

# Contents

<b>Glossary</b>	<b>xi</b>
<b>Acronyms</b>	<b>xiii</b>
<b>1 Introduction and Literature Review</b>	<b>1</b>
1.1 Cancer Research in the Post-Genomic Era . . . . .	1
1.1.1 Cancer is a Global Health Issue . . . . .	2
1.1.1.1 The Genetics and Molecular Biology of Cancers . . . . .	3
1.1.2 The Genomics Revolution in Cancer Research . . . . .	3
1.1.2.1 High-Throughput Technologies . . . . .	4
1.1.2.2 Bioinformatics and Genomic Data . . . . .	5
1.1.3 Genomics Projects . . . . .	5
1.1.3.1 The Cancer Genome Project . . . . .	6
1.1.3.2 The Cancer Genome Atlas Project . . . . .	6
1.1.4 Genomic Cancer Medicine . . . . .	8
1.1.4.1 Cancer Genes and Driver Mutations . . . . .	8
1.1.4.2 Precision Cancer Medicine . . . . .	9
1.1.4.3 Molecular Diagnostics and Pan-Cancer Medicine . . . . .	9
1.1.4.4 Targeted Therapeutics and Pharmacogenomics . . . . .	10
1.1.5 Systems and Network Biology . . . . .	11
1.2 Synthetic Lethal Cancer Medicine . . . . .	12
1.2.1 Synthetic Lethal Genetic Interactions . . . . .	12
1.2.2 Synthetic Lethal Concepts in Genetics . . . . .	14
1.2.3 Synthetic Lethality in Model Systems . . . . .	14
1.2.3.1 Synthetic Lethal Pathways and Networks . . . . .	15
1.2.3.2 Evolution of Synthetic Lethality . . . . .	15
1.2.4 Synthetic Lethality in Cancer . . . . .	16
1.2.5 Clinical Impact of Synthetic Lethality in Cancer . . . . .	18
1.2.6 High-throughput Screening for Synthetic Lethality . . . . .	19
1.2.6.1 Synthetic Lethal Screens . . . . .	21
1.2.7 Computational Prediction of Synthetic Lethality . . . . .	22
1.2.7.1 Bioinformatics Approaches to Genetic Interactions . . . . .	22
1.2.7.2 Comparative Genomics . . . . .	23
1.2.7.3 Analysis and Modelling of Protein Data . . . . .	26
1.2.7.4 Differential Gene Expression . . . . .	28
1.2.7.5 Data Mining and Machine Learning . . . . .	29

1.2.7.6	Mutual Exclusivity and Bimodality . . . . .	31
1.2.7.7	Rationale for Further Development . . . . .	33
1.3	E-cadherin as a Synthetic Lethal Target . . . . .	33
1.3.1	The <i>CDH1</i> gene and its Biological Functions . . . . .	33
1.3.2	Hereditary Diffuse Gastric (and Lobular Breast) Cancer . . . . .	34
1.3.3	Cell Line Models of <i>CDH1</i> Null Mutations . . . . .	35
1.4	Summary and Research Direction of Thesis . . . . .	36
1.4.1	Thesis Aims . . . . .	37
<b>2</b>	<b>Methods and Resources</b>	<b>38</b>
2.1	Bioinformatics Resources for Genomics Research . . . . .	38
2.1.1	Public Data and Software Packages . . . . .	38
2.1.1.1	Cancer Genome Atlas Data . . . . .	39
2.1.1.2	Reactome and Annotation Data . . . . .	40
2.2	Data Handling . . . . .	40
2.2.1	Normalisation . . . . .	40
2.2.2	Sample Triage . . . . .	40
2.2.3	Metagenes and the Singular Value Decomposition . . . . .	41
2.2.4	Candidate Triage and Integration with Screen Data . . . . .	43
2.3	Techniques . . . . .	43
2.3.1	Statistical Procedures and Tests . . . . .	44
2.3.2	Gene Set Over-representation Analysis . . . . .	45
2.3.3	Clustering . . . . .	45
2.3.4	Heatmap . . . . .	45
2.3.5	Modelling and Simulations . . . . .	46
2.3.5.1	Receiver Operating Characteristic Curves . . . . .	47
2.3.6	Resampling Analysis . . . . .	47
2.4	Pathway Structure Methods . . . . .	48
2.4.1	Network and Graph Analysis . . . . .	48
2.4.2	Sourcing Graph Structure Data . . . . .	49
2.4.3	Constructing Pathway Subgraphs . . . . .	49
2.4.4	Network Analysis Metrics . . . . .	50
2.5	Implementation . . . . .	51
2.5.1	Computational Resources and Linux Utilities . . . . .	51
2.5.2	R Language and Packages . . . . .	52
2.5.3	High Performance and Parallel Computing . . . . .	55
<b>3</b>	<b>Methods Developed During Thesis</b>	<b>57</b>
3.1	A Synthetic Lethal Detection Methodology . . . . .	57
3.2	Synthetic Lethal Simulation and Modelling . . . . .	59
3.2.1	A Model of Synthetic Lethality in Expression Data . . . . .	60
3.2.2	Simulation Procedure . . . . .	64
3.3	Detecting Simulated Synthetic Lethal Partners . . . . .	67
3.3.1	Binomial Simulation of Synthetic Lethality . . . . .	67
3.3.2	Multivariate Normal Simulation of Synthetic Lethality . . . . .	69
3.3.2.1	Multivariate Normal Simulation with Correlated Genes . . . . .	71

3.3.2.2	Specificity with Query-Correlated Pathways . . . . .	79
3.4	Graph Structure Methods . . . . .	81
3.4.1	Upstream and Downstream Gene Detection . . . . .	81
3.4.1.1	Permutation Analysis for Statistical Significance . . . . .	82
3.4.2	Simulating Gene Expression from Graph Structures . . . . .	83
3.5	Customised Functions and Packages Developed . . . . .	87
3.5.1	Synthetic Lethal Interaction Prediction Tool . . . . .	87
3.5.2	Data Visualisation . . . . .	88
3.5.3	Extensions to the iGraph Package . . . . .	89
3.5.3.1	Sampling Simulated Data from Graph Structures . . . . .	89
3.5.3.2	Plotting Directed Graph Structures . . . . .	89
3.5.3.3	Computing Information Centrality . . . . .	91
3.5.3.4	Testing Pathway Structure with Permutation Testing . . . . .	91
3.5.3.5	Metapackage to Install iGraph Functions . . . . .	92
<b>4</b>	<b>Synthetic Lethal Analysis of Gene Expression Data</b>	<b>93</b>
4.1	Synthetic Lethal Genes in Breast Cancer . . . . .	94
4.1.1	Synthetic Lethal Pathways in Breast Cancer . . . . .	95
4.1.2	Expression Profiles of Synthetic Lethal Partners . . . . .	97
4.1.2.1	Subgroup Pathway Analysis . . . . .	100
4.2	Comparing Synthetic Lethal Gene Candidates . . . . .	102
4.2.1	Primary siRNA Screen Candidates . . . . .	102
4.2.2	Comparison with Correlation . . . . .	102
4.2.3	Comparison with Primary Screen Viability . . . . .	105
4.2.4	Comparison with Secondary siRNA Screen Validation . . . . .	107
4.2.5	Comparison to Primary Screen at Pathway Level . . . . .	108
4.2.5.1	Resampling Genes for Pathway Enrichment . . . . .	110
4.2.6	Integrating Synthetic Lethal Pathways and Screens . . . . .	115
4.3	Synthetic Lethal Pathway Metagenes . . . . .	116
4.4	Replication in Stomach Cancer . . . . .	118
4.5	Discussion . . . . .	119
4.5.1	Strengths of the SLIPT Methodology . . . . .	119
4.5.2	Synthetic Lethal Pathways for E-cadherin . . . . .	120
4.5.3	Replication and Validation . . . . .	122
4.5.3.1	Integration with siRNA Screening . . . . .	122
4.5.3.2	Replication across Tissues . . . . .	123
4.6	Summary . . . . .	123
<b>5</b>	<b>Synthetic Lethal Pathway Structure</b>	<b>125</b>
5.1	Synthetic Lethal Genes in Reactome Pathways . . . . .	125
5.1.1	The PI3K/AKT Pathway . . . . .	126
5.1.2	The Extracellular Matrix . . . . .	128
5.1.3	G Protein Coupled Receptors . . . . .	131
5.1.4	Gene Regulation and Translation . . . . .	131
5.2	Network Analysis of Synthetic Lethal Genes . . . . .	133
5.2.1	Gene Connectivity and Vertex Degree . . . . .	134

5.2.2	Gene Importance and Centrality . . . . .	135
5.2.2.1	Information Centrality . . . . .	135
5.2.2.2	PageRank Centrality . . . . .	137
5.3	Relationships between Synthetic Lethal Genes . . . . .	138
5.3.1	Detecting Upstream or Downstream Synthetic Lethality . . . . .	139
5.3.2	Resampling for Synthetic Lethal Pathway Structure . . . . .	141
5.4	Discussion . . . . .	143
5.5	Summary . . . . .	145
<b>6</b>	<b>Simulation and Modelling of Synthetic Lethal Pathways</b>	<b>147</b>
6.1	Synthetic Lethal Detection Methods . . . . .	148
6.1.1	Performance of SLIPT and $\chi^2$ across Quantiles . . . . .	148
6.1.1.1	Correlated Query Genes affects Specificity . . . . .	152
6.1.2	Alternative Synthetic Lethal Detection Strategies . . . . .	154
6.1.2.1	Correlation for Synthetic Lethal Detection . . . . .	154
6.1.2.2	Testing for Bimodality with BiSEp . . . . .	156
6.2	Simulations with Graph Structures . . . . .	157
6.2.1	Performance over Graph Structures . . . . .	158
6.2.1.1	Simple Graph Structures . . . . .	158
6.2.1.2	Constructed Graph Structures . . . . .	160
6.2.2	Performance with Inhibitions . . . . .	163
6.2.3	Synthetic Lethality across Graph Structures . . . . .	168
6.2.4	Performance within a Large Simulated Datasets . . . . .	171
6.3	Simulations in More Complex Graph Structures . . . . .	175
6.3.1	Simulations over Pathway-based Graphs . . . . .	176
6.3.2	Pathway Structures in a Large Simulated Datasets . . . . .	179
6.4	Discussion . . . . .	182
6.4.1	Simulation Procedure . . . . .	182
6.4.2	Comparing Methods with Simulated Data . . . . .	183
6.4.3	Design and Performance of SLIPT . . . . .	184
6.4.4	Simulations from Graph Structures . . . . .	186
6.5	Summary . . . . .	187
<b>7</b>	<b>Discussion</b>	<b>188</b>
7.1	Synthetic Lethality and <i>CDH1</i> Biology . . . . .	188
7.1.1	Established Functions of <i>CDH1</i> . . . . .	189
7.1.2	The Molecular Role of <i>CDH1</i> in Cancer . . . . .	189
7.2	Significance . . . . .	190
7.2.1	Synthetic Lethality in the Genomic Era . . . . .	190
7.2.2	Clinical Interventions based on Synthetic Lethality . . . . .	192
7.3	Future Directions . . . . .	193
7.4	Conclusions . . . . .	195
	<b>Bibliography</b>	<b>197</b>

<b>A</b>	<b>Sample Quality</b>	<b>222</b>
A.1	Sample Correlation . . . . .	222
A.2	Replicate Samples in TCGA Breast Cancer Data . . . . .	225
<b>B</b>	<b>Software Used for Thesis</b>	<b>229</b>
<b>C</b>	<b>Mutation Analysis in Breast Cancer</b>	<b>238</b>
C.1	Synthetic Lethal Genes and Pathways . . . . .	238
C.2	Synthetic Lethal Expression Profiles . . . . .	239
C.3	Comparison to Primary Screen . . . . .	242
C.3.1	Resampling Analysis . . . . .	244
C.4	Compare SLIPT genes . . . . .	246
<b>D</b>	<b>Metagene Analysis</b>	<b>248</b>
D.1	Pathway Signature Expression . . . . .	248
D.2	Synthetic Lethal Reactome Metagenes . . . . .	252
<b>E</b>	<b>Intrinsic Subtyping</b>	<b>253</b>
<b>F</b>	<b>Stomach Expression Analysis</b>	<b>255</b>
F.1	Synthetic Lethal Genes and Pathways . . . . .	255
F.2	Comparison to Primary Screen . . . . .	259
F.2.1	Resampling Analysis . . . . .	261
F.3	Metagene Analysis . . . . .	263
<b>G</b>	<b>Synthetic Lethal Genes in Pathways</b>	<b>264</b>
<b>H</b>	<b>Network Analysis for Mutation SLIPT</b>	<b>271</b>
<b>I</b>	<b>Pathway Structure for Mutation SLIPT</b>	<b>274</b>
<b>J</b>	<b>Performance of SLIPT and <math>\chi^2</math></b>	<b>276</b>
J.1	Correlated Query Genes affects Specificity . . . . .	282
<b>K</b>	<b>Simulations on Graph Structures</b>	<b>288</b>
K.0.1	Simulations from Inhibiting Graph Structures . . . . .	289
K.1	Simulation across Graph Structures . . . . .	292
K.2	Simulations from Complex Graph Structures . . . . .	296
K.2.1	Simulations from Complex Inhibiting Graphs . . . . .	299
K.3	Simulations from Pathway Graph Structures . . . . .	305

# List of Figures

1.1	Synthetic genetic interactions . . . . .	13
1.2	Synthetic lethality in cancer . . . . .	17
2.1	Read count density . . . . .	42
2.2	Read count sample mean . . . . .	42
3.1	Framework for synthetic lethal prediction . . . . .	58
3.2	Synthetic lethal prediction adapted for mutation . . . . .	59
3.3	A model of synthetic lethal gene expression . . . . .	61
3.4	Modelling synthetic lethal gene expression . . . . .	62
3.5	Synthetic lethality with multiple genes . . . . .	63
3.6	Simulating gene function . . . . .	65
3.7	Simulating synthetic lethal gene function . . . . .	65
3.8	Simulating synthetic lethal gene expression . . . . .	66
3.9	Performance of binomial simulations . . . . .	68
3.10	Comparison of statistical performance . . . . .	68
3.11	Performance of multivariate normal simulations . . . . .	70
3.12	Simulating expression with correlated gene blocks . . . . .	72
3.13	Simulating expression with correlated gene blocks . . . . .	73
3.14	Synthetic lethal prediction across simulations . . . . .	75
3.15	Performance with correlations . . . . .	76
3.16	Comparison of statistical performance with correlation structure . . . . .	77
3.17	Performance with query correlations . . . . .	78
3.18	Statistical evaluation of directional criteria . . . . .	79
3.19	Performance of directional criteria . . . . .	80
3.20	Simulated graph structures . . . . .	84
3.21	Simulating expression from a graph structure . . . . .	85
3.22	Simulating expression from graph structure with inhibitions . . . . .	86
3.23	Demonstration of violin plots with custom features . . . . .	90
3.24	Demonstration of annotated heatmap . . . . .	90
3.25	Simulating graph structures . . . . .	91
4.1	Synthetic lethal expression profiles of analysed samples . . . . .	98
4.2	Comparison of SLIPT with siRNA . . . . .	103
4.3	Comparison of SLIPT and siRNA genes with correlation . . . . .	103
4.4	Comparison of SLIPT and siRNA genes with correlation . . . . .	105
4.5	Comparison of SLIPT and siRNA genes with screen viability . . . . .	106

