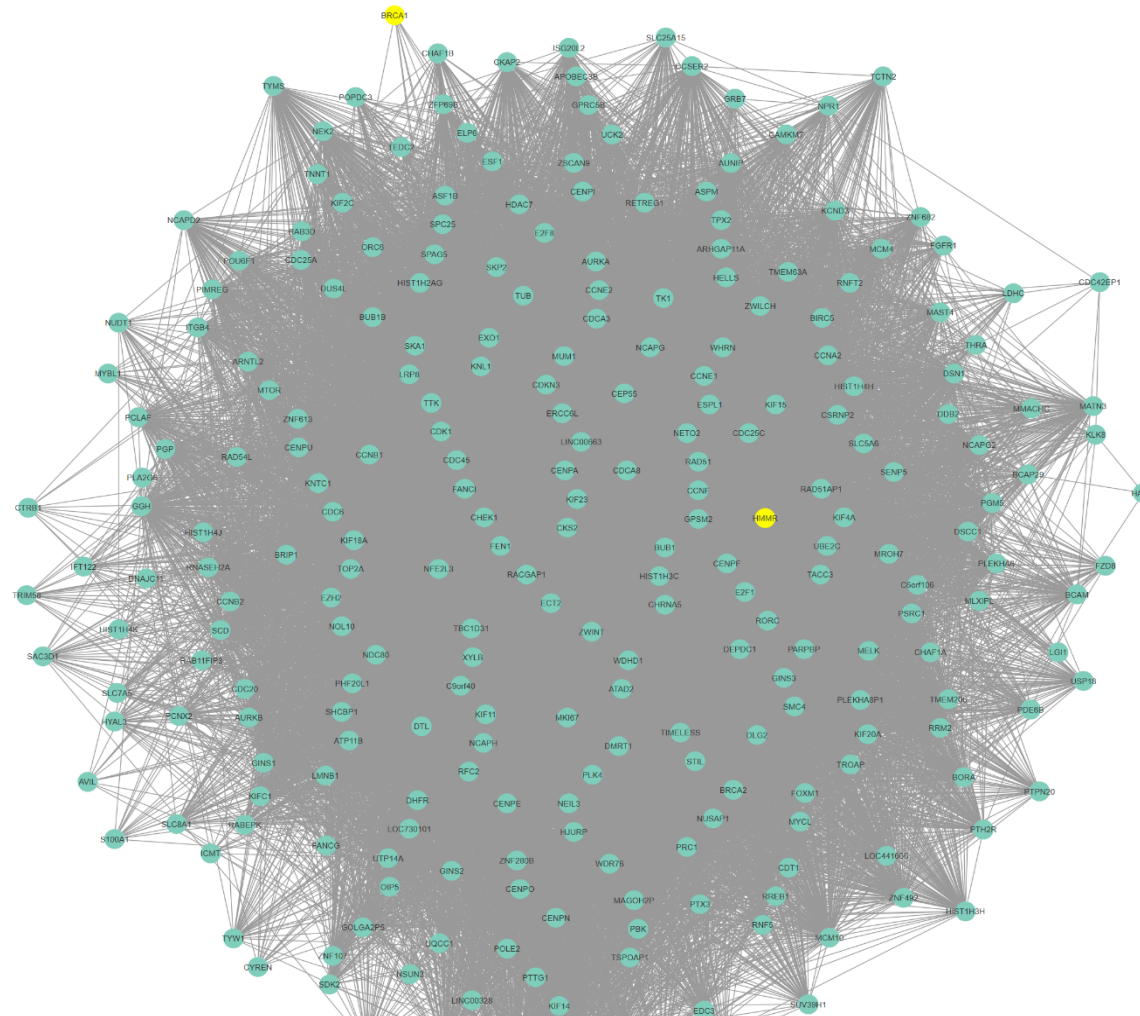


Figure 95. Expanded network of HMMR in Ovarian tumour samples.

BRCA1 was found to be co-expressed with HMMR in this expanded view. Both HMMR and BRCA1 are highlighted in yellow.

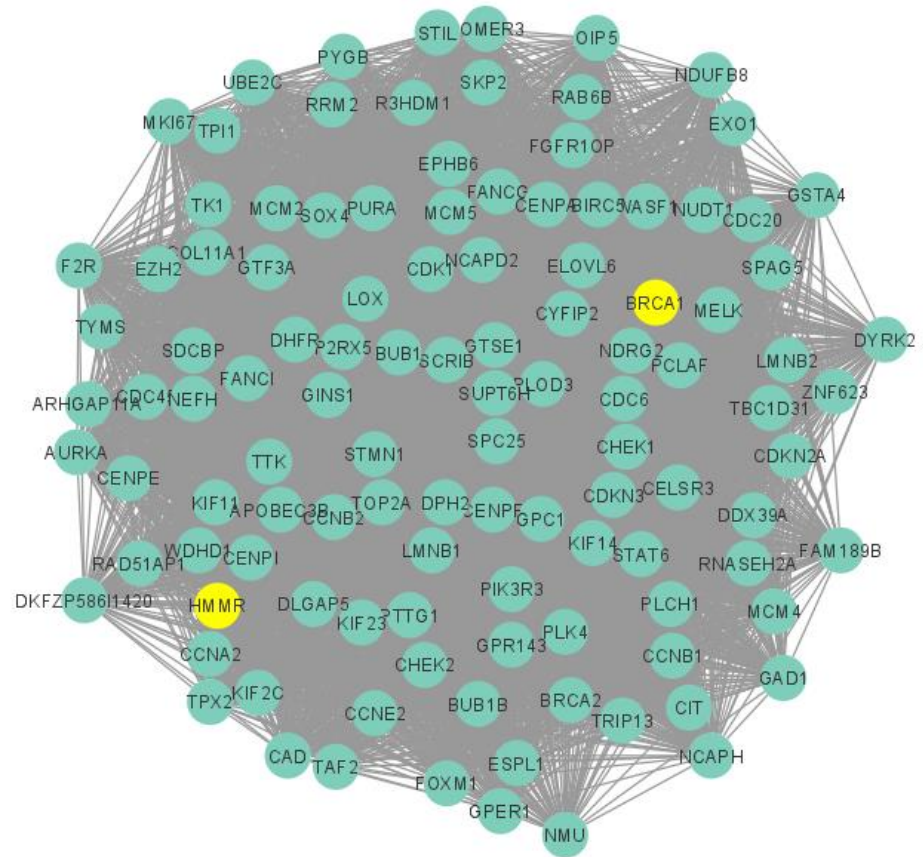


Figure 96. Expanded network of HMMR in Prostate tumour samples.

BRCA1 was found to be co-expressed with HMMR in this expanded view. Both HMMR and BRCA1 are highlighted in yellow.

Table 72. Expression of CD44 in cancer subtypes

Subtype	Log2 FC	P-Value
Breast Cancer Histological		
ILC	0.19	P<0.01
IDC	0.16	P<0.01
DCIS	0.14	P=0.08
Breast Cancer Molecular		
Normal Like	0.37	P<0.01
Luminal A	0.36	P<0.01
Luminal B	0.09	P=0.47
HER2	0.37	P<0.01
Basal	0.69	P<0.01
Ovarian Cancer Tissue location		
Ovarian	-0.24	P<0.01
Peritoneum	-0.01	P=0.98
Fallopian	-0.03	P=0.93
Ovarian Cancer Epithelial		
Serous	-0.40	P<0.01
Endometrioid	-0.01	P=0.96
Mucinous	0.20	P=0.23
Clear Cell	-0.23	P=0.13
Prostate Cancer Gleason Grade Group		
Group 1	-0.39	P<0.01
Group 2&3	-0.38	P<0.01
Group 4	-0.43	P<0.01
Group 5	-0.37	P<0.01

Prostate Cancer Gleason Score		
Score 4	-0.59	P=0.14
Score 5	-0.37	P<0.01
Score 6	-0.37	P<0.01
Score 7	-0.38	P<0.01
Score 8	-0.43	P<0.01
Score 9	-0.41	P<0.01

The expression of CD44 was observed to be significantly upregulated (upregulated DEG) in breast cancer invasive subtypes (ILC, IDC, HER2, and basal). CD44 was significantly downregulated in ovarian and prostate cancer subtypes

IDC: Invasive Ductal Carcinoma, ILC: Invasive Lobular Carcinoma, DCIS: Ductal Carcinoma in-situ

Table 73. Expression of BRCA1 in cancer subtypes

Subtype	Log FC	P-Value
Histological		
IDC	0.34	P<0.01
ILC	0.32	P<0.01
DCIS	0.21	P<0.01
Molecular		
Normal Like	0.08	P=0.50
Luminal A	0.35	P<0.01
Luminal B	0.45	P<0.01
HER2	0.45	P<0.01
Basal	0.54	P<0.01
Tissue location		
Ovarian	0.38	P<0.01
Fallopian	0.69	P<0.01
Peritoneum	-0.08	P=0.71
Epithelial		
Serous	0.36	P<0.01
Endometrioid	0.54	P<0.01
Mucinous	0.61	P<0.01
Clear Cell	0.71	P<0.01
Gleason Grade Group		
Group 1	0.10	P<0.01
Group 2&3	0.13	P<0.01
Group 4	0.07	P=0.21
Group 5	0.16	P<0.01

Gleason Score		
Score 4	0.36	P=0.02
Score 5	0.03	P=0.69
Score 6	0.09	P<0.01
Score 7	0.13	P<0.01
Score 8	0.07	P=0.21
Score 9	0.16	P<0.01

The expression of BRCA1 was observed to be significantly upregulated in nearly all breast, ovarian and prostate cancer subtypes.

IDC: Invasive Ductal Carcinoma, ILC: Invasive Lobular Carcinoma, DCIS: Ductal Carcinoma in-situ

Table 74. Expression of CXCR4 in cancer subtypes

Subtype	Log FC	P-Value
Histological		
IDC	1.15	P<0.01
ILC	1.19	P<0.01
DCIS	1.04	P<0.01
Molecular		
Normal Like	1.33	P<0.01
Luminal A	1.41	P<0.01
Luminal B	1.39	P<0.01
HER2	1.71	P<0.01
Basal	2.01	P<0.01
Tissue location		
Ovarian	1.09	P<0.01
Fallopian	1.54	P<0.01
Peritoneum	-0.25	P=0.54
Epithelial		
Serous	1.13	P<0.01
Endometrioid	1.18	P<0.01
Mucinous	-0.03	P=0.91
Clear Cell	0.71	P<0.01
Gleason Grade Group		
Group 1	-0.72	P<0.01
Group 2&3	-0.65	P<0.01
Group 4	-0.36	P=0.24
Group 5	-0.56	P<0.01

Gleason Score		
Score 4	-1.34	P=0.09
Score 5	-0.80	P=0.03
Score 6	-0.68	P<0.01
Score 7	-0.65	P<0.01
Score 8	-0.36	P=0.24
Score 9	-0.60	P<0.01

The expression of CXCR4 was observed to be significantly upregulated in all breast cancer subtypes and ovarian subtypes that are considered more invasive (ovarian, fallopian, serous, endometrioid, and clear cell). In prostate cancer CXCR4 expression was identified as significantly downregulated.

IDC: Invasive Ductal Carcinoma, ILC: Invasive Lobular Carcinoma, DCIS: Ductal Carcinoma in-situ

Table 75. Expression of TGF- β in cancer subtypes

Subtype	Log FC	P-Value
Histological		
IDC	0.20	P<0.01
ILC	0.10	P=0.25
DCIS	0.29	P<0.01
Molecular		
Normal Like	0.28	P<0.01
Luminal A	0.32	P<0.01
Luminal B	0.30	P<0.01
HER2	0.21	P<0.01
Basal	0.19	P<0.01
Tissue location		
Ovarian	-0.14	P=0.03
Fallopian	-0.06	P=0.85
Peritoneum	-0.12	P=0.61
Epithelial		
Serous	0.13	P=0.03
Endometrioid	0.07	P=0.38
Mucinous	0.10	P=0.33
Clear Cell	0.25	P<0.01
Gleason Grade Group		
Group 1	0.02	P=0.82
Group 2&3	0.003	P=0.95
Group 4	0.16	P=0.24
Group 5	0.03	P=0.69

Gleason Score		
Score 4	0.29	P=0.23
Score 5	-0.11	P=0.35
Score 6	0.04	P=0.42
Score 7	0.06	P=0.09
Score 8	0.12	P=0.16
Score 9	0.02	P=0.72

The expression of TGF- β was observed to be significantly upregulated in all breast cancer subtypes with the exception of ILC. In ovarian subtypes TGF- β was not significant except the clear-cell. This was similarly observed in prostate cancer subtypes with TGF- β not significant.

IDC: Invasive Ductal Carcinoma, ILC: Invasive Lobular Carcinoma, DCIS: Ductal Carcinoma in-situ

Table 76. Expression of HAS1 in cancer subtypes

Subtype	Log FC	P-Value
Histological		
IDC	-0.08	P<0.01
ILC	-0.03	P=0.56
DCIS	-0.06	P=0.38
Molecular		
Normal Like	-0.36	P<0.01
Luminal A	-0.37	P<0.01
Luminal B	-0.45	P<0.01
HER2	-0.39	P<0.01
Basal	-0.25	P<0.01
Tissue location		
Ovarian	-1.65	P<0.01
Fallopian	-1.88	P<0.01
Peritoneum	-0.32	P=0.02
Epithelial		
Serous	-1.49	P<0.01
Endometrioid	-1.55	P<0.01
Mucinous	-1.49	P<0.01
Clear Cell	-1.56	P<0.01
Gleason Grade Group		
Group 1	0.12	P=0.01
Group 2&3	0.09	P=0.02
Group 4	0.11	P=0.23
Group 5	0.06	P=0.30

Gleason Score		
Score 4	0.42	P=0.10
Score 5	0.002	P=0.99
Score 6	0.12	P=0.01
Score 7	0.09	P=0.02
Score 8	0.12	P=0.23
Score 9	0.07	P=0.26

The expression of HAS1 was observed to be significantly downregulated in nearly all breast cancer subtypes. This was similarly observed in ovarian cancer subtypes with also significantly downregulated. In Prostate cancer it was not found to be significantly expressed.

IDC: Invasive Ductal Carcinoma, ILC: Invasive Lobular Carcinoma, DCIS: Ductal Carcinoma in-situ

Table 77. Expression of HAS2 in cancer subtypes

Subtype	Log FC	P-Value
Histological		
IDC	-0.02	P=0.75
ILC	-0.01	P=0.85
DCIS	0.04	P=0.65
Molecular		
Normal Like	-0.02	P=0.92
Luminal A	0.01	P=0.92
Luminal B	-0.16	P=0.25
HER2	-0.05	P=0.72
Basal	0.25	P=0.05
Tissue location		
Ovarian	0.37	P<0.01
Fallopian	0.07	P=0.85
Peritoneum	-0.34	P=0.08
Epithelial		
Serous	0.48	P<0.01
Endometrioid	0.24	P=0.09
Mucinous	0.10	P=0.60
Clear Cell	0.23	P=0.17
Gleason Grade Group		
Group 1	-0.04	P=0.45
Group 2&3	-0.08	P=0.09
Group 4	0.15	P=0.22
Group 5	0.05	P=0.53

Gleason Score		
Score 4	0.04	P=0.92
Score 5	-0.16	P=0.33
Score 6	-0.03	P=0.62
Score 7	-0.08	P=0.09
Score 8	0.15	P=0.22
Score 9	-0.05	P=0.51

The expression of HAS2 was not observed to be significant in breast or prostate cancers and was only significant in ovarian tissue and serous subtypes.

IDC: Invasive Ductal Carcinoma, ILC: Invasive Lobular Carcinoma, DCIS: Ductal Carcinoma in-situ

Table 78. Expression of TNF- α in cancer subtypes

Subtype	Log FC	P-Value
Histological		
IDC	0.13	P<0.01
ILC	0.12	P=0.03
DCIS	0.15	P=0.03
Molecular		
Normal Like	0.05	P=0.69
Luminal A	0.06	P=0.57
Luminal B	0.06	P=0.55
HER2	0.12	P=0.24
Basal	0.51	P<0.01
Tissue location		
Ovarian	0.70	P<0.01
Fallopian	0.55	P=0.04
Peritoneum	-0.02	P=0.96
Epithelial		
Serous	0.76	P<0.01
Endometrioid	0.57	P<0.01
Mucinous	0.36	P<0.01
Clear Cell	0.66	P<0.01
Gleason Grade Group		
Group 1	0.07	P=0.07
Group 2&3	0.09	P<0.01
Group 4	0.07	P=0.38
Group 5	0.08	P=0.10

Gleason Score		
Score 4	0.25	P=0.24
Score 5	-0.04	P=0.72
Score 6	0.07	P=0.07
Score 7	0.09	P<0.01
Score 8	0.07	P=0.37
Score 9	0.09	P=0.04

The expression of TNF- α was significantly upregulated in invasive IDC and basal breast cancers, ovarian tissue and all ovarian epithelial subtypes, and Gleason Grade group 2 and 3.

IDC: Invasive Ductal Carcinoma, ILC: Invasive Lobular Carcinoma, DCIS: Ductal Carcinoma in-situ

Table 79. Expression of MMP9 in cancer subtypes

Subtype	Log FC	P-Value
Histological		
IDC	1.49	P<0.01
ILC	1.36	P<0.01
DCIS	1.45	P<0.01
Molecular		
Normal Like	0.92	P<0.01
Luminal A	1.39	P<0.01
Luminal B	1.98	P<0.01
HER2	1.79	P<0.01
Basal	2.40	P<0.01
Tissue location		
Ovarian	0.90	P<0.01
Fallopian	1.84	P<0.01
Peritoneum	-0.26	P=0.61
Epithelial		
Serous	1.12	P<0.01
Endometrioid	1.23	P<0.01
Mucinous	0.35	P=0.28
Clear Cell	0.56	P=0.04
Gleason Grade Group		
Group 1	0.30	P=0.01
Group 2&3	0.30	P<0.01
Group 4	0.51	P<0.01
Group 5	0.61	P<0.01

Gleason Score		
Score 4	1.27	P=0.06
Score 5	0.10	P=0.80
Score 6	0.31	P=0.02
Score 7	0.30	P<0.01
Score 8	0.51	P=0.04
Score 9	0.58	P<0.01

The expression of MMP9 was significantly upregulated in nearly all breast cancer subtypes. In ovarian cancer clear cell and mucinous were not significant. In prostate cancer MMP9 was significantly upregulated in higher Gleason grade groups and Gleason scores.

IDC: Invasive Ductal Carcinoma, ILC: Invasive Lobular Carcinoma, DCIS: Ductal Carcinoma in-situ

CDC20-BUBR1, Breast, Ovarian and Prostate Histological graded Tumours

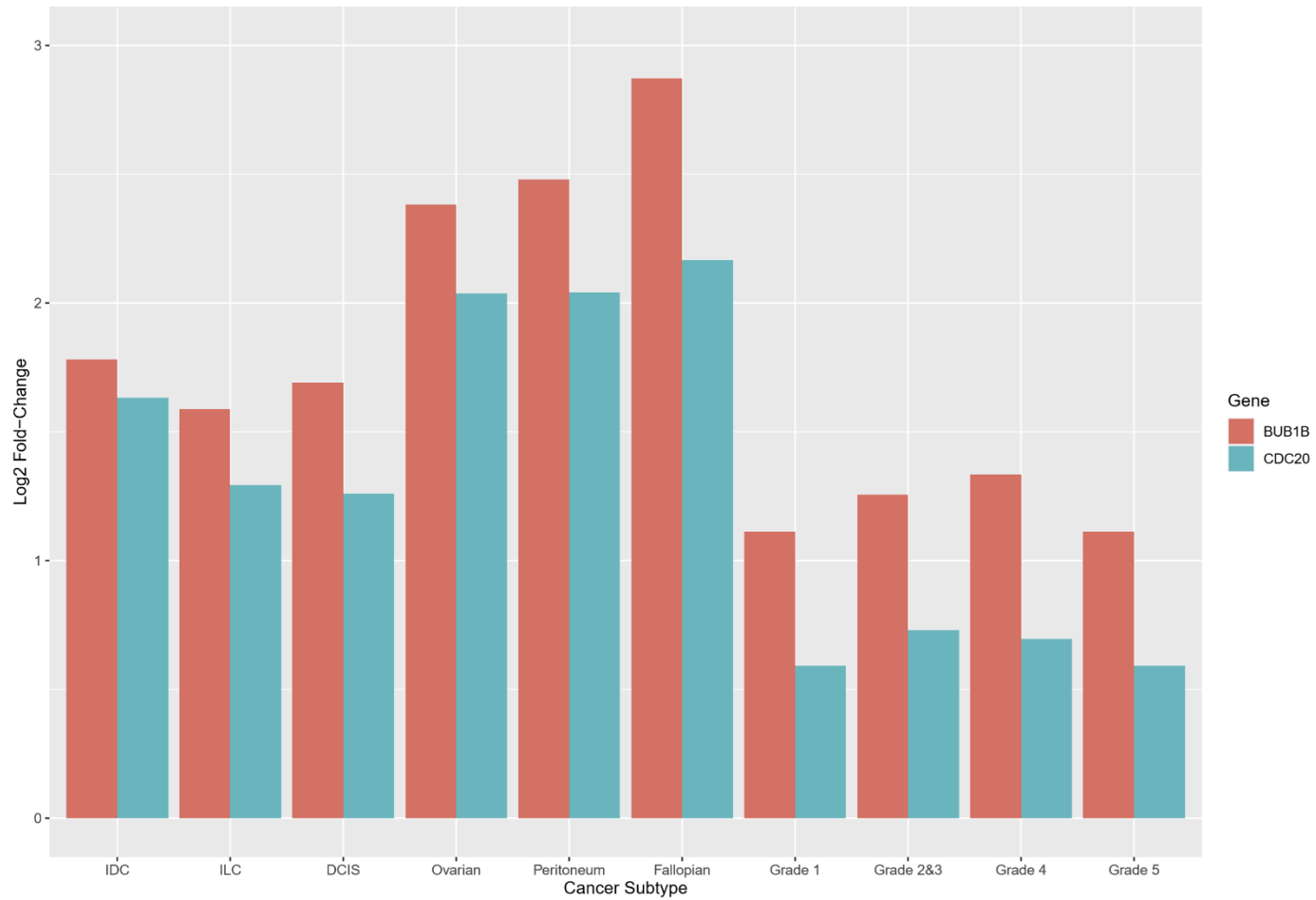


Figure 97. Expression of CDC20 and BUBR1 in breast histological, ovarian tissue, and prostate Gleason grade group subtypes.

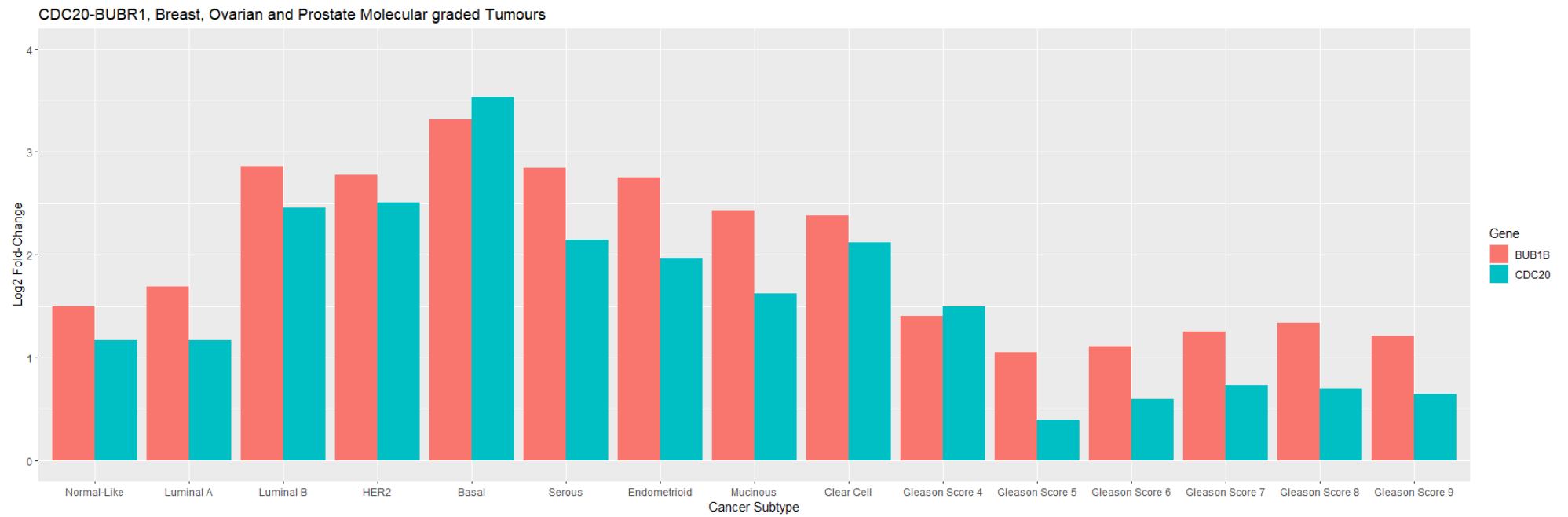


Figure 98. Expression of CDC20 and BUBR1 in breast molecular, ovarian epithelial, and prostate Gleason score subtypes.

Table 80. Expression of CEP135 in cancer subtypes

Subtype	Log FC	P-Value
Histological		
ILC	0.10	P<0.01
IDC	0.11	P<0.01
DCIS	0.03	P=0.53
Molecular		
Normal Like	0.47	P<0.01
Luminal A	0.51	P<0.01
Luminal B	0.57	P<0.01
HER2	0.52	P<0.01
Basal	0.51	P<0.01
Tissue location		
Ovarian	-0.38	P<0.01
Peritoneum	-0.02	P=0.91
Fallopian	-0.25	P=0.16
Epithelial		
Serous	-0.36	P<0.01
Endometrioid	-0.39	P<0.01
Mucinous	-0.62	P<0.01
Clear Cell	-0.55	P<0.01
Gleason Grade Group		
Group 1	-0.21	P<0.01
Group 2&3	-0.22	P<0.01
Group 4	-0.20	P<0.01
Group 5	-0.17	P<0.01

Gleason Score		
Score 4	-0.12	P=0.52
Score 5	-0.13	P=0.12
Score 6	-0.22	P<0.01
Score 7	-0.22	P<0.01
Score 8	-0.20	P<0.01
Score 9	-0.18	P<0.01

The expression of CEP135 was significantly upregulated in breast histological invasive subtypes and molecular subtypes. In ovarian cancer CEP135 was significantly downregulated this was also identified in prostate cancer with CEP135 being significantly downregulated in higher Gleason grades and scores.

IDC: Invasive Ductal Carcinoma, ILC: Invasive Lobular Carcinoma, DCIS: Ductal Carcinoma in-situ

Table 81. Expression of CENPJ in cancer subtypes

Subtype	Log FC	P-Value
Histological		
ILC	0.11	P<0.01
IDC	0.10	P<0.01
DCIS	0.05	P=0.38
Molecular		
Normal Like	0.16	P=0.02
Luminal A	0.11	P=0.04
Luminal B	0.25	P<0.01
HER2	0.32	P<0.01
Basal	0.42	P<0.01
Tissue location		
Ovarian	0.21	P<0.01
Peritoneum	0.32	P=0.91
Fallopian	0.32	P=0.05
Epithelial		
Serous	0.21	P<0.01
Endometrioid	0.29	P<0.01
Mucinous	0.09	P=0.35
Clear Cell	0.11	P=0.20

The expression of CENPJ was significantly upregulated in invasive histological breast cancers and highest expression observed in the more aggressive molecular subtypes. For ovarian cancer subtypes, the highest expression of CENPJ was observed in invasive subtypes also with these being serous and endometrioid.

IDC: Invasive Ductal Carcinoma, ILC: Invasive Lobular Carcinoma, DCIS: Ductal Carcinoma in-situ

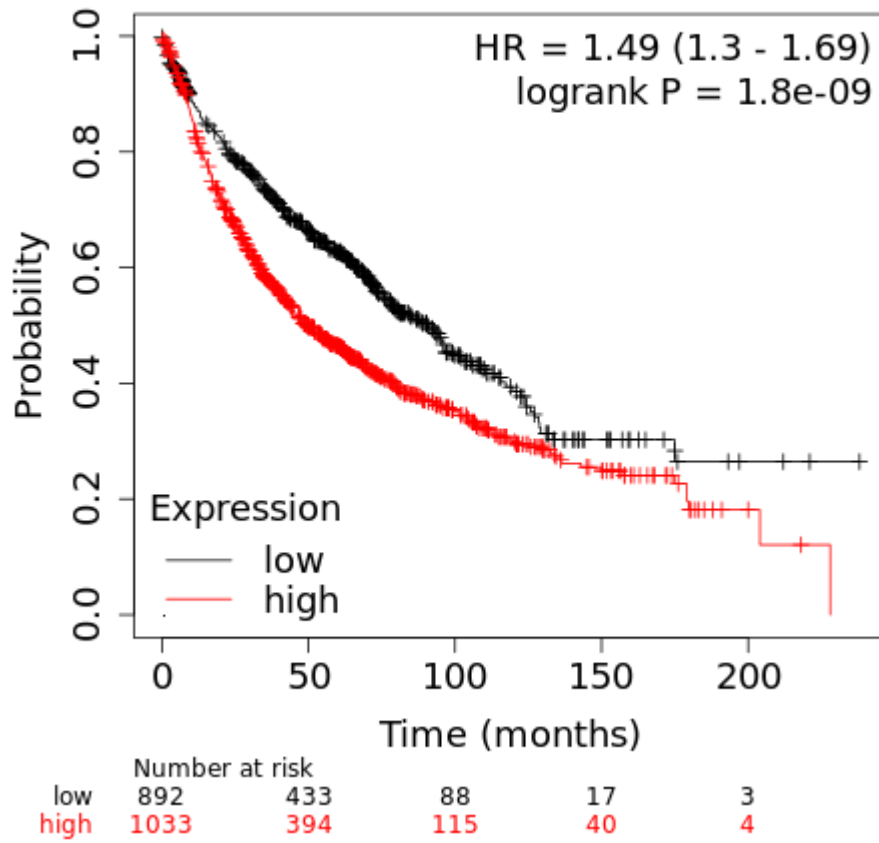


Figure 99. Kaplan-Meier OS survival curves of HSC in lung cancer primary tumour samples.

Red shows high expression and black, low expression. Figure from K-M Plotter web tool.