4.6	Comparison of SLIPT genes with siRNA screen viability . . . . .	106
4.7	Resampled intersection of SLIPT and siRNA candidate genes . . . . .	111
5.1	Synthetic lethality in the PI3K cascade . . . . .	127
5.2	Synthetic lethality in Elastic Fibre Formation . . . . .	129
5.3	Synthetic lethality in Fibrin Clot Formation . . . . .	130
5.4	Synthetic lethality in the GPCRs . . . . .	132
5.5	Synthetic lethality and vertex degree . . . . .	134
5.6	Synthetic lethality and centrality . . . . .	136
5.7	Synthetic lethality and PageRank . . . . .	138
5.8	Structure of synthetic lethality resampling . . . . .	140
6.1	Performance of $\chi^2$ and SLIPT across quantiles . . . . .	150
6.2	Performance of $\chi^2$ and SLIPT across quantiles with more genes . . . . .	151
6.3	Performance of $\chi^2$ and SLIPT across quantiles with query correlation . . . . .	152
6.4	Performance of $\chi^2$ and SLIPT across quantiles with query correlation and more genes . . . . .	153
6.5	Performance of negative correlation and SLIPT . . . . .	155
6.6	Simple graph structures . . . . .	158
6.7	Performance of simulations on a simple graph . . . . .	159
6.8	Performance of simulations is similar in simple graphs . . . . .	161
6.9	Performance of simulations on a pathway . . . . .	162
6.10	Performance of simulations on a simple graph with inhibition . . . . .	164
6.11	Performance is higher on a simple inhibiting graph . . . . .	165
6.12	Performance of simulations on a constructed graph with inhibition . . . . .	166
6.13	Performance is affected by inhibition in graphs . . . . .	168
6.14	Detection of synthetic lethality within a graph structure . . . . .	170
6.15	Performance of simulations including a simple graph . . . . .	172
6.16	Performance on a simple graph improves with more genes . . . . .	174
6.17	Performance on an inhibiting graph improves with more genes . . . . .	175
6.18	Performance of simulations on the PI3K cascade . . . . .	178
6.19	Performance of simulations including the PI3K cascade . . . . .	180
6.20	Performance on pathways improves with more genes . . . . .	181
A.1	Correlation profiles of removed samples . . . . .	223
A.2	Correlation analysis and sample removal . . . . .	224
A.3	Replicate excluded samples . . . . .	225
A.4	Replicate samples with all remaining . . . . .	226
A.5	Replicate samples with some excluded . . . . .	227
C.1	Synthetic lethal expression profiles of analysed samples . . . . .	240
C.2	Comparison of mtSLIPT to short interfering RNA (siRNA) . . . . .	242
C.3	Compare mtSLIPT and siRNA genes with correlation . . . . .	246
C.4	Compare mtSLIPT and siRNA genes with correlation . . . . .	246
C.5	Compare mtSLIPT and siRNA genes with siRNA viability . . . . .	247
D.1	Pathway metagene expression profiles . . . . .	250

D.2	Expression profiles for estrogen receptor related genes . . . . .	251
F.1	Synthetic lethal expression profiles of stomach samples . . . . .	257
F.2	Comparison of SLIPT in stomach to siRNA . . . . .	259
G.1	Synthetic lethality in the PI3K/AKT pathway . . . . .	264
G.2	Synthetic lethality in the PI3K/AKT pathway in cancer . . . . .	265
G.3	Synthetic lethality in the Extracellular Matrix . . . . .	266
G.4	Synthetic lethality in the GPCR Downstream . . . . .	267
G.5	Synthetic lethality in the Translation Elongation . . . . .	268
G.6	Synthetic lethality in the Nonsense-mediated Decay . . . . .	269
G.7	Synthetic lethality in the 3' UTR . . . . .	270
H.1	Synthetic lethality and vertex degree . . . . .	271
H.2	Synthetic lethality and centrality . . . . .	272
H.3	Synthetic lethality and PageRank . . . . .	272
I.1	Structure of synthetic lethality resampling . . . . .	274
J.1	Performance of $\chi^2$ and SLIPT across quantiles . . . . .	276
J.2	Performance of $\chi^2$ and SLIPT across quantiles . . . . .	278
J.3	Performance of $\chi^2$ and SLIPT across quantiles with more genes . . . . .	280
J.4	Performance of $\chi^2$ and SLIPT across quantiles with query correlation . . . . .	282
J.5	Performance of $\chi^2$ and SLIPT across quantiles with query correlation . . . . .	284
J.6	Performance of $\chi^2$ and SLIPT across quantiles with query correlation and more genes . . . . .	286
K.1	Performance of simulations on a simple graph . . . . .	288
K.2	Performance of simulations on an inhibiting graph . . . . .	289
K.3	Performance of simulations on a constructed graph with inhibition . . . . .	290
K.4	Performance of simulations on a constructed graph with inhibition . . . . .	291
K.5	Detection of synthetic lethality within a graph structure . . . . .	292
K.6	Detection of synthetic lethality within an inhibiting graph . . . . .	294
K.7	Detection of synthetic lethality within an inhibiting graph . . . . .	295
K.8	Performance of simulations on a branching graph . . . . .	296
K.9	Performance of simulations on a complex graph . . . . .	297
K.10	Performance of simulations on a large graph . . . . .	298
K.11	Performance of simulations on a branching graph with inhibition . . . . .	299
K.12	Performance of simulations on a branching graph with inhibition . . . . .	300
K.13	Performance of simulations on a complex graph with inhibition . . . . .	301
K.14	Performance of simulations on a complex graph with inhibition . . . . .	302
K.15	Performance of simulations on a large constructed graph with inhibition . . . . .	303
K.16	Performance of simulations on a large constructed graph with inhibition . . . . .	304
K.17	Performance of simulations on the $G_{\alpha i}$ signalling pathway . . . . .	305
K.18	Performance of simulations including the $G_{\alpha i}$ signalling pathway . . . . .	306



# List of Tables

1.1	Methods for predicting genetic interactions . . . . .	23
1.2	Methods for predicting synthetic lethality in cancer . . . . .	23
1.3	Methods used by Wu <i>et al.</i> (2014) . . . . .	25
2.1	Excluded samples by batch and clinical characteristics. . . . .	41
2.2	Computers used during thesis . . . . .	51
2.3	Linux utilities and applications used during thesis . . . . .	52
2.4	R installations used during thesis . . . . .	53
2.5	R Packages used during thesis . . . . .	53
2.6	R packages developed during thesis . . . . .	55
4.1	Candidate synthetic lethal gene partners of <i>CDH1</i> from SLIPT . . . . .	95
4.2	Pathways for <i>CDH1</i> partners from SLIPT . . . . .	96
4.3	Pathways for clusters of <i>CDH1</i> partners from SLIPT . . . . .	101
4.4	ANOVA for synthetic lethality and correlation with <i>CDH1</i> . . . . .	104
4.5	Comparison of Synthetic Lethal Interaction Prediction Tool (SLIPT) genes against secondary siRNA screen . . . . .	108
4.6	Pathways for <i>CDH1</i> partners from SLIPT and siRNA . . . . .	109
4.7	Pathways for <i>CDH1</i> partners from SLIPT . . . . .	112
4.8	Pathways for <i>CDH1</i> partners from SLIPT and siRNA primary screen .	113
4.9	Examples of candidate metagenes synthetic lethal for <i>CDH1</i> from SLIPT	117
5.1	ANOVA for synthetic lethality and vertex degree . . . . .	135
5.2	ANOVA for synthetic lethality and information centrality . . . . .	136
5.3	ANOVA for synthetic lethality and PageRank centrality . . . . .	137
5.4	Resampling for pathway structure of synthetic lethal detection methods	142
B.1	Complete list of R packages used during this thesis . . . . .	229
C.1	Candidate synthetic lethal gene partners of <i>CDH1</i> from mtSLIPT . . .	238
C.2	Pathways for <i>CDH1</i> partners from mtSLIPT . . . . .	239
C.3	Pathways for clusters of <i>CDH1</i> partners from mtSLIPT . . . . .	241
C.4	Pathways for <i>CDH1</i> partners from mtSLIPT and siRNA . . . . .	243
C.5	Pathways for <i>CDH1</i> partners from mtSLIPT . . . . .	244
C.6	Pathways for <i>CDH1</i> partners from mtSLIPT and siRNA primary screen	245
D.1	Candidate synthetic lethal metagenes against <i>CDH1</i> from mtSLIPT . .	252

E.1	Comparison of intrinsic subtypes . . . . .	253
F.1	Synthetic lethal gene partners of <i>CDH1</i> from SLIPT in stomach cancer	255
F.2	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer . . . . .	256
F.3	Pathways for clusters of <i>CDH1</i> partners in stomach SLIPT . . . . .	258
F.4	Pathways for <i>CDH1</i> partners from SLIPT and siRNA . . . . .	260
F.5	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer . . . . .	261
F.6	Pathways for <i>CDH1</i> partners from SLIPT in stomach and siRNA . . . .	262
F.7	Synthetic lethal metagenes against <i>CDH1</i> in stomach cancer . . . . .	263
H.1	ANOVA for synthetic lethality and vertex degree . . . . .	273
H.2	ANOVA for synthetic lethality and information centrality . . . . .	273
H.3	ANOVA for synthetic lethality and PageRank centrality . . . . .	273
I.1	Resampling for pathway structure of synthetic lethal detection methods	275

# Glossary

allele	A gene variant with a specific sequence and phenotype.
driver mutation	A <a href="#">mutation</a> which promotes cancer growth.
E-cadherin	Epithelial cadherin (calcium-dependent adhesion), a cell-adhesion protein encoded by <i>CDH1</i> .
essential	A gene which is required to be functional or expressed for a cell or organism to be viable, grow or develop.
functional redundancy	Genes which perform a common function, also known as genetic redundancy.
gene expression	A measure of the relative expression of each gene from the mRNA extracted from (pooled) cells.
genome	All of the DNA sequence in the genome.
genomic	The use of data from all genes in the genome.
germline mutation	A <a href="#">mutation</a> that occurred in germline cells and is passed between generation.
graph or network	A mathematical structure modelling or depicting the relationships between elements.
hereditary	A trait or disease which has a genetic cause and is inherited from family members.
intrinsic subtype	Distinguishing cancer by molecular and genetic features.
metagene	A consistent signal of expression for a collection of genes such as a biological pathway, derived from singular value decomposition.

microarray	A high-throughput technique to measure presence or abundance of nucleic acid sequences from binding to probes.
mutant	A variant or dysfunctional phenotype arising from a <a href="#">mutation</a> in a gene.
mutation	A change in DNA sequence that disrupts gene function.
non-oncogene addiction	The dependence of a cancer cell on functioning non-mutant genes.
oncogene	A gene that potentially causes cancer, typically by over-expression or mutant gene variants.
pathway	A series of biomolecules that produces a particular product or biological function.
RNA-Seq	The generation of transcriptome data from sequencing RNA.
somatic mutation	A <a href="#">mutation</a> that occurs in somatic cells, during a patient's lifespan.
sporadic cancer	Cancers which do occur in patients with a family history or carry a high-risk genetic variant.
synthetic lethal	Genetic interactions where inactivation of multiple genes is inviable (or deleterious) which are viable if inactivated separately.
tumour suppressor	A gene potentially causes cancer, typically by disruption of functions which protect the cell from cancer.
wild-type	A natural phenotype of a trait or the normally functional <a href="#">allele</a> which encodes it.

# Acronyms

ANOVA	Analysis of Variance.
DNA	Deoxyribonucleic Acid.
FDR	False Discovery Rate.
GPCR	G Protein Coupled Receptor.
HDGC	Hereditary Diffuse Gastric Cancer.
mRNA	Messenger RNA.
mtSLIPT	Synthetic Lethal Interaction Prediction Tool (against mutation).
PAM50	Prediction Analysis of Microarray 50.
PI3K	Phosphoinositide 3-kinase.
RNA	Ribonucleic Acid.
RNAi	RNA Interference.
siRNA	Short Interfering RNA.
SLIPT	Synthetic Lethal Interaction Prediction Tool.
TCGA	The Cancer Genome Atlas (genomics project).

## Chapter 4

# Synthetic Lethal Analysis of Gene Expression Data

Having developed a statistical [synthetic lethal](#) detection methodology, [SLIPT](#), it was next applied to publicly available cancer [gene expression](#) datasets. The analysis presented in this Chapter focuses on breast cancer for which [The Cancer Genome Atlas \(TCGA\) expression](#) data ([Koboldt \*et al.\*, 2012](#)) from a patient cohort and [siRNA](#) screen data, from experiments conducted in MCF10A cells ([Telford \*et al.\*, 2015](#)), were available. Stomach cancer data ([Bass \*et al.\*, 2014](#)) was used to replicate findings in an independent dataset, with this cancer chosen because it also occurs in syndromic [hereditary diffuse gastric cancer \(HDGC\)](#) patients. The [TCGA](#) data also has the advantages of having other clinical and molecular profiles, including [somatic mutation](#) across many of the same samples, in addition to a considerable sample size for [RNA-Seq expression](#) data generated with common [TCGA](#) procedures to minimise batch effects.

Synthetic lethal candidate partners for *CDH1* were identified at both the gene and [pathways](#) level. [SLIPT](#) gene candidates were analysed by cluster analysis for common [expression](#) profiles across samples and relationships with clinical factors and [mutations](#) in key breast cancer genes. These genes were also compared to the gene candidates from primary and secondary (validation) screens conducted by [Telford \*et al.\* \(2015\)](#) on isogenic cell lines. For comparison, the [SLIPT](#) methodology was also applied using [mutation](#) data for *CDH1* against [expression](#) of candidate partners (as described in [Section 3.1](#)) which may better represent the null [mutations](#) in [HDGC](#) patients and the experimental cell model ([Chen \*et al.\*, 2014](#)). Pathways were analysed by over-representation analysis (with resampling for comparisons with [siRNA](#) data) and supported by a [metagene](#) analysis of [pathways](#) gene signatures. The [pathways](#) [metagene](#)

[expression](#) profiles were used to replicate known relationships between clinical and molecular characteristics for breast cancer and to demonstrate application of [SLIPT](#) directly on [metagenes](#) to detect [synthetic lethal pathways](#).

## 4.1 Synthetic Lethal Genes in Breast Cancer

The [SLIPT](#) methodology (as described in Section 3.1) was applied to the normalised [TCGA](#) breast cancer [gene expression](#) dataset ( $n = 1168$ ). As shown in Table 4.1, the most significant genes had strong evidence of [expression](#)-based association with *CDH1* (high  $\chi^2$  values) with fewer samples exhibiting low [expression](#) of both genes than expected statistically. Eukaryotic translation genes were among the highest scoring gene candidates, including initiation factors, elongation factors, and ribosomal proteins. These are clearly necessary for cancer cells to grow and proliferate, with sustained [gene expression](#) needed to maintain growth signalling [pathways](#) and resist apoptosis or immune factors, translation genes may be subject to [non-oncogene addiction](#) for *CDH1*-deficient cells.

While these are among the strongest [synthetic lethal](#) candidates, translational genes are crucial to the viability of healthy cells and dosing for a selective [synthetic lethal](#) effect against these may be difficult compared to other biological functions which may also be supported among the [SLIPT](#) candidate genes. Furthermore, few known biological functions of *CDH1* were among the strongest SL candidates, so the remaining candidate genes may also be informative since they are likely to contain these expected functions in addition to novel relationships for *CDH1*. Thus further [pathways](#) level analyses were also conducted to examine biological functions over-represented among synthetic candidate genes and to identify [synthetic lethal pathways](#).

The modified [mtSLIPT](#) methodology (as described in Section 3.1) was also applied to the normalised [TCGA](#) breast cancer [gene expression](#) dataset, against [somatic](#) loss of function [mutations](#) in *CDH1*. As shown in Appendix Table C.1, the most significant genes also had strong evidence of [expression](#) associated with *CDH1* [mutations](#) (high  $\chi^2$  values) with fewer samples with *CDH1* exhibiting low [expression](#) each candidate gene than expected statistically. These genes were not as strongly supported as the [expression](#) analysis (in Table 4.1), however, nor were as many genes detected. This is perhaps unsurprising due to the lower sample size with matching [somatic mutation](#) data and the lower frequency of *CDH1* [mutations](#) compared to low [expression](#) defined by  $1/3$  quantiles.

Table 4.1: Candidate [synthetic lethal](#) gene partners of *CDH1* from SLIPT

Gene	Observed*	Expected*	$\chi^2$ value	p-value	p-value (False discovery rate (FDR))
<i>TRIP10</i>	62	130	162	$5.65 \times 10^{-34}$	$1.84 \times 10^{-31}$
<i>EEF1B2</i>	56	130	158	$3.10 \times 10^{-33}$	$9.45 \times 10^{-31}$
<i>GBGT1</i>	61	131	156	$1.08 \times 10^{-32}$	$3.14 \times 10^{-30}$
<i>ELN</i>	81	130	149	$3.46 \times 10^{-31}$	$8.82 \times 10^{-29}$
<i>TSPAN4</i>	78	130	146	$1.63 \times 10^{-30}$	$3.79 \times 10^{-28}$
<i>GLIPR2</i>	72	130	146	$1.68 \times 10^{-30}$	$3.86 \times 10^{-28}$
<i>RPS20</i>	73	131	145	$1.89 \times 10^{-30}$	$4.28 \times 10^{-28}$
<i>RPS27A</i>	80	130	143	$5.53 \times 10^{-30}$	$1.18 \times 10^{-27}$
<i>EEF1A1P9</i>	63	130	141	$1.91 \times 10^{-29}$	$3.74 \times 10^{-27}$
<i>C1R</i>	73	130	141	$2.05 \times 10^{-29}$	$3.97 \times 10^{-27}$
<i>LYL1</i>	73	130	140	$2.99 \times 10^{-29}$	$5.74 \times 10^{-27}$
<i>RPLP2</i>	71	130	139	$4.88 \times 10^{-29}$	$9.07 \times 10^{-27}$
<i>C10orf10</i>	73	130	138	$6.72 \times 10^{-29}$	$1.20 \times 10^{-26}$
<i>DULLARD</i>	74	131	138	$9.29 \times 10^{-29}$	$1.61 \times 10^{-26}$
<i>PPM1F</i>	64	130	136	$1.61 \times 10^{-28}$	$2.65 \times 10^{-26}$
<i>OBFC2A</i>	69	130	136	$2.49 \times 10^{-28}$	$3.93 \times 10^{-26}$
<i>RPL11</i>	70	130	136	$2.56 \times 10^{-28}$	$3.97 \times 10^{-26}$
<i>RPL18A</i>	70	130	135	$3.08 \times 10^{-28}$	$4.70 \times 10^{-26}$
<i>MFNG</i>	76	131	133	$7.73 \times 10^{-28}$	$1.12 \times 10^{-25}$
<i>RPS17</i>	77	131	133	$8.94 \times 10^{-28}$	$1.29 \times 10^{-25}$
<i>MGAT1</i>	73	130	132	$1.44 \times 10^{-27}$	$2.03 \times 10^{-25}$
<i>RPS12</i>	72	130	128	$8.57 \times 10^{-27}$	$1.12 \times 10^{-24}$
<i>C10orf54</i>	73	130	127	$1.37 \times 10^{-26}$	$1.75 \times 10^{-24}$
<i>LOC286367</i>	72	130	126	$2.20 \times 10^{-26}$	$2.70 \times 10^{-24}$
<i>GMFG</i>	70	130	126	$2.20 \times 10^{-26}$	$2.70 \times 10^{-24}$

Strongest candidate [synthetic lethal](#) partners for *CDH1* by SLIPT in TCGA breast cancer expression data

\* Observed and expected numbers of samples which had low [expression](#) of both genes

The [mtSLIPT](#) candidates had more genes involved in cell and gene regulation, particularly [DNA](#) and [RNA](#) binding factors. The strongest candidates also included microtubule (*KIF12*), microfibril (*MFAP4*), and cell adhesion (*TENC1*) genes consistent with the established cytoskeletal role of *CDH1*. The elastin gene (*ELN*) was notably strongly supported by both [expression](#) and [mutation SLIPT](#) analysis of *CDH1* supporting interactions with extracellular proteins and the tumour microenvironment.

#### 4.1.1 Synthetic Lethal Pathways in Breast Cancer

Translational [pathways](#) were strongly over-represented in [SLIPT](#) partners, as shown in Table 4.2. These include ribosomal subunits, initiation, peptide elongation, and termination. Regulatory processes involving [mRNA](#) including 3' untranslated region (UTR) binding, L13a-mediated translational silencing, and nonsense-mediated decay were also implicated. These are consistent with protein translation being subject to “[non-oncogene addiction](#)” (Luo *et al.*, 2009), as a core process that is dysregulated to sustain cancer proliferation and survival (Gao and Roux, 2015).



Immune [pathways](#), including the adaptive immune system and responses to infectious diseases were also strongly implicated as [synthetic lethal](#) with loss of [E-cadherin](#). This is consistent with the alterations of immune response being a hallmark of cancer [Hanahan and Weinberg \(2000\)](#), since evading the immune system is necessary for cancer survival. Either of these systems are potential means to target *CDH1* deficient cells, although these were not detected in an isolated cell line experimental screen ([Telford et al., 2015](#)) and the differences between the findings in patient data are described in more detail in Section 4.2.5.

Table 4.2: Pathways for *CDH1* partners from SLIPT

Pathways Over-represented	Pathway Size	SL Genes	p-value (FDR)
Eukaryotic Translation Elongation	86	81	$1.3 \times 10^{-207}$
Peptide chain elongation	83	78	$5.6 \times 10^{-201}$
Eukaryotic Translation Termination	83	77	$1.2 \times 10^{-196}$
Viral <a href="#">mRNA</a> Translation	81	76	$1.2 \times 10^{-196}$
Formation of a pool of free 40S subunits	93	81	$3.7 \times 10^{-194}$
Nonsense Mediated Decay independent of the Exon Junction Complex	88	77	$5.3 \times 10^{-187}$
L13a-mediated translational silencing of Ceruloplasmin <a href="#">expression</a>	103	82	$9.6 \times 10^{-183}$
3' -UTR-mediated translational regulation	103	82	$9.6 \times 10^{-183}$
GTP hydrolysis and joining of the 60S ribosomal subunit	104	82	$1.9 \times 10^{-181}$
Nonsense-Mediated Decay	103	80	$6.2 \times 10^{-176}$
Nonsense Mediated Decay enhanced by the Exon Junction Complex	103	80	$6.2 \times 10^{-176}$
Adaptive Immune System	412	167	$6.5 \times 10^{-174}$
Eukaryotic Translation Initiation	111	82	$5.7 \times 10^{-173}$
Cap-dependent Translation Initiation	111	82	$5.7 \times 10^{-173}$
SRP-dependent cotranslational protein targeting to membrane	104	79	$2.0 \times 10^{-171}$
Translation	141	91	$6.1 \times 10^{-170}$
Infectious disease	347	146	$1.6 \times 10^{-166}$
Influenza Infection	117	81	$1.9 \times 10^{-163}$
Influenza Viral <a href="#">RNA</a> Transcription and Replication	108	77	$1.9 \times 10^{-160}$
Influenza Life Cycle	112	77	$2.5 \times 10^{-156}$

Gene set over-representation analysis (hypergeometric test) for Reactome [pathways](#) in SLIPT partners for *CDH1*.

It is also notable that the [pathways](#) over-represented in SLIPT candidate genes have strongly significant over-representation of Reactome [pathways](#) based on the hypergeometric test (as described in Section 2.3.2). Even after adjusting stringently for multiple testing, biologically related [pathways](#) were supported together. These [pathways](#) are further supported by testing for [synthetic lethality](#) against *CDH1* [mutations](#) (mtSLIPT) with many of these [pathways](#) also among the most strongly supported in this analysis (shown in Appendix Table C.2). This [mutation](#)-based analysis more closely represents the null *CDH1* [mutations](#) in HDGC ([Guilford et al., 1998](#)) and the experimental MCF10A cell model ([Chen et al., 2014](#)). There was still support for trans-

lational and immune [pathways](#) not detected in the isolated experimental system. [G protein coupled receptors](#) also among the most strongly supported [pathways](#), supporting the experimental findings of [Telford et al. \(2015\)](#) for these intracellular signalling [pathways](#) already being targeted for other diseases.

### 4.1.2 Expression Profiles of Synthetic Lethal Partners

Due to the sheer number of gene candidates, investigations proceeded into correlation structure and [pathways](#) over-representation. These analyses also examined [expression](#) patterns of [synthetic lethal](#) gene candidates. This serves to explore the functional similarity of the [synthetic lethal](#) partners of *CDH1*, with the eventual aim to assess their utility as drug targets. As shown in Figure 4.1 (which clusters *CDH1* lowly expressing samples separately), there were several large clusters of genes among the [expression](#) profiles of the *CDH1* [synthetic lethal](#) candidate partners. The clustering suggests co-regulation of genes or [pathways](#) correlation between partner gene candidates. A number of candidates from an experimental [RNA interference \(RNAi\)](#) screen study performed by [Telford et al. \(2015\)](#) were also identified by this approach. In addition, novel gene candidates were also identified, which had not been observed affect viability in isogenic cell line experiments.

In these [expression](#) profiles, a gene with a moderate or high signal across samples exhibiting low *CDH1* [expression](#) would represent a potential drug target. However, it appears that several molecular subtypes of cancer have elevation of different clusters of [synthetic lethal](#) candidates in samples with low *CDH1*. This clustering suggests that different targets (or combinations) could be effective in different patients, suggesting potential utility for stratification. In particular, estrogen receptor negative, basal-like subtype, and “normal-like” tumours ([Dai et al., 2015](#); [Eroles et al., 2012](#); [Parker et al., 2009](#)) have elevation of genes specific to particular clusters, indicative of some [synthetic lethal](#) interactions being specific to a particular molecular subtype or genetic background. Thus [synthetic lethal](#) drug therapy against these subtypes may be ineffective if it were designed against genes in another cluster.

A similar correlation structure was observed among the candidates tested against *CDH1* [mutation \(mtSLIPT\)](#), as shown in Appendix Figure C.1. This clustering analysis similarly identified several major clusters of putative [synthetic lethal](#) partner genes. In this case, many partner genes had consistently high [expression](#) across most of the (predominantly lobular subtype) *CDH1* breast cancer samples. However, a major exception to this in the *CDH1* [expression](#) analysis were the normal tissue samples which

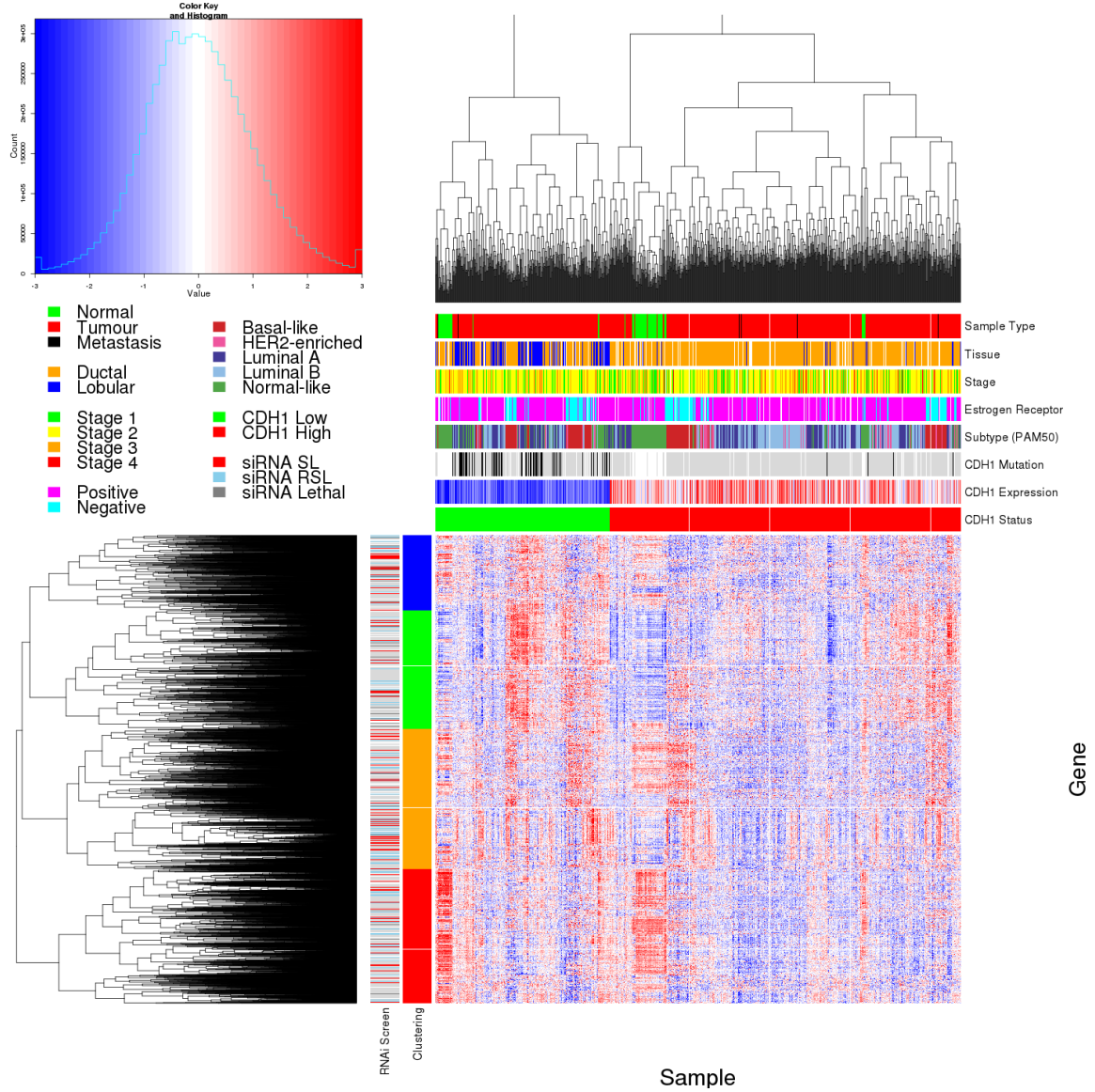


Figure 4.1: **Synthetic lethal expression profiles of analysed samples.** Gene expression profile heatmap (correlation distance, complete linkage) of all samples (separated by the  $1/3$  quantile of *CDH1* expression) analysed in TCGA breast cancer dataset for gene expression of 5165 candidate partners of E-cadherin (*CDH1*) from SLIPT prediction (with FDR adjusted  $p < 0.05$ ). Deeply clustered, inter-correlated genes form several main groups, each containing genes that were SL candidates or lethal in an siRNA screen (Telford *et al.*, 2015). Screen results for synthetic lethal (SL), the reverse effect (RSL), or lethal cell viability are shown as reported by Telford *et al.* (2015). Clusters had different sample groups highly expressing the synthetic lethal candidates in *CDH1* low samples, notably ‘normal-like’, ‘basal-like’, and estrogen receptor negative samples have elevated expression in one or more distinct clusters showing complexity and variation among candidate synthetic lethal partners. *CDH1* low samples also contained most of samples with *CDH1* mutations (shown in black). Negative values for mutation and screen data are shown in light grey, with missing data in white.

were excluded from the [mutation](#) data (as they were not tested for tumour-specific genotypes). This supports [synthetic lethal](#) interventions being more applicable to *CDH1* [mutant](#) tumours. There was still considerable correlation structure, particularly among *CDH1* [wild-type](#) samples, sufficient to distinguish gene clusters. In contrast to the [expression](#) analysis the (predominantly ductal *CDH1* [wild-type](#)) basal-like subtype and estrogen receptor negative samples had depleted [expression](#) among most candidate [synthetic lethal](#) partners. This is consistent with [synthetic lethal](#) interventions only being effective in lobular estrogen receptor positive breast cancers in which they are a more common, as recurrent ([driver](#)) [mutation](#). However, the remaining samples are still informative for [synthetic lethal](#) analysis (by [SLIPT](#)) as it requires highly expressing *CDH1* samples for comparison.