OS: Overall Survival

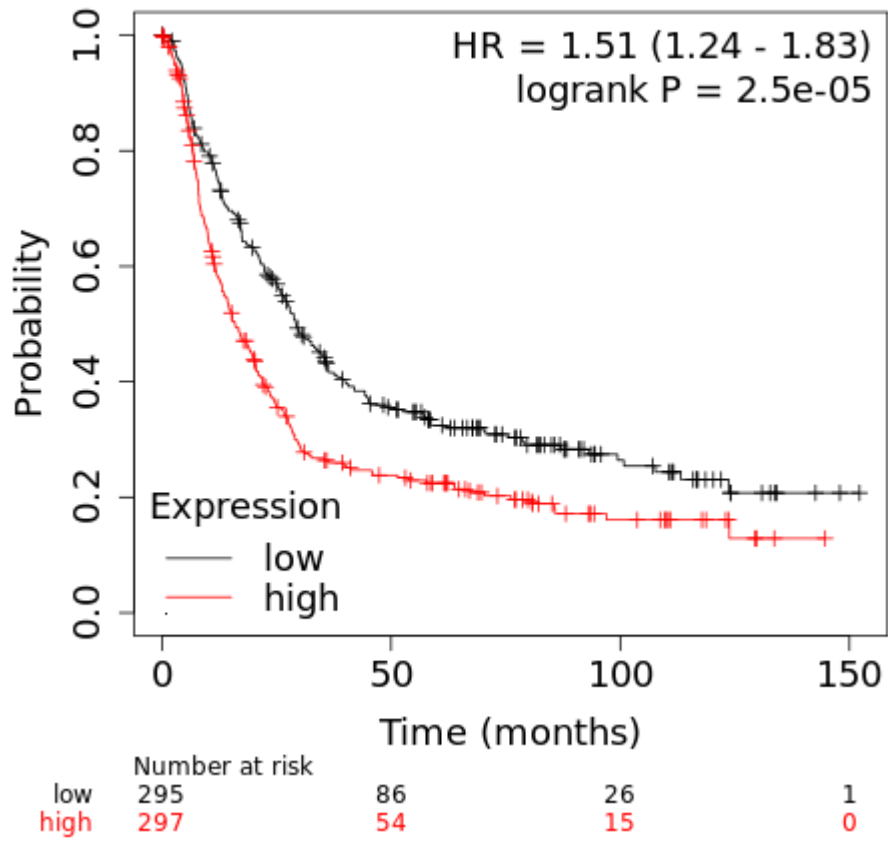


Figure 100. Kaplan-Meier OS survival curves of HSC in gastric cancer primary tumour samples.

Red shows high expression and black, low expression. Figure from K-M Plotter web tool.

OS: Overall Survival

References

- Abdel-Rahman, M. A., Mahfouz, M. and Habashy, H. O. (2022) 'RRM2 expression in different molecular subtypes of breast cancer and its prognostic significance', *Diagnostic Pathology*, 17(1), pp. 1.
- Abiko, K., Matsumura, N., Hamanishi, J., Horikawa, N., Murakami, R., Yamaguchi, K., Yoshioka, Y., Baba, T., Konishi, I. and Mandai, M. (2015) 'IFN-gamma from lymphocytes induces PD-L1 expression and promotes progression of ovarian cancer', *Br J Cancer*, 112(9), pp. 1501-9.
- Agalliu, I., Gern, R., Leanza, S. and Burk, R. D. (2009) 'Associations of high-grade prostate cancer with BRCA1 and BRCA2 founder mutations', *Clinical cancer research : an official journal of the American Association for Cancer Research*, 15(3), pp. 1112-1120.
- Agnoli, C., Grioni, S., Pala, V., Allione, A., Matullo, G., Gaetano, C. D., Tagliabue, G., Sieri, S. and Krogh, V. (2017) 'Biomarkers of inflammation and breast cancer risk: a case-control study nested in the EPIC-Varese cohort', *Scientific reports*, 7(1), pp. 12708-12708.
- Aguirre-Gamboa, R., Gomez-Rueda, H., Martinez-Ledesma, E., Martinez-Torteya, A., Chacolla-Huaringa, R., Rodriguez-Barrientos, A., Tamez-Pena, J. G. and Trevino, V. (2013) 'SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis', *PLoS One*, 8(9), pp. e74250.
- Ahmed, M. and Rahman, N. (2006) 'ATM and breast cancer susceptibility', *Oncogene*, 25(43), pp. 5906-11.
- Ahn, J., Urist, M. and Prives, C. (2004) 'The Chk2 protein kinase', *DNA Repair (Amst)*, 3(8-9), pp. 1039-47.
- Allott, E. H., Masko, E. M. and Freedland, S. J. (2013) 'Obesity and Prostate Cancer: Weighing the Evidence', *European Urology*, 63(5), pp. 800-809.
- Alvarez, J. G. B. and Otterson, G. A. (2019) 'Agents to treat BRAF-mutant lung cancer', *Drugs in context*, 8, pp. 212566-212566.
- Anders, C. K. and Carey, L. A. (2009) 'Biology, metastatic patterns, and treatment of patients with triple-negative breast cancer', *Clin Breast Cancer*, 9 Suppl 2, pp. S73-81.
- Andersen, C. L., Sikora, M. J., Boisen, M. M., Ma, T., Christie, A., Tseng, G., Park, Y., Luthra, S., Chandran, U., Haluska, P., Mantia-Smaldone, G. M., Odunsi, K., McLean, K., Lee, A. V., Elishaev, E., Edwards, R. P. and Oesterreich, S. (2017) 'Active Estrogen Receptor-alpha Signaling in Ovarian Cancer Models and Clinical Specimens', *Clinical cancer research : an official journal of the American Association for Cancer Research*, 23(14), pp. 3802-3812.
- Ang, M., Borg, M., O'Callaghan, M. E. and for the South Australian Prostate Cancer Clinical Outcomes, C. (2020) 'Survival outcomes in men with a positive family history of prostate cancer: a registry based study', *BMC Cancer*, 20(1), pp. 894.
- Angèle, S., Falconer, A., Edwards, S. M., Dörk, T., Bremer, M., Moullan, N., Chapot, B., Muir, K., Houlston, R., Norman, A. R., Bullock, S., Hope, Q., Meitz, J., Dearnaley, D., Dowe, A., Southgate, C., Ardern-Jones, A., Easton, D. F., Eeles, R. A. and Hall, J. (2004) 'ATM polymorphisms as risk factors for prostate cancer development', *Br J Cancer*, 91(4), pp. 783-7.
- Antoniou, A., Pharoah, P. D., Narod, S., Risch, H. A., Eyfjord, J. E., Hopper, J. L., Loman, N., Olsson, H., Johannsson, O., Borg, A., Pasini, B., Radice, P., Manoukian, S., Eccles, D. M., Tang, N., Olah, E.,

- Anton-Culver, H., Warner, E., Lubinski, J., Gronwald, J., Gorski, B., Tulinius, H., Thorlacius, S., Eerola, H., Nevanlinna, H., Syrjakoski, K., Kallioniemi, O. P., Thompson, D., Evans, C., Peto, J., Lalloo, F., Evans, D. G. and Easton, D. F. (2003) 'Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies', *Am J Hum Genet*, 72(5), pp. 1117-30.
- Arendt, L. M., McCready, J., Keller, P. J., Baker, D. D., Naber, S. P., Seewaldt, V. and Kuperwasser, C. (2013) 'Obesity promotes breast cancer by CCL2-mediated macrophage recruitment and angiogenesis', *Cancer Res*, 73(19), pp. 6080-93.
- Arquint, C. and Nigg, E. A. (2014) 'STIL microcephaly mutations interfere with APC/C-mediated degradation and cause centriole amplification', *Curr Biol*, 24(4), pp. 351-60.
- Ascierto, P. A., Kirkwood, J. M., Grob, J.-J., Simeone, E., Grimaldi, A. M., Maio, M., Palmieri, G., Testori, A., Marincola, F. M. and Mozzillo, N. (2012) 'The role of BRAF V600 mutation in melanoma', *Journal of translational medicine*, 10, pp. 85-85.
- Aubrey, B. J., Kelly, G. L., Janic, A., Herold, M. J. and Strasser, A. (2018) 'How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression?', *Cell Death & Differentiation*, 25(1), pp. 104-113.
- Aulmann, S., Braun, L., Mietzsch, F., Longerich, T., Penzel, R., Schirmacher, P. and Sinn, H. P. (2012) 'Transitions between flat epithelial atypia and low-grade ductal carcinoma in situ of the breast', *Am J Surg Pathol*, 36(8), pp. 1247-52.
- Auner, V., Kriegshäuser, G., Tong, D., Horvat, R., Reinthaller, A., Mustea, A. and Zeillinger, R. (2009) 'KRAS mutation analysis in ovarian samples using a high sensitivity biochip assay', *BMC Cancer*, 9(1), pp. 111.
- Ayob, A. Z. and Ramasamy, T. S. (2018) 'Cancer stem cells as key drivers of tumour progression', *Journal of Biomedical Science*, 25(1), pp. 20.
- Babaier, A. and Ghatage, P. (2020) 'Mucinous Cancer of the Ovary: Overview and Current Status', *Diagnostics (Basel, Switzerland)*, 10(1), pp. 52.
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S. and Soboleva, A. (2013) 'NCBI GEO: archive for functional genomics data sets--update', *Nucleic Acids Res*, 41(Database issue), pp. D991-5.
- Barzi, A., Lenz, A. M., Labonte, M. J. and Lenz, H. J. (2013) 'Molecular pathways: Estrogen pathway in colorectal cancer', *Clin Cancer Res*, 19(21), pp. 5842-8.
- Basakran, N. S. (2015) 'CD44 as a potential diagnostic tumor marker', *Saudi medical journal*, 36(3), pp. 273-279.
- Basto, R., Brunk, K., Vinadogrova, T., Peel, N., Franz, A., Khodjakov, A. and Raff, J. W. (2008) 'Centrosome amplification can initiate tumorigenesis in flies', *Cell*, 133(6), pp. 1032-42.
- Bauer, K. R., Brown, M., Cress, R. D., Parise, C. A. and Caggiano, V. (2007) 'Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry', *Cancer*, 109(9), pp. 1721-8.
- Bayliss, R., Sardon, T., Vernos, I. and Conti, E. (2003) 'Structural basis of Aurora-A activation by TPX2 at the mitotic spindle', *Mol Cell*, 12(4), pp. 851-62.

- Bechis, S. K., Carroll, P. R. and Cooperberg, M. R. (2011) 'Impact of age at diagnosis on prostate cancer treatment and survival', *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 29(2), pp. 235-241.
- Beehler, G. P., Sekhon, M., Baker, J. A., Teter, B. E., McCann, S. E., Rodabaugh, K. J. and Moysich, K. B. (2006) 'Risk of Ovarian Cancer Associated with BMI Varies by Menopausal Status', *The Journal of Nutrition*, 136(11), pp. 2881-2886.
- Begg, C. B., Rice, M. S., Zabor, E. C. and Tworoger, S. S. (2017) 'Examining the common aetiology of serous ovarian cancers and basal-like breast cancers using double primaries', *Br J Cancer*, 116(8), pp. 1088-1091.
- Bellance, C., Khan, J. A., Meduri, G., Guiochon-Mantel, A., Lombes, M. and Loosfelt, H. (2013) 'Progesterone receptor isoforms PRA and PRB differentially contribute to breast cancer cell migration through interaction with focal adhesion kinase complexes', *Mol Biol Cell*, 24(9), pp. 1363-74.
- Belsches, A. P., Haskell, M. D. and Parsons, S. J. (1997) 'Role of c-Src tyrosine kinase in EGF-induced mitogenesis', *Front Biosci*, 2, pp. d501-18.
- Beltran, M., Tavares, M., Justin, N., Khandelwal, G., Ambrose, J., Foster, B. M., Worlock, K. B., Tvardovskiy, A., Kunzelmann, S., Herrero, J., Bartke, T., Gamblin, S. J., Wilson, J. R. and Jenner, R. G. (2019) 'G-tract RNA removes Polycomb repressive complex 2 from genes', *Nature Structural & Molecular Biology*, 26(10), pp. 899-909.
- Beral, V., Gaitskell, K., Hermon, C., Moser, K., Reeves, G. and Peto, R. (2015) 'Menopausal hormone use and ovarian cancer risk: individual participant meta-analysis of 52 epidemiological studies', *Lancet*, 385(9980), pp. 1835-42.
- Berkemeyer, S., Lemke, D. and Hense, H. W. (2016) 'Incidence and Mortality Trends in German Women with Breast Cancer Using Age, Period and Cohort 1999 to 2008.(Report)', 11(3).
- Bertheau, P., Lehmann-Che, J., Varna, M., Dumay, A., Poirot, B., Porcher, R., Turpin, E., Plassa, L. F., de Roquancourt, A., Bourstyn, E., de Cremoux, P., Janin, A., Giacchetti, S., Espie, M. and de The, H. (2013) 'p53 in breast cancer subtypes and new insights into response to chemotherapy', *Breast*, 22 Suppl 2, pp. S27-9.
- Bickenson, A. F. (2012) 'A stabilizing influence for K-fibres', *Nature Reviews Molecular Cell Biology*, 13(1), pp. 3-3.
- Bjornstrom, L. and Sjoberg, M. (2005) 'Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes', *Mol Endocrinol*, 19(4), pp. 833-42.
- Blanco, A., de la Hoya, M., Osorio, A., Diez, O., Miramar, M. D., Infante, M., Martinez-Bouzas, C., Torres, A., Lasa, A., Llorca, G., Brunet, J., Grana, B., Perez Segura, P., Garcia, M. J., Gutierrez-Enriquez, S., Carracedo, A., Tejada, M. I., Velasco, E. A., Calvo, M. T., Balmana, J., Benitez, J., Caldes, T. and Vega, A. (2013) 'Analysis of PALB2 gene in BRCA1/BRCA2 negative Spanish hereditary breast/ovarian cancer families with pancreatic cancer cases', *PLoS One*, 8(7), pp. e67538.
- Bogliolo, S., Cassani, C., Gardella, B., Musacchi, V., Babilonti, L., Venturini, P. L., Ferrero, S. and Spinillo, A. (2015) 'Oxaliplatin for the treatment of ovarian cancer', *Expert Opin Investig Drugs*, 24(9), pp. 1275-86.
- Bombonati, A. and Sgroi, D. C. (2011) 'The molecular pathology of breast cancer progression', *J Pathol*, 223(2), pp. 307-17.

- Bonkhoff, H. (2018) 'Estrogen receptor signaling in prostate cancer: Implications for carcinogenesis and tumor progression', *Prostate*, 78(1), pp. 2-10.
- Bonome, T., Levine, D. A., Shih, J., Randonovich, M., Pise-Masison, C. A., Bogomolnii, F., Ozbun, L., Brady, J., Barrett, J. C., Boyd, J. and Birrer, M. J. (2008) 'A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer', *Cancer Res*, 68(13), pp. 5478-86.
- Borg, A., Haile, R. W., Malone, K. E., Capanu, M., Diep, A., Torngren, T., Teraoka, S., Begg, C. B., Thomas, D. C., Concannon, P., Mellemkjaer, L., Bernstein, L., Tellhed, L., Xue, S., Olson, E. R., Liang, X., Dolle, J., Borresen-Dale, A. L. and Bernstein, J. L. (2010) 'Characterization of BRCA1 and BRCA2 deleterious mutations and variants of unknown clinical significance in unilateral and bilateral breast cancer: the WECARE study', *Hum Mutat*, 31(3), pp. E1200-40.
- Bourguignon, L. Y. W., Wong, G., Earle, C. A. and Xia, W. (2011) 'Interaction of low molecular weight hyaluronan with CD44 and toll-like receptors promotes the actin filament-associated protein 110-actin binding and MyD88-NFκB signaling leading to proinflammatory cytokine/chemokine production and breast tumor invasion', *Cytoskeleton (Hoboken, N.J.)*, 68(12), pp. 671-693.
- Bousquenaud, M., Fico, F., Solinas, G., Rüegg, C. and Santamaria-Martínez, A. (2018) 'Obesity promotes the expansion of metastasis-initiating cells in breast cancer', *Breast Cancer Res*, 20(1), pp. 104.
- Bowen, N. J., Walker, L. D., Matyunina, L. V., Logani, S., Totten, K. A., Benigno, B. B. and McDonald, J. F. (2009) 'Gene expression profiling supports the hypothesis that human ovarian surface epithelia are multipotent and capable of serving as ovarian cancer initiating cells', *BMC Med Genomics*, 2, pp. 71.
- Brandt, A., Bermejo, J. L., Sundquist, J. and Hemminki, K. (2008) 'Age of onset in familial cancer', *Annals of oncology : official journal of the European Society for Medical Oncology*, 19(12), pp. 2084-2088.
- Brandt, A., Bermejo, J. L., Sundquist, J. and Hemminki, K. (2010) 'Age of onset in familial breast cancer as background data for medical surveillance', *British journal of cancer*, 102(1), pp. 42-47.
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. and Jemal, A. (2018) 'Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries', *CA: A Cancer Journal for Clinicians*, 68(6), pp. 394-424.
- Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., Gaasterland, T., Glenisson, P., Holstege, F. C. P., Kim, I. F., Markowitz, V., Matese, J. C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J. and Vingron, M. (2001) 'Minimum information about a microarray experiment (MIAME)—toward standards for microarray data', *Nature Genetics*, 29(4), pp. 365-371.
- Brisken, C. (2013) 'Progesterone signalling in breast cancer: a neglected hormone coming into the limelight', *Nat Rev Cancer*, 13(6), pp. 385-96.
- Brouckaert, O., Wildiers, H., Floris, G. and Neven, P. (2012) 'Update on triple-negative breast cancer: prognosis and management strategies', *Int J Womens Health*, 4, pp. 511-20.
- Brown, J. and Frumovitz, M. (2014) 'Mucinous tumors of the ovary: current thoughts on diagnosis and management', *Current oncology reports*, 16(6), pp. 389-389.
- Burke, L. J., Sevcik, J., Gambino, G., Tudini, E., Mucaki, E. J., Shirley, B. C., Whiley, P., Parsons, M. T., De Leener, K., Gutiérrez-Enríquez, S., Santamariña, M., Caputo, S. M., Santana Dos Santos, E.,

Soukupova, J., Janatova, M., Zemankova, P., Lhotova, K., Stolarova, L., Borecka, M., Moles-Fernández, A., Manoukian, S., Bonanni, B., Consortium, E., Edwards, S. L., Blok, M. J., van Overeem Hansen, T., Rossing, M., Diez, O., Vega, A., Claes, K. B. M., Goldgar, D. E., Rouleau, E., Radice, P., Peterlongo, P., Rogan, P. K., Caligo, M., Spurdle, A. B. and Brown, M. A. (2018) 'BRCA1 and BRCA2 5' noncoding region variants identified in breast cancer patients alter promoter activity and protein binding', *Human mutation*, 39(12), pp. 2025-2039.

Burstein, M. D., Tsimelzon, A., Poage, G. M., Covington, K. R., Contreras, A., Fuqua, S. A., Savage, M. I., Osborne, C. K., Hilsenbeck, S. G., Chang, J. C., Mills, G. B., Lau, C. C. and Brown, P. H. (2015) 'Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer', *Clin Cancer Res*, 21(7), pp. 1688-98.

Butler, D. E., Marlein, C., Walker, H. F., Frame, F. M., Mann, V. M., Simms, M. S., Davies, B. R., Collins, A. T. and Maitland, N. J. (2017) 'Inhibition of the PI3K/AKT/mTOR pathway activates autophagy and compensatory Ras/Raf/MEK/ERK signalling in prostate cancer', *Oncotarget*, 8(34), pp. 56698-56713.

Cai, H., Li, H., Li, J., Li, X., Li, Y., Shi, Y. and Wang, D. (2016) 'Sonic hedgehog signaling pathway mediates development of hepatocellular carcinoma', *Tumour Biol*.

Cai, J., Yang, J. and Jones, D. P. (1998) 'Mitochondrial control of apoptosis: the role of cytochrome c', *Biochim Biophys Acta*, 1366(1-2), pp. 139-49.

Cai, S., O'Connell, C. B., Khodjakov, A. and Walczak, C. E. (2009) 'Chromosome congression in the absence of kinetochore fibres', *Nature Cell Biology*, 11(7), pp. 832-838.

Cai, Y., Mei, J., Xiao, Z., Xu, B., Jiang, X., Zhang, Y. and Zhu, Y. (2019a) 'Identification of five hub genes as monitoring biomarkers for breast cancer metastasis in silico', *Hereditas*, 156(1), pp. 20.

Cai, Y., Sheng, Z., Chen, Y. and Wang, J. (2019b) 'LncRNA HMMR-AS1 promotes proliferation and metastasis of lung adenocarcinoma by regulating MiR-138/sirt6 axis', *Aging*, 11(10), pp. 3041-3054.

Campbell, I. G., Russell, S. E., Choong, D. Y., Montgomery, K. G., Ciavarella, M. L., Hooi, C. S., Cristiano, B. E., Pearson, R. B. and Phillips, W. A. (2004) 'Mutation of the PIK3CA gene in ovarian and breast cancer', *Cancer Res*, 64(21), pp. 7678-81.

Campbell, P. J. and Getz, G. and Korbel, J. O. and Stuart, J. M. and Jennings, J. L. and Stein, L. D. and Perry, M. D. and Nahal-Bose, H. K. and Ouellette, B. F. F. and Li, C. H. and Rheinbay, E. and Nielsen, G. P. and Sgroi, D. C. and Wu, C.-L. and Faquin, W. C. and Deshpande, V. and Boutros, P. C. and Lazar, A. J. and Hoadley, K. A. and Louis, D. N. and Dursi, L. J. and Yung, C. K. and Bailey, M. H. and Saksena, G. and Raine, K. M. and Buchhalter, I. and Kleinheinz, K. and Schlesner, M. and Zhang, J. and Wang, W. and Wheeler, D. A. and Ding, L. and Simpson, J. T. and O'Connor, B. D. and Yakneen, S. and Ellrott, K. and Miyoshi, N. and Butler, A. P. and Royo, R. and Shorser, S. I. and Vazquez, M. and Rausch, T. and Tiao, G. and Waszak, S. M. and Rodriguez-Martin, B. and Shringarpure, S. and Wu, D.-Y. and Demidov, G. M. and Delaneau, O. and Hayashi, S. and Imoto, S. and Habermann, N. and Segre, A. V. and Garrison, E. and Cafferkey, A. and Alvarez, E. G. and Heredia-Genestar, J. M. and Muyas, F. and Drechsel, O. and Bruzos, A. L. and Temes, J. and Zamora, J. and Baez-Ortega, A. and Kim, H.-L. and Mashl, R. J. and Ye, K. and DiBiase, A. and Huang, K.-I. and Letunic, I. and McLellan, M. D. and Newhouse, S. J. and Shmaya, T. and Kumar, S. and Wedge, D. C. and Wright, M. H. and Yellapantula, V. D. and Gerstein, M. and Khurana, E. and Marques-Bonet, T. and Navarro, A. and Bustamante, C. D. and Siebert, R. and Nakagawa, H. and Easton, D. F. and Ossowski, S. and Tubio, J. M. C. and De La Vega, F. M. and Estivill, X. and Yuen, D. and Mihaiescu, G. L. and Omberg, L. and Ferretti, V. and Sabarinathan, R. and Pich, O. and Gonzalez-Perez, A. and Taylor-Weiner, A. and Fittall, M. W. and Demeulemeester, J. and Tarabichi,

M. and Roberts, N. D. and Van Loo, P. and Cortés-Ciriano, I. and Urban, L. and Park, P. and Zhu, B. and Pitkänen, E. and Li, Y. and Saini, N. and Klimczak, L. J. and Weischenfeldt, J. and Sidiropoulos, N. and Alexandrov, L. B. and Rabionet, R. and Escaramis, G. and Bosio, M. and Holik, A. Z. and Susak, H. and Prasad, A. and Erkek, S. and Calabrese, C. and Raeder, B. and Harrington, E. and Mayes, S. and Turner, D. and Juul, S. and Roberts, S. A. and Song, L. and Koster, R. and Mirabello, L. and Hua, X. and Tanskanen, T. J. and Tojo, M. and Chen, J. and Aaltonen, L. A. and Rättsch, G. and Schwarz, R. F. and Butte, A. J. and Brazma, A. and Chanock, S. J. and Chatterjee, N. and Stegle, O. and Harismendy, O. and Bova, G. S. and Gordenin, D. A. and Haan, D. and Sieverling, L. and Feuerbach, L. and Chalmers, D. and Joly, Y. and Knoppers, B. and Molnár-Gábor, F. and Phillips, M. and Thorogood, A. and Townend, D. and Goldman, M. and Fonseca, N. A. and Xiang, Q. and Craft, B. and Piñeiro-Yáñez, E. and Muñoz, A. and Petryszak, R. and Füllgrabe, A. and Al-Shahrour, F. and Keays, M. and Haussler, D. and Weinstein, J. and Huber, W. and Valencia, A. and Papatheodorou, I. and Zhu, J. and Fan, Y. and Torrents, D. and Bieg, M. and Chen, K. and Chong, Z. and Cibulskis, K. and Eils, R. and Fulton, R. S. and Gelpi, J. L. and Gonzalez, S. and Gut, I. G. and Hach, F. and Heinold, M. and Hu, T. and Huang, V. and Hutter, B. and Jäger, N. and Jung, J. and Kumar, Y. and Lalansingh, C. and Leshchiner, I. and Livitz, D. and Ma, E. Z. and Maruvka, Y. E. and Milovanovic, A. and Nielsen, M. M. and Paramasivam, N. and Pedersen, J. S. and Puiggròs, M. and Sahinalp, S. C. and Sarrafi, I. and Stewart, C. and Stobbe, M. D. and Wala, J. A. and Wang, J. and Wendl, M. and Werner, J. and Wu, Z. and Xue, H. and Yamaguchi, T. N. and Yellapantula, V. and Davis-Dusenbery, B. N. and Grossman, R. L. and Kim, Y. and Heinold, M. C. and Hinton, J. and Jones, D. R. and Menzies, A. and Stebbings, L. and Hess, J. M. and Rosenberg, M. and Dunford, A. J. and Gupta, M. and Imielinski, M. and Meyerson, M. and Beroukhi, R. and Reimand, J. and Dhingra, P. and Favero, F. and Dentro, S. and Wintersinger, J. and Rudneva, V. and Park, J. W. and Hong, E. P. and Heo, S. G. and Kahles, A. and Lehmann, K.-V. and Soulette, C. M. and Shiraishi, Y. and Liu, F. and He, Y. and Demircioğlu, D. and Davidson, N. R. and Greger, L. and Li, S. and Liu, D. and Stark, S. G. and Zhang, F. and Amin, S. B. and Bailey, P. and Chateigner, A. and Frenkel-Morgenstern, M. and Hou, Y. and Huska, M. R. and Kilpinen, H. and Lamaze, F. C. and Li, C. and Li, X. and Li, X. and Liu, X. and Marin, M. G. and Markowski, J. and Nandi, T. and Ojesina, A. I. and Pan-Hammarström, Q. and Park, P. J. and Peadarallu, C. S. and Su, H. and Tan, P. and Teh, B. T. and Wang, J. and Xiong, H. and Ye, C. and Yung, C. and Zhang, X. and Zheng, L. and Zhu, S. and Awadalla, P. and Creighton, C. J. and Wu, K. and Yang, H. and Göke, J. and Zhang, Z. and Brooks, A. N. and Martincorena, I. and Rubio-Perez, C. and Juul, M. and Schumacher, S. and Shapira, O. and Tamborero, D. and Mularoni, L. and Hornshøj, H. and Deu-Pons, J. and Muiños, F. and Bertl, J. and Guo, Q. and The, I. T. P.-C. A. o. W. G. C. (2020) 'Pan-cancer analysis of whole genomes', *Nature*, 578(7793), pp. 82-93.

Campos, M. P., Cohen, M., Von Eeuw, E., Velculescu, V., Kujak, J. L., Conklin, D., Hallberg, D., Slamon, D. J., Elvin, J. and Konecny, G. E. (2018) 'BRAF Mutations Occur Infrequently in Ovarian Cancer but Suggest Responsiveness to BRAF and MEK Inhibition', *JCO Precision Oncology*, (2), pp. 1-6.

Cancer Genome Atlas Research, N., Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C. and Stuart, J. M. (2013) 'The Cancer Genome Atlas Pan-Cancer analysis project', *Nature genetics*, 45(10), pp. 1113-1120.

Cancer Genome Atlas Research Network, Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C. and Stuart, J. M. (2013) 'The Cancer Genome Atlas Pan-Cancer analysis project', *Nature genetics*, 45(10), pp. 1113-1120.

Caputo, F., Santini, C., Bardasi, C., Cerma, K., Casadei-Gardini, A., Spallanzani, A., Andrikou, K., Cascinu, S. and Gelsomino, F. (2019) 'BRAF-Mutated Colorectal Cancer: Clinical and Molecular Insights', *International journal of molecular sciences*, 20(21), pp. 5369.

Carlson, M. 2016a. *hgu95av2.db: Affymetrix Human Genome U95 Set annotation data (chip hgu95av2)*. Bioconductor.

- Carlson, M. 2016b. *hgu133a.db: Affymetrix Human Genome U133 Set annotation data (chip hgu133a)* . Bioconductor.
- Carlson, M. 2016c. *hgu133plus2.db: Affymetrix Human Genome U133 Plus 2.0 Array annotation data (chip hgu133plus2)* . Bioconductor.
- Carlson, M. 2016d. *u133x3p.db: Affymetrix Human X3P Array annotation data (chip u133x3p)* . Bioconductor.
- Carneiro, B. A. and El-Deiry, W. S. (2020) 'Targeting apoptosis in cancer therapy', *Nature Reviews Clinical Oncology*, 17(7), pp. 395-417.
- Castiel, A., Danieli, M. M., David, A., Moshkovitz, S., Aplan, P. D., Kirsch, I. R., Brandeis, M., Krämer, A. and Izraeli, S. (2011) 'The Stil protein regulates centrosome integrity and mitosis through suppression of Chfr', *Journal of cell science*, 124(Pt 4), pp. 532-539.
- Castro, E. and Eeles, R. (2012) 'The role of BRCA1 and BRCA2 in prostate cancer', *Asian journal of andrology*, 14(3), pp. 409-414.
- Cathro, H. P. and Stoler, M. H. (2002) 'Expression of cytokeratins 7 and 20 in ovarian neoplasia', *Am J Clin Pathol*, 117(6), pp. 944-51.
- Chan, G. K., Jablonski, S. A., Sudakin, V., Hittle, J. C. and Yen, T. J. (1999) 'Human BUBR1 is a mitotic checkpoint kinase that monitors CENP-E functions at kinetochores and binds the cyclosome/APC', *The Journal of cell biology*, 146(5), pp. 941-954.
- Chan, K. K., Matchett, K. B., McEnhill, P. M., Dakir, E. H., McMullin, M. F., El-Tanani, Y., Patterson, L., Faheem, A., Rudland, P. S., McCarron, P. A. and El-Tanani, M. (2015) 'Protein deregulation associated with breast cancer metastasis', *Cytokine & Growth Factor Reviews*, 26(4), pp. 415-423.
- Chandra, A., Lan, S., Zhu, J., Siclari, V. A. and Qin, L. (2013) 'Epidermal growth factor receptor (EGFR) signaling promotes proliferation and survival in osteoprogenitors by increasing early growth response 2 (EGR2) expression', *J Biol Chem*, 288(28), pp. 20488-98.
- Chandrasekaran, S., Marshall, J. R., Messing, J. A., Hsu, J. W. and King, M. R. (2014) 'TRAIL-mediated apoptosis in breast cancer cells cultured as 3D spheroids', *PLoS One*, 9(10), pp. e111487.
- Chang, F., Lee, J. T., Navolanic, P. M., Steelman, L. S., Shelton, J. G., Blalock, W. L., Franklin, R. A. and McCubrey, J. A. (2003) 'Involvement of PI3K/Akt pathway in cell cycle progression, apoptosis, and neoplastic transformation: a target for cancer chemotherapy', *Leukemia*, 17(3), pp. 590-603.
- Chappell, W. H., Lehmann, B. D., Terrian, D. M., Abrams, S. L., Steelman, L. S. and McCubrey, J. A. (2012) 'p53 expression controls prostate cancer sensitivity to chemotherapy and the MDM2 inhibitor Nutlin-3', *Cell cycle (Georgetown, Tex.)*, 11(24), pp. 4579-4588.
- Charitou, T., Bryan, K. and Lynn, D. J. (2016) 'Using biological networks to integrate, visualize and analyze genomics data', *Genetics Selection Evolution*, 48(1), pp. 27.
- Chatterjee, S. J. and McCaffrey, L. (2014) 'Emerging role of cell polarity proteins in breast cancer progression and metastasis', *Breast cancer (Dove Medical Press)*, 6, pp. 15-27.
- Cheab, B., Auguste, A. and Leary, A. (2015) 'The PI3K/Akt/mTOR pathway in ovarian cancer: therapeutic opportunities and challenges', *Chinese journal of cancer*, 34(1), pp. 4-16.

- Cheang, M. C. U., Voduc, D., Bajdik, C., Leung, S., McKinney, S., Chia, S. K., Perou, C. M. and Nielsen, T. O. (2008) 'Basal-Like Breast Cancer Defined by Five Biomarkers Has Superior Prognostic Value than Triple-Negative Phenotype', *Clinical Cancer Research*, 14(5), pp. 1368.
- Chen, C., Chen, Y., Hu, L.-K., Jiang, C.-C., Xu, R.-F. and He, X.-Z. (2018) 'The performance of the new prognostic grade and stage groups in conservatively treated prostate cancer', *Asian journal of andrology*, 20(4), pp. 366-371.
- Chen, D. T., Nasir, A., Culhane, A., Venkataramu, C., Fulp, W., Rubio, R., Wang, T., Agrawal, D., McCarthy, S. M., Gruidl, M., Bloom, G., Anderson, T., White, J., Quackenbush, J. and Yeatman, T. (2010) 'Proliferative genes dominate malignancy-risk gene signature in histologically-normal breast tissue', *Breast Cancer Res Treat*, 119(2), pp. 335-46.
- Chen, E., Zhu, H., Yang, Y., Wang, L., Zhang, J., Han, Y. and Liu, X. (2020) 'Analysis of expression and prognosis of KLK7 in ovarian cancer', *Open Medicine*, 15(1), pp. 932-939.
- Chen, H., Mohan, P., Jiang, J., Nemirovsky, O., He, D., Fleisch, M. C., Niederacher, D., Pilarski, L. M., Lim, C. J. and Maxwell, C. A. (2014) 'Spatial regulation of Aurora A activity during mitotic spindle assembly requires RHAMM to correctly localize TPX2', *Cell Cycle*, 13(14), pp. 2248-2261.
- Chen, H., Zhou, L., Wu, X., Li, R., Wen, J., Sha, J. and Wen, X. (2016) 'The PI3K/AKT pathway in the pathogenesis of prostate cancer', *Front Biosci (Landmark Ed)*, 21, pp. 1084-91.
- Chen, J. (2016) 'The Cell-Cycle Arrest and Apoptotic Functions of p53 in Tumor Initiation and Progression', *Cold Spring Harbor perspectives in medicine*, 6(3), pp. a026104-a026104.
- Chen, S., Dai, X., Gao, Y., Shen, F., Ding, J. and Chen, Q. (2017) 'The positivity of estrogen receptor and progesterone receptor may not be associated with metastasis and recurrence in epithelial ovarian cancer', *Scientific Reports*, 7(1), pp. 16922.
- Chen, S. and Parmigiani, G. (2007) 'Meta-analysis of BRCA1 and BRCA2 penetrance', *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 25(11), pp. 1329-1333.
- Chen, S., Xu, Y., Yuan, X., Bublely, G. J. and Balk, S. P. (2006) 'Androgen receptor phosphorylation and stabilization in prostate cancer by cyclin-dependent kinase 1', *Proc Natl Acad Sci U S A*, 103(43), pp. 15969-74.
- Chen, W.-x., Yang, L.-g., Xu, L.-y., Cheng, L., Qian, Q., Sun, L. and Zhu, Y.-l. (2019) 'Bioinformatics analysis revealing prognostic significance of RRM2 gene in breast cancer', *Bioscience Reports*, 39(4), pp. BSR20182062.
- Chen, Y.-C., Page, J. H., Chen, R. and Giovannucci, E. (2008) 'Family history of prostate and breast cancer and the risk of prostate cancer in the PSA era', *The Prostate*, 68(14), pp. 1582-1591.
- Chen, Z., Song, H., Zeng, X., Quan, M. and Gao, Y. (2021) 'Screening and discrimination of optimal prognostic genes for pancreatic cancer based on a prognostic prediction model', *G3 Genes/Genomes/Genetics*, 11(11), pp. jkab296.
- Chene, G., Ouellet, V., Rahimi, K., Barres, V., Meunier, L., De Ladurantaye, M., Provencher, D. and Mes-Masson, A. M. (2015) 'Expression of Stem Cell Markers in Preinvasive Tubal Lesions of Ovarian Carcinoma', *BioMed research international*, 2015, pp. 808531-808531.
- Chong, T., Sarac, A., Yao, C. Q., Liao, L., Lyttle, N., Boutros, P. C., Bartlett, J. M. S. and Spears, M. (2018) 'Deregulation of the spindle assembly checkpoint is associated with paclitaxel resistance in ovarian cancer', *Journal of ovarian research*, 11(1), pp. 27-27.

- Choo, J. R. and Nielsen, T. O. (2010) 'Biomarkers for Basal-like Breast Cancer', *Cancers*, 2(2), pp. 1040-1065.
- Cleary, M. P. and Grossmann, M. E. (2009) 'Minireview: Obesity and breast cancer: the estrogen connection', *Endocrinology*, 150(6), pp. 2537-2542.
- Cleveland, D. W., Mao, Y. and Sullivan, K. F. (2003) 'Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling', *Cell*, 112(4), pp. 407-21.
- Conneely, O. M., Mulac-Jericevic, B. and Arnett-Mansfield, R. (2007) 'Progesterone signaling in mammary gland development', *Ernst Schering Found Symp Proc*, (1), pp. 45-54.
- Cooke, P. S., Nanjappa, M. K., Ko, C., Prins, G. S. and Hess, R. A. (2017) 'Estrogens in Male Physiology', *Physiological reviews*, 97(3), pp. 995-1043.
- Cooper, J. A. and Qian, H. (2008) 'A mechanism for SRC kinase-dependent signaling by noncatalytic receptors', *Biochemistry*, 47(21), pp. 5681-5688.
- Correa, A. F., Ruth, K. J., Al-Saleem, T., Pei, J., Dulaimi, E., Kister, D., Collins, M., Abbosh, P. H., Slifker, M. J., Ross, E., Uzzo, R. G. and Testa, J. R. (2020) 'Overall tumor genomic instability: an important predictor of recurrence-free survival in patients with localized clear cell renal cell carcinoma', *Cancer biology & therapy*, 21(5), pp. 424-431.
- Cotrim, D. P., Ribeiro, A. R. G., Paixão, D., de Queiroz Soares, D. C., Jbili, R., Pandolfi, N. C., Cezana, C., de Cássia Mauro, C., Mantoan, H., Bovolim, G., de Brot, L., Torrezan, G. T., Carraro, D. M., Baiocchi, G., da Cruz Formiga, M. N. and da Costa, A. A. B. A. (2019) 'Prevalence of BRCA1 and BRCA2 pathogenic and likely pathogenic variants in non-selected ovarian carcinoma patients in Brazil', *BMC cancer*, 19(1), pp. 4-4.
- Creekmore, A. L., Silkworth, W. T., Cimini, D., Jensen, R. V., Roberts, P. C. and Schmelz, E. M. (2011) 'Changes in gene expression and cellular architecture in an ovarian cancer progression model', *PloS one*, 6(3), pp. e17676-e17676.
- Creighton, C. J. (2007) 'A gene transcription signature associated with hormone independence in a subset of both breast and prostate cancers', *BMC Genomics*, 8(1), pp. 199.
- Cuadros, M., Lopez, F., Blanco, A. and Concha, A. (2011) 'Identifying breast cancer biomarkers' (Accessed).
- Cuzick, J., Swanson, G. P., Fisher, G., Brothman, A. R., Berney, D. M., Reid, J. E., Mesher, D., Speights, V. O., Stankiewicz, E., Foster, C. S., Møller, H., Scardino, P., Warren, J. D., Park, J., Younus, A., Flake, D. D., 2nd, Wagner, S., Gutin, A., Lanchbury, J. S. and Stone, S. (2011) 'Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study', *Lancet Oncol*, 12(3), pp. 245-55.
- D'Assoro, A. B., Lingle, W. L. and Salisbury, J. L. (2002) 'Centrosome amplification and the development of cancer', *Oncogene*, 21(40), pp. 6146-6153.
- Dai, L., Song, K. and Di, W. (2020) 'Adipocytes: active facilitators in epithelial ovarian cancer progression?', *Journal of Ovarian Research*, 13(1), pp. 115.
- Davey, R. A. and Grossmann, M. (2016) 'Androgen Receptor Structure, Function and Biology: From Bench to Bedside', *The Clinical biochemist. Reviews*, 37(1), pp. 3-15.
- Davis, S. and Meltzer, P. S. (2007) 'GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor', *Bioinformatics*, 23(14), pp. 1846-1847.

- Dawany, N. B., Dampier, W. N. and Tozeren, A. (2011) 'Large-scale integration of microarray data reveals genes and pathways common to multiple cancer types', *International Journal of Cancer*, 128(12), pp. 2881-2891.
- Dawany, N. B. and Tozeren, A. (2010) 'Asymmetric microarray data produces gene lists highly predictive of research literature on multiple cancer types', *BMC bioinformatics*, 11, pp. 483-483.
- Dedeurwaerder, S., Desmedt, C., Calonne, E., Singhal, S. K., Haibe-Kains, B., Defrance, M., Michiels, S., Volkmar, M., Deplus, R., Luciani, J., Lallemand, F., Larsimont, D., Toussaint, J., Haussy, S., Rothe, F., Rouas, G., Metzger, O., Majjaj, S., Saini, K., Putmans, P., Hames, G., van Baren, N., Coulie, P. G., Piccart, M., Sotiriou, C. and Fuks, F. (2011) 'DNA methylation profiling reveals a predominant immune component in breast cancers', *EMBO Mol Med*, 3(12), pp. 726-41.
- Delpeuch, A., Leveque, D., Rob, L. and Bergerat, J. P. (2011) 'Off-label use of oxaliplatin in patients with metastatic breast cancer', *Anticancer Res*, 31(5), pp. 1765-7.
- Denkert, C., Budczies, J., Darb-Esfahani, S., Gyorffy, B., Sehouli, J., Konsgen, D., Zeillinger, R., Weichert, W., Noske, A., Buckendahl, A. C., Muller, B. M., Dietel, M. and Lage, H. (2009) 'A prognostic gene expression index in ovarian cancer - validation across different independent data sets', *J Pathol*, 218(2), pp. 273-80.
- Desai, A., Xu, J., Aysola, K., Qin, Y., Okoli, C., Hariprasad, R., Chinemerem, U., Gates, C., Reddy, A., Danner, O., Franklin, G., Ngozi, A., Cantuaria, G., Singh, K., Grizzle, W., Landen, C., Partridge, E. E., Rice, V. M., Reddy, E. S. P. and Rao, V. N. (2014) 'Epithelial ovarian cancer: An overview', *World journal of translational medicine*, 3(1), pp. 1-8.
- Desrichard, A., Bidet, Y., Uhrhammer, N. and Bignon, Y. J. (2011) 'CHEK2 contribution to hereditary breast cancer in non-BRCA families', *Breast Cancer Res*, 13(6), pp. R119.
- Devarajan, E., Sahin, A. A., Chen, J. S., Krishnamurthy, R. R., Aggarwal, N., Brun, A. M., Sapino, A., Zhang, F., Sharma, D., Yang, X. H., Tora, A. D. and Mehta, K. (2002) 'Down-regulation of caspase 3 in breast cancer: a possible mechanism for chemoresistance', *Oncogene*, 21(57), pp. 8843-51.
- Dezső, Z., Oestreicher, J., Weaver, A., Santiago, S., Agoulnik, S., Chow, J., Oda, Y. and Funahashi, Y. (2014) 'Gene Expression Profiling Reveals Epithelial Mesenchymal Transition (EMT) Genes Can Selectively Differentiate Eribulin Sensitive Breast Cancer Cells', *PLOS ONE*, 9(8), pp. e106131.
- Dhillon, A. S., Hagan, S., Rath, O. and Kolch, W. (2007) 'MAP kinase signalling pathways in cancer', *Oncogene*, 26(22), pp. 3279-90.
- Di, Y., Chen, D., Yu, W. and Yan, L. (2019) 'Bladder cancer stage-associated hub genes revealed by WGCNA co-expression network analysis', *Hereditas*, 156(1), pp. 7.
- Di Zazzo, E., Galasso, G., Giovannelli, P., Di Donato, M. and Castoria, G. (2018) 'Estrogens and Their Receptors in Prostate Cancer: Therapeutic Implications', *Frontiers in oncology*, 8, pp. 2-2.
- Dolezal, S., Hester, S., Kirby, P. S., Nairn, A., Pierce, M. and Abbott, K. L. (2014) 'Elevated levels of glycosylphosphatidylinositol (GPI) anchored proteins in plasma from human cancers detected by C. septicum alpha toxin', *Cancer biomarkers : section A of Disease markers*, 14(1), pp. 55-62.
- Dolle, J. M., Daling, J. R., White, E., Brinton, L. A., Doody, D. R., Porter, P. L. and Malone, K. E. (2009) 'Risk factors for triple-negative breast cancer in women under the age of 45 years', *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 18(4), pp. 1157-1166.

- Dong, M., How, T., Kirkbride, K. C., Gordon, K. J., Lee, J. D., Hempel, N., Kelly, P., Moeller, B. J., Marks, J. R. and Globe, G. C. (2007) 'The type III TGF-beta receptor suppresses breast cancer progression', *J Clin Invest*, 117(1), pp. 206-17.
- Dong, X., Wang, L., Taniguchi, K., Wang, X., Cunningham, J. M., McDonnell, S. K., Qian, C., Marks, A. F., Slager, S. L., Peterson, B. J., Smith, D. I., Chevillat, J. C., Blute, M. L., Jacobsen, S. J., Schaid, D. J., Tindall, D. J., Thibodeau, S. N. and Liu, W. (2003) 'Mutations in CHEK2 associated with prostate cancer risk', *Am J Hum Genet*, 72(2), pp. 270-80.
- Dossus, L. and Benusiglio, P. R. (2015) 'Lobular breast cancer: incidence and genetic and non-genetic risk factors', *Breast Cancer Res*, 17, pp. 37.
- du Bois, A., Reuss, A., Pujade-Lauraine, E., Harter, P., Ray-Coquard, I. and Pfisterer, J. (2009) 'Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: A combined exploratory analysis of 3 prospectively randomized phase 3 multicenter trials', *Cancer*, 115(6), pp. 1234-1244.
- Duggan, M. A., Anderson, W. F., Altekruse, S., Penberthy, L. and Sherman, M. E. (2016) 'The Surveillance, Epidemiology, and End Results (SEER) Program and Pathology: Toward Strengthening the Critical Relationship', *The American journal of surgical pathology*, 40(12), pp. e94-e102.
- Dumay, A., Feugeas, J. P., Wittmer, E., Lehmann-Che, J., Bertheau, P., Espié, M., Plassa, L. F., Cottu, P., Marty, M., André, F., Sotiriou, C., Pusztai, L. and De Thé, H. (2013) 'Distinct tumor protein p53 mutants in breast cancer subgroups', *International Journal of Cancer*, 132(5), pp. 1227-1231.
- Duraker, N., Hot, S., Akan, A. and Nayır, P. Ö. (2020) 'A Comparison of the Clinicopathological Features, Metastasis Sites and Survival Outcomes of Invasive Lobular, Invasive Ductal and Mixed Invasive Ductal and Lobular Breast Carcinoma', *European journal of breast health*, 16(1), pp. 22-31.
- Duronio, R. J. and Xiong, Y. (2013) 'Signaling pathways that control cell proliferation', *Cold Spring Harbor perspectives in biology*, 5(3), pp. a008904-a008904.
- D'Alterio, C., Scala, S., Sozzi, G., Roz, L. and Bertolini, G. (2020) 'Paradoxical effects of chemotherapy on tumor relapse and metastasis promotion', *Seminars in Cancer Biology*, 60, pp. 351-361.
- Ecke, T. H., Schlechte, H. H., Schiemenz, K., Sachs, M. D., Lenk, S. V., Rudolph, B. D. and Loening, S. A. (2010) 'TP53 gene mutations in prostate cancer progression', *Anticancer Res*, 30(5), pp. 1579-86.
- Eder, A. M., Sui, X., Rosen, D. G., Nolden, L. K., Cheng, K. W., Lahad, J. P., Kango-Singh, M., Lu, K. H., Warneke, C. L., Atkinson, E. N., Bedrosian, I., Keyomarsi, K., Kuo, W.-I., Gray, J. W., Yin, J. C. P., Liu, J., Halder, G. and Mills, G. B. (2005) 'Atypical PKC α contributes to poor prognosis through loss of apical-basal polarity and cyclin E overexpression in ovarian cancer', *Proceedings of the National Academy of Sciences of the United States of America*, 102(35), pp. 12519-12524.
- Edgar, R., Domrachev, M. and Lash, A. E. (2002) 'Gene Expression Omnibus: NCBI gene expression and hybridization array data repository', *Nucleic acids research*, 30(1), pp. 207.
- Ehdaivand, S. (2012) *WHO classification of ovarian neoplasms*. website: PathologyOutlines.com. Available at: <https://www.pathologyoutlines.com/topic/ovarytumorwhoclassif.html> (Accessed: 2019).

- Elgaaen, B. V., Olstad, O. K., Sandvik, L., Odegaard, E., Sauer, T., Staff, A. C. and Gautvik, K. M. (2012) 'ZNF385B and VEGFA are strongly differentially expressed in serous ovarian carcinomas and correlate with survival', *PLoS One*, 7(9), pp. e46317.
- Eliyatkın, N., Yalçın, E., Zengel, B., Aktaş, S. and Vardar, E. (2015) 'Molecular Classification of Breast Carcinoma: From Traditional, Old-Fashioned Way to A New Age, and A New Way', *The journal of breast health*, 11(2), pp. 59-66.
- Elizabeth, A. M. and Robert, L. S. (2009) 'Biological determinants of endocrine resistance in breast cancer', *Nature Reviews Cancer*, 9(9), pp. 631.
- Elledge, R. M., Lock-Lim, S., Allred, D. C., Hilsenbeck, S. G. and Corder, L. (1995) 'p53 mutation and tamoxifen resistance in breast cancer', *Clin Cancer Res*, 1(10), pp. 1203-8.
- Elledge, R. M. and Osborne, C. K. (1997) 'Oestrogen receptors and breast cancer', *BMJ*, 314(7098), pp. 1843.
- Englund, E., Canesin, G., Papadakos, K. S., Vishnu, N., Persson, E., Reitsma, B., Anand, A., Jacobsson, L., Helczynski, L., Mulder, H., Bjartell, A. and Blom, A. M. (2017) 'Cartilage oligomeric matrix protein promotes prostate cancer progression by enhancing invasion and disrupting intracellular calcium homeostasis', *Oncotarget*, 8(58), pp. 98298-98311.
- Epstein, J. I., Zelefsky, M. J., Sjoberg, D. D., Nelson, J. B., Egevad, L., Magi-Galluzzi, C., Vickers, A. J., Parwani, A. V., Reuter, V. E., Fine, S. W., Eastham, J. A., Wiklund, P., Han, M., Reddy, C. A., Ciezki, J. P., Nyberg, T. and Klein, E. A. (2016) 'A Contemporary Prostate Cancer Grading System: A Validated Alternative to the Gleason Score', *European urology*, 69(3), pp. 428-435.
- Erickson, B. K., Conner, M. G. and Landen, C. N., Jr. (2013) 'The role of the fallopian tube in the origin of ovarian cancer', *American journal of obstetrics and gynecology*, 209(5), pp. 409-414.
- Ewertz, M., Jensen, M. B., Gunnarsdottir, K. A., Hojris, I., Jakobsen, E. H., Nielsen, D., Stenbygaard, L. E., Tange, U. B. and Cold, S. (2011) 'Effect of obesity on prognosis after early-stage breast cancer', *J Clin Oncol*, 29(1), pp. 25-31.
- Ezzati, M., Abdullah, A., Shariftabrizi, A., Hou, J., Kopf, M., Stedman, J. K., Samuelson, R. and Shahabi, S. (2014) 'Recent Advancements in Prognostic Factors of Epithelial Ovarian Carcinoma', *International Scholarly Research Notices*, 2014, pp. 953509.
- Fallahpour, S., Navaneelan, T., De, P. and Borgo, A. (2017) 'Breast cancer survival by molecular subtype: a population-based analysis of cancer registry data', *CMAJ Open*, 5(3), pp. E734-e739.
- Feng, Y., Spezia, M., Huang, S., Yuan, C., Zeng, Z., Zhang, L., Ji, X., Liu, W., Huang, B., Luo, W., Liu, B., Lei, Y., Du, S., Vuppapapati, A., Luu, H. H., Haydon, R. C., He, T.-C. and Ren, G. (2018) 'Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis', *Genes & diseases*, 5(2), pp. 77-106.
- Ferlay, J., Ervik, M., Lam, F., Colombet, M., Mery, L., Piñeros, M., Znaor, A., Soerjomataram, I. and Bray, F. (2020) *Global Cancer Observatory: Cancer Today*. <https://gco.iarc.fr/today>. France: International Agency for Research on Cancer (Accessed: 5/5 2021).
- Fernandez-Cuesta, L., Anaganti, S., Hainaut, P. and Olivier, M. (2011) 'p53 status influences response to tamoxifen but not to fulvestrant in breast cancer cell lines', *International Journal of Cancer*, 128(8), pp. 1813-1821.
- Ferreira, M., Francisco, S., Soares, A. R., Nobre, A., Pinheiro, M., Reis, A., Neto, S., Rodrigues, A. J., Sousa, N., Moura, G. and Santos, M. A. S. (2021) 'Integration of segmented regression analysis

with weighted gene correlation network analysis identifies genes whose expression is remodeled throughout physiological aging in mouse tissues', *Aging (Albany NY)*, 13(14), pp. 18150-18190.

Foley, J. P., Lam, D., Jiang, H., Liao, J., Cheong, N., McDevitt, T. M., Zaman, A., Wright, J. R. and Savani, R. C. (2012) 'Toll-like receptor 2 (TLR2), transforming growth factor- β , hyaluronan (HA), and receptor for HA-mediated motility (RHAMM) are required for surfactant protein A-stimulated macrophage chemotaxis', *The Journal of biological chemistry*, 287(44), pp. 37406-37419.

Freedland, S. J. and Aronson, W. J. (2004) 'Examining the relationship between obesity and prostate cancer', *Reviews in urology*, 6(2), pp. 73-81.

Fu, J. and Zhang, C. 2010. Coordination of Cell Cycle Events by Ran GTPase.

Fu, Y., Zhou, Q.-Z., Zhang, X.-L., Wang, Z.-Z. and Wang, P. (2019) 'Identification of Hub Genes Using Co-Expression Network Analysis in Breast Cancer as a Tool to Predict Different Stages', *Medical science monitor : international medical journal of experimental and clinical research*, 25, pp. 8873-8890.

Fujita, K. and Nonomura, N. (2019) 'Role of Androgen Receptor in Prostate Cancer: A Review', *The world journal of men's health*, 37(3), pp. 288-295.

Fulda, S. and Debatin, K. M. (2006) 'Extrinsic versus intrinsic apoptosis pathways in anticancer chemotherapy', *Oncogene*, 25(34), pp. 4798-811.

Fuqua, S. A. W., Gu, G. and Rechoum, Y. (2014) 'Estrogen receptor (ER) α mutations in breast cancer: hidden in plain sight', *Breast cancer research and treatment*, 144(1), pp. 11-19.

Fuster, J. J., Ouchi, N., Gokce, N. and Walsh, K. (2016) 'Obesity-Induced Changes in Adipose Tissue Microenvironment and Their Impact on Cardiovascular Disease', *Circulation research*, 118(11), pp. 1786-1807.

Gail, P. R., Ian, D. D., Stephen, N. B. and Wayne, D. T. (2010) 'Breast and prostate cancer: more similar than different', *Nature Reviews Cancer*, 10(3), pp. 205.

Gaillard, S., Gay, L. M., Steiner, M., Andreano, K., Davidson, B. A., Havrilesky, L. J., Secord, A. A., Valea, F. A., Colon-Otero, G., Zajchowski, D. A., Chang, C.-Y., McDonnell, D. P., Berchuck, A. and Elvin, J. A. (2018) 'Assessment of activating estrogen receptor 1 (ESR1) mutations in gynecologic malignancies', *Journal of Clinical Oncology*, 36(15_suppl), pp. 5590-5590.

Gallardo, M. E., Moreno-Loshuertos, R., Lopez, C., Casqueiro, M., Silva, J., Bonilla, F., Rodriguez de Cordoba, S. and Enriquez, J. A. (2006) 'm.6267G>A: a recurrent mutation in the human mitochondrial DNA that reduces cytochrome c oxidase activity and is associated with tumors', *Hum Mutat*, 27(6), pp. 575-82.

Gallardo-Alvarado, L. N., Tusié-Luna, M. T., Tussíé-Luna, M. I., Díaz-Chávez, J., Segura, Y. X., Bargallo-Rocha, E., Villarreal, C., Herrera-Montalvo, L. A., Herrera-Medina, E. M. and Cantu-de Leon, D. F. (2019) 'Prevalence of germline mutations in the TP53 gene in patients with early-onset breast cancer in the Mexican population', *BMC Cancer*, 19(1), pp. 118.

Galvin, K. E., Ye, H., Erstad, D. J., Feddersen, R. and Wetmore, C. (2008) 'Gli1 induces G2/M arrest and apoptosis in hippocampal but not tumor-derived neural stem cells', *Stem Cells*, 26(4), pp. 1027-36.

Garcez, P. P., Diaz-Alonso, J., Crespo-Enriquez, I., Castro, D., Bell, D. and Guillemot, F. (2015) 'Cenpj/CPAP regulates progenitor divisions and neuronal migration in the cerebral cortex downstream of Ascl1', *Nature communications*, 6, pp. 6474-6474.

- Garziera, M., Cecchin, E., Canzonieri, V., Sorio, R., Giorda, G., Scalone, S., De Mattia, E., Roncato, R., Gagno, S., Poletto, E., Romanato, L., Sartor, F., Polesel, J. and Toffoli, G. (2018) 'Identification of Novel Somatic TP53 Mutations in Patients with High-Grade Serous Ovarian Cancer (HGSOC) Using Next-Generation Sequencing (NGS)', *International journal of molecular sciences*, 19(5), pp. 1510.
- Gatti-Mays, M. E., Balko, J. M., Gameiro, S. R., Bear, H. D., Prabhakaran, S., Fukui, J., Disis, M. L., Nanda, R., Gulley, J. L., Kalinsky, K., Abdul Sater, H., Sparano, J. A., Cescon, D., Page, D. B., McArthur, H., Adams, S. and Mittendorf, E. A. (2019) 'If we build it they will come: targeting the immune response to breast cancer', *npj Breast Cancer*, 5(1), pp. 37.
- Gautier, L., Cope, L., Bolstad, B. M. and Irizarry, R. A. (2004) 'affyanalysis of Affymetrix GeneChip data at the probe level', *Bioinformatics*, 20(3), pp. 307-315.
- Gemoll, T., Auer, G., Ried, T. and Habermann, J. K. (2015) 'Genetic Instability and Disease Prognostication', *Recent results in cancer research. Fortschritte der Krebsforschung. Progres dans les recherches sur le cancer*, 200, pp. 81-94.
- George, S. H., Greenaway, J., Milea, A., Clary, V., Shaw, S., Sharma, M., Virtanen, C. and Shaw, P. A. (2011) 'Identification of abrogated pathways in fallopian tube epithelium from BRCA1 mutation carriers', *J Pathol*, 225(1), pp. 106-17.
- George, S. H. L., Garcia, R. and Slomovitz, B. M. (2016) 'Ovarian Cancer: The Fallopian Tube as the Site of Origin and Opportunities for Prevention', *Frontiers in oncology*, 6, pp. 108-108.
- Giovannelli, P., Di Donato, M., Galasso, G., Di Zazzo, E., Bilancio, A. and Migliaccio, A. (2018) 'The Androgen Receptor in Breast Cancer', *Frontiers in endocrinology*, 9, pp. 492-492.
- Giri, V. N. and Beebe-Dimmer, J. L. (2016) 'Familial prostate cancer', *Seminars in oncology*, 43(5), pp. 560-565.
- Giussani, M., Merlino, G., Cappelletti, V., Tagliabue, E. and Daidone, M. G. (2015) 'Tumor-extracellular matrix interactions: Identification of tools associated with breast cancer progression', *Semin Cancer Biol*, 35, pp. 3-10.
- Gokulnath, P., de Cristofaro, T., Manipur, I., Di Palma, T., Soriano, A. A., Guarracino, M. R. and Zannini, M. (2020) 'Long Non-Coding RNA HAND2-AS1 Acts as a Tumor Suppressor in High-Grade Serous Ovarian Carcinoma', *Int J Mol Sci*, 21(11).
- Gong, C., Fujino, K., Monteiro, L. J., Gomes, A. R., Drost, R., Davidson-Smith, H., Takeda, S., Khoo, U. S., Jonkers, J., Sproul, D. and Lam, E. W. F. (2015) 'FOXA1 repression is associated with loss of BRCA1 and increased promoter methylation and chromatin silencing in breast cancer', *Oncogene*, 34(39), pp. 5012-5024.
- Goodman, M. T. and Shvetsov, Y. B. (2009) 'Incidence of ovarian, peritoneal, and fallopian tube carcinomas in the United States, 1995-2004', *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 18(1), pp. 132-139.
- Gordetsky, J. and Epstein, J. (2016) 'Grading of prostatic adenocarcinoma: current state and prognostic implications', *Diagnostic pathology*, 11, pp. 25-25.
- Gottardo, M. F., Capobianco, C. S., Sidabra, J. E., Garona, J., Perera, Y., Perea, S. E., Alonso, D. F. and Farina, H. G. (2020) 'Preclinical efficacy of CIGB-300, an anti-CK2 peptide, on breast cancer metastatic colonization', *Scientific Reports*, 10(1), pp. 14689.
- Goymer, P. (2008) 'Why do we need hubs?', *Nature Reviews Genetics*, 9(9), pp. 651-651.