The *CDH1* [mutant](#) samples (as shown in Figure 4.1) were predominantly among the low *CDH1* expressing samples, clustering throughout them with similar expression profiles to other samples exhibiting low *CDH1* expression. Thus the molecular profiles of *CDH1* low samples were indistinguishable from *CDH1* [mutant](#) samples, with the exception of normal samples (that do not have [somatic mutation](#) data available). Conversely, many of the *CDH1* [mutant](#) samples (in Appendix Figure C.1) had among the lowest *CDH1* [expression](#), and some of the [synthetic lethal](#) partners were also highly expressed in low expressing *CDH1* [wild-type](#) samples, despite these not being considered as “inactivated” by [mtSLIPT](#) analysis.

Together these results support the use of low *CDH1* [expression](#) as a strategy for detecting *CDH1* inactivation. This has the benefit of increasing sample size (including samples such as normal tissue which do not have [somatic mutation](#) data available) and increasing the expected number of mutually inactive (low-low) samples for the directional criteria of (mt)[SLIPT](#) which enables it to better distinguish significant deviations below this (as discussed in Section 6.1). This also circumvents the assumption that all (detected) [mutations](#) are inactivating (although synonymous [mutations](#) were excluded from the analysis), which may not be the case for several highly expressing *CDH1* [mutant](#) samples that do not cluster together in Figure 4.1 or Appendix Figure C.1. One of these exhibits among the lowest [expression](#) for many predicted [synthetic lethal](#) partners and would not be vulnerable to inactivation of these genes. As such, correctly genotyping inactivating [mutations](#) will be [essential](#) in clinical practice for [synthetic lethal](#) targeting of [tumour suppressor](#) genes, particularly for other genes such as *TP53* where oncogenic and [tumour suppressor mutations](#) (with different molecular consequences) are both commons. Using [expression](#) as a measure of gene

function also avoids the assumptions that **mutations** are **somatic**, rather than **germline mutation**, and that gene inactivation occurs by detectable **mutations**, rather than other mechanisms such as epigenetic changes. These may also account for some of the lowly expressing *CDH1* **wild-type** samples clustering with similar profiles to **mutant** samples.

#### 4.1.2.1 Subgroup Pathway Analysis

Synthetic lethal gene candidates for *CDH1* from **SLIPT** analysis of **RNA-Seq gene expression** data were also used for **pathways** over-representation analyses (as described in Section 2.3.2). The correlation structure in the **expression** of candidates **synthetic lethal** genes in *CDH1* low tumours (lowest  $1/3^{\text{rd}}$  quantile of **expression**) was examined for distinct biological **pathways** in subgroups of genes elevated in different clusters of samples. These genes were highly expressed in different samples with their clinical factors including estrogen receptor status and **intrinsic subtypes**, from the **Prediction Analysis of Microarray 50 (PAM50)** procedure (Parker *et al.*, 2009) shown in Figure 4.1.

As shown by the most over-represented **pathways** in Table 4.3, each correlated cluster of candidate **synthetic lethal** partners of *CDH1* contains functionally different genes. Cluster 1 contains genes with less evidence of over-represented **pathways** than other clusters, corresponding to less correlation between genes within the cluster, and to it being a relatively small group. While there is some indication that collagen biosynthesis, microfibril elastic fibres, extracellular matrix, and metabolic **pathways** may be over-represented in Cluster 1, these results are mainly based on small **pathways** containing few **synthetic lethal** genes. Genes in Cluster 2 exhibited low **expression** in normal tissue samples compared to tumour samples (as shown in Figure 4.1) and show compelling evidence of over-representation of post-transcriptional gene regulation and protein translation processes. Similarly, Cluster 3 has over-representation of immune signalling **pathways** (including chemokines, secondary messenger, and TCR signalling) and downstream intracellular signalling cascades such as **GPCR** and  $G_{\alpha i}$  signalling events. While **pathways** over-representation was weaker among genes in Cluster 4, they contained intracellular signalling **pathways** and were highly expressed in normal samples (in contrast to Cluster 2). Cluster 4 also involved extracellular factors and stimuli such as extracellular matrix, platelet activation, ligand receptors, and retinoic acid signalling.

Based on these results, potential **synthetic lethal** partners of *CDH1* include processes known to be dysregulated in cancer, such as translational, cytoskeletal, and immune processes. Intracellular signalling cascades such as the **GPCRs** and extracel-

Table 4.3: Pathways for clusters of *CDH1* partners from SLIPT

Pathways Over-represented in Cluster 1	Pathway Size	Cluster Genes	p-value (FDR)
Collagen formation	67	10	$4.0 \times 10^{-11}$
Extracellular matrix organisation	238	21	$1.8 \times 10^{-9}$
Collagen biosynthesis and modifying enzymes	56	8	$1.8 \times 10^{-9}$
Uptake and actions of bacterial toxins	22	5	$9.5 \times 10^{-9}$
Elastic fibre formation	37	6	$1.9 \times 10^{-8}$
Muscle contraction	62	7	$2.4 \times 10^{-7}$
Fatty acid, triacylglycerol, and ketone body metabolism	117	10	$4.9 \times 10^{-7}$
XBP1(S) activates chaperone genes	51	6	$6.6 \times 10^{-7}$
IRE1alpha activates chaperones	54	6	$1.2 \times 10^{-6}$
Neurotoxicity of clostridium toxins	10	3	$1.3 \times 10^{-6}$
Retrograde neurotrophin signalling	10	3	$1.3 \times 10^{-6}$
Assembly of collagen fibrils and other multimeric structures	40	5	$1.9 \times 10^{-6}$
Collagen degradation	58	6	$2.0 \times 10^{-6}$
Arachidonic acid metabolism	41	5	$2.1 \times 10^{-6}$
Synthesis of PA	26	4	$3.0 \times 10^{-6}$
Signalling by NOTCH	80	7	$3.3 \times 10^{-6}$
Signalling to RAS	27	4	$3.7 \times 10^{-6}$
Integrin cell surface interactions	82	7	$4.2 \times 10^{-6}$
Pathways Over-represented in Cluster 2	Pathway Size	Cluster Genes	p-value (FDR)
Eukaryotic Translation Elongation	86	75	$1.1 \times 10^{-181}$
Viral mRNA Translation	81	72	$9.8 \times 10^{-179}$
Peptide chain elongation	83	72	$1.9 \times 10^{-175}$
Eukaryotic Translation Termination	83	72	$1.9 \times 10^{-175}$
Formation of a pool of free 40S subunits	93	75	$1.9 \times 10^{-171}$
Nonsense Mediated Decay independent of the Exon Junction Complex	88	72	$9.9 \times 10^{-168}$
L13a-mediated translational silencing of Ceruloplasmin expression	103	75	$3.0 \times 10^{-159}$
3'-UTR-mediated translational regulation	103	75	$3.0 \times 10^{-159}$
Nonsense-Mediated Decay	103	75	$3.0 \times 10^{-159}$
Nonsense Mediated Decay enhanced by the Exon Junction Complex	103	75	$3.0 \times 10^{-159}$
SRP-dependent cotranslational protein targeting to membrane	104	75	$3.2 \times 10^{-158}$
GTP hydrolysis and joining of the 60S ribosomal subunit	104	75	$3.2 \times 10^{-158}$
Eukaryotic Translation Initiation	111	75	$4.5 \times 10^{-151}$
Cap-dependent Translation Initiation	111	75	$4.5 \times 10^{-151}$
Influenza Infection	117	75	$1.4 \times 10^{-145}$
Influenza Viral RNA Transcription and Replication	108	72	$5.7 \times 10^{-145}$
Translation	141	81	$8.0 \times 10^{-143}$
Influenza Life Cycle	112	72	$2.3 \times 10^{-141}$
Pathways Over-represented in Cluster 3	Pathway Size	Cluster Genes	p-value (FDR)
Adaptive Immune System	412	90	$6.1 \times 10^{-61}$
Chemokine receptors bind chemokines	52	27	$6.7 \times 10^{-56}$
Generation of second messenger molecules	29	21	$6.5 \times 10^{-55}$
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	64	29	$6.5 \times 10^{-55}$
TCR signalling	62	27	$8.9 \times 10^{-51}$
Peptide ligand-binding receptors	161	40	$1.5 \times 10^{-45}$
Translocation of ZAP-70 to Immunological synapse	16	14	$3.1 \times 10^{-43}$
Costimulation by the CD28 family	51	22	$4.0 \times 10^{-43}$
PD-1 signalling	21	15	$4.0 \times 10^{-41}$
Class A/1 (Rhodopsin-like receptors)	258	50	$6.7 \times 10^{-41}$
Phosphorylation of CD3 and TCR zeta chains	18	14	$1.3 \times 10^{-40}$
Interferon gamma signalling	74	24	$5.0 \times 10^{-39}$
GPCR ligand binding	326	57	$1.8 \times 10^{-38}$
Cytokine Signalling in Immune system	268	48	$8.9 \times 10^{-37}$
Downstream TCR signalling	45	18	$1.8 \times 10^{-35}$
G <sub>αi</sub> signalling events	167	33	$2.2 \times 10^{-33}$
Cell surface interactions at the vascular wall	99	21	$1.3 \times 10^{-26}$
Interferon Signalling	164	28	$1.7 \times 10^{-26}$
Pathways Over-represented in Cluster 4	Pathway Size	Cluster Genes	p-value (FDR)
Extracellular matrix organisation	238	48	$8.0 \times 10^{-41}$
Class A/1 (Rhodopsin-like receptors)	258	47	$2.8 \times 10^{-36}$
GPCR ligand binding	326	54	$2.1 \times 10^{-34}$
G <sub>αs</sub> signalling events	83	22	$1.4 \times 10^{-31}$
GPCR downstream signalling	472	68	$1.1 \times 10^{-29}$
Haemostasis	423	61	$3.3 \times 10^{-29}$
Platelet activation, signalling and aggregation	180	31	$7.1 \times 10^{-28}$
Binding and Uptake of Ligands by Scavenger Receptors	40	14	$9.9 \times 10^{-27}$
RA biosynthesis pathways	22	11	$2.5 \times 10^{-26}$
Response to elevated platelet cytosolic Ca <sup>2+</sup>	82	19	$3.0 \times 10^{-26}$
Developmental Biology	420	57	$3.5 \times 10^{-26}$
G <sub>αi</sub> signalling events	167	28	$7.3 \times 10^{-26}$
Platelet degranulation	77	18	$1.6 \times 10^{-25}$
Gastrin-CREB signalling pathways via PKC and MAPK	171	28	$2.5 \times 10^{-25}$
Muscle contraction	62	16	$4.7 \times 10^{-25}$
G <sub>αq</sub> signalling events	150	25	$3.2 \times 10^{-24}$
Retinoid metabolism and transport	34	12	$5.0 \times 10^{-24}$
Phase 1 - Functionalisation of compounds	67	16	$6.5 \times 10^{-24}$

Pathway over-representation analysis for Reactome pathways with the number of genes in each pathways (Pathway Size), number of genes within the pathways identified (Cluster Genes), and the pathways over-representation p-value (adjusted by FDR) from the hypergeometric test.



lular stimuli for these [pathways](#) were also implicated in potential [synthetic lethality](#) with *CDH1*.

Similar translational, cytoskeletal, and immune processes were identified among [SLIPT](#) partners with respect to *CDH1* [mutation](#), shown in Appendix Table C.3. While [GPCR](#) signalling was replicated in [mtSLIPT](#) analysis, there was also stronger over-representation for NOTCH, ERBB2, and PI3K/AKT signalling in [mutation](#) analysis consistent with these signals being important for proliferation of *CDH1* deficient tumours. The [GPCR](#) and [PI3K/AKT pathways](#) are of particular interest as [pathways](#) with oncogenic [mutations](#) that can be targeted and downstream effects on translation (a strongly supported process across analyses). Extracellular matrix [pathways](#) (e.g., elastic fibre formation) were also supported across analyses (in Table 4.3 and Appendix Table C.3) consistent with the established cell-cell signalling role of *CDH1* and the importance of the tumour microenvironment for cancer proliferation.

## 4.2 Comparing Synthetic Lethal Gene Candidates

### 4.2.1 Primary siRNA Screen Candidates

Gene candidates were compared between computational ([SLIPT](#) in [TCGA](#) breast cancer data) and experimental (the primary [siRNA](#) screen performed by [Telford et al. \(2015\)](#)) approaches in Figure 4.2. The number of genes detected by both methods did not produce a significant overlap but these may be difficult to compare due to vast differences between the detection methods. There were similar issues in the comparison of [mtSLIPT](#) genes tested against *CDH1* [mutations](#) (in Appendix Figure C.2), despite excluding genes not tested by both methods in either test. However, these intersecting genes may still be functionally informative or amenable to drug triage as they were replicated across both methods and [pathways](#) over-representation differed between the Sections of the Venn diagram (as shown in Figure 4.2).