- Gradishar, W. J. (2012) 'Taxanes for the treatment of metastatic breast cancer', *Breast cancer : basic and clinical research*, 6, pp. 159-171.
- Green, W. J. F., Ball, G., Hulman, G., Johnson, C., Van Schalwyk, G., Ratan, H. L., Soria, D., Garibaldi, J. M., Parkinson, R., Hulman, J., Rees, R. and Powe, D. G. (2016) 'KI67 and DLX2 predict increased risk of metastasis formation in prostate cancer—a targeted molecular approach', *British Journal of Cancer*, 115(2), pp. 236-242.
- Gregory, C. W., Hamil, K. G., Kim, D., Hall, S. H., Pretlow, T. G., Mohler, J. L. and French, F. S. (1998) 'Androgen receptor expression in androgen-independent prostate cancer is associated with increased expression of androgen-regulated genes', *Cancer Res*, 58(24), pp. 5718-24.
- Grimes, T., Potter, S. S. and Datta, S. (2019) 'Integrating gene regulatory pathways into differential network analysis of gene expression data', *Scientific Reports*, 9(1), pp. 5479.
- Grimm, S. L., Hartig, S. M. and Edwards, D. P. (2016) 'Progesterone Receptor Signaling Mechanisms', *Journal of Molecular Biology*, 428(19), pp. 3831-3849.
- Grindstad, T., Richardsen, E., Andersen, S., Skjefstad, K., Rakaee Khanehkenari, M., Donnem, T., Ness, N., Nordby, Y., Bremnes, R. M., Al-Saad, S. and Busund, L.-T. (2018) 'Progesterone Receptors in Prostate Cancer: Progesterone receptor B is the isoform associated with disease progression', *Scientific reports*, 8(1), pp. 11358-11358.
- Grist, E., Parry, M., Mendes, L., Lall, S., Kudahetti, S. C., Santos Vidal, S., Atako, N. B., Anjum, M., Ishaq, S., Todorovic, T., Huang, H.-C., Waugh, D., Sweeney, C., Clarke, N., James, N. D., Berney, D., Parmar, M. K. B., Davicioni, E., Brown, L. C. and Attard, G. (2020) 'Multiregion expression profiling of prostate cancer from men randomized in the STAMPEDE trial: Stage I results of a multistage biomarker analysis', *Journal of Clinical Oncology*, 38(6_suppl), pp. 153-153.
- Gronwald, J., Jauch, A., Cybulski, C., Schoell, B., Bohm-Steuer, B., Lener, M., Grabowska, E., Gorski, B., Jakubowska, A., Domagala, W., Chosia, M., Scott, R. J. and Lubinski, J. (2005) 'Comparison of genomic abnormalities between BRCA1 and sporadic breast cancers studied by comparative genomic hybridization', *Int J Cancer*, 114(2), pp. 230-6.
- Gruss, O. J., Carazo-Salas, R. E., Schatz, C. A., Guarguaglini, G., Kast, J., Wilm, M., Le Bot, N., Vernos, I., Karsenti, E. and Mattaj, I. W. (2001) 'Ran Induces Spindle Assembly by Reversing the Inhibitory Effect of Importin α on TPX2 Activity', *Cell*, 104(1), pp. 83-93.
- Gu, S., Chen, L., Hong, Q., Yan, T., Zhuang, Z., Wang, Q., Jin, W., Zhu, H. and Wu, J. (2011) 'PEA3 activates CXCR4 transcription in MDA-MB-231 and MCF7 breast cancer cells', *Acta Biochimica et Biophysica Sinica*, 43(10), pp. 771-778.
- Guan, H., Guo, Y., Liu, L., Ye, R., Liang, W., Li, H., Xiao, H. and Li, Y. (2018) 'INAVA promotes aggressiveness of papillary thyroid cancer by upregulating MMP9 expression', *Cell Biosci*, 8, pp. 26.
- Guille, A., Chaffanet, M. and Birnbaum, D. (2013) 'Signaling pathway switch in breast cancer', *Cancer Cell Int*, 13(1), pp. 66.
- Guo, Y. J., Pan, W. W., Liu, S. B., Shen, Z. F., Xu, Y. and Hu, L. L. (2020) 'ERK/MAPK signalling pathway and tumorigenesis (Review)', *Exp Ther Med*, 19(3), pp. 1997-2007.
- Göbel, C., Özden, C., Schroeder, C., Hube-Magg, C., Kluth, M., Möller-Koop, C., Neubauer, E., Hirsch, A., Jacobsen, F., Simon, R., Sauter, G., Michl, U., Pehrke, D., Hülndt, H., Graefen, M., Schlomm, T. and Luebke, A. M. (2018) 'Upregulation of centromere protein F is linked to aggressive prostate cancers', *Cancer management and research*, 10, pp. 5491-5504.

Haldar, S., Negrini, M., Monne, M., Sabbioni, S. and Croce, C. M. (1994) 'Down-regulation of bcl-2 by p53 in breast cancer cells', *Cancer Res*, 54(8), pp. 2095-7.

Hale, V., Weischer, M. and Park, J. Y. (2014) 'CHEK2 (*) 1100delC Mutation and Risk of Prostate Cancer', *Prostate cancer*, 2014, pp. 294575-294575.

Hall, P., Ploner, A., Bjohle, J., Huang, F., Lin, C. Y., Liu, E. T., Miller, L. D., Nordgren, H., Pawitan, Y., Shaw, P., Skoog, L., Smeds, J., Wedren, S., Ohd, J. and Bergh, J. (2006) 'Hormone-replacement therapy influences gene expression profiles and is associated with breast-cancer prognosis: a cohort study', *BMC Med*, 4, pp. 16.

Hamilton, S. R., Fard, S. F., Paiwand, F. F., Tolg, C., Veisheh, M., Wang, C., McCarthy, J. B., Bissell, M. J., Koropatnick, J. and Turley, E. A. (2007) 'The Hyaluronan Receptors CD44 and Rhamm (CD168) Form Complexes with ERK1,2 That Sustain High Basal Motility in Breast Cancer Cells', *Journal of Biological Chemistry*, 282(22), pp. 16667-16680.

Hammarsten, P., Josefsson, A., Thysell, E., Lundholm, M., Hägglöf, C., Iglesias-Gato, D., Flores-Morales, A., Stattin, P., Egevad, L., Granfors, T., Wikström, P. and Bergh, A. (2019) 'Immunoreactivity for prostate specific antigen and Ki67 differentiates subgroups of prostate cancer related to outcome', *Modern Pathology*, 32(9), pp. 1310-1319.

Hanahan, D. and Weinberg, R. A. (2011) 'Hallmarks of cancer: the next generation', *Cell*, 144(5), pp. 646-74.

Hannele, E., Bing, X., Jenni, N., Johanna, S., Kirsi, S., Arto, M., Anne, K., Katri, P., Sanna-Maria, K., Katrin, R., Alexander, M., Qing, S., Guilan, L., Henna, M., Daphne, W. B., Daniel, A. H., Mervi, G., Mervi, R., Arja, J.-V., Aki, M., Juha, K., Lauri, A. A., Veli-Matti, K., Vesa, K., Ylermi, S., Ronny, I. D., David, M. L. and Robert, W. (2007) 'A recurrent mutation in PALB2 in Finnish cancer families', *Nature*, 446(7133), pp. 316.

Harkness, T. A. A. (2018) 'Activating the Anaphase Promoting Complex to Enhance Genomic Stability and Prolong Lifespan', *International journal of molecular sciences*, 19(7), pp. 1888.

Harrison, M. R., Holen, K. D. and Liu, G. (2009) 'Beyond taxanes: a review of novel agents that target mitotic tubulin and microtubules, kinases, and kinesins', *Clinical advances in hematology & oncology : H&O*, 7(1), pp. 54-64.

Harrod, A., Fulton, J., Nguyen, V. T. M., Periyasamy, M., Ramos-Garcia, L., Lai, C. F., Metodieva, G., de Giorgio, A., Williams, R. L., Santos, D. B., Gomez, P. J., Lin, M. L., Metodiev, M. V., Stebbing, J., Castellano, L., Magnani, L., Coombes, R. C., Buluwela, L. and Ali, S. (2017) 'Genomic modelling of the ESR1 Y537S mutation for evaluating function and new therapeutic approaches for metastatic breast cancer', *Oncogene*, 36(16), pp. 2286-2296.

Hashmi, A. A., Mudassir, G., Irfan, M., Hussain, Z. F., Hashmi, S. K., Asif, H., Nisar, L., Naeem, M. and Faridi, N. (2019) 'Prognostic Significance of High Androgen Receptor Expression in Prostatic Acinar Adenocarcinoma', *Asian Pacific journal of cancer prevention : APJCP*, 20(3), pp. 893-896.

Hastie, T., Tibshirani, R., Narasimhan, B. and Chu, G. 2018. *impute: Imputation for microarray data* . <https://www.bioconductor.org/packages/release/bioc/html/impute.html>: Bioconductor.

Hellems, P., van Dam, P. A., Weyler, J., van Oosterom, A. T., Buytaert, P. and Van Marck, E. (1995) 'Prognostic value of bcl-2 expression in invasive breast cancer', *Br J Cancer*, 72(2), pp. 354-60.

- Hempel, N., How, T., Dong, M., Murphy, S. K., Fields, T. A. and Blobe, G. C. (2007) 'Loss of Betaglycan Expression in Ovarian Cancer: Role in Motility and Invasion', *Cancer Research*, 67(11), pp. 5231.
- Hendrix, N. D., Wu, R., Kuick, R., Schwartz, D. R., Fearon, E. R. and Cho, K. R. (2006) 'Fibroblast growth factor 9 has oncogenic activity and is a downstream target of Wnt signaling in ovarian endometrioid adenocarcinomas', *Cancer Res*, 66(3), pp. 1354-62.
- Hennigs, A., Riedel, F., Gondos, A., Sinn, P., Schirmacher, P., Marmé, F., Jäger, D., Kauczor, H.-U., Stieber, A., Lindel, K., Debus, J., Golatta, M., Schütz, F., Sohn, C., Heil, J. and Schneeweiss, A. (2016) 'Prognosis of breast cancer molecular subtypes in routine clinical care: A large prospective cohort study', *BMC cancer*, 16(1), pp. 734-734.
- Hernandez-Cueto, A., Hernandez-Cueto, D., Antonio-Andres, G., Mendoza-Marin, M., Jimenez-Gutierrez, C., Sandoval-Mejia, A. L., Mora-Campos, R., Gonzalez-Bonilla, C., Vega, M. I., Bonavida, B. and Huerta-Yepez, S. (2014) 'Death receptor 5 expression is inversely correlated with prostate cancer progression', *Molecular medicine reports*, 10(5), pp. 2279-2286.
- Hevey, D. (2018) 'Network analysis: a brief overview and tutorial', *Health Psychology and Behavioral Medicine*, 6(1), pp. 301-328.
- Higgins, M., Obaidi, I. and McMorrow, T. (2019) 'Primary cilia and their role in cancer', *Oncology letters*, 17(3), pp. 3041-3047.
- Hinck, L. and Silberstein, G. B. (2005) 'Key stages in mammary gland development: the mammary end bud as a motile organ', *Breast Cancer Res*, 7(6), pp. 245-51.
- Hisham, M., Russell, I. A., Rory, S., Oscar, M. R., Theresa, E. H., Gerard, A. T., Aurelien, A. A. S., Stephen, N. B., Alejandra, B., Amel, S., Suraj, M., James, H., Michelle, P., Ganesh, V. R., Gordon, D. B., Clive, D. s., Jessica, L. L. R., Grace, S., Rosalind, L., Charles, M. P., John, S., Carlos, C., Wayne, D. T. and Jason, S. C. (2015) 'Progesterone receptor modulates ER α action in breast cancer', *Nature*, 523(7560).
- Hochegger, H., Takeda, S. and Hunt, T. (2008) 'Cyclin-dependent kinases and cell-cycle transitions: does one fit all?', *Nature Reviews Molecular Cell Biology*, 9(11), pp. 910-916.
- Hodgson, A. and Turashvili, G. (2020) 'Pathology of Hereditary Breast and Ovarian Cancer', *Frontiers in Oncology*, 10, pp. 2026.
- Holland, A. J., Lan, W. and Cleveland, D. W. (2010) 'Centriole duplication: A lesson in self-control', *Cell cycle (Georgetown, Tex.)*, 9(14), pp. 2731-2736.
- Hollander, M. C. and Fornace, A. J. (2002) 'Genomic instability, centrosome amplification, cell cycle checkpoints and Gadd45a', *Oncogene*, 21(40), pp. 6228-6233.
- Hollis, R. L., Thomson, J. P., Stanley, B., Churchman, M., Meynert, A. M., Rye, T., Bartos, C., Iida, Y., Croy, I., Mackean, M., Nussey, F., Okamoto, A., Semple, C. A., Gourley, C. and Herrington, C. S. (2020) 'Molecular stratification of endometrioid ovarian carcinoma predicts clinical outcome', *Nature Communications*, 11(1), pp. 4995.
- Howe, G. R., Hirohata, T., Hislop, T. G., Iscovich, J. M., Yuan, J. M., Katsouyanni, K., Lubin, F., Marubini, E., Modan, B., Rohan, T. and et al. (1990) 'Dietary factors and risk of breast cancer: combined analysis of 12 case-control studies', *J Natl Cancer Inst*, 82(7), pp. 561-9.

- Howlader, N., Cronin, K. A., Kurian, A. W. and Andridge, R. (2018) 'Differences in Breast Cancer Survival by Molecular Subtypes in the United States', *Cancer Epidemiol Biomarkers Prev*, 27(6), pp. 619-626.
- Howlader, N., Noone, A., Krapcho, M., Miller, D., Brest, A., Yu, M., Ruhl, J., Tatalovich, Z., Mariotto, A., Lewis, D., Chen, H., Feuer, E. and Cronin, K. e. 2020. SEER Cancer Statistics Review, 1975-2017. National Cancer Institute. Bethesda.
- Huang, C.-H., Mandelker, D., Gabelli, S. B. and Amzel, L. M. (2008) 'Insights into the oncogenic effects of PIK3CA mutations from the structure of p110alpha/p85alpha', *Cell cycle (Georgetown, Tex.)*, 7(9), pp. 1151-1156.
- Huang, X., Hatcher, R., York, J. P. and Zhang, P. (2005) 'Securin and separase phosphorylation act redundantly to maintain sister chromatid cohesion in mammalian cells', *Molecular biology of the cell*, 16(10), pp. 4725-4732.
- Huber, M. A., Azoitei, N., Baumann, B., Grünert, S., Sommer, A., Pehamberger, H., Kraut, N., Beug, H. and Wirth, T. (2004) 'NF-kappaB is essential for epithelial-mesenchymal transition and metastasis in a model of breast cancer progression', *The Journal of clinical investigation*, 114(4), pp. 569-581.
- Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oles, A. K., Pages, H., Reyes, A., Shannon, P., Smyth, G. K., Tenenbaum, D., Waldron, L. and Morgan, M. (2015) 'Orchestrating high-throughput genomic analysis with Bioconductor', *Nat Meth*, 12(2), pp. 115-121.
- Huijts, P. E., Hollestelle, A., Balliu, B., Houwing-Duistermaat, J. J., Meijers, C. M., Blom, J. C., Ozturk, B., Krol-Warmerdam, E. M., Wijnen, J., Berns, E. M., Martens, J. W., Seynaeve, C., Kiemeny, L. A., van der Heijden, H. F., Tollenaar, R. A., Devilee, P. and van Asperen, C. J. (2014) 'CHEK2*1100delC homozygosity in the Netherlands--prevalence and risk of breast and lung cancer', *Eur J Hum Genet*, 22(1), pp. 46-51.
- Humphrey, P. A. (2004) 'Gleason grading and prognostic factors in carcinoma of the prostate', *Modern Pathology*, 17(3), pp. 292-306.
- Iczkowski, K. (2019) *Grading (Gleason)*. <http://www.pathologyoutlines.com/topic/prostategrading.html>: PathologyOutlines.com (Accessed: 2019).
- Iczkowski, K. A. (2010) 'Cell adhesion molecule CD44: its functional roles in prostate cancer', *American journal of translational research*, 3(1), pp. 1-7.
- Ihemelandu, C. U., Leffall, L. D., Jr., Dewitty, R. L., Naab, T. J., Mezgebe, H. M., Makambi, K. H., Adams-Campbell, L. and Frederick, W. A. (2007) 'Molecular breast cancer subtypes in premenopausal African-American women, tumor biologic factors and clinical outcome', *Ann Surg Oncol*, 14(10), pp. 2994-3003.
- Intgen 2005. Expression Project for Oncology (expo). *In*: Intgen (ed.).
- Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U. and Speed, T. P. (2003) 'Exploration, normalization, and summaries of high density oligonucleotide array probe level data', *Biostatistics*, 4(2), pp. 249-264.

- Iversen, L., Fielding, S., Lidegaard, Ø., Mørch, L. S., Skovlund, C. W. and Hannaford, P. C. (2018) 'Association between contemporary hormonal contraception and ovarian cancer in women of reproductive age in Denmark: prospective, nationwide cohort study', *BMJ*, 362, pp. k3609.
- Jackson, S. P. and Bartek, J. (2009) 'The DNA-damage response in human biology and disease', *Nature*, 461(7267), pp. 1071-8.
- Jafarian, A. H., Mirshekar Nasirabadi, K., Etemad, S., Jafaripour, M., Darijani, M., Sheikhi, M., Ayatollahi, H., Shakeri, S., Shams, S. F. and Davari, S. (2018) 'Molecular Status of BRAF Mutation in Prostate Adenocarcinoma: The Analysis of 100 Cases in North-East of IRAN', *Iranian journal of pathology*, 13(4), pp. 415-421.
- Jain, S., Lyons, C. A., Walker, S. M., McQuaid, S., Hynes, S. O., Mitchell, D. M., Pang, B., Logan, G. E., McCavigan, A. M., O'Rourke, D., McArt, D. G., McDade, S. S., Mills, I. G., Prise, K. M., Knight, L. A., Steele, C. J., Medlow, P. W., Berge, V., Katz, B., Loblaw, D. A., Harkin, D. P., James, J. A., O'Sullivan, J. M., Kennedy, R. D. and Waugh, D. J. (2018) 'Validation of a Metastatic Assay using biopsies to improve risk stratification in patients with prostate cancer treated with radical radiation therapy', *Ann Oncol*, 29(1), pp. 215-222.
- Jakowlew, S. B. (2006) 'Transforming growth factor-beta in cancer and metastasis', *Cancer Metastasis Rev*, 25(3), pp. 435-57.
- Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E. and Forman, D. (2011) 'Global cancer statistics', *CA Cancer J Clin*, 61(2), pp. 69-90.
- Jeong, J., Kim, W., Kim, L. K., VanHouten, J. and Wysolmerski, J. J. (2017) 'HER2 signaling regulates HER2 localization and membrane retention', *PLOS ONE*, 12(4), pp. e0174849.
- Jeyasekharan, A. D., Liu, Y., Hattori, H., Pisupati, V., Jonsdottir, A. B., Rajendra, E., Lee, M., Sundaramoorthy, E., Schlachter, S., Kaminski, C. F., Ofir-Rosenfeld, Y., Sato, K., Savill, J., Ayoub, N. and Venkitaraman, A. R. (2013) 'A cancer-associated BRCA2 mutation reveals masked nuclear export signals controlling localization', *Nature structural & molecular biology*, 20(10), pp. 1191-1198.
- Jia, D., Nagaoka, Y., Katsumata, M. and Orsulic, S. (2018) 'Inflammation is a key contributor to ovarian cancer cell seeding', *Scientific Reports*, 8(1), pp. 12394.
- Johnson, W. E., Li, C. and Rabinovic, A. (2007) 'Adjusting batch effects in microarray expression data using empirical Bayes methods', *Biostatistics*, 8(1), pp. 118-27.
- Joukov, V., Groen, A. C., Prokhorova, T., Gerson, R., White, E., Rodriguez, A., Walter, J. C. and Livingston, D. M. (2006) 'The BRCA1/BARD1 heterodimer modulates ran-dependent mitotic spindle assembly', *Cell*, 127(3), pp. 539-52.
- Ju, W., Yoo, B. C., Kim, I. J., Kim, J. W., Kim, S. C. and Lee, H. P. (2009) 'Identification of genes with differential expression in chemoresistant epithelial ovarian cancer using high-density oligonucleotide microarrays', *Oncol Res*, 18(2-3), pp. 47-56.
- Jung, Y., Jung, W. and Koo, J. (2016) 'BRAF mutation in breast cancer by BRAF V600E mutation-specific antibody', *International Journal of Clinical Pathology*, 9(2).
- Kaaks, R., Lukanova, A. and Sommersberg, B. (2000) 'Plasma androgens, IGF-1, body size, and prostate cancer risk: a synthetic review', *Prostate Cancer Prostatic Dis*, 3(3), pp. 157-172.

- Kaku, T., Ogawa, S., Kawano, Y., Ohishi, Y., Kobayashi, H., Hirakawa, T. and Nakano, H. (2003) 'Histological classification of ovarian cancer', *Medical Electron Microscopy*, 36(1), pp. 9-17.
- Kalluri, R. and Weinberg, R. A. (2009) 'The basics of epithelial-mesenchymal transition', *J Clin Invest*, 119(6), pp. 1420-8.
- Kao, K. J., Chang, K. M., Hsu, H. C. and Huang, A. T. (2011) 'Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization', *BMC Cancer*, 11, pp. 143.
- Karpf, A. R. and Jones, D. A. (2002) 'Reactivating the expression of methylation silenced genes in human cancer', *Oncogene*, 21(35), pp. 5496-5503.
- Kasai, K., Inaguma, S., Yoneyama, A., Yoshikawa, K. and Ikeda, H. (2008) 'SCL/TAL1 Interrupting Locus Derepresses GLI1 from the Negative Control of Suppressor-of-Fused in Pancreatic Cancer Cell', *Cancer Research*, 68(19), pp. 7723.
- Kassambara, A., Kosinski, M. and Biecek, P. 2019. survminer: Drawing Survival Curves using 'ggplot2'.
- Kaur, H., Salles, D. C., Murali, S., Hicks, J. L., Nguyen, M., Pritchard, C. C., De Marzo, A. M., Lanchbury, J. S., Trock, B. J., Isaacs, W. B., Timms, K. M., Antonarakis, E. S. and Lotan, T. L. (2020) 'Genomic and Clinicopathologic Characterization of ATM-deficient Prostate Cancer', *Clin Cancer Res*, 26(18), pp. 4869-4881.
- Kawai, T. and Akira, S. (2006) 'TLR signaling', *Cell Death & Differentiation*, 13(5), pp. 816-825.
- Kazanets, A., Shorstova, T., Hilmi, K., Marques, M. and Witcher, M. (2016) 'Epigenetic silencing of tumor suppressor genes: Paradigms, puzzles, and potential', *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1865(2), pp. 275-288.
- Kensler, K. H., Regan, M. M., Heng, Y. J., Baker, G. M., Pyle, M. E., Schnitt, S. J., Hazra, A., Kammler, R., Thürlimann, B., Colleoni, M., Viale, G., Brown, M. and Tamimi, R. M. (2019) 'Prognostic and predictive value of androgen receptor expression in postmenopausal women with estrogen receptor-positive breast cancer: results from the Breast International Group Trial 1–98', *Breast Cancer Research*, 21(1), pp. 30.
- Key, T. J., Verkasalo, P. K. and Banks, E. (2001) 'Epidemiology of breast cancer', *Lancet Oncology*, 2(3), pp. 133-140.
- Khaidar, N. G., Lane, D., Matte, I., Rancourt, C. and Piché, A. (2012) 'Targeted ovarian cancer treatment: the TRAILS of resistance', *American journal of cancer research*, 2(1), pp. 75-92.
- Khan, J. A., Bellance, C., Guiochon-Mantel, A., Lombes, M. and Loosfelt, H. (2012) 'Differential regulation of breast cancer-associated genes by progesterone receptor isoforms PRA and PRB in a new bi-inducible breast cancer cell line', *PLoS One*, 7(9), pp. e45993.
- Kim, S. Y. (2009) 'Effects of sample size on robustness and prediction accuracy of a prognostic gene signature', *BMC Bioinformatics*, 10, pp. 147.
- Kleylein-Sohn, J., Westendorf, J., Le Clech, M., Habedanck, R., Stierhof, Y. D. and Nigg, E. A. (2007) 'Plk4-induced centriole biogenesis in human cells', *Dev Cell*, 13(2), pp. 190-202.
- Kluska, A., Balabas, A., Piatkowska, M., Czarny, K., Paczkowska, K., Nowakowska, D., Mikula, M. and Ostrowski, J. (2017) 'PALB2 mutations in BRCA1/2-mutation negative breast and ovarian cancer patients from Poland', *BMC medical genomics*, 10(1), pp. 14-14.

- Koschützki, D. and Schreiber, F. (2008) 'Centrality analysis methods for biological networks and their application to gene regulatory networks', *Gene regulation and systems biology*, 2, pp. 193-201.
- Koti, M., Gooding, R. J., Nuin, P., Haslehurst, A., Crane, C., Weberpals, J., Childs, T., Bryson, P., Dharsee, M., Evans, K., Feilotter, H. E., Park, P. C. and Squire, J. A. (2013) 'Identification of the IGF1/PI3K/NF kappaB/ERK gene signalling networks associated with chemotherapy resistance and treatment response in high-grade serous epithelial ovarian cancer', *BMC Cancer*, 13, pp. 549.
- Kotsopoulos, J., Sopik, V., Rosen, B., Fan, I., McLaughlin, J. R., Risch, H., Sun, P., Narod, S. A. and Akbari, M. R. (2017) 'Frequency of germline PALB2 mutations among women with epithelial ovarian cancer', *Fam Cancer*, 16(1), pp. 29-34.
- Kretschmer, C., Sterner-Kock, A., Siedentopf, F., Schoenegg, W., Schlag, P. M. and Kemmner, W. (2011) 'Identification of early molecular markers for breast cancer', *Mol Cancer*, 10(1), pp. 15.
- Kriplani, D. and Patel, M. M. (2013) 'Immunohistochemistry: A diagnostic aid in differentiating primary epithelial ovarian tumors and tumors metastatic to the ovary', *South Asian journal of cancer*, 2(4), pp. 254-258.
- Kryvenko, O. N. and Epstein, J. I. (2016) 'Changes in prostate cancer grading: Including a new patient-centric grading system', *Prostate*, 76(5), pp. 427-33.
- Kumar, P. and Aggarwal, R. (2016) 'An overview of triple-negative breast cancer', *Arch Gynecol Obstet*, 293(2), pp. 247-69.
- Kung, P. P., Martinez, R., Zhu, Z., Zager, M., Blasina, A., Rymer, I., Hallin, J., Xu, M., Carroll, C., Chionis, J., Wells, P., Kozminski, K., Fan, J., Guicherit, O., Huang, B., Cui, M., Liu, C., Huang, Z., Sistla, A., Yang, J. and Murray, B. W. (2014) 'Chemogenetic evaluation of the mitotic kinesin CENP-E reveals a critical role in triple-negative breast cancer', *Mol Cancer Ther*, 13(8), pp. 2104-15.
- Kuo, K.-T., Mao, T.-L., Jones, S., Veras, E., Ayhan, A., Wang, T.-L., Glas, R., Slamon, D., Velculescu, V. E., Kuman, R. J. and Shih, I.-M. (2009) 'Frequent activating mutations of PIK3CA in ovarian clear cell carcinoma', *The American journal of pathology*, 174(5), pp. 1597-1601.
- Kuo, W. P., Jenssen, T. K., Butte, A. J., Ohno-Machado, L. and Kohane, I. S. (2002) 'Analysis of matched mRNA measurements from two different microarray technologies', *Bioinformatics*, 18(3), pp. 405-12.
- Kurman, R. J. and Shih, I.-M. (2010) 'The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory', *The American journal of surgical pathology*, 34(3), pp. 433-443.
- Kyo, S., Ishikawa, N., Nakamura, K. and Nakayama, K. (2020) 'The fallopian tube as origin of ovarian cancer: Change of diagnostic and preventive strategies', *Cancer Medicine*, 9(2), pp. 421-431.
- Köllermann, J., Albrecht, H., Schlomm, T., Huland, H., Graefen, M., Bokemeyer, C., Simon, R., Sauter, G. and Wilczak, W. (2010) 'Activating BRAF gene mutations are uncommon in hormone refractory prostate cancer in Caucasian patients', *Oncology letters*, 1(4), pp. 729-732.
- Labidi-Galy, S. I., Papp, E., Hallberg, D., Niknafs, N., Adleff, V., Noe, M., Bhattacharya, R., Novak, M., Jones, S., Phallen, J., Hruban, C. A., Hirsch, M. S., Lin, D. I., Schwartz, L., Maire, C. L., Tille, J.-C., Bowden, M., Ayhan, A., Wood, L. D., Scharpf, R. B., Kurman, R., Wang, T.-L., Shih, I.-M., Karchin, R., Drapkin, R. and Velculescu, V. E. (2017) 'High grade serous ovarian carcinomas originate in the fallopian tube', *Nature communications*, 8(1), pp. 1093-1093.

Lacey, J. J. V., Mink, P. J., Lubin, J. H., Sherman, M. E., Troisi, R., Hartge, P., Schatzkin, A. and Schairer, C. (2002) 'Menopausal Hormone Replacement Therapy and Risk of Ovarian Cancer', *JAMA*, 288(3), pp. 334-341.

Lafront, C., Germain, L., Weidmann, C. and Audet-Walsh, É. (2020) 'A Systematic Study of the Impact of Estrogens and Selective Estrogen Receptor Modulators on Prostate Cancer Cell Proliferation', *Scientific Reports*, 10(1), pp. 4024.

Lan, Y., Zhang, Y., Wang, J., Lin, C., Ittmann, M. M. and Wang, F. (2008) 'Aberrant expression of Cks1 and Cks2 contributes to prostate tumorigenesis by promoting proliferation and inhibiting programmed cell death', *International journal of cancer*, 123(3), pp. 543-551.

Langdon, S. P., Herrington, C. S., Hollis, R. L. and Gourley, C. (2020) 'Estrogen Signaling and Its Potential as a Target for Therapy in Ovarian Cancer', *Cancers (Basel)*, 12(6).

Larue, L. and Bellacosa, A. (2005) 'Epithelial-mesenchymal transition in development and cancer: role of phosphatidylinositol 3' kinase/AKT pathways', *Oncogene*, 24(50), pp. 7443-54.

Lawrenson, K. and Iversen, E. S. and Tyrer, J. and Weber, R. P. and Concannon, P. and Hazelett, D. J. and Li, Q. and Marks, J. R. and Berchuck, A. and Lee, J. M. and Aben, K. K. H. and Anton-Culver, H. and Antonenkova, N. and Australian Cancer, S. and Australian Ovarian Cancer Study, G. and Bandera, E. V. and Bean, Y. and Beckmann, M. W. and Bisogna, M. and Bjorge, L. and Bogdanova, N. and Brinton, L. A. and Brooks-Wilson, A. and Bruinsma, F. and Butzow, R. and Campbell, I. G. and Carty, K. and Chang-Claude, J. and Chenevix-Trench, G. and Chen, A. and Chen, Z. and Cook, L. S. and Cramer, D. W. and Cunningham, J. M. and Cybulski, C. and Plisiecka-Halasa, J. and Dennis, J. and Dicks, E. and Doherty, J. A. and Dörk, T. and du Bois, A. and Eccles, D. and Easton, D. T. and Edwards, R. P. and Eilber, U. and Ekici, A. B. and Fasching, P. A. and Fridley, B. L. and Gao, Y.-T. and Gentry-Maharaj, A. and Giles, G. G. and Glasspool, R. and Goode, E. L. and Goodman, M. T. and Gronwald, J. and Harter, P. and Hasmad, H. N. and Hein, A. and Heitz, F. and Hildebrandt, M. A. T. and Hillemanns, P. and Hogdall, E. and Hogdall, C. and Hosono, S. and Jakubowska, A. and Paul, J. and Jensen, A. and Karlan, B. Y. and Kjaer, S. K. and Kelemen, L. E. and Kellar, M. and Kelley, J. L. and Kiemeny, L. A. and Krakstad, C. and Lambrechts, D. and Lambrechts, S. and Le, N. D. and Lee, A. W. and Cannioto, R. and Leminen, A. and Lester, J. and Levine, D. A. and Liang, D. and Lissowska, J. and Lu, K. and Lubinski, J. and Lundvall, L. and Massuger, L. F. A. G. and Matsuo, K. and McGuire, V. and McLaughlin, J. R. and Nevanlinna, H. and McNeish, I. and Menon, U. and Modugno, F. and Moysich, K. B. and Narod, S. A. and Nedergaard, L. and Ness, R. B. and Noor Azmi, M. A. and Odunsi, K. and Olson, S. H. and Orlow, I. and Orsulic, S. and Pearce, C. L. and Pejovic, T. and Peltari, L. M. and Permuth-Wey, J. and Phelan, C. M. and Pike, M. C. and Poole, E. M. and Ramus, S. J. and Risch, H. A. and Rosen, B. and Rossing, M. A. and Rothstein, J. H. and Rudolph, A. and Runnebaum, I. B. and Rzepecka, I. K. and Salvesen, H. B. and Budzilowska, A. and Sellers, T. A. and Shu, X.-O. and Shvetsov, Y. B. and Siddiqui, N. and Sieh, W. and Song, H. and Southey, M. C. and Sucheston, L. and Tangen, I. L. and Teo, S.-H. and Terry, K. L. and Thompson, P. J. and Timorek, A. and Tworoger, S. S. and Van Nieuwenhuysen, E. and Vergote, I. and Vierkant, R. A. and Wang-Gohrke, S. and Walsh, C. and Wentzensen, N. and Whittemore, A. S. and Wicklund, K. G. and Wilkens, L. R. and Woo, Y.-L. and Wu, X. and Wu, A. H. and Yang, H. and Zheng, W. and Ziogas, A. and Coetzee, G. A. and Freedman, M. L. and Monteiro, A. N. A. and Moes-Sosnowska, J. and Kupryjanczyk, J. and Pharoah, P. D. and Gayther, S. A. and Schildkraut, J. M. (2015) 'Common variants at the CHEK2 gene locus and risk of epithelial ovarian cancer', *Carcinogenesis*, 36(11), pp. 1341-1353.

Leda, M., Holland, A. J. and Goryachev, A. B. (2018) 'Autoamplification and Competition Drive Symmetry Breaking: Initiation of Centriole Duplication by the PLK4-STIL Network', *iScience*, 8, pp. 222-235.

Lee, K., Lee, A., Song, B. J. and Kang, C. S. (2011) 'Expression of AIB1 protein as a prognostic factor in breast cancer', *World J Surg Oncol*, 9, pp. 139.

Lee, S., Stewart, S., Nagtegaal, I., Luo, J., Wu, Y., Colditz, G., Medina, D. and Allred, D. C. (2012) 'Differentially expressed genes regulating the progression of ductal carcinoma in situ to invasive breast cancer', *Cancer Res*, 72(17), pp. 4574-86.

Author (2017) *sva: surrogate variable analysis* (Version R package version 3.24.4).

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. and Storey, J. D. (2012) 'The sva package for removing batch effects and other unwanted variation in high-throughput experiments', *Bioinformatics (Oxford, England)*, 28(6), pp. 882-883.

Lehmann, B. D., Bauer, J. A., Chen, X., Sanders, M. E., Chakravarthy, A. B., Shyr, Y. and Pietenpol, J. A. (2011) 'Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies.(Research article)(Report)', *Journal of Clinical Investigation*, 121(7), pp. 1.

Leinonen, R., Sugawara, H., Shumway, M. and International Nucleotide Sequence Database, C. (2011) 'The sequence read archive', *Nucleic acids research*, 39(Database issue), pp. D19-D21.

Leitzmann, M. F., Koebnick, C., Danforth, K. N., Brinton, L. A., Moore, S. C., Hollenbeck, A. R., Schatzkin, A. and Lacey, J. V., Jr. (2009) 'Body mass index and risk of ovarian cancer', *Cancer*, 115(4), pp. 812-822.

Lengyel, E. (2010) 'Ovarian cancer development and metastasis', *The American journal of pathology*, 177(3), pp. 1053-1064.

Li, J., Li, M., Chen, P. and Ba, Q. (2017) 'High expression of PALB2 predicts poor prognosis in patients with advanced breast cancer', *FEBS open bio*, 8(1), pp. 56-63.

Li, J., Zou, C., Bai, Y., Wazer, D. E., Band, V. and Gao, Q. (2006a) 'DSS1 is required for the stability of BRCA2', *Oncogene*, 25(8), pp. 1186-1194.

Li, S. Y., Rong, M., Grieu, F. and Iacopetta, B. (2006b) 'PIK3CA mutations in breast cancer are associated with poor outcome', *Breast Cancer Res Treat*, 96(1), pp. 91-5.

Li, T., Gao, X., Han, L., Yu, J. and Li, H. (2018) 'Identification of hub genes with prognostic values in gastric cancer by bioinformatics analysis', *World Journal of Surgical Oncology*, 16(1), pp. 114.

Li, Y., Deng, G., Qi, Y., Zhang, H., Jiang, H., Geng, R., Ye, Z., Liu, B. and Chen, Q. (2020) 'Downregulation of LUZP2 Is Correlated with Poor Prognosis of Low-Grade Glioma', *BioMed research international*, 2020, pp. 9716720-9716720.

Li, Z., Cui, J., Yu, Q., Wu, X., Pan, A. and Li, L. (2016) 'Evaluation of CCND1 amplification and CyclinD1 expression: diffuse and strong staining of CyclinD1 could have same predictive roles as CCND1 amplification in ER positive breast cancers', *Am J Transl Res*, 8(1), pp. 142-53.

Liesecke, F., De Craene, J.-O., Besseau, S., Courdavault, V., Clastre, M., Vergès, V., Papon, N., Giglioli-Guivarc'h, N., Glévarec, G., Pichon, O. and Dugé de Bernonville, T. (2019) 'Improved gene co-expression network quality through expression dataset down-sampling and network aggregation', *Scientific Reports*, 9(1), pp. 14431.

Lin, P.-H., Aronson, W. and Freedland, S. J. (2015) 'Nutrition, dietary interventions and prostate cancer: the latest evidence', *BMC medicine*, 13, pp. 3-3.

- Lin, Y., Fukuchi, J., Hiipakka, R. A., Kokontis, J. M. and Xiang, J. (2007) 'Up-regulation of Bcl-2 is required for the progression of prostate cancer cells from an androgen-dependent to an androgen-independent growth stage', *Cell Res*, 17(6), pp. 531-6.
- Liou, G. Y. and Storz, P. (2010) 'Reactive oxygen species in cancer', *Free Radic Res*, 44(5), pp. 479-96.
- Lisowska, K. M., Olbryt, M., Dudaladava, V., Pamula-Pilat, J., Kujawa, K., Grzybowska, E., Jarzab, M., Student, S., Rzepecka, I. K., Jarzab, B. and Kupryjanczyk, J. (2014) 'Gene expression analysis in ovarian cancer - faults and hints from DNA microarray study', *Front Oncol*, 4, pp. 6.
- Liu, Y., Chen, T.-Y., Yang, Z.-Y., Fang, W., Wu, Q. and Zhang, C. (2020) 'Identification of hub genes in papillary thyroid carcinoma: robust rank aggregation and weighted gene co-expression network analysis', *Journal of Translational Medicine*, 18(1), pp. 170.
- Liu, Y., Gu, H.-Y., Zhu, J., Niu, Y.-M., Zhang, C. and Guo, G.-L. (2019a) 'Identification of Hub Genes and Key Pathways Associated With Bipolar Disorder Based on Weighted Gene Co-expression Network Analysis', *Frontiers in Physiology*, 10, pp. 1081.
- Liu, Y., Yi, Y., Wu, W., Wu, K. and Zhang, W. (2019b) 'Bioinformatics prediction and analysis of hub genes and pathways of three types of gynecological cancer', *Oncol Lett*, 18(1), pp. 617-628.
- Liu, Z., Zhao, X., Zhang, L. and Pei, B. (2019c) 'Cytochrome C inhibits tumor growth and predicts favorable prognosis in clear cell renal cell carcinoma', *Oncol Lett*, 18(6), pp. 6026-6032.
- Lo, H. W., Hsu, S. C. and Hung, M. C. (2006) 'EGFR signaling pathway in breast cancers: from traditional signal transduction to direct nuclear translocalization', *Breast Cancer Res Treat*, 95(3), pp. 211-8.
- Lohiya, V., Aragon-Ching, J. B. and Sonpavde, G. (2016) 'Role of Chemotherapy and Mechanisms of Resistance to Chemotherapy in Metastatic Castration-Resistant Prostate Cancer', *Clinical Medicine Insights. Oncology*, 10(Suppl 1), pp. 57-66.
- Lokeshwar, V. B., Mirza, S. and Jordan, A. (2014) 'Targeting hyaluronic acid family for cancer chemoprevention and therapy', *Advances in cancer research*, 123, pp. 35-65.
- Long, T., Liu, Z., Zhou, X., Yu, S., Tian, H. and Bao, Y. (2019) 'Identification of differentially expressed genes and enriched pathways in lung cancer using bioinformatics analysis', *Molecular medicine reports*, 19(3), pp. 2029-2040.
- Lu, D., Bai, X., Zou, Q., Gan, Z. and Lv, Y. (2020) 'Identification of the association between HMMR expression and progression of hepatocellular carcinoma via construction of a co-expression network', *Oncol Lett*, 20(3), pp. 2645-2654.
- Lue, Y., Wang, C., Lydon, J. P., Leung, A., Li, J. and Swerdloff, R. S. (2013) 'Functional role of progestin and the progesterone receptor in the suppression of spermatogenesis in rodents', *Andrology*, 1(2), pp. 308-17.
- Luo, Z. and Farnham, P. J. (2020) 'Genome-wide analysis of HOXC4 and HOXC6 regulated genes and binding sites in prostate cancer cells', *PLoS one*, 15(2), pp. e0228590-e0228590.
- Lánczky, A. and Gyórfy, B. (2021) 'Web-Based Survival Analysis Tool Tailored for Medical Research (KMplot): Development and Implementation', *J Med Internet Res*, 23(7), pp. e27633.

- Ma, M., Li, L., Chen, H. and Feng, Y. (2019) 'Oxytocin Inhibition of Metastatic Colorectal Cancer by Suppressing the Expression of Fibroblast Activation Protein- α ', *Frontiers in neuroscience*, 13, pp. 1317-1317.
- Ma, X. J., Dahiya, S., Richardson, E., Erlander, M. and Sgroi, D. C. (2009) 'Gene expression profiling of the tumor microenvironment during breast cancer progression', *Breast Cancer Res*, 11(1), pp. R7.
- Mackay, H. J., Brady, M. F., Oza, A. M., Reuss, A., Pujade-Lauraine, E., Swart, A. M., Siddiqui, N., Colombo, N., Bookman, M. A., Pfisterer, J. and du Bois, A. (2010) 'Prognostic Relevance of Uncommon Ovarian Histology in Women With Stage III/IV Epithelial Ovarian Cancer', *International Journal of Gynecologic Cancer*, 20(6), pp. 945.
- Mah, N., Thelin, A., Lu, T., Nikolaus, S., Kühbacher, T., Gurbuz, Y., Eickhoff, H., Klöppel, G., Lehrach, H., Mellgård, B., Costello, C. M. and Schreiber, S. (2004) 'A comparison of oligonucleotide and cDNA-based microarray systems', *Physiological Genomics*, 16(3), pp. 361-370.
- Maier, H. J., Wirth, T. and Beug, H. (2010) 'Epithelial-mesenchymal transition in pancreatic carcinoma', *Cancers (Basel)*, 2(4), pp. 2058-83.
- Maire, V., Nemati, F., Richardson, M., Vincent-Salomon, A., Tesson, B., Rigaille, G., Gravier, E., Marty-Prouvost, B., De Koning, L., Lang, G., Gentien, D., Dumont, A., Barillot, E., Marangoni, E., Decaudin, D., Roman-Roman, S., Pierre, A., Cruzalegui, F., Depil, S., Tucker, G. C. and Dubois, T. (2013) 'Polo-like kinase 1: a potential therapeutic option in combination with conventional chemotherapy for the management of patients with triple-negative breast cancer', *Cancer Res*, 73(2), pp. 813-23.
- Maistro, S., Teixeira, N., Encinas, G., Katayama, M. L., Niewiadonski, V. D., Cabral, L. G., Ribeiro, R. M., Gaburo Junior, N., de Gouvea, A. C., Carraro, D. M., Sabino, E. C., Diz, M. D., Chammas, R., de Bock, G. H. and Folgueira, M. A. (2016) 'Germline mutations in BRCA1 and BRCA2 in epithelial ovarian cancer patients in Brazil', *BMC Cancer*, 16(1), pp. 934.
- Mak, P., Leav, I., Pursell, B., Bae, D., Yang, X., Taglienti, C. A., Gouvin, L. M., Sharma, V. M. and Mercurio, A. M. (2010) 'ERbeta impedes prostate cancer EMT by destabilizing HIF-1alpha and inhibiting VEGF-mediated snail nuclear localization: implications for Gleason grading', *Cancer Cell*, 17(4), pp. 319-32.
- Maleki, F., Ovens, K., McQuillan, I. and Kusalik, A. J. (2019) 'Size matters: how sample size affects the reproducibility and specificity of gene set analysis', *Human Genomics*, 13(1), pp. 42.
- Malkin, D. (2011) 'Li-fraumeni syndrome', *Genes & cancer*, 2(4), pp. 475-484.
- Mao, Y., Desai, A. and Cleveland, D. W. (2005) 'Microtubule capture by CENP-E silences BubR1-dependent mitotic checkpoint signaling', *Journal of Cell Biology*, 170(6), pp. 873-880.
- Maravei, D. V., Trbovich, A. M., Perez, G. I., Tilly, K. I., Banach, D., Talanian, R. V., Wong, W. W. and Tilly, J. L. (1997) 'Cleavage of cytoskeletal proteins by caspases during ovarian cell death: evidence that cell-free systems do not always mimic apoptotic events in intact cells', *Cell Death & Differentiation*, 4(8), pp. 707-712.
- Marhaba, R. and Zöller, M. (2004) 'CD44 in Cancer Progression: Adhesion, Migration and Growth Regulation', *Journal of Molecular Histology*, 35(3), pp. 211-231.
- Martin, A. M. and Weber, B. L. (2000) 'Genetic and hormonal risk factors in breast cancer', *J Natl Cancer Inst*, 92(14), pp. 1126-35.

Marumoto, T., Hirota, T., Morisaki, T., Kunitoku, N., Zhang, D., Ichikawa, Y., Sasayama, T., Kuninaka, S., Mimori, T., Tamaki, N., Kimura, M., Okano, Y. and Saya, H. (2002) 'Roles of aurora-A kinase in mitotic entry and G2 checkpoint in mammalian cells', *Genes Cells*, 7(11), pp. 1173-82.

Maréchal, A. and Zou, L. (2013) 'DNA damage sensing by the ATM and ATR kinases', *Cold Spring Harbor perspectives in biology*, 5(9), pp. a012716.

Matsuoka, T. and Yashiro, M. (2014) 'The Role of PI3K/Akt/mTOR Signaling in Gastric Carcinoma', *Cancers (Basel)*, 6(3), pp. 1441-63.

Maxwell, C. A. and Benítez, J. and Gómez-Baldó, L. and Osorio, A. and Bonifaci, N. and Fernández-Ramires, R. and Costes, S. V. and Guinó, E. and Chen, H. and Evans, G. J. R. and Mohan, P. and Català, I. and Petit, A. and Aguilar, H. and Villanueva, A. and Aytes, A. and Serra-Musach, J. and Rennert, G. and Lejbkowitz, F. and Peterlongo, P. and Manoukian, S. and Peissel, B. and Ripamonti, C. B. and Bonanni, B. and Viel, A. and Allavena, A. and Bernard, L. and Radice, P. and Friedman, E. and Kaufman, B. and Laitman, Y. and Dubrovsky, M. and Milgrom, R. and Jakubowska, A. and Cybulski, C. and Gorski, B. and Jaworska, K. and Durda, K. and Sukiennicki, G. and Lubiński, J. and Shugart, Y. Y. and Domchek, S. M. and Letrero, R. and Weber, B. L. and Hogervorst, F. B. L. and Rookus, M. A. and Collee, J. M. and Devilee, P. and Ligtenberg, M. J. and Lujt, R. B. v. d. and Aalfs, C. M. and Waisfisz, Q. and Wijnen, J. and Roozendaal, C. E. P. v. and Hebon and Embrace and Easton, D. F. and Peock, S. and Cook, M. and Oliver, C. and Frost, D. and Harrington, P. and Evans, D. G. and Lalloo, F. and Eeles, R. and Izatt, L. and Chu, C. and Eccles, D. and Douglas, F. and Brewer, C. and Nevanlinna, H. and Heikkinen, T. and Couch, F. J. and Lindor, N. M. and Wang, X. and Godwin, A. K. and Caligo, M. A. and Lombardi, G. and Loman, N. and Karlsson, P. and Ehrencrona, H. and Wachenfeldt, A. v. and Swe, B. and Barkardottir, R. B. and Hamann, U. and Rashid, M. U. and Lasa, A. and Caldés, T. and Andrés, R. and Schmitt, M. and Assmann, V. and Stevens, K. and Offit, K. and Curado, J. and Tilgner, H. and Guigó, R. and Aiza, G. and Brunet, J. and Castellsagué, J. and Martrat, G. and Urruticoechea, A. and Blanco, I. and Tihomirova, L. and Goldgar, D. E. and Buys, S. and John, E. M. and Miron, A. and Southey, M. and Daly, M. B. and Bcfr and Schmutzler, R. K. and Wappenschmidt, B. and Meindl, A. and Arnold, N. and Deissler, H. and Varon-Mateeva, R. and Sutter, C. and Niederacher, D. and Imyamitov, E. and Sinilnikova, O. M. and Stoppa-Lyonne, D. and Mazoyer, S. and Verny-Pierre, C. and Castera, L. and de Pauw, A. and Bignon, Y.-J. and Uhrhammer, N. and Peyrat, J.-P. and Vennin, P. and Fert Ferrer, S. and Collonge-Rame, M.-A. and Mortemousque, I. and Collaborators, G. S. and Spurdle, A. B. and Beesley, J. and Chen, X. and Healey, S. and kConFab and Barcellos-Hoff, M. H. and Vidal, M. and Gruber, S. B. and Lázaro, C. and Capellá, G. and McGuffog, L. and Nathanson, K. L. and Antoniou, A. C. and Chenevix-Trench, G. and Fleisch, M. C. and Moreno, V. and Pujana, M. A. (2011) 'Interplay between BRCA1 and RHMAMM regulates epithelial apicobasal polarization and may influence risk of breast cancer', *PLoS biology*, 9(11), pp. e1001199-e1001199.

Maxwell, C. A., Keats, J. J., Belch, A. R., Pilarski, L. M. and Reiman, T. (2005) 'Receptor for hyaluronan-mediated motility correlates with centrosome abnormalities in multiple myeloma and maintains mitotic integrity', *Cancer Res*, 65(3), pp. 850-60.

Mazzu, Y. Z., Armenia, J., Nandakumar, S., Chakraborty, G., Yoshikawa, Y., Jehane, L. E., Lee, G.-S. M., Atiq, M., Khan, N., Schultz, N. and Kantoff, P. W. (2020) 'Ribonucleotide reductase small subunit M2 is a master driver of aggressive prostate cancer', *Molecular oncology*, 14(8), pp. 1881-1897.

McCain, J. (2013) 'The MAPK (ERK) Pathway: Investigational Combinations for the Treatment Of BRAF-Mutated Metastatic Melanoma', *P t*, 38(2), pp. 96-108.

McMullin, R. P., Wittner, B. S., Yang, C., Denton-Schneider, B. R., Hicks, D., Singavarapu, R., Moulis, S., Lee, J., Akbari, M. R., Narod, S. A., Aldape, K. D., Steeg, P. S., Ramaswamy, S. and Sgroi, D. C. (2014) 'A BRCA1 deficient-like signature is enriched in breast cancer brain metastases and

predicts DNA damage-induced poly (ADP-ribose) polymerase inhibitor sensitivity', *Breast Cancer Res*, 16(2), pp. R25.

Meinhold-Heerlein, I., Fotopoulou, C., Harter, P., Kurzeder, C., Mustea, A., Wimberger, P., Hauptmann, S. and Sehouli, J. (2016) 'The new WHO classification of ovarian, fallopian tube, and primary peritoneal cancer and its clinical implications', *Arch Gynecol Obstet*, 293(4), pp. 695-700.

Menard, S., Fortis, S., Castiglioni, F., Agresti, R. and Balsari, A. (2001) 'HER2 as a prognostic factor in breast cancer', *Oncology*, 61 Suppl 2, pp. 67-72.

Mendoza, M. C., Er, E. E. and Blenis, J. (2011) 'The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation', *Trends in biochemical sciences*, 36(6), pp. 320-328.

Meng, F., Zhang, L., Ren, Y. and Ma, Q. (2020) 'Transcriptome analysis reveals key signature genes involved in the oncogenesis of lung cancer', *Cancer Biomark*.

Meng, Y., Wang, W., Kang, J., Wang, X. and Sun, L. (2017) 'Role of the PI3K/AKT signalling pathway in apoptotic cell death in the cerebral cortex of streptozotocin-induced diabetic rats', *Experimental and therapeutic medicine*, 13(5), pp. 2417-2422.

Menzl, I., Lebeau, L., Pandey, R., Hassounah, N. B., Li, F. W., Nagle, R., Weihs, K. and McDermott, K. M. (2014) 'Loss of primary cilia occurs early in breast cancer development', *Cilia*, 3, pp. 7-7.

Metcalfe, K., Lubinski, J., Lynch, H. T., Ghadirian, P., Foulkes, W. D., Kim-Sing, C., Neuhausen, S., Tung, N., Rosen, B., Gronwald, J., Ainsworth, P., Sweet, K., Eisen, A., Sun, P. and Narod, S. A. (2010) 'Family history of cancer and cancer risks in women with BRCA1 or BRCA2 mutations', *J Natl Cancer Inst*, 102(24), pp. 1874-8.

Metzger-Filho, O., Michiels, S., Bertucci, F., Catteau, A., Salgado, R., Galant, C., Fumagalli, D., Singhal, S. K., Desmedt, C., Ignatiadis, M., Haussy, S., Finetti, P., Birnbaum, D., Saini, K. S., Berlière, M., Veys, I., de Azambuja, E., Bozovic, I., Peyro-Saint-Paul, H., Larsimont, D., Piccart, M. and Sotiriou, C. (2013) 'Genomic grade adds prognostic value in invasive lobular carcinoma', *Annals of Oncology*, 24(2), pp. 377-384.

Meyniel, J. P., Cottu, P. H., Decraene, C., Stern, M. H., Couturier, J., Lebigot, I., Nicolas, A., Weber, N., Fourchette, V., Alran, S., Rapinat, A., Gentien, D., Roman-Roman, S., Mignot, L. and Sastre-Garau, X. (2010) 'A genomic and transcriptomic approach for a differential diagnosis between primary and secondary ovarian carcinomas in patients with a previous history of breast cancer', *BMC Cancer*, 10, pp. 222.

Michael, A., Catherine, A. B., Judith, A. B., David, B., Heather, B., Cherry, J. M., Allan, P. D., Kara, D., Selina, S. D., Janan, T. E., Midori, A. H., David, P. H., Laurie, I.-T., Andrew, K., Suzanna, L., John, C. M., Joel, E. R., Martin, R., Gerald, M. R. and Gavin, S. (2000) 'Gene Ontology: tool for the unification of biology', *Nature Genetics*, 25(1), pp. 25.

Mikuła-Pietrasik, J., Witucka, A., Pakuła, M., Uruski, P., Begier-Krasińska, B., Niklas, A., Tykarski, A. and Książek, K. (2019) 'Comprehensive review on how platinum- and taxane-based chemotherapy of ovarian cancer affects biology of normal cells', *Cellular and molecular life sciences : CMLS*, 76(4), pp. 681-697.

Miller, T. W., Rexer, B. N., Garrett, J. T. and Arteaga, C. L. (2011) 'Mutations in the phosphatidylinositol 3-kinase pathway: role in tumor progression and therapeutic implications in breast cancer', *Breast Cancer Res*, 13(6), pp. 224.

Miller, W. L. and Flück, C. E. (2014) 'CHAPTER 13 - Adrenal cortex and its disorders', in Sperling, M.A. (ed.) *Pediatric Endocrinology (Fourth Edition)*: W.B. Saunders, pp. 471-532.e1.

- Mills, R. C., 3rd (2017) 'Breast Cancer Survivors, Common Markers of Inflammation, and Exercise: A Narrative Review', *Breast cancer : basic and clinical research*, 11, pp. 1178223417743976-1178223417743976.
- Milunović-Jevtić, A., Mooney, P., Sulerud, T., Bisht, J. and Gatlin, J. C. (2016) 'Centrosomal clustering contributes to chromosomal instability and cancer', *Current Opinion in Biotechnology*, 40, pp. 113-118.
- Mishra, A. P., Hartford, S. A., Sahu, S., Klarmann, K., Chittela, R. K., Biswas, K., Jeon, A. B., Martin, B. K., Burkett, S., Southon, E., Reid, S., Albaugh, M. E., Karim, B., Tessarollo, L., Keller, J. R. and Sharan, S. K. (2022) 'BRCA2-DSS1 interaction is dispensable for RAD51 recruitment at replication-induced and meiotic DNA double strand breaks', *Nature Communications*, 13(1), pp. 1751.
- Misra, S., Hascall, V. C., Markwald, R. R. and Ghatak, S. (2015) 'Interactions between Hyaluronan and Its Receptors (CD44, RHAMM) Regulate the Activities of Inflammation and Cancer', *Frontiers in immunology*, 6, pp. 201-201.
- Mittal, R. D., Mandal, R. K., Singh, A. and Srivastava, P. (2015) 'Death receptor 4 variants enhanced prostate cancer risk in North Indian population', *Tumour Biol*, 36(7), pp. 5655-61.
- Mizushima, T. and Miyamoto, H. (2019) 'The Role of Androgen Receptor Signaling in Ovarian Cancer', *Cells*, 8(2), pp. 176.
- Moerkens, M., Zhang, Y., Wester, L., van de Water, B. and Meerman, J. H. (2014) 'Epidermal growth factor receptor signalling in human breast cancer cells operates parallel to estrogen receptor alpha signalling and results in tamoxifen insensitive proliferation', *BMC Cancer*, 14, pp. 283.
- Mogami, T., Yokota, N., Asai-Sato, M., Yamada, R., Koizume, S., Sakuma, Y., Yoshihara, M., Nakamura, Y., Takano, Y., Hirahara, F., Miyagi, Y. and Miyagi, E. (2013) 'Annexin A4 Is Involved in Proliferation, Chemo-Resistance and Migration and Invasion in Ovarian Clear Cell Adenocarcinoma Cells', *PLOS ONE*, 8(11), pp. e80359.
- Mok, S. C., Bonome, T., Vathipadiekal, V., Bell, A., Johnson, M. E., Wong, K. K., Park, D. C., Hao, K., Yip, D. K., Donninger, H., Ozbun, L., Samimi, G., Brady, J., Randonovich, M., Pise-Masison, C. A., Barrett, J. C., Wong, W. H., Welch, W. R., Berkowitz, R. S. and Birrer, M. J. (2009) 'A gene signature predictive for outcome in advanced ovarian cancer identifies a survival factor: microfibril-associated glycoprotein 2', *Cancer Cell*, 16(6), pp. 521-32.
- Monninkhof, E. M., van der Schouw, Y. T. and Peeters, P. H. (1999) 'Early age at menopause and breast cancer: are leaner women more protected? A prospective analysis of the Dutch DOM cohort', *Breast Cancer Res Treat*, 55(3), pp. 285-91.
- Monzon, F. A., Lyons-Weiler, M., Buturovic, L. J., Rigl, C. T., Henner, W. D., Sciulli, C., Dumur, C. I., Medeiros, F. and Anderson, G. G. (2009) 'Multicenter validation of a 1,550-gene expression profile for identification of tumor tissue of origin', *J Clin Oncol*, 27(15), pp. 2503-8.
- Moorman, P. G., Calingaert, B., Palmieri, R. T., Iversen, E. S., Bentley, R. C., Halabi, S., Berchuck, A. and Schildkraut, J. M. (2008) 'Hormonal risk factors for ovarian cancer in premenopausal and postmenopausal women', *American journal of epidemiology*, 167(9), pp. 1059-1069.
- Mortensen, M. M., Hoyer, S., Lynnerup, A. S., Orntoft, T. F., Sorensen, K. D., Borre, M. and Dyrskjot, L. (2015) 'Expression profiling of prostate cancer tissue delineates genes associated with recurrence after prostatectomy', *Sci Rep*, 5, pp. 16018.