### 4.2.2 Comparison with Correlation

Another potential means to triage drug target candidates is by correlation of [expression](#) profiles with *CDH1*. Correlation with *CDH1* was compared to [SLIPT](#) and [siRNA](#) results in Figure 4.3. The genes not detected by [SLIPT](#) (including [siRNA](#) candidates) included genes with non-significant [SLIPT](#) p-values. As expected, these genes were distributed around a correlation of zero. Genes with higher correlation with *CDH1* (either direction) were more significant, although there were exceptions to this trend and larger positive correlations than negative correlations. The majority of [SLIPT](#)

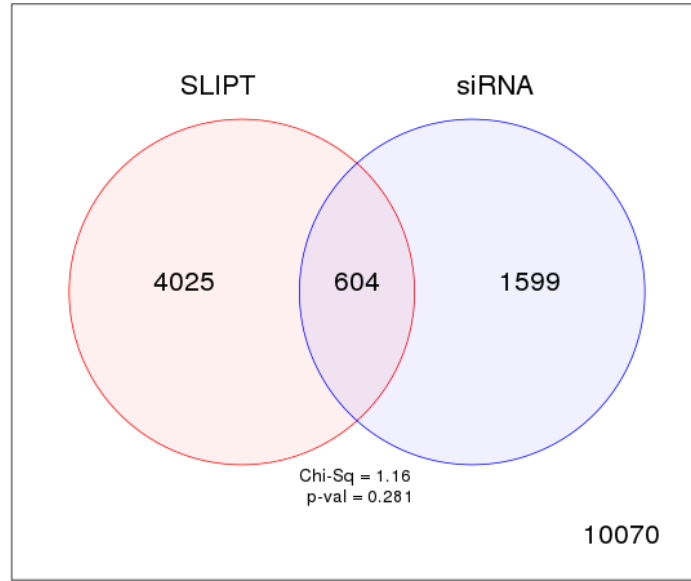


Figure 4.2: **Comparison of SLIPT with siRNA.** Testing the overlap of gene candidates for *E-cadherin* synthetic lethal partners between computational (SLIPT) and experimental screening (siRNA) approaches. The  $\chi^2$  test suggests that the overlap is no more than would be expected by chance ( $p = 0.281$ ). Only genes tested by both methods were included.

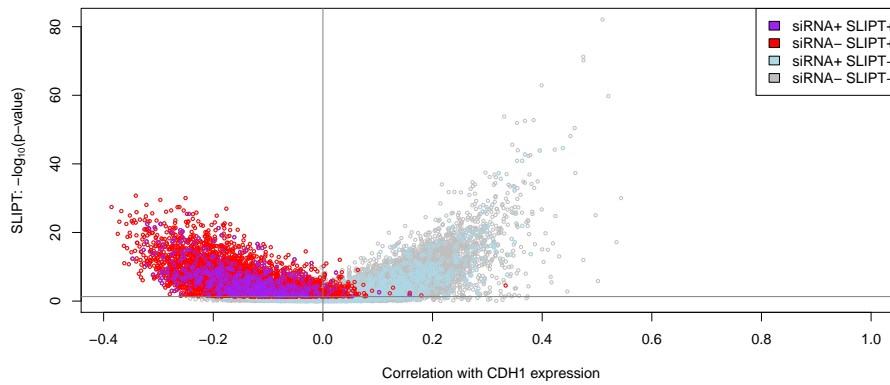


Figure 4.3: **Comparison of SLIPT and siRNA genes with correlation.** The  $\chi^2$  p-values for genes tested by SLIPT (in TCGA breast cancer) expression analysis were compared against Pearson correlation of gene expression with *CDH1*. Genes detected by SLIPT or siRNA are coloured according to the legend.



candidates had negative correlations, particularly genes detected by both approaches, although these were typically weak correlations and are unlikely to be sufficient to detect such genes on their own. This is supported by simulation results in Section 6.1.

There were not strong positive correlations with *CDH1* among siRNA candidates, consistent with previous findings that co-expression was not predictive of synthetic lethality (Jerby-Arnon *et al.*, 2014; Lu *et al.*, 2015). Negative correlation may not be indicative of synthetic lethality either as many siRNA candidates also had positive correlations. The SLIPT methodology has shown to detect genes with both positive and negative correlations, although it does appear to preferentially detect negatively correlated genes to some extent. These findings were replicated with the mtSLIPT approach against *CDH1* mutation (in Appendix Figure C.3), although the range of the  $\chi^2$  p-values differs due to lower sample size for mutation analysis.

The apparent tendency for genes detected by SLIPT or siRNA to have negative correlations with *CDH1* expression was not due to the smaller number of genes in these groups. The distribution of *CDH1* correlations differed across these gene groups (as shown by Figure 4.4 and Appendix Figure C.4) and tended to be lower in SLIPT candidates (as supported by analysis of variance (ANOVA) in Table 4.4). However, these are relatively weak correlations and further triage of gene candidates by correlation is not suitable. The genes detected both SLIPT and siRNA did not differ from SLIPT genes and the number of positively correlated SLIPT genes was very small. The use of correlation itself is also less effective than SLIPT to predict synthetic lethal partners in the first place (as shown in Section 6.1.2.1).

Table 4.4: ANOVA for synthetic lethality and correlation with *CDH1*

	DF	Sum Squares	Mean Squares	F-value	p-value
siRNA	1	0.027	0.027	2.8209	0.09306
SLIPT	1	134.603	134.603	14115.9824	<0.0001
siRNA×SLIPT	1	0.000	0.000	0.0073	0.93212

Analysis of variance for correlation with *CDH1* against synthetic lethal detection approaches (with an interaction term). Only genes tested by both methods were included in this analysis.

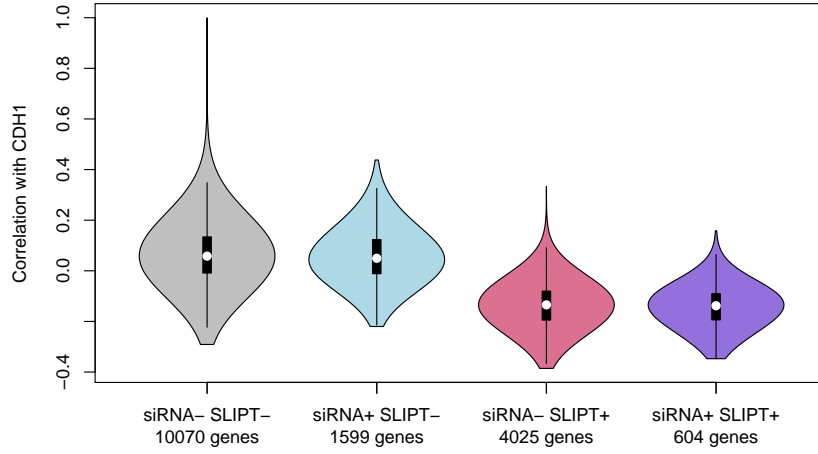


Figure 4.4: **Comparison of SLIPT and siRNA genes with correlation.** Genes detected as candidate **synthetic lethal** partners by SLIPT (in TCGA breast cancer) **expression** analysis and experimental screening (with siRNA) were compared against Pearson correlation of **gene expression** with *CDH1*. There were significant differences in correlation between gene groups (as shown in Table 4.4).

### 4.2.3 Comparison with Primary Screen Viability

A similar comparison of SLIPT results was made with the viability ratio (*CDH1*<sup>-/-</sup> mutant to wild-type) of MCF10A cells in the primary siRNA screen performed by Telford *et al.* (2015). The significance and viability thresholds used for SLIPT and siRNA detection of **synthetic lethal** candidate partners of *CDH1* are shown in Figure 4.5. Not all of the genes below the viability thresholds were necessarily selected to be candidate partners, however, as additional criteria were used in each case: directional criteria as for SLIPT (in Section 3.1) and minimum **wild-type** viability for siRNA (Telford *et al.*, 2015).

There does not appear to be a clear relationship between SLIPT and siRNA candidates. The genes detected by one approach but not the other were numerous in Figure 4.2 and Appendix Figure C.2. These genes detected by one approach are not necessarily near the thresholds for the other. In this respect, the SLIPT approach with patient data and the siRNA cell line experiments are independent means to identify **synthetic lethal** candidates. While genes detected by both approaches were not necessarily more strongly supported by either, the genes with a viability closer to 1 (no **synthetic**

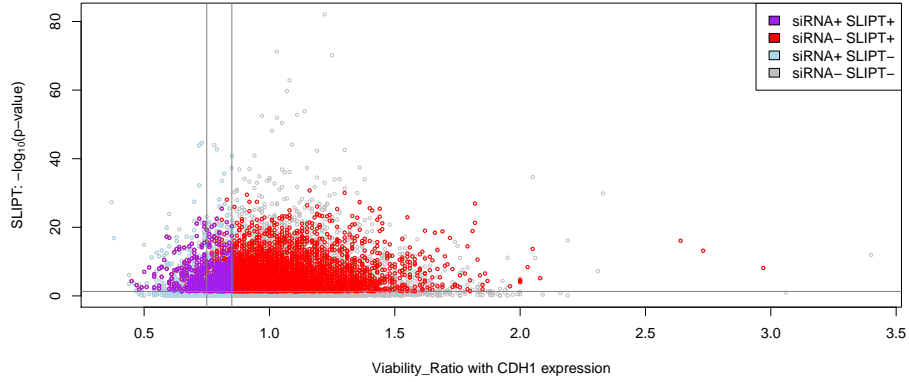


Figure 4.5: **Comparison of SLIPT and siRNA genes with screen viability.** The  $\chi^2$  p-values (log-scale) for genes tested by SLIPT (in TCGA breast cancer) were compared against the viability ratio of *CDH1* mutant and wild-type cells in the primary siRNA screen. Genes detected by SLIPT or siRNA are coloured according to the legend. Lines show the thresholds of significance with SLIPT and of viability used by Telford *et al.* (2015).

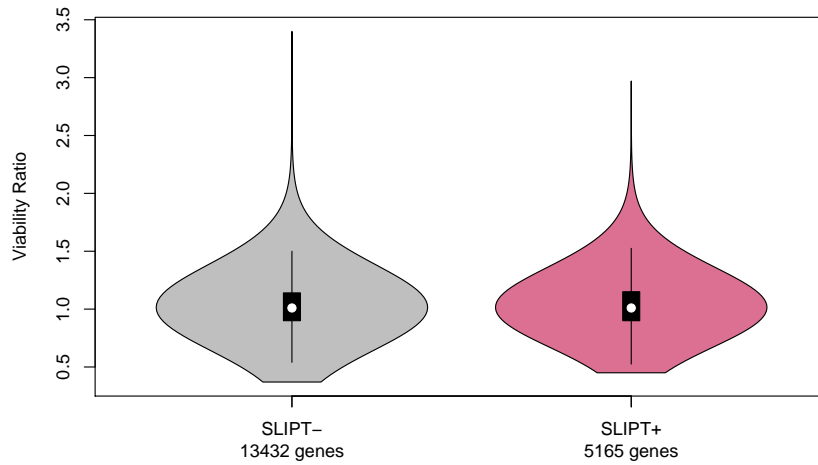


Figure 4.6: **Comparison of SLIPT genes with siRNA screen viability.** Genes detected as candidate synthetic lethal partners by SLIPT (in TCGA breast cancer) expression analysis were compared against the viability ratio of *CDH1* mutant and wild-type cells in the primary siRNA screen. There were clear no differences in viability between genes detected by SLIPT and those not detected. The genes identified by SLIPT had a higher viability ratio (by t-test:  $t = 2.1553$ ,  $p = 0.03117$ ), although the effect size was relatively small (mean SLIPT- 1.029, mean SLIPT+ 1.037).

lethal effect) in siRNA included those with more significant SLIPT p-values, whereas more extreme viability ratios tended to be less significant (as shown by Figure 4.5). However, it should be noted that genes with more moderate viability ratios were more common and SLIPT was capable (despite adjusting for multiple testing) of detecting significant genes with extreme viability ratios, particularly those considerably lower than 1. Lower viability ratios were used by Telford *et al.* (2015) to detect synthetic lethal candidates in the primary screen. However, there was little support for SLIPT candidates differing with respect to viability ratio (as shown in Figures 4.6 and C.5) and the vast majority of SLIPT candidate genes did not have different viability in the primary screen to genes not identified by SLIPT.

#### 4.2.4 Comparison with Secondary siRNA Screen Validation

It should be noted that genes with a lower viability ratio were not necessarily the most strongly supported by experimental screening. The primary screen (with 4 pooled siRNAs for each gene) has been used for the majority of comparisons in this thesis because the genomes-wide panel of target genes screened enables a large number of genes to be compared with SLIPT results from gene expression and somatic mutation analysis. A secondary screen was also performed by Telford *et al.* (2015) on the isogenic MCF10A breast cell lines to validate the individual (i.e., non-pooled) siRNAs separately, with the strongest candidates being those exhibiting synthetic lethal viability ratios replicated across independently targeting siRNAs. The strongest candidates from the primary screen were subject to a further secondary screen for validation by independent replication with 4 gene knockdowns with different targeting siRNAs. This was performed for the top 500 candidates (with the lowest viability ratio) from the primary screen: 482 of these genes were also tested by SLIPT in breast cancer.

The secondary screen results show that SLIPT candidate genes were more significantly ( $p = 7.49 \times 10^{-3}$  by Fisher's exact test) more likely to be validated with detection by more independently targeting siRNAs in the secondary screen. Gene detected by SLIPT are thus informative of more robust partner genes, in addition to providing support that these interactions are consistent with expression profiles from heterogeneous patient samples across genetic backgrounds. As shown in Table 4.5, there is significant association between SLIPT candidates and stronger validations of siRNA candidates. Since there were more SLIPT- genes among those not validated and more SLIPT+ genes among those validated with several siRNAs, this supports the use of SLIPT as a synthetic lethal discovery procedure which may augment such screening experiments.

Table 4.5: Comparison of SLIPT genes against secondary siRNA screen

		Secondary Screen					
siRNAs*		0/4	1/4	2/4	3/4	4/4	Total
SLIPT+	Observed	70	46	31	8	2	157
	Expected	85	44	10	4	2	
SLIPT−	Observed	190	90	31	10	4	325
	Expected	175	91	42	12	4	
Total		280	136	52	18	6	482

\* Number of siRNAs (targeting the same gene) to successfully reproduce synthetic lethality in MCF10A cells (Telford *et al.*, 2015)

While the individual genes detected by either approach do not necessarily match (and are potentially false-positives), the biological functions important in *CDH1* deficient cancers and potential mechanisms for specific targeting of them can be further supported by pathways analysis of the gene detected by either method. The genes detected by both approaches may therefore be more informative at the pathways level, where it is less likely for a pathways to be consistently detected by chance. As the SLIPT candidates differ from the siRNA candidates (in addition to those detected by both approaches which were more likely to be validated), they can provide information about additional mechanisms by which *CDH1* deficient cancers proliferate, and vulnerabilities that may be exploited against them by using the synthetic lethal pathways.