- Moyer, T. C. and Holland, A. J. (2019) 'PLK4 promotes centriole duplication by phosphorylating STIL to link the procentriole cartwheel to the microtubule wall', *eLife*, 8, pp. e46054.
- Muga, S. V. D., Timpson, P., Cubells, L., Evans, R., Hayes, T. E., Rentero, C., Hegemann, A., Reverter, M., Leschner, J., Pol, A., Tebar, F., Daly, R. J., Enrich, C. and Grewal, T. (2008) 'Annexin A6 inhibits Ras signalling in breast cancer cells', *Oncogene*, 28(3), pp. 363.
- Mukhtar, E., Adhami, V. M. and Mukhtar, H. (2014) 'Targeting microtubules by natural agents for cancer therapy', *Molecular cancer therapeutics*, 13(2), pp. 275-284.
- Mulac-Jericevic, B., Mullinax, R. A., DeMayo, F. J., Lydon, J. P. and Conneely, O. M. (2000) 'Subgroup of reproductive functions of progesterone mediated by progesterone receptor- B isoform', *Science*, 289(5485), pp. 1751-1754.
- Muller, A., Homey, B., Soto, H., Ge, N., Catron, D., Buchanan, M. E., McClanahan, T., Murphy, E., Yuan, W., Wagner, S. N., Barrera, J. L., Mohar, A., Verastegui, E. and Zlotnik, A. (2001) 'Involvement of chemokine receptors in breast cancer metastasis', *Nature*, 410(6824), pp. 50-6.
- Muralidhar, G. G. and Barbolina, M. V. (2013) 'Chemokine receptors in epithelial ovarian cancer', *International journal of molecular sciences*, 15(1), pp. 361-376.
- Muranen, T. A., Greco, D., Fagerholm, R., Kilpivaara, O., Kampjarvi, K., Aittomaki, K., Blomqvist, C., Heikkila, P., Borg, A. and Nevanlinna, H. (2011) 'Breast tumors from CHEK2 1100delC-mutation carriers: genomic landscape and clinical implications', *Breast Cancer Res*, 13(5), pp. R90.
- Musgrove, E. A. and Sutherland, R. L. (2009) 'Biological determinants of endocrine resistance in breast cancer', *Nature Reviews Cancer*, 9(9), pp. 631-643.
- Myers, M. B., Banda, M., McKim, K. L., Wang, Y., Powell, M. J. and Parsons, B. L. (2016) 'Breast Cancer Heterogeneity Examined by High-Sensitivity Quantification of PIK3CA, KRAS, HRAS, and BRAF Mutations in Normal Breast and Ductal Carcinomas', *Neoplasia*, 18(4), pp. 253-63.
- Nacu, Ş., Critchley-Thorne, R., Lee, P. and Holmes, S. (2007) 'Gene expression network analysis and applications to immunology', *Bioinformatics*, 23(7), pp. 850-858.
- Nakagawa, T., Kollmeyer, T. M., Morlan, B. W., Anderson, S. K., Bergstralh, E. J., Davis, B. J., Asmann, Y. W., Klee, G. G., Ballman, K. V. and Jenkins, R. B. (2008) 'A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy', *PLoS One*, 3(5), pp. e2318.
- National Cancer Institute, N. (2020) *Cancer Stat Facts*. Surveillance, Epidemiology, and End Results (SEER) Program. <https://seer.cancer.gov/statfacts/>: National Cancer Institute (Accessed: February 2020).
- Nedvetzki, S., Gonen, E., Assayag, N., Reich, R., Williams, R. O., Thurmond, R. L., Huang, J.-F., Neudecker, B. A., Wang, F.-S., Turley, E. A. and Naor, D. (2004) 'RHAMM, a receptor for hyaluronan-mediated motility, compensates for CD44 in inflamed CD44-knockout mice: A different interpretation of redundancy', *Proceedings of the National Academy of Sciences*, 101(52), pp. 18081.
- Neff, R. T., Senter, L. and Salani, R. (2017) 'BRCA mutation in ovarian cancer: testing, implications and treatment considerations', *Therapeutic advances in medical oncology*, 9(8), pp. 519-531.
- Negishi, Y., Iwabuchi, H., Sakunaga, H., Sakamoto, M., Okabe, K., Sato, H. and Asano, G. (1993) 'Serum and tissue measurements of CA72-4 in ovarian cancer patients', *Gynecol Oncol*, 48(2), pp. 148-54.

- Negri, E., Pelucchi, C., Franceschi, S., Montella, M., Conti, E., Dal Maso, L., Parazzini, F., Tavani, A., Carbone, A. and La Vecchia, C. (2003) 'Family history of cancer and risk of ovarian cancer', *Eur J Cancer*, 39(4), pp. 505-10.
- Negrini, S., Gorgoulis, V. G. and Halazonetis, T. D. (2010) 'Genomic instability--an evolving hallmark of cancer', *Nat Rev Mol Cell Biol*, 11(3), pp. 220-8.
- Nevanlinna, H. and Bartek, J. (2006) 'The CHEK2 gene and inherited breast cancer susceptibility', *Oncogene*, 25(43), pp. 5912-9.
- Newton, K. and Dixit, V. M. (2012) 'Signaling in innate immunity and inflammation', *Cold Spring Harb Perspect Biol*, 4(3).
- Nickols, N. G., Nazarian, R., Zhao, S. G., Tan, V., Uzunangelov, V., Xia, Z., Baertsch, R., Neeman, E., Gao, A. C., Thomas, G. V., Howard, L., De Hoedt, A. M., Stuart, J., Goldstein, T., Chi, K., Gleave, M. E., Graff, J. N., Beer, T. M., Drake, J. M., Evans, C. P., Aggarwal, R., Foye, A., Feng, F. Y., Small, E. J., Aronson, W. J., Freedland, S. J., Witte, O. N., Huang, J., Alumkal, J. J., Reiter, R. E. and Rettig, M. B. (2019) 'MEK-ERK signaling is a therapeutic target in metastatic castration resistant prostate cancer', *Prostate Cancer and Prostatic Diseases*, 22(4), pp. 531-538.
- Nigg, E. A. and Raff, J. W. (2009) 'Centrioles, Centrosomes, and Cilia in Health and Disease', *Cell*, 139(4), pp. 663-678.
- Nyberg, T., Frost, D., Barrowdale, D., Evans, D. G., Bancroft, E., Adlard, J., Ahmed, M., Barwell, J., Brady, A. F., Brewer, C., Cook, J., Davidson, R., Donaldson, A., Eason, J., Gregory, H., Henderson, A., Izatt, L., Kennedy, M. J., Miller, C., Morrison, P. J., Murray, A., Ong, K.-R., Porteous, M., Pottinger, C., Rogers, M. T., Side, L., Snape, K., Walker, L., Tischkowitz, M., Eeles, R., Easton, D. F. and Antoniou, A. C. (2020) 'Prostate Cancer Risks for Male BRCA1 and BRCA2 Mutation Carriers: A Prospective Cohort Study', *European Urology*, 77(1), pp. 24-35.
- Oba, S., Sato, M. A., Takemasa, I., Monden, M., Matsubara, K. and Ishii, S. (2003) 'A Bayesian missing value estimation method for gene expression profile data', *Bioinformatics*, 19(16), pp. 2088-96.
- Obermeier, K., Sachsenweger, J., Friedl, T. W. P., Pospiech, H., Winqvist, R. and Wiesmüller, L. (2016) 'Heterozygous PALB2 c.1592delT mutation channels DNA double-strand break repair into error-prone pathways in breast cancer patients', *Oncogene*, 35(29), pp. 3796-3806.
- Obr, A. E. and Edwards, D. P. (2012) 'The biology of progesterone receptor in the normal mammary gland and in breast cancer', *Mol Cell Endocrinol*, 357(1-2), pp. 4-17.
- Obr, A. E., Grimm, S. L., Bishop, K. A., Pike, J. W., Lydon, J. P. and Edwards, D. P. (2013) 'Progesterone receptor and Stat5 signaling cross talk through RANKL in mammary epithelial cells', *Mol Endocrinol*, 27(11), pp. 1808-24.
- Ohashi, A., Otori, M. and Iwai, K. (2016) 'Motor activity of centromere-associated protein-E contributes to its localization at the center of the midbody to regulate cytokinetic abscission', *Oncotarget*, 7(48), pp. 79964-79980.
- Ohta, M., Ashikawa, T., Nozaki, Y., Kozuka-Hata, H., Goto, H., Inagaki, M., Oyama, M. and Kitagawa, D. (2014) 'Direct interaction of Plk4 with STIL ensures formation of a single procentriole per parental centriole', *Nature Communications*, 5(1), pp. 5267.
- Oldenburg, R. A., Kroeze-Jansema, K., Kraan, J., Morreau, H., Klijn, J. G., Hoogerbrugge, N., Ligtenberg, M. J., van Asperen, C. J., Vasen, H. F., Meijers, C., Meijers-Heijboer, H., de Bock, T. H.,

- Cornelisse, C. J. and Devilee, P. (2003) 'The CHEK2*1100delC variant acts as a breast cancer risk modifier in non-BRCA1/BRCA2 multiple-case families', *Cancer Res*, 63(23), pp. 8153-7.
- Olivier, M., Hollstein, M. and Hainaut, P. (2010) 'TP53 mutations in human cancers: origins, consequences, and clinical use', *Cold Spring Harbor perspectives in biology*, 2(1), pp. a001008-a001008.
- Ovcaricek, T., Frkovic, S. G., Matos, E., Mozina, B. and Borstnar, S. (2011) 'Triple negative breast cancer - prognostic factors and survival', *Radiol Oncol*, 45(1), pp. 46-52.
- Ozaki, T. and Nakagawara, A. (2011) 'Role of p53 in Cell Death and Human Cancers', *Cancers*, 3(1), pp. 994-1013.
- Pagès, H., Carlson, M., Falcon, S. and Li, N. (2018) *AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor* . <https://bioconductor.org/packages/AnnotationDbi> (Accessed).
- Pakkanen, S., Wahlfors, T., Siltanen, S., Patrikainen, M., Matikainen, M. P., Tammela, T. L. J. and Schleutker, J. (2009) 'PALB2 variants in hereditary and unselected Finnish prostate cancer cases', *Journal of negative results in biomedicine*, 8, pp. 12-12.
- Papa, A. and Pandolfi, P. P. (2019) 'The PTEN⁻PI3K Axis in Cancer', *Biomolecules*, 9(4), pp. 153.
- Papadakos, K. S., Darlix, A., Jacot, W. and Blom, A. M. (2019) 'High Levels of Cartilage Oligomeric Matrix Protein in the Serum of Breast Cancer Patients Can Serve as an Independent Prognostic Marker', *Frontiers in oncology*, 9, pp. 1141-1141.
- Parise, C. A., Bauer, K. R. and Caggiano, V. (2010) 'Variation in breast cancer subtypes with age and race/ethnicity', *Crit Rev Oncol Hematol*, 76(1), pp. 44-52.
- Park, S. H., Cheung, L. W., Wong, A. S. and Leung, P. C. (2008) 'Estrogen regulates Snail and Slug in the down-regulation of E-cadherin and induces metastatic potential of ovarian cancer cells through estrogen receptor alpha', *Mol Endocrinol*, 22(9), pp. 2085-98.
- Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, N., Coulson, R., Farne, A., Holloway, E., Kolesnykov, N., Lilja, P., Lukk, M., Mani, R., Rayner, T., Sharma, A., William, E., Sarkans, U. and Brazma, A. (2007) 'ArrayExpress--a public database of microarray experiments and gene expression profiles', *Nucleic Acids Res*, 35(Database issue), pp. D747-50.
- Paruthiyil, S., Parmar, H., Kerekatte, V., Cunha, G. R., Firestone, G. L. and Leitman, D. C. (2004) 'Estrogen receptor beta inhibits human breast cancer cell proliferation and tumor formation by causing a G2 cell cycle arrest', *Cancer Res*, 64(1), pp. 423-8.
- Patwardhan, D., Mani, S., Passemard, S., Gressens, P. and El Ghouzzi, V. (2018) 'STIL balancing primary microcephaly and cancer', *Cell death & disease*, 9(2), pp. 65-65.
- Paul, M. W., Sidhu, A., Liang, Y., van Rossum-Fikkert, S. E., Odijk, H., Zelensky, A. N., Kanaar, R. and Wyman, C. (2021) 'Role of BRCA2 DNA-binding and C-terminal domain in its mobility and conformation in DNA repair', *eLife*, 10, pp. e67926.
- Pauty, J., Couturier, A. M., Rodrigue, A., Caron, M.-C., Coulombe, Y., Delleire, G. and Masson, J.-Y. (2017) 'Cancer-causing mutations in the tumor suppressor PALB2 reveal a novel cancer mechanism using a hidden nuclear export signal in the WD40 repeat motif', *Nucleic Acids Research*, 45(5), pp. 2644-2657.
- Pawitan, Y., Bjohle, J., Amler, L., Borg, A. L., Eghazi, S., Hall, P., Han, X., Holmberg, L., Huang, F., Klaar, S., Liu, E. T., Miller, L., Nordgren, H., Ploner, A., Sandelin, K., Shaw, P. M., Smeds, J., Skoog,

- L., Wedren, S. and Bergh, J. (2005) 'Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts', *Breast Cancer Res*, 7(6), pp. R953-64.
- Pe'er, D. and Hacohen, N. (2011) 'Principles and strategies for developing network models in cancer', *Cell*, 144(6), pp. 864-73.
- Pearson, H. B., Li, J., Meniel, V. S., Fennell, C. M., Waring, P., Montgomery, K. G., Rebello, R. J., Macpherson, A. A., Koushyar, S., Furic, L., Cullinane, C., Clarkson, R. W., Smalley, M. J., Simpson, K. J., Pheese, T. J., Shepherd, P. R., Humbert, P. O., Sansom, O. J. and Phillips, W. A. (2018) 'Identification of Pik3ca Mutation as a Genetic Driver of Prostate Cancer That Cooperates with Pten Loss to Accelerate Progression and Castration-Resistant Growth', *Cancer Discov*, 8(6), pp. 764-779.
- Pease, J. C. and Tirnauer, J. S. (2011) 'Mitotic spindle misorientation in cancer--out of alignment and into the fire', *J Cell Sci*, 124(Pt 7), pp. 1007-16.
- Pepe, P. and Pennisi, M. (2015) 'Gleason score stratification according to age at diagnosis in 1028 men', *Contemporary oncology (Poznan, Poland)*, 19(6), pp. 471-473.
- Perlmutter, M. A. and Lepor, H. (2007) 'Androgen deprivation therapy in the treatment of advanced prostate cancer', *Reviews in urology*, 9 Suppl 1(Suppl 1), pp. S3-S8.
- Petrucelli, N., Daly, M. and Tuya, P. (2016) *BRCA1- and BRCA2-Associated Hereditary Breast and Ovarian Cancer*
- Pharoah, P. D., Day, N. E., Duffy, S., Easton, D. F. and Ponder, B. A. (1997) 'Family history and the risk of breast cancer: a systematic review and meta-analysis', *Int J Cancer*, 71(5), pp. 800-9.
- Phi, L. T. H., Sari, I. N., Yang, Y.-G., Lee, S.-H., Jun, N., Kim, K. S., Lee, Y. K. and Kwon, H. Y. (2018) 'Cancer Stem Cells (CSCs) in Drug Resistance and their Therapeutic Implications in Cancer Treatment', *Stem cells international*, 2018, pp. 5416923-5416923.
- Pilié, P. G., Johnson, A. M., Hanson, K. L., Dayno, M. E., Kapron, A. L., Stoffel, E. M. and Cooney, K. A. (2017) 'Germline genetic variants in men with prostate cancer and one or more additional cancers', *Cancer*, 123(20), pp. 3925-3932.
- Pinder, S. E., Duggan, C., Ellis, I. O., Cuzick, J., Forbes, J. F., Bishop, H., Fentiman, I. S. and George, W. D. (2010) 'A new pathological system for grading DCIS with improved prediction of local recurrence: results from the UKCCCR/ANZ DCIS trial', *Br J Cancer*, 103(1), pp. 94-100.
- Pinton, G., Nilsson, S. and Moro, L. (2018) 'Targeting estrogen receptor beta (ER β) for treatment of ovarian cancer: importance of KDM6B and SIRT1 for ER β expression and functionality', *Oncogenesis*, 7(2), pp. 15-15.
- Plasilova, M. L., Hayse, B., Killelea, B. K., Horowitz, N. R., Chagpar, A. B. and Lannin, D. R. (2016) 'Features of triple-negative breast cancer: Analysis of 38,813 cases from the national cancer database', *Medicine (Baltimore)*, 95(35), pp. e4614.
- Podo, F., Buydens, L. M. C., Degani, H., Hilhorst, R., Klipp, E., Gribbestad, I. S., Van Huffel, S., van Laarhoven, H. W. M., Luts, J., Monleon, D., Postma, G. J., Schneiderhan-Marra, N., Santoro, F., Wouters, H., Russnes, H. G., Sørli, T., Tagliabue, E., Børresen-Dale, A.-L. and Consortium, F. (2010) 'Triple-negative breast cancer: present challenges and new perspectives', *Molecular oncology*, 4(3), pp. 209-229.

- Pokhriyal, R., Hariprasad, R., Kumar, L. and Hariprasad, G. (2019) 'Chemotherapy Resistance in Advanced Ovarian Cancer Patients', *Biomarkers in cancer*, 11, pp. 1179299X19860815-1179299X19860815.
- Pontius, J., Wagner, L. and Schuler, G. 2002. UniGene: A Unified View of the Transcriptome.
- Prakash, R., Zhang, Y., Feng, W. and Jasin, M. (2015) 'Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins', *Cold Spring Harbor perspectives in biology*, 7(4), pp. a016600-a016600.
- Prins, G. S. and Korach, K. S. (2008) 'The role of estrogens and estrogen receptors in normal prostate growth and disease', *Steroids*, 73(3), pp. 233-244.
- Prueitt, R. L., Yi, M., Hudson, R. S., Wallace, T. A., Howe, T. M., Yfantis, H. G., Lee, D. H., Stephens, R. M., Liu, C. G., Calin, G. A., Croce, C. M. and Ambis, S. (2008) 'Expression of microRNAs and protein-coding genes associated with perineural invasion in prostate cancer', *Prostate*, 68(11), pp. 1152-64.
- Pusztai, L., Mazouni, C., Anderson, K., Wu, Y. and Symmans, W. F. (2006) 'Molecular Classification of Breast Cancer: Limitations and Potential', *The Oncologist*, 11(8), pp. 868-877.
- Qin, L., Liu, Z., Chen, H. and Xu, J. (2009) 'The steroid receptor coactivator-1 regulates twist expression and promotes breast cancer metastasis', *Cancer Res*, 69(9), pp. 3819-27.
- Author (2017) *R: A language and environment for statistical computing*. (Version 3.4.0). Available at: <https://www.R-project.org/>.
- Radisky, E. S. and Radisky, D. C. (2010) 'Matrix metalloproteinase-induced epithelial-mesenchymal transition in breast cancer', *Journal of mammary gland biology and neoplasia*, 15(2), pp. 201-212.
- Raffo, A. J., Perlman, H., Chen, M. W., Day, M. L., Streitman, J. S. and Buttyan, R. (1995) 'Overexpression of bcl-2 protects prostate cancer cells from apoptosis in vitro and confers resistance to androgen depletion in vivo', *Cancer Res*, 55(19), pp. 4438-45.
- Rahman, N., Seal, S., Thompson, D., Kelly, P., Renwick, A., Elliott, A., Reid, S., Spanova, K., Barfoot, R., Chagtai, T., Jayatilake, H., McGuffog, L., Hanks, S., Evans, D. G., Eccles, D., Easton, D. F. and Stratton, M. R. (2007) 'PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene', *Nat Genet*, 39(2), pp. 165-7.
- Rakha, E. A., El-Sayed, M. E., Green, A. R., Lee, A. H., Robertson, J. F. and Ellis, I. O. (2007) 'Prognostic markers in triple-negative breast cancer', *Cancer*, 109(1), pp. 25-32.
- Ramalingam, P. (2016) 'Morphologic, Immunophenotypic, and Molecular Features of Epithelial Ovarian Cancer', *Oncology (Williston Park)*, 30(2), pp. 166-76.
- Rao, C. V., Yamada, H. Y., Yao, Y. and Dai, W. (2009) 'Enhanced genomic instabilities caused by deregulated microtubule dynamics and chromosome segregation: a perspective from genetic studies in mice', *Carcinogenesis*, 30(9), pp. 1469-1474.
- Rawla, P. (2019) 'Epidemiology of Prostate Cancer', *World journal of oncology*, 10(2), pp. 63-89.
- Reddy, K. B., Krueger, J. S., Kondapaka, S. B. and Diglio, C. A. (1999) 'Mitogen-activated protein kinase (MAPK) regulates the expression of progelatinase B (MMP-9) in breast epithelial cells', *Int J Cancer*, 82(2), pp. 268-73.

Reed, B. G. and Carr, B. R. (2000) 'The Normal Menstrual Cycle and the Control of Ovulation', in Feingold, K.R., Anawalt, B., Boyce, A., Chrousos, G., de Herder, W.W., Dhatariya, K., Dungan, K., Grossman, A., Hershman, J.M., Hofland, J., Kalra, S., Kaltsas, G., Koch, C., Kopp, P., Korbonits, M., Kovacs, C.S., Kuohung, W., Laferrère, B., McGee, E.A., McLachlan, R., Morley, J.E., New, M., Purnell, J., Sahay, R., Singer, F., Stratakis, C.A., Trence, D.L. and Wilson, D.P. (eds.) *Endotext*. South Dartmouth (MA): MDText.com, Inc.

Copyright © 2000-2021, MDText.com, Inc.

Renwick, A., Thompson, D., Seal, S., Kelly, P., Chagtai, T., Ahmed, M., North, B., Jayatilake, H., Barfoot, R., Spanova, K., McGuffog, L., Evans, D. G., Eccles, D., Easton, D. F., Stratton, M. R. and Rahman, N. (2006) 'ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles', *Nat Genet*, 38(8), pp. 873-5.

Ressler, S., Mlineritsch, B. and Greil, R. (2010) 'Triple Negative Breast Cancer', *an international journal for oncology and haematology professionals*, 3(4), pp. 185-189.

Rexhepi, M., Trajkovska, E., Ismaili, H., Besimi, F. and Rufati, N. (2017) 'Primary Fallopian Tube Carcinoma: A Case Report and Literature Review', *Open access Macedonian journal of medical sciences*, 5(3), pp. 344-348.

Rezaei-Tavirani, M., Rezaei-Taviran, S., Mansouri, M., Rostami-Nejad, M. and Rezaei-Tavirani, M. (2017) 'Protein-Protein Interaction Network Analysis for a Biomarker Panel Related to Human Esophageal Adenocarcinoma', *Asian Pacific journal of cancer prevention : APJCP*, 18(12), pp. 3357-3363.

Ricci, F., Affatato, R., Carrassa, L. and Damia, G. (2018) 'Recent Insights into Mucinous Ovarian Carcinoma', *International journal of molecular sciences*, 19(6), pp. 1569.

Richardson, A. L., Wang, Z. C., De Nicolo, A., Lu, X., Brown, M., Miron, A., Liao, X., Iglehart, J. D., Livingston, D. M. and Ganesan, S. (2006) 'X chromosomal abnormalities in basal-like human breast cancer', *Cancer Cell*, 9(2), pp. 121-32.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. and Smyth, G. K. (2015) 'limma powers differential expression analyses for RNA-sequencing and microarray studies', *Nucleic Acids Research*, 43(7), pp. e47-e47.

Rivera, E. and Gomez, H. (2010) 'Chemotherapy resistance in metastatic breast cancer: the evolving role of ixabepilone', *Breast Cancer Research*, 12(2), pp. S2.

Roberts, N. J., Jiao, Y., Yu, J., Kopelovich, L., Petersen, G. M., Bondy, M. L., Gallinger, S., Schwartz, A. G., Syngal, S., Cote, M. L., Axilbund, J., Schulick, R., Ali, S. Z., Eshleman, J. R., Velculescu, V. E., Goggins, M., Vogelstein, B., Papadopoulos, N., Hruban, R. H., Kinzler, K. W. and Klein, A. P. (2012) 'ATM mutations in patients with hereditary pancreatic cancer', *Cancer Discov*, 2(1), pp. 41-6.

Ross, R. K., Paganini-Hill, A., Wan, P. C. and Pike, M. C. (2000) 'Effect of hormone replacement therapy on breast cancer risk: estrogen versus estrogen plus progestin', *J Natl Cancer Inst*, 92(4), pp. 328-32.

Rostamizadeh, L., Fakhrjou, A., Montazeri, V., Estiar, M. A., Naghavi-Behzad, M., Hosseini, S., Sakhinia, M. and Sakhinia, E. (2013) 'Bcl-2 gene expression in human breast cancers in iran', *Asian Pac J Cancer Prev*, 14(7), pp. 4209-14.

Rothblum-Oviatt, C., Wright, J., Lefton-Greif, M. A., McGrath-Morrow, S. A., Crawford, T. O. and Lederman, H. M. (2016) 'Ataxia telangiectasia: a review', *Orphanet Journal of Rare Diseases*, 11(1), pp. 159.

Roy, S., Chakravarty, D., Cortez, V., De Mukhopadhyay, K., Bandyopadhyay, A., Ahn, J. M., Raj, G. V., Tekmal, R. R., Sun, L. and Vadlamudi, R. K. (2012) 'Significance of PELP1 in ER-negative breast cancer metastasis', *Mol Cancer Res*, 10(1), pp. 25-33.

Royer, C. and Lu, X. (2011) 'Epithelial cell polarity: a major gatekeeper against cancer?', *Cell death and differentiation*, 18(9), pp. 1470-1477.

Rubin, I. and Yarden, Y. (2001) 'The basic biology of HER2', *Ann Oncol*, 12 Suppl 1, pp. S3-8.

Ruiz de Sabando, A., Urrutia Lafuente, E., García-Amigot, F., Alonso Sánchez, A., Morales Garofalo, L., Moreno, S., Ardanaz, E. and Ramos-Arroyo, M. A. (2019) 'Genetic and clinical characterization of BRCA-associated hereditary breast and ovarian cancer in Navarra (Spain)', *BMC Cancer*, 19(1), pp. 1145.

Sabatier, R., Finetti, P., Adelaide, J., Guille, A., Borg, J. P., Chaffanet, M., Lane, L., Birnbaum, D. and Bertucci, F. (2011) 'Down-regulation of ECRG4, a candidate tumor suppressor gene, in human breast cancer', *PLoS One*, 6(11), pp. e27656.

Saha Roy, S. and Vadlamudi, R. K. (2012) 'Role of estrogen receptor signaling in breast cancer metastasis', *Int J Breast Cancer*, 2012, pp. 654698.

Saha, S., Dey, S. and Nath, S. (2021) 'Steroid Hormone Receptors: Links With Cell Cycle Machinery and Breast Cancer Progression', *Frontiers in Oncology*, 11, pp. 460.

Salazar, N., Castellan, M., Shirodkar, S. S. and Lokeshwar, B. L. (2013) 'Chemokines and chemokine receptors as promoters of prostate cancer growth and progression', *Critical reviews in eukaryotic gene expression*, 23(1), pp. 77-91.

Sandoval, N., Platzer, M., Rosenthal, A., Dork, T., Bendix, R., Skawran, B., Stuhmann, M., Wegner, R. D., Sperling, K., Banin, S., Shiloh, Y., Baumer, A., Bernthaler, U., Sennefelder, H., Brohm, M., Weber, B. H. and Schindler, D. (1999) 'Characterization of ATM gene mutations in 66 ataxia telangiectasia families', *Hum Mol Genet*, 8(1), pp. 69-79.

Sanz, D. J., Acedo, A., Infante, M., Durán, M., Pérez-Cabornero, L., Esteban-Cardenosa, E., Lastra, E., Pagani, F., Miner, C. and Velasco, E. A. (2010) 'A High Proportion of DNA Variants of BRCA1 & BRCA2 Is Associated with Aberrant Splicing in Breast/Ovarian Cancer Patients', *Clinical Cancer Research*, 16(6), pp. 1957.

Sardar, H. S., Luczak, V. G., Lopez, M. M., Lister, B. C. and Gilbert, S. P. (2010) 'Mitotic kinesin CENP-E promotes microtubule plus-end elongation', *Current biology : CB*, 20(18), pp. 1648-1653.

Satake, H., Tamura, K., Furihata, M., Anchi, T., Sakoda, H., Kawada, C., Iiyama, T., Ashida, S. and Shuin, T. (2010) 'The ubiquitin-like molecule interferon-stimulated gene 15 is overexpressed in human prostate cancer', *Oncol Rep*, 23(1), pp. 11-6.

Savagner, P. 2001. Leaving the neighborhood: molecular mechanisms involved during epithelial-mesenchymal transition. *Bioessays*.

Sawyer, E. and Roylance, R. and Petridis, C. and Brook, M. N. and Nowinski, S. and Papouli, E. and Fletcher, O. and Pinder, S. and Hanby, A. and Kohut, K. and Gorman, P. and Caneppele, M. and Peto, J. and Dos Santos Silva, I. and Johnson, N. and Swann, R. and Dwek, M. and Perkins, K. A. and Gillett, C. and Houlston, R. and Ross, G. and De Ieso, P. and Southey, M. C. and Hopper, J. L. and Provenzano, E. and Apicella, C. and Wesseling, J. and Cornelissen, S. and Keeman, R. and Fasching, P. A. and Jud, S. M. and Ekici, A. B. and Beckmann, M. W. and Kerin, M. J. and Marme, F. and Schneeweiss, A. and Sohn, C. and Burwinkel, B. and Guenel, P. and Truong, T. and Laurent-Puig, P. and Kerbrat, P. and Bojesen, S. E. and Nordestgaard, B. G. and Nielsen, S. F. and Flyger, H.

and Milne, R. L. and Perez, J. I. and Menendez, P. and Benitez, J. and Brenner, H. and Dieffenbach, A. K. and Arndt, V. and Stegmaier, C. and Meindl, A. and Lichtner, P. and Schmutzler, R. K. and Lochmann, M. and Brauch, H. and Fischer, H. P. and Ko, Y. D. and Nevanlinna, H. and Muranen, T. A. and Aittomaki, K. and Blomqvist, C. and Bogdanova, N. V. and Dork, T. and Lindblom, A. and Margolin, S. and Mannermaa, A. and Kataja, V. and Kosma, V. M. and Hartikainen, J. M. and Chenevix-Trench, G. and Investigators, K. and Lambrechts, D. and Weltens, C. and Van Limbergen, E. and Hatse, S. and Chang-Claude, J. and Rudolph, A. and Seibold, P. and Flesch-Janys, D. and Radice, P. and Peterlongo, P. and Bonanni, B. and Volorio, S. and Giles, G. G. and Severi, G. and Baglietto, L. and McLean, C. A. and Haiman, C. A. and Henderson, B. E. and Schumacher, F. and Le Marchand, L. and Simard, J. and Goldberg, M. S. and Labreche, F. and Dumont, M. and Kristensen, V. and Winqvist, R. and Pylkas, K. and Jukkola-Vuorinen, A. and Kauppila, S. and Andrulis, I. L. and Knight, J. A. and Glendon, G. and Mulligan, A. M. and Devillee, P. and Tollenaar, R. A. and Seynaeve, C. M. and Kriege, M. and Figueroa, J. and Chanock, S. J. and Sherman, M. E. and Hooning, M. J. and Hollestelle, A. and van den Ouweland, A. M. and van Deurzen, C. H. and Li, J. and Czene, K. and Humphreys, K. and Cox, A. and Cross, S. S. and Reed, M. W. and Shah, M. and Jakubowska, A. and Lubinski, J. and Jaworska-Bieniek, K. and Durda, K. and Swerdlow, A. and Ashworth, A. and Orr, N. and Schoemaker, M. and Couch, F. J. and Hallberg, E. and Gonzalez-Neira, A. and Pita, G. and Alonso, M. R. and Tessier, D. C. and Vincent, D. and Bacot, F. and Bolla, M. K. and Wang, Q. and Dennis, J. and Michailidou, K. and Dunning, A. M. and Hall, P. and Easton, D. and Pharoah, P. and Schmidt, M. K. and Tomlinson, I. and Garcia-Closas, M. (2014) 'Genetic predisposition to in situ and invasive lobular carcinoma of the breast', *PLoS Genet*, 10(4), pp. e1004285.

Sboner, A., Demichelis, F., Calza, S., Pawitan, Y., Setlur, S. R., Hoshida, Y., Perner, S., Adami, H. O., Fall, K., Mucci, L. A., Kantoff, P. W., Stampfer, M., Andersson, S. O., Varenhorst, E., Johansson, J. E., Gerstein, M. B., Golub, T. R., Rubin, M. A. and Andren, O. (2010) 'Molecular sampling of prostate cancer: a dilemma for predicting disease progression', *BMC Med Genomics*, 3, pp. 8.

Scarpin, K. M., Graham, J. D., Mote, P. A. and Clarke, C. L. (2009) 'Progesterone action in human tissues: regulation by progesterone receptor (PR) isoform expression, nuclear positioning and coregulator expression', *Nucl Recept Signal*, 7, pp. e009.

Schairer, C., Lubin, J., Troisi, R., Sturgeon, S., Brinton, L. and Hoover, R. (2000) 'Menopausal estrogen and estrogen-progestin replacement therapy and breast cancer risk', *Jama*, 283(4), pp. 485-91.

Schildkraut, J. M., Cooper, G. S., Halabi, S., Calingaert, B., Hartge, P. and Whittemore, A. S. (2001) 'Age at natural menopause and the risk of epithelial ovarian cancer', *Obstet Gynecol*, 98(1), pp. 85-90.

Schnitt, S. J. (2003) 'The diagnosis and management of pre-invasive breast disease: flat epithelial atypia--classification, pathologic features and clinical significance', *Breast Cancer Res*, 5(5), pp. 263-8.

Schoemaker, M. J., Nichols, H. B., Wright, L. B., Brook, M. N., Jones, M. E., O'Brien, K. M., Adami, H. O., Baglietto, L., Bernstein, L., Bertrand, K. A., Boutron-Ruault, M. C., Braaten, T., Chen, Y., Connor, A. E., Dorransoro, M., Dossus, L., Eliassen, A. H., Giles, G. G., Hankinson, S. E., Kaaks, R., Key, T. J., Kirsh, V. A., Kitahara, C. M., Koh, W. P., Larsson, S. C., Linet, M. S., Ma, H., Masala, G., Merritt, M. A., Milne, R. L., Overvad, K., Ozasa, K., Palmer, J. R., Peeters, P. H., Riboli, E., Rohan, T. E., Sadakane, A., Sund, M., Tamimi, R. M., Trichopoulou, A., Ursin, G., Vatten, L., Visvanathan, K., Weiderpass, E., Willett, W. C., Wolk, A., Yuan, J. M., Zeleniuch-Jacquotte, A., Sandler, D. P. and Swerdlow, A. J. (2018) 'Association of Body Mass Index and Age With Subsequent Breast Cancer Risk in Premenopausal Women', *JAMA Oncol*, 4(11), pp. e181771.

Schwab, C. L., English, D. P., Roque, D. M. and Santin, A. D. (2014) 'Taxanes: their impact on gynecologic malignancy', *Anti-cancer drugs*, 25(5), pp. 522-535.

Schwertfeger, K. L., Cowman, M. K., Telmer, P. G., Turley, E. A. and McCarthy, J. B. (2015) 'Hyaluronan, Inflammation, and Breast Cancer Progression', *Frontiers in Immunology*, 6, pp. 236.

Schüler-Toprak, S., Weber, F., Skrzypczak, M., Ortmann, O. and Treeck, O. (2018) 'Estrogen receptor β is associated with expression of cancer associated genes and survival in ovarian cancer', *BMC Cancer*, 18(1), pp. 981.

Shafer, M. E. R., Nguyen, A. H. T., Tremblay, M., Viala, S., Béland, M., Bertos, N. R., Park, M. and Bouchard, M. (2017) 'Lineage Specification from Prostate Progenitor Cells Requires Gata3-Dependent Mitotic Spindle Orientation', *Stem cell reports*, 8(4), pp. 1018-1031.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) 'Cytoscape: a software environment for integrated models of biomolecular interaction networks', *Genome Research*, 13(11), pp. 2498-2504.

Shayesteh, L., Lu, Y., Kuo, W.-L., Baldocchi, R., Godfrey, T., Collins, C., Pinkel, D., Powell, B., Mills, G. B. and Gray, J. W. (1999) 'PIK3CA is implicated as an oncogene in ovarian cancer', *Nature Genetics*, 21(1), pp. 99-102.

Shen, Z., Luo, H., Li, S., Sheng, B., Zhao, M., Zhu, H. and Zhu, X. (2017) 'Correlation between estrogen receptor expression and prognosis in epithelial ovarian cancer: a meta-analysis', *Oncotarget*, 8(37), pp. 62400-62413.

Shi, L. and Campbell, G. and Jones, W. D. and Campagne, F. and Wen, Z. and Walker, S. J. and Su, Z. and Chu, T. M. and Goodsaid, F. M. and Pusztai, L. and Shaughnessy, J. D., Jr. and Oberthuer, A. and Thomas, R. S. and Paules, R. S. and Fielden, M. and Barlogie, B. and Chen, W. and Du, P. and Fischer, M. and Furlanello, C. and Gallas, B. D. and Ge, X. and Megherbi, D. B. and Symmans, W. F. and Wang, M. D. and Zhang, J. and Bitter, H. and Brors, B. and Bushel, P. R. and Bylesjo, M. and Chen, M. and Cheng, J. and Chou, J. and Davison, T. S. and Delorenzi, M. and Deng, Y. and Devanarayan, V. and Dix, D. J. and Dopazo, J. and Dorff, K. C. and Elloumi, F. and Fan, J. and Fan, S. and Fan, X. and Fang, H. and Gonzaludo, N. and Hess, K. R. and Hong, H. and Huan, J. and Irizarry, R. A. and Judson, R. and Juraeva, D. and Lababidi, S. and Lambert, C. G. and Li, L. and Li, Y. and Li, Z. and Lin, S. M. and Liu, G. and Lobenhofer, E. K. and Luo, J. and Luo, W. and McCall, M. N. and Nikolsky, Y. and Pennello, G. A. and Perkins, R. G. and Philip, R. and Popovici, V. and Price, N. D. and Qian, F. and Scherer, A. and Shi, T. and Shi, W. and Sung, J. and Thierry-Mieg, D. and Thierry-Mieg, J. and Thodima, V. and Trygg, J. and Vishnuvajjala, L. and Wang, S. J. and Wu, J. and Wu, Y. and Xie, Q. and Yousef, W. A. and Zhang, L. and Zhang, X. and Zhong, S. and Zhou, Y. and Zhu, S. and Arasappan, D. and Bao, W. and Lucas, A. B. and Berthold, F. and Brennan, R. J. and Bunes, A. and Catalano, J. G. and Chang, C. and Chen, R. and Cheng, Y. and Cui, J. and Czika, W. and Demichelis, F. and Deng, X. and Dosymbekov, D. and Eils, R. and Feng, Y. and Fostel, J. and Fulmer-Smentek, S. and Fuscoe, J. C. and Gatto, L. and Ge, W. and Goldstein, D. R. and Guo, L. and Halbert, D. N. and Han, J. and Harris, S. C. and Hatzis, C. and Herman, D. and Huang, J. and Jensen, R. V. and Jiang, R. and Johnson, C. D. and Jurman, G. and Kahlert, Y. and Khuder, S. A. and Kohl, M. and Li, J. and Li, M. and Li, Q. Z. and Li, S. and Liu, J. and Liu, Y. and Liu, Z. and Meng, L. and Madera, M. and Martinez-Murillo, F. and Medina, I. and Meehan, J. and Miclaus, K. and Moffitt, R. A. and Montaner, D. and Mukherjee, P. and Mulligan, G. J. and Neville, P. and Nikolskaya, T. and Ning, B. and Page, G. P. and Parker, J. and Parry, R. M. and Peng, X. and Peterson, R. L. and Phan, J. H. and Quanz, B. and Ren, Y. and Riccadonna, S. and Roter, A. H. and Samuelson, F. W. and Schumacher, M. M. and Shambaugh, J. D. and Shi, Q. and Shippy, R. and Si, S. and Smalter, A. and Sotiriou, C. and Soukup, M. and Staedtler, F. and Steiner, G. and Stokes, T. H. and Sun, Q. and Tan, P. Y. and Tang, R. and Tezak, Z. and Thorn, B. and Tsyganova, M. and Turpaz, Y. and Vega, S. C. and Visintainer, R. and von Frese, J. and Wang, C. and Wang, E. and Wang, J. and Wang, W. and

- Westermann, F. and Willey, J. C. and Woods, M. and Wu, S. and Xiao, N. and Xu, J. and Xu, L. and Yang, L. and Zeng, X. and Zhang, M. and Zhao, C. and Puri, R. K. and Scherf, U. and Tong, W. and Wolfinger, R. D. (2010) 'The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models', *Nat Biotechnol*, 28(8), pp. 827-38.
- Shigeishi, H., Biddle, A., Gammon, L., Emich, H., Rodini, C. O., Gemenetzidis, E., Fazil, B., Sugiyama, M., Kamata, N. and Mackenzie, I. C. (2013) 'Maintenance of stem cell self-renewal in head and neck cancers requires actions of GSK3 β influenced by CD44 and RHAMM', *STEM CELLS*, 31(10), pp. 2073-2083.
- Shiloh, Y. and Ziv, Y. (2013) 'The ATM protein kinase: regulating the cellular response to genotoxic stress, and more', *Nat Rev Mol Cell Biol*, 14(4), pp. 197-210.
- Shiozawa, Y., Nie, B., Pienta, K. J., Morgan, T. M. and Taichman, R. S. (2013) 'Cancer stem cells and their role in metastasis', *Pharmacology & therapeutics*, 138(2), pp. 285-293.
- Sieben, N. L., Macropoulos, P., Roemen, G. M., Kolkman-Uljee, S. M., Jan Fleuren, G., Houmadi, R., Diss, T., Warren, B., Al Adnani, M., De Goeij, A. P., Krausz, T. and Flanagan, A. M. (2004) 'In ovarian neoplasms, BRAF, but not KRAS, mutations are restricted to low-grade serous tumours', *J Pathol*, 202(3), pp. 336-40.
- Signoretti, S. and Loda, M. (2001) 'Estrogen receptor beta in prostate cancer: brake pedal or accelerator?', *The American journal of pathology*, 159(1), pp. 13-16.
- Sikirzhytski, V., Renda, F., Tikhonenko, I., Magidson, V., McEwen, B. F. and Khodjakov, A. (2018) 'Microtubules assemble near most kinetochores during early prometaphase in human cells', *Journal of Cell Biology*, 217(8), pp. 2647-2659.
- Singletary, K. W. and Gapstur, S. M. (2001) 'Alcohol and breast cancer: review of epidemiologic and experimental evidence and potential mechanisms', *Jama*, 286(17), pp. 2143-51.
- Skliris, G. P., Munot, K., Bell, S. M., Carder, P. J., Lane, S., Horgan, K., Lansdown, M. R., Parkes, A. T., Hanby, A. M., Markham, A. F. and Speirs, V. (2003) 'Reduced expression of oestrogen receptor beta in invasive breast cancer and its re-expression using DNA methyl transferase inhibitors in a cell line model', *J Pathol*, 201(2), pp. 213-20.
- Southey, M. C., Winship, I. and Nguyen-Dumont, T. (2016) 'PALB2: research reaching to clinical outcomes for women with breast cancer', *Hereditary cancer in clinical practice*, 14, pp. 9-9.
- Stacklies, W., Redestig, H. and Wright, K. 2018. pcaMethods – a Bioconductor package providing PCA methods for incomplete data.
<https://www.bioconductor.org/packages/release/bioc/html/pcaMethods.html>: Bioconductor.
- Stanbrough, M., Buble, G. J., Ross, K., Golub, T. R., Rubin, M. A., Penning, T. M., Febbo, P. G. and Balk, S. P. (2006) 'Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer', *Cancer Res*, 66(5), pp. 2815-25.
- Stark, T., Livas, L. and Kyprianou, N. (2015) 'Inflammation in prostate cancer progression and therapeutic targeting', *Translational andrology and urology*, 4(4), pp. 455-463.
- Steffen, D., Paul, T. S., Ewan, B. and Wolfgang, H. (2009) 'Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt', *Nature Protocols*, 4(8), pp. 1184.
- Stekhoven, D. J. and Bühlmann, P. (2012) 'MissForest—non-parametric missing value imputation for mixed-type data', *Bioinformatics*, 28(1), pp. 112-118.

- Stemke-Hale, K., Gonzalez-Angulo, A. M., Lluch, A., Neve, R. M., Kuo, W. L., Davies, M., Carey, M., Hu, Z., Guan, Y., Sahin, A., Symmans, W. F., Pusztai, L., Nolden, L. K., Horlings, H., Berns, K., Hung, M. C., van de Vijver, M. J., Valero, V., Gray, J. W., Bernardis, R., Mills, G. B. and Hennessy, B. T. (2008) 'An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer', *Cancer Res*, 68(15), pp. 6084-91.
- Stinson, S., Lackner, M. R., Adai, A. T., Yu, N., Kim, H. J., O'Brien, C., Spoerke, J., Jhunjunwala, S., Boyd, Z., Januario, T., Newman, R. J., Yue, P., Bourgon, R., Modrusan, Z., Stern, H. M., Warming, S., de Sauvage, F. J., Amler, L., Yeh, R. F. and Dornan, D. (2011) 'TRPS1 targeting by miR-221/222 promotes the epithelial-to-mesenchymal transition in breast cancer', *Sci Signal*, 4(177), pp. ra41.
- Storz, P. (2005) 'Reactive oxygen species in tumor progression', *Front Biosci*, 10, pp. 1881-96.
- Stretch, C., Khan, S., Asgarian, N., Eisner, R., Vaisipour, S., Damaraju, S., Graham, K., Bathe, O. F., Steed, H., Greiner, R. and Baracos, V. E. (2013) 'Effects of sample size on differential gene expression, rank order and prediction accuracy of a gene signature', *PloS one*, 8(6), pp. e65380-e65380.
- Stumpff, J., Ghule, P. N., Shimamura, A., Stein, J. L. and Greenblatt, M. (2014) 'Spindle microtubule dysfunction and cancer predisposition', *Journal of cellular physiology*, 229(12), pp. 1881-1883.
- Sun, J., Zhang, K., Cai, Z., Li, K., Zhao, C., Fan, C. and Wang, J. (2019) 'Identification of critical pathways and hub genes in TP53 mutation prostate cancer by bioinformatics analysis', *Biomarkers in Medicine*, 13(10), pp. 831-840.
- Sun, L., Carr, A. L., Li, P., Lee, J., McGregor, M. and Li, L. (2014) 'Characterization of the human oncogene SCL/TAL1 interrupting locus (Stil) mediated Sonic hedgehog (Shh) signaling transduction in proliferating mammalian dopaminergic neurons', *Biochemical and Biophysical Research Communications*, 449(4), pp. 444-448.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F. (2021) 'Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries', *CA Cancer J Clin*, 71(3), pp. 209-249.
- Symmans, W. F., Hatzis, C., Sotiriou, C., Andre, F., Peintinger, F., Regitnig, P., Daxenbichler, G., Desmedt, C., Domont, J., Marth, C., Delalogue, S., Bauernhofer, T., Valero, V., Booser, D. J., Hortobagyi, G. N. and Pusztai, L. (2010) 'Genomic index of sensitivity to endocrine therapy for breast cancer', *J Clin Oncol*, 28(27), pp. 4111-9.
- Takeshita, T., Asaoka, M., Katsuta, E., Photiadis, S. J., Narayanan, S., Yan, L. and Takabe, K. (2019) 'High expression of polo-like kinase 1 is associated with TP53 inactivation, DNA repair deficiency, and worse prognosis in ER positive Her2 negative breast cancer', *American journal of translational research*, 11(10), pp. 6507-6521.
- Tan, A. S., Yeong, J. P. S., Lai, C. P. T., Ong, C. H. C., Lee, B., Lim, J. C. T., Thike, A. A., Iqbal, J., Dent, R. A., Lim, E. H. and Tan, P. H. (2019) 'The role of Ki-67 in Asian triple negative breast cancers: a novel combinatory panel approach', *Virchows Arch*, 475(6), pp. 709-725.
- Tan, H., Yi, L., Rote, N. S., Hurd, W. W. and Mesiano, S. (2012) 'Progesterone receptor-A and -B have opposite effects on proinflammatory gene expression in human myometrial cells: implications for progesterone actions in human pregnancy and parturition', *J Clin Endocrinol Metab*, 97(5), pp. E719-30.
- Tan, M. and Yu, D. (2007) 'Molecular mechanisms of erbB2-mediated breast cancer chemoresistance', *Adv Exp Med Biol*, 608, pp. 119-29.

Tan, M. H. E., Li, J., Xu, H. E., Melcher, K. and Yong, E.-I. (2015) 'Androgen receptor: structure, role in prostate cancer and drug discovery', *Acta pharmacologica Sinica*, 36(1), pp. 3-23.

Tang, C. J., Lin, S. Y., Hsu, W. B., Lin, Y. N., Wu, C. T., Lin, Y. C., Chang, C. W., Wu, K. S. and Tang, T. K. (2011) 'The human microcephaly protein STIL interacts with CPAP and is required for procentriole formation', *Embo j*, 30(23), pp. 4790-804.

Tang, W., Li, J., Chang, X., Jia, L., Tang, Q., Wang, Y., Zheng, Y., Sun, L. and Feng, Z. (2020) 'Construction of a novel prognostic-predicting model correlated to ovarian cancer', *Bioscience reports*, 40(8), pp. BSR20201261.

Tarhriz, V., Bandehpour, M., Dastmalchi, S., Ouladsahebmadarek, E., Zarredar, H. and Eyvazi, S. (2019) 'Overview of CD24 as a new molecular marker in ovarian cancer', *J Cell Physiol*, 234(3), pp. 2134-2142.

Thakkar, J. P., McCarthy, B. J. and Villano, J. L. (2014) 'Age-specific cancer incidence rates increase through the oldest age groups', *Am J Med Sci*, 348(1), pp. 65-70.

The Cancer Genome Atlas Network and Koboldt, D. C. and Fulton, R. S. and McLellan, M. D. and Schmidt, H. and Kalicki-Veizer, J. and McMichael, J. F. and Fulton, L. L. and Dooling, D. J. and Ding, L. and Mardis, E. R. and Wilson, R. K. and Ally, A. and Balasundaram, M. and Butterfield, Y. S. N. and Carlsen, R. and Carter, C. and Chu, A. and Chuah, E. and Chun, H.-J. E. and Coope, R. J. N. and Dhalla, N. and Guin, R. and Hirst, C. and Hirst, M. and Holt, R. A. and Lee, D. and Li, H. I. and Mayo, M. and Moore, R. A. and Mungall, A. J. and Pleasance, E. and Gordon Robertson, A. and Schein, J. E. and Shafiei, A. and Sipahimalani, P. and Slobodan, J. R. and Stoll, D. and Tam, A. and Thiessen, N. and Varhol, R. J. and Wye, N. and Zeng, T. and Zhao, Y. and Birol, I. and Jones, S. J. M. and Marra, M. A. and Cherniack, A. D. and Saksena, G. and Onofrio, R. C. and Pho, N. H. and Carter, S. L. and Schumacher, S. E. and Tabak, B. and Hernandez, B. and Gentry, J. and Nguyen, H. and Crenshaw, A. and Ardlie, K. and Beroukhim, R. and Winckler, W. and Getz, G. and Gabriel, S. B. and Meyerson, M. and Chin, L. and Park, P. J. and Kucherlapati, R. and Hoadley, K. A. and Todd Auman, J. and Fan, C. and Turman, Y. J. and Shi, Y. and Li, L. and Topal, M. D. and He, X. and Chao, H.-H. and Prat, A. and Silva, G. O. and Iglesia, M. D. and Zhao, W. and Usary, J. and Berg, J. S. and Adams, M. and Booker, J. and Wu, J. and Gulabani, A. and Bodenheimer, T. and Hoyle, A. P. and Simons, J. V. and Soloway, M. G. and Mose, L. E. and Jefferys, S. R. and Balu, S. and Parker, J. S. and Neil Hayes, D. and Perou, C. M. and Malik, S. and Mahurkar, S. and Shen, H. and Weisenberger, D. J. and Triche Jr, T. and Lai, P. H. and Bootwalla, M. S. and Maglinte, D. T. and Berman, B. P. and Van Den Berg, D. J. and Baylin, S. B. and Laird, P. W. and Creighton, C. J. and Donehower, L. A. and Noble, M. and Voet, D. and Gehlenborg, N. and DiCara, D. and Zhang, J. and Zhang, H. and Wu, C.-J. and Yingchun Liu, S. and Lawrence, M. S. and Zou, L. and Sivachenko, A. and Lin, P. and Stojanov, P. and Jing, R. and Cho, J. and Sinha, R. and Park, R. W. and Nazaire, M.-D. and Robinson, J. and Thorvaldsdottir, H. and Mesirov, J. and Reynolds, S. and Kreisberg, R. B. and Bernard, B. and Bressler, R. and Erkkila, T. and Lin, J. and Thorsson, V. and Zhang, W. and Shmulevich, I. and Ciriello, G. and Weinhold, N. and Schultz, N. and Gao, J. and Cerami, E. and Gross, B. and Jacobsen, A. and Sinha, R. and Arman Aksoy, B. and Antipin, Y. and Reva, B. and Shen, R. and Taylor, B. S. and Ladanyi, M. and Sander, C. and Anur, P. and Spellman, P. T. and Lu, Y. and Liu, W. and Verhaak, R. R. G. and Mills, G. B. and Akbani, R. and Zhang, N. and Broom, B. M. and Casasent, T. D. and Wakefield, C. and Unruh, A. K. and Baggerly, K. and Coombes, K. and Weinstein, J. N. and Haussler, D. and Benz, C. C. and Stuart, J. M. and Benz, S. C. and Zhu, J. and Szeto, C. C. and Scott, G. K. and Yau, C. and Paull, E. O. and Carlin, D. and Wong, C. and Sokolov, A. and Thusberg, J. and Mooney, S. and Ng, S. and Goldstein, T. C. and Ellrott, K. and Grifford, M. and Wilks, C. and Ma, S. and Craft, B. and Yan, C. and Hu, Y. and Meerzaman, D. and Gastier-Foster, J. M. and Bowen, J. and Ramirez, N. C. and Black, A. D. and Pyatt, R. E. and White, P. and Zmuda, E. J. and Frick, J. and Lichtenberg, T. M. and Brookens, R. and George, M. M. and Gerken, M. A. and Harper, H. A. and Leraas, K. M. and Wise, L. J. and Tabler, T. R. and McAllister, C. and Barr, T. and Hart-Kothari, M. and Tarvin, K. and Saller, C. and Sandusky, G. and Mitchell, C. and Iacocca, M. V.

and Brown, J. and Rabeno, B. and Czerwinski, C. and Petrelli, N. and Dolzhansky, O. and Abramov, M. and Voronina, O. and Potapova, O. and Marks, J. R. and Suchorska, W. M. and Murawa, D. and Kycler, W. and Ibbs, M. and Korski, K. and Spychała, A. and Murawa, P. and Brzeziński, J. J. and Perz, H. and Łażniak, R. and Teresiak, M. and Tatka, H. and Leporowska, E. and Bogusz-Czerniewicz, M. and Malicki, J. and Mackiewicz, A. and Wiznerowicz, M. and Van Le, X. and Kohl, B. and Viet Tien, N. and Thorp, R. and Van Bang, N. and Sussman, H. and Duc Phu, B. and Hajek, R. and Phi Hung, N. and Viet The Phuong, T. and Quyet Thang, H. and Zaki Khan, K. and Penny, R. and Mallery, D. and Curley, E. and Shelton, C. and Yena, P. and Ingle, J. N. and Couch, F. J. and Lingle, W. L. and King, T. A. and Maria Gonzalez-Angulo, A. and Dyer, M. D. and Liu, S. and Meng, X. and Patangan, M. and Genome sequencing centres: Washington University in St. L. and Genome characterization centres, B. C. C. A. and Broad, I. and Brigham and Women's, H. and Harvard Medical, S. and University of North Carolina, C. H. and University of Southern California/Johns, H. and Genome data analysis: Baylor College of, M. and Institute for Systems, B. and Memorial Sloan-Kettering Cancer, C. and Oregon, H. and Science, U. and The University of Texas, M. D. A. C. C. and University of California, S. C. B. I. and Nci and Biospecimen core resource: Nationwide Children's Hospital Biospecimen Core, R. and Tissue source sites, A.-I. and Christiana and Careline and Duke University Medical, C. and The Greater Poland Cancer, C. and ILSbio and International Genomics, C. and Mayo, C. and Mskcc and Center, M. D. A. C. (2012) 'Comprehensive molecular portraits of human breast tumours', *Nature*, 490(7418), pp. 61-70.

Thermo Fisher Scientific (2017). <https://www.affymetrix.com/site/mainPage.affx>: Affymetrix (Accessed: September 2017).

Therneau, T. 2020. A Package for Survival Analysis in R.

The Gene Ontology Consortium (2018) 'The Gene Ontology Resource: 20 years and still GOing strong', *Nucleic Acids Research*, 47(D1), pp. D330-D338.

Tholander, B., Koliadi, A., Botling, J., Dahlstrand, H., Von Heideman, A., Ahlström, H., Öberg, K. and Ullenhag, G. J. (2020) 'Complete response with combined BRAF and MEK inhibition in BRAF mutated advanced low-grade serous ovarian carcinoma', *Upsala journal of medical sciences*, 125(4), pp. 325-329.

Thompson, D., Duedal, S., Kirner, J., McGuffog, L., Last, J., Reiman, A., Byrd, P., Taylor, M. and Easton, D. F. (2005) 'Cancer risks and mortality in heterozygous ATM mutation carriers', *J Natl Cancer Inst*, 97(11), pp. 813-22.

Thorstenson, Y. R., Roxas, A., Kroiss, R., Jenkins, M. A., Yu, K. M., Bachrich, T., Muhr, D., Wayne, T. L., Chu, G., Davis, R. W., Wagner, T. M. U. and Oefner, P. J. (2003) 'Contributions of **ATM** Mutations to Familial Breast and Ovarian Cancer', *Cancer Research*, 63(12), pp. 3325.

Tischkowitz, M. and Xia, B. (2010) 'PALB2/FANCN: recombining cancer and Fanconi anemia', *Cancer Res*, 70(19), pp. 7353-9.

Toft, D. J. and Cryns, V. L. (2011) 'Minireview: Basal-like breast cancer: from molecular profiles to targeted therapies', *Molecular endocrinology (Baltimore, Md.)*, 25(2), pp. 199-211.

Tolg, C., Hamilton, S. R., Morningstar, L., Zhang, J., Zhang, S., Esguerra, K. V., Telmer, P. G., Luyt, L. G., Harrison, R., McCarthy, J. B. and Turley, E. A. (2010) 'RHAMM promotes interphase microtubule instability and mitotic spindle integrity through MEK1/ERK1/2 activity', *J Biol Chem*, 285(34), pp. 26461-74.