#### 4.2.5 Comparison to Primary Screen at Pathway Level

These pathways over-representation analyses (performed as described in Section 2.3.2) correspond to genes separated into SLIPT or siRNA screen candidates unique to either method, or detected by both (Table 4.6). The SLIPT-specific gene candidates were involved most strongly with translational and immune regulatory pathways, although extracellular matrix pathways were also supported. These pathways were largely consistent with those identified in Table 4.2 and in the clustering analysis (Table 4.3). The genes detected only by the siRNA screen had over-representation of cell signalling pathways, including many containing genes known to be involved in cancer (e.g., MAPK, PDGF, ERBB2, and FGFR), with the detection of Class A GPCRs supporting the independent analyses by Telford *et al.* (2015). The intersection of com-

Table 4.6: Pathways for *CDH1* partners from SLIPT and siRNA

Predicted only by SLIPT (4025 genes)	Pathway Size	Genes Identified	p-value (FDR)
Eukaryotic Translation Elongation	80	75	$1.5 \times 10^{-182}$
Peptide chain elongation	77	72	$2.9 \times 10^{-176}$
Viral mRNA Translation	75	70	$4.9 \times 10^{-172}$
Eukaryotic Translation Termination	76	70	$5.9 \times 10^{-170}$
Formation of a pool of free 40S subunits	87	74	$9.5 \times 10^{-166}$
Nonsense Mediated Decay independent of the Exon Junction Complex	81	70	$1.2 \times 10^{-160}$
L13a-mediated translational silencing of Ceruloplasmin expression	97	75	$3.8 \times 10^{-155}$
3' -UTR-mediated translational regulation	97	75	$3.8 \times 10^{-155}$
GTP hydrolysis and joining of the 60S ribosomal subunit	98	75	$6.0 \times 10^{-154}$
Nonsense-Mediated Decay	96	73	$5.2 \times 10^{-150}$
Nonsense Mediated Decay enhanced by the Exon Junction Complex	96	73	$5.2 \times 10^{-150}$
SRP-dependent cotranslational protein targeting to membrane	97	73	$7.8 \times 10^{-149}$
Eukaryotic Translation Initiation	105	75	$4.7 \times 10^{-146}$
Cap-dependent Translation Initiation	105	75	$4.7 \times 10^{-146}$
Translation	133	83	$4.0 \times 10^{-142}$
Influenza Viral RNA Transcription and Replication	102	71	$2.9 \times 10^{-137}$
Influenza Infection	111	74	$3.7 \times 10^{-137}$
Influenza Life Cycle	106	71	$2.3 \times 10^{-133}$
Infectious disease	326	125	$4.2 \times 10^{-120}$
Extracellular matrix organisation	189	77	$5.4 \times 10^{-95}$

Detected only by siRNA screen (1599 genes)	Pathway Size	Genes Identified	p-value (FDR)
Class A/1 (Rhodopsin-like receptors)	282	44	$1.3 \times 10^{-27}$
GPCR ligand binding	363	52	$5.8 \times 10^{-26}$
G <sub>αs</sub> signalling events	159	26	$6.7 \times 10^{-23}$
Gastrin-CREB signalling pathways via PKC and MAPK	180	27	$2.0 \times 10^{-21}$
G <sub>αi</sub> signalling events	184	27	$5.3 \times 10^{-21}$
Downstream signal transduction	146	23	$7.6 \times 10^{-21}$
Signalling by PDGF	172	25	$4.0 \times 10^{-20}$
Peptide ligand-binding receptors	175	25	$8.5 \times 10^{-20}$
Signalling by ERBB2	146	22	$1.3 \times 10^{-19}$
DAP12 interactions	159	23	$2.6 \times 10^{-19}$
DAP12 signalling	149	22	$2.7 \times 10^{-19}$
Organelle biogenesis and maintenance	264	33	$5.5 \times 10^{-19}$
Signalling by NGF	266	33	$8.2 \times 10^{-19}$
Downstream signalling of activated FGFR1	134	20	$1.1 \times 10^{-18}$
Downstream signalling of activated FGFR2	134	20	$1.1 \times 10^{-18}$
Downstream signalling of activated FGFR3	134	20	$1.1 \times 10^{-18}$
Downstream signalling of activated FGFR4	134	20	$1.1 \times 10^{-18}$
Signalling by FGFR	146	21	$1.3 \times 10^{-18}$
Signalling by FGFR1	146	21	$1.3 \times 10^{-18}$
Signalling by FGFR2	146	21	$1.3 \times 10^{-18}$

Intersection of SLIPT and siRNA screen (604 genes)	Pathway Size	Genes Identified	p-value (FDR)
Visual phototransduction	54	9	$6.9 \times 10^{-10}$
G <sub>αs</sub> signalling events	48	7	$1.6 \times 10^{-7}$
Retinoid metabolism and transport	24	5	$1.7 \times 10^{-7}$
Acyl chain remodelling of PS	10	3	$6.5 \times 10^{-6}$
Transcriptional regulation of white adipocyte differentiation	51	6	$6.5 \times 10^{-6}$
Chemokine receptors bind chemokines	22	4	$6.5 \times 10^{-6}$
Signalling by NOTCH4	11	3	$6.9 \times 10^{-6}$
Defective EXT2 causes exostoses 2	11	3	$6.9 \times 10^{-6}$
Defective EXT1 causes exostoses 1, TRPS2 and CHDS	11	3	$6.9 \times 10^{-6}$
Platelet activation, signalling and aggregation	146	12	$6.9 \times 10^{-6}$
Phase 1 - Functionalisation of compounds	41	5	$1.3 \times 10^{-5}$
Amine ligand-binding receptors	13	3	$1.7 \times 10^{-5}$
Acyl chain remodelling of PE	14	3	$2.4 \times 10^{-5}$
Signalling by GPCR	300	23	$2.4 \times 10^{-5}$
Molecules associated with elastic fibres	29	4	$2.6 \times 10^{-5}$
DAP12 interactions	128	10	$2.6 \times 10^{-5}$
Cytochrome P <sub>450</sub> - arranged by substrate type	30	4	$3.2 \times 10^{-5}$
GPCR ligand binding	147	11	$3.8 \times 10^{-5}$
Acyl chain remodelling of PC	16	3	$4.0 \times 10^{-5}$
Response to elevated platelet cytosolic Ca <sup>2+</sup>	66	6	$4.2 \times 10^{-5}$

putational and experimental [synthetic lethal](#) partners of *CDH1* had stronger evidence for over-representation of [GPCR pathways](#) and more specific subclasses, such as visual phototransduction ( $p = 6.9 \times 10^{-10}$ ) and  $G_{\alpha s}$  signalling events ( $p = 1.7 \times 10^{-7}$ ), than other signalling [pathways](#).

The [pathways](#) analysis for [mtSLIPT](#) against *CDH1* [mutations](#) (in Table C.4) had similar results to [SLIPT](#), particularly for [mtSLIPT-specific pathways](#). The specific [pathways](#) composition of the intersection of these analyses differed from [SLIPT](#) against low *CDH1* [expression](#). However, signalling [pathways](#) were also detected, including [GPCRs](#), [NOTCH](#), [EERB2](#), [PDGF](#), and [SCF-KIT](#). These findings indicate the signalling [pathways](#) are among the most suitable vulnerability to exploit in targeting *CDH1* deficient tumours as they can be detected in both a patient cohort (with [TCGA expression](#) data) and experimentally tested by inhibition with [siRNA](#) or drugs ([Telford et al., 2015](#)). However, it is possible that the [siRNA](#) screen, that was conducted in an isolated experimental system of MCF10A cell lines, was pre-disposed to preferentially detect kinase signalling [pathways](#) (which are amenable to pharmacological inhibition and clinical application). Nevertheless, the other [pathways](#) identified by [SLIPT](#) may still be informative of the role of *CDH1* loss of function in cancers or mechanisms by which further gene loss leads to specific inviability.

#### 4.2.5.1 Resampling Genes for Pathway Enrichment

A high number of significantly over-represented [pathways](#) were detected between [SLIPT](#) in [TCGA expression](#) data and [siRNA](#) genes despite relatively few genes being detected by both approaches. These strongly supported [pathways](#) are not unexpected, since [synthetic lethal pathways](#) are more robustly conserved ([Dixon et al., 2008](#)) and the computational approach using patient samples from complex tumour micro-environment has considerably different strengths to an experimental screen ([Telford et al., 2015](#)) based on genetically homogenous cell line models in an isolated laboratory environment. For instance, it is unlikely for immune signalling to be detected in an isolated cell culture system.

While many [pathways](#) were highly over-represented in the genes detected by both [SLIPT](#) and [siRNA](#), some of these [pathways](#) were also highly over-represented in the [siRNA](#) candidate genes and these may not reflect the results of [SLIPT](#) in expression data. A resampling approach (as described in Section 2.3.6) was used to assess whether [SLIPT](#) and the [siRNA](#) candidate genes had more frequently over-represented [pathways](#) than expected by chance. This resampling procedure tests whether [pathways](#) were over-represented in the genes detected by both approaches beyond that expected from

any subset of **siRNA** candidates. Thus resampling can determine whether **SLIPT** independently supports these putative **synthetic lethal pathways** (without assuming an underlying test statistic distribution).

A resampling approach is also applicable to testing whether the number of genes detected by each approach significantly intersected. As shown in Figure 4.7, resampling did not find evidence of significant depletion or over-representation for experimental **synthetic lethal** candidate genes in the computationally predicted **synthetic lethal** partners of *CDH1*, and thus the observed overlap may be due to chance. This is consistent with previous findings (as shown in Figure 4.2) and does not preclude **pathways** relationships being supported by resampling.

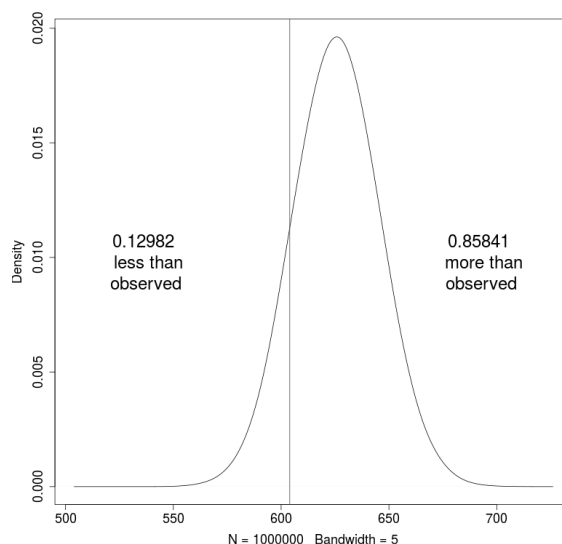


Figure 4.7: **Resampled intersection of SLIPT and siRNA candidate genes.** Resampling analysis of intersect size from genes detected by **SLIPT** and **siRNA** screening approaches over 1 million replicates. The proportion of expected intersection sizes for random samples below or above the observed intersection size respectively, lacking significant over-representation or depletion of **siRNA** screen candidates within the **SLIPT** predictions for *CDH1*.

A permutation analysis was performed to resample the genes tested by both approaches to investigate whether the observed **pathways** over-representation could have occurred in a randomly selected sample of genes from the experimental candidates, that is, whether the **pathways** predictions from **SLIPT** could be expected by chance (as described in Sections 2.2.4 and 2.3.6). While the number of **siRNA** candidate genes also detected by **SLIPT** was not statistically significant ( $p = 0.281$ ), this may be due



to the vastly different limitations of the approaches and the correlation structure of [gene expression](#) not being independent (as assumed for multiple testing procedures). The intersection may still be functionally relevant to *CDH1*-deficient cancers, such as the [pathways](#) data in Table 4.6. The resampling analysis for [pathways](#) was compared to the [pathways](#) over-representation for SLIPT predicted [synthetic lethal](#) partners in Table 4.7. Similarly, the [pathways](#) resampling for intersection between SLIPT predictions and experimental screen candidates was compared to [pathways](#) over-representation in Table 4.8 for intersection with [siRNA](#) data.

The [pathways](#) resampling approach for SLIPT-specific gene candidates (Table 4.7) largely recapitulates the parametric gene set over-representation analysis for all SLIPT genes, detecting evidence of [synthetic lethal pathways](#) for *CDH1* in translational, immune, and cell signalling [pathways](#) including  $G_{\alpha i}$  signalling, [GPCR](#) downstream signalling, and chemokine receptor binding. While the immune and signal transduction [pathways](#) were not significantly over-represented in the resampling analysis, the results for the two approaches were largely consistent for translation and post-transcriptional

Table 4.7: Pathways for *CDH1* partners from SLIPT

Reactome Pathway	Over-representation	Permutation
Eukaryotic Translation Elongation	$1.3 \times 10^{-207}$	$< 1.241 \times 10^{-5}$
Peptide chain elongation	$5.6 \times 10^{-201}$	$< 1.241 \times 10^{-5}$
Viral <a href="#">mRNA</a> Translation	$1.2 \times 10^{-196}$	$< 1.241 \times 10^{-5}$
Eukaryotic Translation Termination	$1.2 \times 10^{-196}$	$< 1.241 \times 10^{-5}$
Formation of a pool of free 40S subunits	$3.7 \times 10^{-194}$	$< 1.241 \times 10^{-5}$
Nonsense Mediated Decay independent of the Exon Junction Complex	$5.3 \times 10^{-187}$	$< 1.241 \times 10^{-5}$
L13a-mediated translational silencing of Ceruloplasmin <a href="#">expression</a>	$9.6 \times 10^{-183}$	$< 1.241 \times 10^{-5}$
3' -UTR-mediated translational regulation	$9.6 \times 10^{-183}$	$< 1.241 \times 10^{-5}$
GTP hydrolysis and joining of the 60S ribosomal subunit	$1.9 \times 10^{-181}$	$< 1.241 \times 10^{-5}$
Nonsense-Mediated Decay	$6.2 \times 10^{-176}$	$< 1.241 \times 10^{-5}$
Nonsense Mediated Decay enhanced by the Exon Junction Complex	$6.2 \times 10^{-176}$	$< 1.241 \times 10^{-5}$
Adaptive Immune System	$6.5 \times 10^{-174}$	0.15753
Eukaryotic Translation Initiation	$5.7 \times 10^{-173}$	$< 1.241 \times 10^{-5}$
Cap-dependent Translation Initiation	$5.7 \times 10^{-173}$	$< 1.241 \times 10^{-5}$
SRP-dependent cotranslational protein targeting to membrane	$2.0 \times 10^{-171}$	$< 1.241 \times 10^{-5}$
Translation	$6.1 \times 10^{-170}$	$< 1.241 \times 10^{-5}$
Infectious disease	$1.6 \times 10^{-166}$	0.23231
Influenza Infection	$1.9 \times 10^{-163}$	$< 1.241 \times 10^{-5}$
Influenza Viral <a href="#">RNA</a> Transcription and Replication	$1.9 \times 10^{-160}$	$< 1.241 \times 10^{-5}$
Influenza Life Cycle	$2.5 \times 10^{-156}$	$< 1.241 \times 10^{-5}$
<i>Extracellular matrix organisation</i>	$1.1 \times 10^{-152}$	0.071761
GPCR ligand binding	$1.1 \times 10^{-143}$	0.55801
Class A/1 (Rhodopsin-like receptors)	$1.5 \times 10^{-142}$	0.58901
<i>GPCR downstream signalling</i>	$7.6 \times 10^{-140}$	0.098357
Haemostasis	$1.9 \times 10^{-134}$	0.27059
Developmental Biology	$2.0 \times 10^{-123}$	0.52737
Metabolism of lipids and lipoproteins	$3.3 \times 10^{-120}$	0.724
Cytokine Signalling in Immune system	$2.6 \times 10^{-119}$	0.39661
Peptide ligand-binding receptors	$3.7 \times 10^{-109}$	0.61102
<b><math>G_{\alpha i}</math> signalling events</b>	$8.9 \times 10^{-100}$	$< 1.241 \times 10^{-5}$

Over-representation (hypergeometric test) and Permutation p-values adjusted for multiple tests across [pathways](#) (FDR). Significant [pathways](#) are marked in bold (FDR < 0.05) and italics (FDR < 0.1).

Table 4.8: Pathways for *CDH1* partners from SLIPT and siRNA primary screen

Reactome Pathway	Over-representation	Permutation
Visual phototransduction	$6.9 \times 10^{-10}$	0.91116
<b>G<sub>as</sub> signalling events</b>	$1.6 \times 10^{-7}$	0.012988
Retinoid metabolism and transport	$1.7 \times 10^{-7}$	0.20487
Transcriptional regulation of white adipocyte differentiation	$6.5 \times 10^{-6}$	0.38197
Acyl chain remodelling of PS	$6.5 \times 10^{-6}$	0.58485
Chemokine receptors bind chemokines	$6.5 \times 10^{-6}$	0.97255
<i>Defective EXT2 causes exostoses 2</i>	$6.9 \times 10^{-6}$	0.056437
<i>Defective EXT1 causes exostoses 1, TRPS2 and CHDS</i>	$6.9 \times 10^{-6}$	0.056437
Signalling by NOTCH4	$6.9 \times 10^{-6}$	0.15497
Platelet activation, signalling and aggregation	$6.9 \times 10^{-6}$	0.53358
Phase 1 - Functionalisation of compounds	$1.3 \times 10^{-5}$	0.24836
Amine ligand-binding receptors	$1.7 \times 10^{-5}$	0.3195
Acyl chain remodelling of PE	$2.4 \times 10^{-5}$	0.7307
Signalling by GPCR	$2.4 \times 10^{-5}$	0.9939
<b>Molecules associated with elastic fibres</b>	$2.6 \times 10^{-5}$	0.0072929
DAP12 interactions	$2.6 \times 10^{-5}$	0.78273
Cytochrome P <sub>450</sub> - arranged by substrate type	$3.2 \times 10^{-5}$	0.87019
GPCR ligand binding	$3.8 \times 10^{-5}$	0.99417
Acyl chain remodelling of PC	$4.0 \times 10^{-5}$	0.65415
Response to elevated platelet cytosolic Ca <sup>2+</sup>	$4.2 \times 10^{-5}$	0.55461
<i>Arachidonic acid metabolism</i>	$4.4 \times 10^{-5}$	0.060298
Defective B4GALT7 causes EDS, progeroid type	$4.9 \times 10^{-5}$	0.15497
Defective B3GAT3 causes JDSSDHD	$4.9 \times 10^{-5}$	0.15497
<b>Elastic fibre formation</b>	$4.9 \times 10^{-5}$	0.0019227
<b>HS-GAG degradation</b>	$6.2 \times 10^{-5}$	0.017747
Bile acid and bile salt metabolism	$6.2 \times 10^{-5}$	0.15497
Netrin-1 signalling	$7.1 \times 10^{-5}$	0.95056
<b>Integration of energy metabolism</b>	$7.1 \times 10^{-5}$	0.0019287
DAP12 signalling	$7.9 \times 10^{-5}$	0.67835
GPCR downstream signalling	$8.1 \times 10^{-5}$	0.88678
<b>Diseases associated with glycosaminoglycan metabolism</b>	$8.7 \times 10^{-5}$	0.017747
<b>Diseases of glycosylation</b>	$8.7 \times 10^{-5}$	0.017747
Signalling by Retinoic Acid	$8.7 \times 10^{-5}$	0.13592
Signalling by Leptin	$8.7 \times 10^{-5}$	0.15497
Signalling by SCF-KIT	$8.7 \times 10^{-5}$	0.73399
Opioid Signalling	$8.7 \times 10^{-5}$	0.99417
Signalling by NOTCH	0.0001	0.26453
Platelet homeostasis	0.0001	0.55912
Signalling by NOTCH1	0.00011	0.13797
Class B/2 (Secretin family receptors)	0.00011	0.4659
Diseases of Immune System	0.00013	0.15497
Diseases associated with the TLR signalling cascade	0.00013	0.15497
A tetrasaccharide linker sequence is required for GAG synthesis	0.00013	0.33566
Nuclear Receptor transcription <a href="#">pathways</a>	0.00016	0.22735
<b>Formation of Fibrin Clot (Clotting Cascade)</b>	0.00016	0.0054639
Syndecan interactions	0.00016	0.3974
Class A/1 (Rhodopsin-like receptors)	0.00016	0.99454
HS-GAG biosynthesis	0.0002	0.37199
Platelet degranulation	0.0002	0.39003
EPH-ephrin mediated repulsion of cells	0.00021	0.6193

Over-representation (hypergeometric test) and Permutation p-values adjusted for multiple tests across [pathways](#) (FDR). Significant [pathways](#) are marked in bold (FDR < 0.05) and italics (FDR < 0.1).

gene regulation, supporting gene set over-representation of the SLIPT-specific pathways in Table 4.7. In particular, some of the most significantly over-represented pathways had higher observed  $\chi^2$  values than any of the 1 million random permutations. Similar pathways were also replicated by permutation analysis for mtSLIPT candidate partners against *CDH1* mutation (shown in Appendix Table C.5). For the genes detected by SLIPT, the permutation approach detected many of the most strongly over-represented pathways.

The permutation approach was also applied to the intersection between computational and experimental candidates. This permutation analysis tested whether for consistent detection of pathways was independent of their pre-existing status as experimental candidates. In contrast to the SLIPT pathways (in Table 4.7), the pathways results for these candidate partners (in Table 4.8) differed considerably between over-representation and resampling analyses.

Namely, many of the over-represented pathways were not significant in the resampling analysis, including visual phototransduction and retinoic acid signalling, and were likely over-represented in the intersection due to over-representation in the siRNA candidates rather than additional support from SLIPT. Of the highest over-represented pathways in the intersection, only  $G_{\alpha s}$  signalling events were supported by both over-representation and resampling analyses. Other pathways supported by both analyses were cytoplasmic elastic fibre formation, associated HS-GAG protein modification pathways, energy metabolism, and the fibrin clotting cascade.

Many of the pathways supported in the intersection by permutation analysis were also replicated in the mtSLIPT analysis of partners tested with *CDH1* mutation (in Table C.6), including  $G_{\alpha s}$ , elastic fibres, HS-GAG, and energy metabolism. While there were differences between the pathways identified by over-representation analysis, those replicated by permutation were highly concordant, supporting the combined use of these pathways approaches to identify synthetic lethal gene functions and targets.

While this indicates that  $G_{\alpha s}$  and GPCR class A/1 signalling events were significantly detected by both approaches, GPCR signalling pathways overall were not. It is likely that GPCRs were primarily over-represented in the intersection with the experimental candidates due to strong over-representation of these pathways in experimental candidates, rather than detection by SLIPT, which may be driven by these more specific constituent pathways.

Several pathways, including some immune functions and neurotransmitters, were supported by the resampling analysis (in Table 4.8 and Appendix Table C.6) when the

initial [pathways](#) over-representation test was not significant. These functions appear to have been detected by both approaches more than expected by chance but must be interpreted with caution since they were still not common enough to be detected in [pathways](#) over-representation analysis.

#### 4.2.6 Integrating Synthetic Lethal Pathways and Screens

Based on these results, it appears that computational and experimental approaches to [synthetic lethal](#) screening for *CDH1* lead to a broader functional characterisation, and many candidate partners, when combined, despite different strengths and limitations. Compared to candidate gene approaches, experimental [genomes](#)-wide screens are an appealing unbiased strategy for identifying [synthetic lethal](#) interactions. Since these screens are costly, laborious, and specific to genetic background, computational analysis can augment candidate triage to either reduce the initial panel of screened genes or prioritise validation.

[GPCR pathways](#) were detected among both computational and experimental [synthetic lethal](#) candidates, with more support in the experimental screen (Table 4.8). The homogeneous cell line model may be more likely to detect particular [pathways](#). For instance, [SLIPT](#) identified immune [pathways](#), not expected to be detected in isolated cell culture. [GPCR](#) signalling was supported in experimental models [Telford et al. \(2015\)](#) with some of these [pathways](#) replicated in varied genetic backgrounds of patient samples. These [pathways](#) require further investigation such as identification of more specific [pathways](#), higher order interactions, and modes of resistance.

The [pathways](#) composition across computational and experimental [synthetic lethal](#) candidates was informative with over-representation (Table 4.6) and was supported by resampling analysis (Table 4.8), despite a modest intersection of genes between them (Figure 4.2). Either approach may be significant for a [pathways](#) in this intersection without being supported by the other: resampling analysis may support [pathways](#) that were not over-represented due to small effect sizes, thus both tests are required to identify candidate [pathways](#). The [pathways](#) detected by both over-representation and resampling are the strongest candidates for further investigation, such as  $G_{\alpha s}$  signalling, a strong candidate in prior analyses with a role in the regulation of translation in cancer [Gao and Roux \(2015\)](#), another function supported by [SLIPT](#) analysis.

The predicted [synthetic lethal](#) partners occurred across functionally distinct [pathways](#), including characterised functions of *CDH1*. This diversity is consistent with the wide ranging role of *CDH1* in cell-cell adhesion, cell signalling, and the cytoskeletal

structure of epithelial tissues. Pathway structure may be relevant to identifying potential drug targets from [gene expression](#) signatures, indicating downstream effector genes and mechanisms leading to cell inviability. These distinct [synthetic lethal](#) gene clusters and [pathways](#) may further lead to the elucidation of drug resistance mechanisms.

### 4.3 Synthetic Lethal Pathway Metagenes

[Metagenes](#) are a one-dimensional summary of the activity for each biological [pathway](#). The direction of [metagenes](#) (derived by the singular value matrix decomposition as described in Section 2.2.3) reflects overall activation of the [pathways](#). This has been verified by examining the expression patterns of previously published gene signatures ([Gatza \*et al.\*, 2011, 2014](#)) in Appendix D. Pathway [metagenes](#) for Reactome [pathways](#) were used for testing [synthetic lethal pathways](#). Since the [metagenes](#) values are higher when the [pathways](#) as a whole is activated, they are suitable for SLIPT analysis using low [metagene](#) levels to represent less activated [pathways](#).

The TCGA breast cancer [expression](#) data was used to generate [pathway metagenes](#) for each collection of genes in a [pathway](#) from the Reactome database ([Croft \*et al.\*, 2014](#)). These metagenes were tested against the expression of *CDH1* by SLIPT to directly detect [synthetic lethal pathways](#). These [synthetic lethal metagenes](#) differed to the over-represented [pathways](#) among [synthetic lethal](#) gene candidates. However, there were some similarities to previous findings, as shown in Table 4.9. In particular, translational [pathways](#) were replicated as observed in Table 4.2. While the specific [pathways](#) differ, immune [pathways](#) (e.g., NF- $\kappa$ B) were also supported by [metagene synthetic lethal](#) analysis.

Signalling [pathways](#) were more strongly supported by mtSLIPT analysis of [metagene pathways expression](#) against *CDH1* [mutation](#), as shown in Table D.1, although these results were generally less statistically significant than [expression](#) analyses. Signalling [pathways](#) detected as [synthetic lethal metagenes](#) include  $G_{\alpha z}$ , insulin-related growth factor (IGF), GABA receptor,  $G_{\alpha s}$ , S6K1 and various toxin responses mediated by GPCRs. Metabolic processes including processing of carbohydrates and fatty acids were also implicated across these analyses.

The [metagene](#) analyses differ more between expression and *CDH1* [mutation](#) than previous analyses, with more specific signalling [pathways](#) identified in the [mutation](#) analysis. This supports the usage of a complete null [mutant](#) model in experimental testing for [synthetic lethality](#) of signalling [pathways](#) against CDH1 inactivation rather than a knockdown in [expression](#). However, low [expression](#) of partners has been used in

Table 4.9: Examples of candidate [metagenes](#) [synthetic lethal](#) for *CDH1* from SLIPT

Pathway	ID	Observed	Expected	$\chi^2$ value	p-value	p-value (FDR)
Glycogen storage diseases	3229121	68	130	176	$6.62 \times 10^{-37}$	$1.82 \times 10^{-34}$
Myoclonic epilepsy of Lafora	3785653	68	130	176	$6.62 \times 10^{-37}$	$1.82 \times 10^{-34}$
Diseases of carbohydrate metabolism	5663084	68	130	176	$6.62 \times 10^{-37}$	$1.82 \times 10^{-34}$
Arachidonic acid metabolism	2142753	81	130	157	$8.13 \times 10^{-33}$	$1.49 \times 10^{-30}$
Translation initiation complex formation	72649	70	130	152	$7.08 \times 10^{-32}$	$1.17 \times 10^{-29}$
Synthesis of 5-eicosatetraenoic acids	2142688	68	130	151	$1.25 \times 10^{-31}$	$1.88 \times 10^{-29}$
SRP-dependent cotranslational protein targeting to membrane	1799339	69	130	150	$2.01 \times 10^{-31}$	$2.76 \times 10^{-29}$
L13a-mediated translational silencing of Ceruloplasmin <a href="#">expression</a>	156827	72	130	148	$5.91 \times 10^{-31}$	$6.44 \times 10^{-29}$
3' -UTR-mediated translational regulation	157279	72	130	148	$5.91 \times 10^{-31}$	$6.44 \times 10^{-29}$
Activation of the <a href="#">mRNA</a> upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S	72662	70	130	147	$1.14 \times 10^{-30}$	$9.28 \times 10^{-29}$
Formation of the ternary complex, and subsequently, the 43S complex	72695	70	130	147	$1.14 \times 10^{-30}$	$9.28 \times 10^{-29}$
Ribosomal scanning and start codon recognition	72702	70	130	147	$1.14 \times 10^{-30}$	$9.28 \times 10^{-29}$
Eukaryotic Translation Elongation	156842	72	130	146	$1.19 \times 10^{-30}$	$9.28 \times 10^{-29}$
Nonsense Mediated Decay independent of the Exon Junction Complex	975956	71	130	146	$1.24 \times 10^{-30}$	$9.28 \times 10^{-29}$
Viral <a href="#">mRNA</a> Translation	192823	70	130	146	$1.51 \times 10^{-30}$	$1.04 \times 10^{-28}$
Eukaryotic Translation Termination	72764	70	130	146	$1.51 \times 10^{-30}$	$1.04 \times 10^{-28}$
NF- $\kappa$ B is activated and signals survival	209560	71	130	145	$1.90 \times 10^{-30}$	$1.19 \times 10^{-28}$
Peptide chain elongation	156902	72	130	145	$1.91 \times 10^{-30}$	$1.19 \times 10^{-28}$
Influenza Life Cycle	168255	70	130	145	$1.95 \times 10^{-30}$	$1.19 \times 10^{-28}$
Formation of a pool of free 40S subunits	72689	73	130	145	$2.01 \times 10^{-30}$	$1.19 \times 10^{-28}$
Nonsense-Mediated Decay	927802	71	130	145	$2.44 \times 10^{-30}$	$1.34 \times 10^{-28}$
Nonsense Mediated Decay enhanced by the Exon Junction Complex	975957	71	130	145	$2.44 \times 10^{-30}$	$1.34 \times 10^{-28}$
GTP hydrolysis and joining of the 60S ribosomal subunit	72706	72	130	145	$2.58 \times 10^{-30}$	$1.37 \times 10^{-28}$
Influenza Viral <a href="#">RNA</a> Transcription and Replication	168273	72	130	144	$4.01 \times 10^{-30}$	$2.07 \times 10^{-28}$
Signalling by NOTCH1 HD Domain Mutants in Cancer	2691230	79	130	143	$5.99 \times 10^{-30}$	$2.82 \times 10^{-28}$

Strongest candidate [synthetic lethal](#) partners for *CDH1* by SLIPT with observed and expected numbers of TCGA breast cancer samples with low [expression](#) of both *CDH1* and the [metagene](#). The most significant 26 pathways are reported to show the diversity of candidate [synthetic lethal metagene](#) and demonstrate the highly significant results of SLIPT when performed on [pathway metagenes](#).

either case to be applicable to dose-dependent pharmacological inhibition and across genes where [mutations](#) have different functional consequences, including variants of unknown significance.