- Tolg, C., McCarthy, J. B., Yazdani, A. and Turley, E. A. (2014) 'Hyaluronan and RHAMM in wound repair and the "cancerization" of stromal tissues', *BioMed research international*, 2014, pp. 103923-103923.
- Tone, A. A., Begley, H., Sharma, M., Murphy, J., Rosen, B., Brown, T. J. and Shaw, P. A. (2008) 'Gene expression profiles of luteal phase fallopian tube epithelium from BRCA mutation carriers resemble high-grade serous carcinoma', *Clin Cancer Res*, 14(13), pp. 4067-78.
- Tothill, R. W., Tinker, A. V., George, J., Brown, R., Fox, S. B., Lade, S., Johnson, D. S., Trivett, M. K., Etemadmoghadam, D., Locandro, B., Traficante, N., Fereday, S., Hung, J. A., Chiew, Y. E., Haviv, I., Gertig, D., DeFazio, A. and Bowtell, D. D. (2008) 'Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome', *Clin Cancer Res*, 14(16), pp. 5198-208.
- Tozbikian, G., Brogi, E., Vallejo, C. E., Giri, D., Murray, M., Catalano, J., Olcese, C., Van Zee, K. J. and Wen, H. Y. (2017) 'Atypical Ductal Hyperplasia Bordering on Ductal Carcinoma In Situ', *International journal of surgical pathology*, 25(2), pp. 100-107.
- Tram, E., Savas, S. and Ozcelik, H. (2013) 'Missense variants of uncertain significance (VUS) altering the phosphorylation patterns of BRCA1 and BRCA2', *PLoS One*, 8(5), pp. e62468.
- Tretiakova, M. S., Wei, W., Boyer, H. D., Newcomb, L. F., Hawley, S., Auman, H., Vakar-Lopez, F., McKenney, J. K., Fazli, L., Simko, J., Troyer, D. A., Hurtado-Coll, A., Thompson, I. M., Jr., Carroll, P. R., Ellis, W. J., Gleave, M. E., Nelson, P. S., Lin, D. W., True, L. D., Feng, Z. and Brooks, J. D. (2016) 'Prognostic value of Ki67 in localized prostate carcinoma: a multi-institutional study of >1000 prostatectomies', *Prostate cancer and prostatic diseases*, 19(3), pp. 264-270.
- Tripathi, A., King, C., de la Morenas, A., Perry, V. K., Burke, B., Antoine, G. A., Hirsch, E. F., Kavanah, M., Mendez, J., Stone, M., Gerry, N. P., Lenburg, M. E. and Rosenberg, C. L. (2008) 'Gene expression abnormalities in histologically normal breast epithelium of breast cancer patients', *Int J Cancer*, 122(7), pp. 1557-66.
- Truong, T. H. and Lange, C. A. (2018) 'Deciphering Steroid Receptor Crosstalk in Hormone-Driven Cancers', *Endocrinology*, 159(12), pp. 3897-3907.
- Tsaousis, G. N., Papadopoulou, E., Apeessos, A., Agiannitopoulos, K., Pepe, G., Kampouri, S., Diamantopoulos, N., Floros, T., Iosifidou, R., Katopodi, O., Koumariou, A., Markopoulos, C., Papazisis, K., Venizelos, V., Xanthakis, I., Xepapadakis, G., Banu, E., Eniu, D. T., Negru, S., Stanculeanu, D. L., Ungureanu, A., Ozmen, V., Tansan, S., Tekinel, M., Yalcin, S. and Nasioulas, G. (2019) 'Analysis of hereditary cancer syndromes by using a panel of genes: novel and multiple pathogenic mutations', *BMC Cancer*, 19(1), pp. 535.
- Tung, C. S., Mok, S. C., Tsang, Y. T., Zu, Z., Song, H., Liu, J., Deavers, M. T., Malpica, A., Wolf, J. K., Lu, K. H., Gershenson, D. M. and Wong, K. K. (2009) 'PAX2 expression in low malignant potential ovarian tumors and low-grade ovarian serous carcinomas', *Mod Pathol*, 22(9), pp. 1243-50.
- Turashvili, G., Bouchal, J., Baumforth, K., Wei, W., Dziechciarkova, M., Ehrmann, J., Klein, J., Fridman, E., Skarda, J., Srovnal, J., Hajduch, M., Murray, P. and Kolar, Z. (2007) 'Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis', *BMC Cancer*, 7, pp. 55-55.
- Turley, R. S., Finger, E. C., Hempel, N., How, T., Fields, T. A. and Blobel, G. C. (2007) 'The type III transforming growth factor-beta receptor as a novel tumor suppressor gene in prostate cancer', *Cancer Res*, 67(3), pp. 1090-8.

- Tzikas, A.-K., Nemes, S. and Linderholm, B. K. (2020) 'A comparison between young and old patients with triple-negative breast cancer: biology, survival and metastatic patterns', *Breast cancer research and treatment*, 182(3), pp. 643-654.
- Valero-Mora, P. M. (2010) 'ggplot2: Elegant Graphics for Data Analysis', *Journal of Statistical Software*, 35(Book Review 1).
- van Diest, P. J., van Der Wall, E. and Baak, J. P. A. (2004) 'Prognostic value of proliferation in invasive breast cancer: a review', *Journal of Clinical Pathology*, 57(7), pp. 675.
- Vara, J. Á. F., Casado, E., de Castro, J., Cejas, P., Belda-Iniesta, C. and González-Barón, M. (2004) 'PI3K/Akt signalling pathway and cancer', *Cancer Treatment Reviews*, 30(2), pp. 193-204.
- Varadarajan, R. and Rusan, N. M. (2018) 'Bridging centrioles and PCM in proper space and time', *Essays in biochemistry*, 62(6), pp. 793-801.
- Vegran, F., Rebucci, M., Chevrier, S., Cadouot, M., Boidot, R. and Lizard-Nacol, S. (2013) 'Only missense mutations affecting the DNA binding domain of p53 influence outcomes in patients with breast carcinoma', *PLoS One*, 8(1), pp. e55103.
- Vercellini, P., Parazzini, F., Bolis, G., Carinelli, S., Dindelli, M., Vendola, N., Luchini, L. and Crosignani, P. G. (1993) 'Endometriosis and ovarian cancer', *American Journal of Obstetrics and Gynecology*, 169(1), pp. 181-182.
- Vesuna, F., Bergman, Y. and Raman, V. (2017) 'Genomic pathways modulated by Twist in breast cancer', *BMC Cancer*, 17(1), pp. 52.
- Vidal, A. C., Howard, L. E., Moreira, D. M., Castro-Santamaria, R., Andriole, G. L., Jr. and Freedland, S. J. (2014) 'Obesity increases the risk for high-grade prostate cancer: results from the REDUCE study', *Cancer Epidemiol Biomarkers Prev*, 23(12), pp. 2936-42.
- Vrtacnik, P., Ostanek, B., Mencej-Bedrac, S. and Marc, J. (2014) 'The many faces of estrogen signaling', *Biochem Med (Zagreb)*, 24(3), pp. 329-42.
- Vu, T. and Claret, F. X. (2012) 'Trastuzumab: updated mechanisms of action and resistance in breast cancer', *Front Oncol*, 2, pp. 62.
- Vulprecht, J., David, A., Tibelius, A., Castiel, A., Konotop, G., Liu, F., Bestvater, F., Raab, M. S., Zentgraf, H., Izraeli, S. and Krämer, A. (2012) 'STIL is required for centriole duplication in human cells', *Journal of Cell Science*, 125(5), pp. 1353.
- Wade, C. A. and Kyprianou, N. (2019) 'Adipose tissue: enabler of prostate cancer aggressive behavior', *Translational andrology and urology*, 8(Suppl 3), pp. S242-S245.
- Walczak, C. E. and Heald, R. (2008) 'Mechanisms of mitotic spindle assembly and function', *Int Rev Cytol*, 265, pp. 111-58.
- Wang, C., Thor, A. D., Moore, D. H., Zhao, Y., Kerschmann, R., Stern, R., Watson, P. H. and Turley, E. A. (1998) 'The overexpression of RHAMM, a hyaluronan-binding protein that regulates ras signaling, correlates with overexpression of mitogen-activated protein kinase and is a significant parameter in breast cancer progression', *Clinical Cancer Research*, 4(3), pp. 567.
- Wang, D.-Y., Done, S. J., Mc Cready, D. R. and Leong, W. L. (2014) 'Validation of the prognostic gene portfolio, ClinicoMolecular Triad Classification, using an independent prospective breast cancer cohort and external patient populations', *Breast Cancer Research*, 16(4).

- Wang, J., Zhang, Y., Dou, Z., Jiang, H., Wang, Y., Gao, X. and Xin, X. (2019a) 'Knockdown of STIL suppresses the progression of gastric cancer by down-regulating the IGF-1/PI3K/AKT pathway', *J Cell Mol Med*, 23(8), pp. 5566-5575.
- Wang, X., Ghareeb, W. M., Zhang, Y., Yu, Q., Lu, X., Huang, Y., Huang, S., Sun, Y. and Chi, P. (2019b) 'Hypermethylated and downregulated MEIS2 are involved in stemness properties and oxaliplatin-based chemotherapy resistance of colorectal cancer', *Journal of Cellular Physiology*, 234(10), pp. 18180-18191.
- Wang, X., Lee, S. O., Xia, S., Jiang, Q., Luo, J., Li, L., Yeh, S. and Chang, C. (2013) 'Endothelial cells enhance prostate cancer metastasis via IL-6-->androgen receptor-->TGF-beta-->MMP-9 signals', *Mol Cancer Ther*, 12(6), pp. 1026-37.
- Wang, Y., Dai, B. and Ye, D. (2015) 'CHEK2 mutation and risk of prostate cancer: a systematic review and meta-analysis', *International journal of clinical and experimental medicine*, 8(9), pp. 15708-15715.
- Warnat, P., Eils, R. and Brors, B. (2005) 'Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes', *BMC bioinformatics*, 6, pp. 265-265.
- Weaver, B. A., Silk, A. D., Putkey, F. R. and Cleveland, D. W. (2006) 'Aneuploidy and chromosomal instability caused by CENP-E heterozygosity contribute to tumorigenicity', *Cancer Research*, 66(8 Supplement), pp. 1325.
- Weaver, B. A. A., Bonday, Z. Q., Putkey, F. R., Kops, G. J. P. L., Silk, A. D. and Cleveland, D. W. (2003) 'Centromere-associated protein-E is essential for the mammalian mitotic checkpoint to prevent aneuploidy due to single chromosome loss', *The Journal of cell biology*, 162(4), pp. 551-563.
- Winter, R. N., Kramer, A., Borkowski, A. and Kyprianou, N. (2001) 'Loss of caspase-1 and caspase-3 protein expression in human prostate cancer', *Cancer Res*, 61(3), pp. 1227-32.
- Winter, W. E., Maxwell, G. L., Tian, C., Carlson, J. W., Ozols, R. F., Rose, P. G., Markman, M., Armstrong, D. K., Muggia, F. and McGuire, W. P. (2007) 'Prognostic Factors for Stage III Epithelial Ovarian Cancer: A Gynecologic Oncology Group Study', *Journal of Clinical Oncology*, 25(24), pp. 3621-3627.
- Wokołorczyk, D., Kluźniak, W., Stempa, K., Rusak, B., Huzarski, T., Gronwald, J., Gliniewicz, K., Kashyap, A., Morawska, S., Dębniak, T., Jakubowska, A., Szwiec, M., Domagała, P., Lubiński, J., Narod, S. A., Akbari, M. R., Cybulski, C., Masojć, B., Gołąb, A., Gliniewicz, B., Sikorski, A., Stojewski, M., Światała, J., Borkowski, T., Borkowski, A., Antczak, A., Wojnar, Ł., Przybyła, J., Sosnowski, M., Małkiewicz, B., Zdrojowy, R., Sikorska-Radek, P., Matych, J., Wilkosz, J., Różański, W., Kiś, J., Bar, K., Bryniarski, P., Paradysz, A., Jersak, K., Niemirowicz, J., Słupski, P., Jarzemski, P., Skrzypczyk, M., Dobruch, J., Puszyński, M., Soczawa, M., Kordowski, M., Życzkowski, M., Borówka, A., Bagińska, J., Krajka, K., Stawicka, M., Haus, O., Janiszewska, H., Stembalska, A., Sąsiadek, M. M. and the Polish Hereditary Prostate Cancer, C. (2021) 'PALB2 mutations and prostate cancer risk and survival', *British Journal of Cancer*, 125(4), pp. 569-575.
- Wolf, B. B., Schuler, M., Echeverri, F. and Green, D. R. (1999) 'Caspase-3 is the primary activator of apoptotic DNA fragmentation via DNA fragmentation factor-45/inhibitor of caspase-activated DNase inactivation', *J Biol Chem*, 274(43), pp. 30651-6.
- Wong, E. (2012) *Breast cancer pathogenesis and histologic vs molecular subtypes*: Eric Wong. Available at: <http://www.pathophys.org/breast-cancer/> (Accessed: 09/12/2016 2016).

- Wong-Brown, M. W., Avery-Kiejda, K. A., Bowden, N. A. and Scott, R. J. (2014) 'Low prevalence of germline PALB2 mutations in Australian triple-negative breast cancer', *Int J Cancer*, 134(2), pp. 301-5.
- Wood, K. W., Lad, L., Luo, L., Qian, X., Knight, S. D., Nevins, N., Brejc, K., Sutton, D., Gilmartin, A. G., Chua, P. R., Desai, R., Schauer, S. P., McNulty, D. E., Annan, R. S., Belmont, L. D., Garcia, C., Lee, Y., Diamond, M. A., Faucette, L. F., Giardiniera, M., Zhang, S., Sun, C.-M., Vidal, J. D., Lichtsteiner, S., Cornwell, W. D., Greshock, J. D., Wooster, R. F., Finer, J. T., Copeland, R. A., Huang, P. S., Morgans, D. J., Jr., Dhanak, D., Bergnes, G., Sakowicz, R. and Jackson, J. R. (2010) 'Antitumor activity of an allosteric inhibitor of centromere-associated protein-E', *Proceedings of the National Academy of Sciences of the United States of America*, 107(13), pp. 5839-5844.
- Wu, S., Zhou, J., Zhang, K., Chen, H., Luo, M., Lu, Y., Sun, Y. and Chen, Y. (2020) 'Molecular Mechanisms of PALB2 Function and Its Role in Breast Cancer Management', *Frontiers in Oncology*, 10, pp. 301.
- Wu, X., Dong, X., Liu, W. and Chen, J. (2006) 'Characterization of CHEK2 mutations in prostate cancer', *Hum Mutat*, 27(8), pp. 742-7.
- Wu, Y. H., Chang, T. H., Huang, Y. F., Huang, H. D. and Chou, C. Y. (2014) 'COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer', *Oncogene*, 33(26), pp. 3432-40.
- Xiao, H., Chen, P., Zeng, G., Xu, D., Wang, X. and Zhang, X. (2019) 'Three novel hub genes and their clinical significance in clear cell renal cell carcinoma', *Journal of Cancer*, 10(27), pp. 6779-6791.
- Xiong, D.-D., Zeng, C.-M., Jiang, L., Luo, D.-Z. and Chen, G. (2019) 'Ki-67/MKI67 as a Predictive Biomarker for Clinical Outcome in Gastric Cancer Patients: an Updated Meta-analysis and Systematic Review involving 53 Studies and 7078 Patients', *Journal of Cancer*, 10(22), pp. 5339-5354.
- Xu, H., Wu, K., Tian, Y., Liu, Q., Han, N., Yuan, X., Zhang, L. and Wu, G. S. (2016) 'CD44 correlates with clinicopathological characteristics and is upregulated by EGFR in breast cancer', *Int J Oncol*, 49(4), pp. 1343-50.
- Xu, J., Lamouille, S. and Derynck, R. (2009) 'TGF-beta-induced epithelial to mesenchymal transition', *Cell Res*, 19(2), pp. 156-72.
- Xu, J. H., Wang, Y. and Xu, D. (2019) 'CKS2 promotes tumor progression and metastasis and is an independent predictor of poor prognosis in epithelial ovarian cancer', *Eur Rev Med Pharmacol Sci*, 23(8), pp. 3225-3234.
- Xu, L., Geman, D. and Winslow, R. L. (2007) 'Large-scale integration of cancer microarray data identifies a robust common cancer signature', *BMC Bioinformatics*, 8(1), pp. 275.
- Xu, L., Tan, A. C., Winslow, R. L. and Geman, D. (2008) 'Merging microarray data from separate breast cancer studies provides a robust prognostic test', *BMC bioinformatics*, 9, pp. 125-125.
- Xue, M., Zhang, K., Mu, K., Xu, J., Yang, H., Liu, Y., Wang, B., Wang, Z., Li, Z., Kong, Q., Li, X., Wang, H., Zhu, J. and Zhuang, T. (2019) 'Regulation of estrogen signaling and breast cancer proliferation by an ubiquitin ligase TRIM56', *Oncogenesis*, 8(5), pp. 30.
- Yalaza, M., İnan, A. and Bozer, M. (2016) 'Male Breast Cancer', *The journal of breast health*, 12(1), pp. 1-8.

- Yan, X., Liu, X.-P., Guo, Z.-X., Liu, T.-Z. and Li, S. (2019) 'Identification of Hub Genes Associated With Progression and Prognosis in Patients With Bladder Cancer', *Frontiers in Genetics*, 10.
- Yancik, R., Wesley, M. N., Ries, L. A., Havlik, R. J., Edwards, B. K. and Yates, J. W. (2001) 'Effect of age and comorbidity in postmenopausal breast cancer patients aged 55 years and older', *Jama*, 285(7), pp. 885-92.
- Yang, H., Hu, L., Liu, Z., Qin, Y., Li, R., Zhang, G., Zhao, B., Bi, C., Lei, Y. and Bai, Y. (2017) 'Inhibition of Gli1-mediated prostate cancer cell proliferation by inhibiting the mTOR/S6K1 signaling pathway', *Oncology letters*, 14(6), pp. 7970-7976.
- Yang, W., Zhao, X., Han, Y., Duan, L., Lu, X., Wang, X., Zhang, Y., Zhou, W., Liu, J., Zhang, H., Zhao, Q., Hong, L. and Fan, D. (2019) 'Identification of hub genes and therapeutic drugs in esophageal squamous cell carcinoma based on integrated bioinformatics strategy', *Cancer Cell International*, 19(1), pp. 142.
- Yao, X., Anderson, K. L. and Cleveland, D. W. (1997) 'The microtubule-dependent motor centromere-associated protein E (CENP-E) is an integral component of kinetochore corona fibers that link centromeres to spindle microtubules', *J Cell Biol*, 139(2), pp. 435-47.
- Yao, Y., Zhou, D., Shi, D., Zhang, H., Zhan, S., Shao, X., Sun, K., Sun, L., Wu, G., Tian, K., Zhu, X. and He, S. (2019) 'GLI1 overexpression promotes gastric cancer cell proliferation and migration and induces drug resistance by combining with the AKT-mTOR pathway', *Biomedicine & Pharmacotherapy*, 111, pp. 993-1004.
- Yen, T. J., Li, G., Schaar, B. T., Szilak, I. and Cleveland, D. W. (1992) 'CENP-E is a putative kinetochore motor that accumulates just before mitosis', *Nature*, 359(6395), pp. 536-9.
- Yeung, T. L., Leung, C. S., Wong, K. K., Gutierrez-Hartmann, A., Kwong, J., Gershenson, D. M. and Mok, S. C. (2017) 'ELF3 is a negative regulator of epithelial-mesenchymal transition in ovarian cancer cells', *Oncotarget*, 8(10), pp. 16951-16963.
- Yin, J., Lin, C., Jiang, M., Tang, X., Xie, D., Chen, J. and Ke, R. (2021) 'CENPL, ISG20L2, LSM4, MRPL3 are four novel hub genes and may serve as diagnostic and prognostic markers in breast cancer', *Scientific Reports*, 11(1), pp. 15610.
- Yoshihara, K., Tajima, A., Yahata, T., Kodama, S., Fujiwara, H., Suzuki, M., Onishi, Y., Hatae, M., Sueyoshi, K., Kudo, Y., Kotera, K., Masuzaki, H., Tashiro, H., Katabuchi, H., Inoue, I. and Tanaka, K. (2010) 'Gene expression profile for predicting survival in advanced-stage serous ovarian cancer across two independent datasets', *PLoS One*, 5(3), pp. e9615.
- Yoshihara, K., Tsunoda, T., Shigemizu, D., Fujiwara, H., Hatae, M., Masuzaki, H., Katabuchi, H., Kawakami, Y., Okamoto, A., Nogawa, T., Matsumura, N., Udagawa, Y., Saito, T., Itamochi, H., Takano, M., Miyagi, E., Sudo, T., Ushijima, K., Iwase, H., Seki, H., Terao, Y., Enomoto, T., Mikami, M., Akazawa, K., Tsuda, H., Moriya, T., Tajima, A., Inoue, I. and Tanaka, K. (2012) 'High-risk ovarian cancer based on 126-gene expression signature is uniquely characterized by downregulation of antigen presentation pathway', *Clin Cancer Res*, 18(5), pp. 1374-85.
- Yu, D. H., Huang, J. Y., Liu, X. P., Ruan, X. L., Chen, C., Hu, W. D. and Li, S. (2020) 'Effects of hub genes on the clinicopathological and prognostic features of lung adenocarcinoma', *Oncol Lett*, 19(2), pp. 1203-1214.
- Yu, Y. P., Landsittel, D., Jing, L., Nelson, J., Ren, B., Liu, L., McDonald, C., Thomas, R., Dhir, R., Finkelstein, S., Michalopoulos, G., Becich, M. and Luo, J. H. (2004) 'Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy', *J Clin Oncol*, 22(14), pp. 2790-9.

- Yuan, T. L. and Cantley, L. C. (2008) 'PI3K pathway alterations in cancer: variations on a theme', *Oncogene*, 27(41), pp. 5497-5510.
- Zambrano-Bigiarini, M. 2018. hydroGOF: Goodness-of-Fit Functions for Comparison of Simulated and Observed Hydrological Time Series. <https://cran.r-project.org/web/packages/hydroGOF/index.html>.
- Zhan, Y., Jiang, L., Jin, X., Ying, S., Wu, Z., Wang, L., Yu, W., Tong, J., Zhang, L., Lou, Y. and Qiu, Y. (2021) 'Inhibiting RRM2 to enhance the anticancer activity of chemotherapy', *Biomedicine & Pharmacotherapy*, 133, pp. 110996.
- Zhang, C., Wang, Y., Feng, Y., Zhang, Y., Ji, B., Wang, S., Sun, Y., Zhu, C., Zhang, D. and Sun, Y. (2016a) 'Gli1 promotes colorectal cancer metastasis in a Foxm1-dependent manner by activating EMT and PI3K-AKT signaling', *Oncotarget*, 7(52), pp. 86134-86147.
- Zhang, H., Zhang, X., Li, X., Meng, W. B., Bai, Z. T., Rui, S. Z., Wang, Z. F., Zhou, W. C. and Jin, X. D. (2018) 'Effect of CCNB1 silencing on cell cycle, senescence, and apoptosis through the p53 signaling pathway in pancreatic cancer', *J Cell Physiol*, 234(1), pp. 619-631.
- Zhang, J., Ding, N., He, Y., Tao, C., Liang, Z., Xin, W., Zhang, Q. and Wang, F. (2021) 'Bioinformatic identification of genomic instability-associated lncRNAs signatures for improving the clinical outcome of cervical cancer by a prognostic model', *Scientific Reports*, 11(1), pp. 20929.
- Zhang, L., Sun, L., Zhang, B. and Chen, L. (2019) 'Identification of Differentially Expressed Genes (DEGs) Relevant to Prognosis of Ovarian Cancer by Use of Integrated Bioinformatics Analysis and Validation by Immunohistochemistry Assay', *Medical science monitor : international medical journal of experimental and clinical research*, 25, pp. 9902-9912.
- Zhang, L. N., Li, J. Y. and Xu, W. (2013) 'A review of the role of Puma, Noxa and Bim in the tumorigenesis, therapy and drug resistance of chronic lymphocytic leukemia', *Cancer Gene Therapy*, 20(1), pp. 1-7.
- Zhang, M., Chen, H., Wang, M., Bai, F. and Wu, K. (2020a) 'Bioinformatics analysis of prognostic significance of COL10A1 in breast cancer', *Biosci Rep*, 40(2).
- Zhang, M., Zhuang, G., Sun, X., Shen, Y., Wang, W., Li, Q. and Di, W. (2017) 'TP53 mutation-mediated genomic instability induces the evolution of chemoresistance and recurrence in epithelial ovarian cancer', *Diagnostic Pathology*, 12(1), pp. 16.
- Zhang, N. and Pati, D. (2017) 'Biology and insights into the role of cohesin protease separate in human malignancies', *Biological Reviews*, 92(4), pp. 2070-2083.
- Zhang, W., Chai, W., Zhu, Z. and Li, X. (2020b) 'Aldehyde oxidase 1 promoted the occurrence and development of colorectal cancer by up-regulation of expression of CD133', *International Immunopharmacology*, 85, pp. 106618.
- Zhang, W. and Liu, H. T. (2002) 'MAPK signal pathways in the regulation of cell proliferation in mammalian cells', *Cell Res*, 12(1), pp. 9-18.
- Zhang, Y., Cao, L., Nguyen, D. and Lu, H. (2016b) 'TP53 mutations in epithelial ovarian cancer', *Translational cancer research*, 5(6), pp. 650-663.
- Zhang, Y. and Chen, Q. (2017) 'Relationship between matrix metalloproteinases and the occurrence and development of ovarian cancer', *Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas*, 50(6), pp. e6104-e6104.

- Zhao, B., Erwin, A. and Xue, B. (2018) 'How many differentially expressed genes: A perspective from the comparison of genotypic and phenotypic distances', *Genomics*, 110(1), pp. 67-73.
- Zhao, J., Zhao, Y., Wang, L., Zhang, J., Karnes, R. J., Kohli, M., Wang, G. and Huang, H. (2016) 'Alterations of androgen receptor-regulated enhancer RNAs (eRNAs) contribute to enzalutamide resistance in castration-resistant prostate cancer', *Oncotarget*, 7(25), pp. 38551-38565.
- Zhao, L., Li, Y., Zhang, Z., Zou, J., Li, J., Wei, R., Guo, Q., Zhu, X., Chu, C., Fu, X., Yue, J. and Li, X. (2020) 'Meta-analysis based gene expression profiling reveals functional genes in ovarian cancer', *Bioscience reports*, 40(11), pp. BSR20202911.
- Zhao, R., Gish, K., Murphy, M., Yin, Y., Notterman, D., Hoffman, W. H., Tom, E., Mack, D. H. and Levine, A. J. (2000) 'Analysis of p53-regulated gene expression patterns using oligonucleotide arrays', *Genes & development*, 14(8), pp. 981-993.
- Zhao, S. G., Chang, S. L., Erho, N., Yu, M., Lehrer, J., Alshalalfa, M., Speers, C., Cooperberg, M. R., Kim, W., Ryan, C. J., Den, R. B., Freedland, S. J., Posadas, E., Sandler, H., Klein, E. A., Black, P., Seiler, R., Tomlins, S. A., Chinnaiyan, A. M., Jenkins, R. B., Davicioni, E., Ross, A. E., Schaeffer, E. M., Nguyen, P. L., Carroll, P. R., Karnes, R. J., Spratt, D. E. and Feng, F. Y. (2017) 'Associations of Luminal and Basal Subtyping of Prostate Cancer With Prognosis and Response to Androgen Deprivation Therapy', *JAMA oncology*, 3(12), pp. 1663-1672.
- Zhao, Y. and Adjei, A. A. (2014) 'The clinical development of MEK inhibitors', *Nat Rev Clin Oncol*, 11(7), pp. 385-400.
- Zheng, G., Yu, H., Kanerva, A., Försti, A., Sundquist, K. and Hemminki, K. (2018) 'Familial risks of ovarian cancer by age at diagnosis, proband type and histology', *PLoS one*, 13(10), pp. e0205000-e0205000.
- Zheng, S., Huang, J., Zhou, K., Zhang, C., Xiang, Q., Tan, Z., Wang, T. and Fu, X. (2011) '17beta-Estradiol enhances breast cancer cell motility and invasion via extra-nuclear activation of actin-binding protein ezrin', *PLoS One*, 6(7), pp. e22439.
- Zhou, J., Du, Y., Lu, Y., Luan, B., Xu, C., Yu, Y. and Zhao, H. (2019) 'CD44 Expression Predicts Prognosis of Ovarian Cancer Patients Through Promoting Epithelial-Mesenchymal Transition (EMT) by Regulating Snail, ZEB1, and Caveolin-1', *Frontiers in oncology*, 9, pp. 802-802.
- Zhou, J., Wu, S.-G., Wang, J., Sun, J.-Y., He, Z.-Y., Jin, X. and Zhang, W.-W. (2018a) 'The Effect of Histological Subtypes on Outcomes of Stage IV Epithelial Ovarian Cancer', *Frontiers in oncology*, 8, pp. 577-577.
- Zhou, J., Yang, T., Liu, L. and Lu, B. (2017) 'Chemotherapy oxaliplatin sensitizes prostate cancer to immune checkpoint blockade therapies via stimulating tumor immunogenicity', *Mol Med Rep*, 16(3), pp. 2868-2874.
- Zhou, Z., Cheng, Y., Jiang, Y., Liu, S., Zhang, M., Liu, J. and Zhao, Q. (2018b) 'Ten hub genes associated with progression and prognosis of pancreatic carcinoma identified by co-expression analysis', *Int J Biol Sci*, 14(2), pp. 124-136.
- Zhu, A., Li, Y., Song, W., Xu, Y., Yang, F., Zhang, W., Yin, Y. and Guan, X. (2016) 'Antiproliferative Effect of Androgen Receptor Inhibition in Mesenchymal Stem-Like Triple-Negative Breast Cancer', *Cell Physiol Biochem*, 38(3), pp. 1003-14.
- Zhu, X., Han, Y., Fang, Z., Wu, W., Ji, M., Teng, F., Zhu, W., Yang, X., Jia, X. and Zhang, C. (2013) 'Progesterone protects ovarian cancer cells from cisplatin-induced inhibitory effects through

progesterone receptor membrane component 1/2 as well as AKT signaling', *Oncol Rep*, 30(5), pp. 2488-94.

Zhu, X., Luo, X., Feng, G., Huang, H., He, Y., Ma, W., Zhang, C., Zeng, M. and Liu, H. (2019) 'CENPE expression is associated with its DNA methylation status in esophageal adenocarcinoma and independently predicts unfavorable overall survival', *PLOS ONE*, 14(2), pp. e0207341.

Zhu, Y., Wu, R., Sangha, N., Yoo, C., Cho, K. R., Shedden, K. A., Katabuchi, H. and Lubman, D. M. (2006) 'Classifications of ovarian cancer tissues by proteomic patterns', *Proteomics*, 6(21), pp. 5846-5856.