These results show an alternative [pathways](#)-based approach to detecting [synthetic lethal](#) gene functions interacting with *CDH1*. The use of [synthetic lethal metagenes](#) replicates support for these [pathways](#) independent of [pathways](#) size (as genes are weighted equally). Having verified that the direction of [metagenes](#) recapitulates the activity of a [pathways](#), these demonstrate that many of the [pathways](#) previously identified (from SLIPT candidate genes) are [synthetic lethal pathways](#), with their activity dependent on [synthetic lethal](#) genes, rather than containing [synthetic lethal](#) genes as inhibitors or peripheral regulators of the [pathways](#).

The [synthetic lethal](#) analysis against low *CDH1* [expression](#) supports prior findings in translational and immune [pathways](#) even if they were not able to detected in an experimental screen (Telford *et al.*, 2015). Together these findings support the role of *CDH1* loss in cancer disrupting cell signalling with wider effects on protein translation and metabolism necessary for the proliferation of cancer cells. This is consistent with

the [GPCR pathways](#), such as  $G_{\alpha s}$  signalling, being supported by [SLIPT](#) gene candidates and the experimental primary [siRNA](#) screen, as shown by resampling in Section 4.2.5.1.

## 4.4 Replication in Stomach Cancer

*CDH1* is also important in stomach cancer biology as a [driver tumour suppressor](#) gene, including as a [germline mutation](#) in many cases of [hereditary](#) diffuse gastric cancer. The [synthetic lethal](#) analysis of genes and [pathways](#) (previously identified for [TCGA](#) breast cancer data) was replicated in [TCGA](#) stomach cancer. The accompanying data for [SLIPT](#) analysis against *CDH1* expression is provided in Appendix F.

While the sample size was lower for [TCGA](#) stomach cancer (particularly for [mutations](#)), the results serve to support the findings in breast cancer in an independent patient cohort and tissue samples. The molecular profiling, including [RNA-Seq expression](#), were performed by [TCGA](#) using the sample procedures as for breast cancer and the findings reported here were performed using data analysis techniques identical to those presented previously. These procedures should ensure as close a comparison as feasible across both of the cancer types most relevant to [HDGC](#) and recurrent *CDH1* mutations.

The strongest [SLIPT](#) genes for stomach cancer (shown in Table F.1) did not necessarily directly correspond to those observed in breast cancer (shown in Appendix Table 4.1). However, several gene functions were replicated in stomach cancer. Together, these gene candidates indicate widespread functions of *CDH1* and strongly detectable [synthetic lethality](#) with many genes, using a strategy that can be applied across cancer types. More specifically, the signalling genes included [GPCR](#) signalling genes, which was one of the most supported [synthetic lethal pathways](#) in breast cancer analysis and the experimental screen (Telford *et al.*, 2015). These findings were further supported by the [pathways](#) over-represented in [SLIPT](#) candidates from [TCGA](#) stomach cancer (shown in Appendix Table F.2) which replicated the translational and immune [pathways](#) observed in [TCGA](#) breast cancer (shown in Table 4.2) and further supported [GPCR](#) signalling [pathways](#), including the class A/1 receptors. The extracellular matrix was also detected at the [pathways](#) level in stomach cancer, including elastic fibres, glycosylation, collagen, and integrin cell-surface interactions. While fewer [pathways](#) were supported by resampling for the intersection of [SLIPT](#) and experimental screen candidate partners in stomach cancer than breast cancer, many of those detected (shown in Appendix Table F.6) replicate those detected in breast cancer (shown in Table 4.8). The [pathways](#) detected by both permutation and over-representation



analysis were more likely to be replicated across stomach and breast cancer than those detected by over-representation alone, supporting the use of this procedure to detect [synthetic lethal pathways](#) applicable across cancer types. The include  $G_{\alpha s}$  signalling and elastic fibre formation as discussed for breast cancer (in Section 4.2.5.1).

## 4.5 Discussion

### 4.5.1 Strengths of the SLIPT Methodology

Synthetic lethal discovery with [SLIPT](#) used established statistical procedures to identify putative partner genes from [gene expression](#) data. Such use of the  $\chi^2$ -value is amenable to [pathways](#) or permutation analyses and could feasibly be applied to other disease gene or pair-wise across the [genomes](#), although previous [genomes](#)-wide approaches were unable to find informative candidate genes for E-cadherin (Lu *et al.*, 2015). [Synthetic lethal](#) discovery in cancer has focused on genes with severe cellular [mutant](#) phenotypes, such as [essential](#) genes or the [oncogenes](#) *TP53* and *AKT* (Lu *et al.*, 2015; Tiong *et al.*, 2014; Wang and Simon, 2013), with other cancer genes, such as *CDH1*, requiring more focused investigations. Prior computational approaches for [synthetic lethal](#) discovery, in cancer, vary widely (Jerby-Arnon *et al.*, 2014; Lu *et al.*, 2015; Tiong *et al.*, 2014; Wappett *et al.*, 2016). There is no consensus as to which approach is more appropriate, and the methods are difficult to compare, as they either do not have a released code implementation or do not make predictions solely from normalised [expression](#) data.

The query-based approach demonstrated by [SLIPT](#) analysis is suitable for wider application on [expression](#) data and for augmenting experimental studies such as high-throughput screens. This approach has identified biologically plausible [synthetic lethal pathways](#) for *CDH1*, triaged candidates from experimental screening (Telford *et al.*, 2015), and replicates genes and [pathways](#) across breast and stomach cancer datasets. In addition, [SLIPT](#) avoids critical assumptions underlying the design of some approaches such as co-expression of synthetic candidates or that interacting gene pairs will have known (annotated) similarities in function.

The DAISY methodology Jerby-Arnon *et al.* (2014), which took a similar query-based approach with the [tumour suppressor](#) *VHL*, has been critiqued for being too stringent (Lu *et al.*, 2015) which impedes [pathways](#) analysis. Since [functional redundancy](#) does not require genes to be expressed at the same time, the [SLIPT](#) approach does not assume co-expression of [synthetic lethal](#) genes which may enrich for [synthetic lethal](#) genes in established coregulated [pathways](#). Rather, the interpretation of



synthetic lethality for SLIPT was similar to other computational methods based on ‘co-loss under-representation’, ‘compensation’, or ‘simultaneous differential expression’ (Lu *et al.*, 2015; Tiong *et al.*, 2014; Wang and Simon, 2013).

Genomics analyses are prone to false-positives and require statistical caution, particularly where working with gene-pairs scale up the number of multiple tests drastically, at the expense of statistical power. Experimental screens for synthetic lethality are also error-prone (Fece de la Cruz *et al.*, 2015; Lord *et al.*, 2015; Lu *et al.*, 2015), especially with false-positives, raising the need for understanding the expected behaviour and number of functional relationships and genetic interactions in the genomes, or in discovery of synthetic lethal partners of a particular query gene. Thus analyses throughout this thesis have focused on querying for partners of a particular gene of interest. Statistical modelling and simulations (in Section 3.3 and Chapter 6) will further support the design decisions underlying SLIPT analysis and its strengths over other approaches.

#### 4.5.2 Synthetic Lethal Pathways for E-cadherin

Specific genes were difficult to replicate across experiments. This is consistent with gene expression profiles for synthetic lethal partners reflecting the complexity of biological pathways which are subject to higher-order interactions and do not consistently compensate for loss of gene function across all samples (Jerby-Arnon *et al.*, 2014; Kelly, 2013; Lu *et al.*, 2015). The predicted synthetic lethal partners of *CDH1* (with FDR correction) were investigated with gene expression profiles and clinical variables to find relationships in gene expression, gene function, and clinical characteristics. The large number of genes detected indicates that synthetic lethal detection is potentially error-prone, and that identifying genes relevant for clinical application will be difficult without a supporting biological pathways rationale. As such, investigations into the genes identified by SLIPT, the correlation structure between them, and those which were validated by experimental screening (Telford *et al.*, 2015) focused at the pathways level throughout this Chapter. Similarly, comparisons across analyses were largely made at the pathways level, including comparisons between expression and mutation, breast and stomach TCGA datasets.

Potential synthetic lethal partners of *CDH1* identified by SLIPT had many distinct functions, with each gene cluster highly expressed in different patient subgroups (Figure 4.1). The expression profiles of the SL partners of *CDH1* predicted from TCGA breast cancer RNA-Seq data (expected to have compensating high or stable expression)

and their corresponding functional enrichment found in subgroups of genes, particularly among *CDH1* low breast tumours. Ductal breast cancers showed higher [expression](#) of [synthetic lethal](#) partners, suggesting treatment may be more effective in this tumour subtype. However, there was consistently low [expression](#) of SL partners in estrogen receptor negative tumours, although this is independent of tumour stage and consistent with poor prognosis in these patients and could inform other treatment strategies or prevent ineffective treatment further impacting quality of life in these patients. These results suggest that [synthetic lethal](#) partner [expression](#) varies between patients; that these different tumour classes would react differently to the same treatment; that treatment of different [pathways](#) and combinations in different patients is the most effective approach to target genes compensating for *CDH1* gene loss; and that the [expression](#) of synthetic partners could be a clinically important biomarker.

The [pathways](#) that [synthetic lethal](#) partners of *CDH1* identified by SLIPT were involved in a diverse range of biological functions and differed to those detected experimentally. This discrepancy may be accounted for by [gene expression](#) analyses detecting both [synthetic lethal](#) partners, as screened for experimentally by Telford *et al.* (2015), and their downstream targets (not detected by siRNA), capturing the wider [pathways](#) and mechanisms involved in [synthetic lethality](#) with *CDH1* inactivation. In particular, GPCR phosphorylation cascades (which regulate [gene expression](#) and translation in cancers (Gao and Roux, 2015)) were predicted to be [synthetic lethal](#) with *CDH1*. The predicted [synthetic lethal](#) partners occurred across functionally distinct [pathways](#), including characterised functions of *CDH1*. The most consistently supported [pathways](#) included elastic fibres in the extracellular matrix, GPCR signalling, and translation presenting vulnerabilities for *CDH1* deficient cancer cells from extracellular stimuli to the core growth mechanisms of a cell.

This diversity in [synthetic lethal](#) functions is consistent with the wide ranging role of *CDH1* in cell-cell adhesion, cell signalling, and the cytoskeletal structure of epithelial tissues. Pathway structure may be relevant to identifying potential drug targets from [gene expression](#) signatures, indicating downstream effector genes and mechanisms leading to cell inviability. Identification of distinct [synthetic lethal](#) gene clusters may further lead to the elucidation of drug resistance mechanisms. While these [pathways](#) are indicative of the main functions of E-cadherin and [synthetic lethal](#) partners, it remains to identify the genes within these [pathways](#) that are the most actionable or supported across SLIPT analysis in patient samples and detected by experiments in preclinical models (Chen *et al.*, 2014; Telford *et al.*, 2015). The specific genes within

key [pathways](#) will be discussed in Chapter 5, along with further investigations into their relation to [pathway](#) structure. While these are important clinical implications, the [synthetic lethal](#) predictions lack enough confidence for direct translation into pre-clinical models or clinical applications leading to a need for statistical modelling and simulation of [synthetic lethality](#) in [genomics expression](#) data.

These [synthetic lethal pathways](#) have potential clinical implications, particularly those supported in pre-clinical models and in patient [expression](#) data. However, further validation of gene candidates will be necessary to ensure that these are able to be reproduced in further pre-clinical studies, they are applicable to tumours *in vivo*, and that effective inhibitory agents can be repurposed or designed against them.

### 4.5.3 Replication and Validation

#### 4.5.3.1 Integration with siRNA Screening

The [pathways](#) composition across computational and experimental [synthetic lethal](#) candidates was informative with over-representation (Table 4.6) and supported by resampling analysis (Table 4.8), despite a modest intersection of genes between them (Figure 4.2). Either approach may be significant for a [pathways](#) in this intersection without being supported by the other: resampling analysis may support [pathways](#) that were not over-represented due to small effect sizes, thus both tests are required for a candidate [pathways](#).

The [pathways](#) detected by both over-representation and resampling are the strongest candidates for further investigation and the [pathway](#) structure analyses in Chapter 5 will focus on these [pathways](#) detected by both over-representation and resampling. Particularly, those replicated across datasets or with [pathways metagenes](#). In addition to GCPR [pathways](#) detected across these analyses, the [phosphoinositide 3-kinase \(PI3K\)](#) cascade will also be investigated in Chapter 5, this signalling [pathways](#) is a well characterised mediator between GCPR receptors and regulation of translation (Gao and Roux, 2015) (both detected throughout this Chapter) and exhibited unexpected behaviour with [pathways](#) the [metagenes](#) (in Section D). This [pathways](#) is activated by protein phosphorylation states and thus inactivation may not be detectable with [expression](#).

The SLIPT approach was shown to be predictive of which [siRNA](#) primary screen candidate partners of *CDH1* were validated in a secondary screen (as shown in Section 4.2.4). These results further support SLIPT for identifying robust [synthetic lethal](#)

candidates which can be validated and as a triage approach for interpreting screening experiments.

#### 4.5.3.2 Replication across Tissues

Furthermore, [synthetic lethal](#) partners identified by [SLIPT](#) were replicated across breast and stomach cancer. These were particularly concordant at the [pathways](#) level, as expected between tissues, since [synthetic lethal pathways](#) have higher conservation between species ([Dixon \*et al.\*, 2008](#)). These findings support gene functions conserved across *CDH1* deficient cancers in breast and stomach tissues, presenting vulnerabilities that could be applied against molecular targets in both cancers. In addition, these analyses serve as a replication across independent patient cohorts from breast and stomach cancers, decreasing the likelihood of the [synthetic lethal pathways](#) detected being false positives or artifacts of either dataset.

Synthetic lethal [pathways](#) were also replicated across [expression](#) analyses of [TCGA](#) patient samples in heterogeneous tumours and homogeneous cell line isolates. This further supports that the subset of [synthetic lethal](#) functions detectable in experimental models ([Chen \*et al.\*, 2014](#); [Telford \*et al.\*, 2015](#)) would be applicable tumours of patients with *CDH1* deficient cancers.

There are many gene functions replicated across breast cancer [gene expression](#) analyses. Many of these were also replicated with [mutation](#) analysis and with stomach cancer or cell line [expression](#) data. These [pathways](#) were more consistent across replication analyses than previous investigations with [TCGA microarray](#) data ([Kelly, 2013](#)).

## 4.6 Summary

We have developed a simple, interpretable, computational approach to predict [synthetic lethal](#) partners from [genomics](#) data. The analyses focus on [gene expression](#) data as it is widely available for applications in other cancers and other disease genes, particularly those with malignant loss of function.

This approach has been applied to robustly detect [synthetic lethal pathways](#) for [E-cadherin](#) (the *CDH1* gene) in [TCGA](#) breast cancer molecular profiles, with comparisons to experimental screening ([Telford \*et al.\*, 2015](#)) in cell lines and replication in [TCGA](#) stomach cancer molecular profiles. The [pathways](#) replicated across several analyses included extracellular matrix [pathways](#) (e.g., elastic fibres formation), cell signalling (including [GPCRs](#)), and core gene regulation and translation processes crucial for the growth and proliferation of cancer cells. These [pathways](#) show evidence of [non-oncogene addiction](#) for *CDH1* deficient cells, which presents vulnerabilities that could

be exploited for specific treatment against *CDH1* mutations in HCGC and sporadic cancers. There was also support for synthetic lethal pathways with *CDH1* in cell adhesion and cytoskeletal processes to which *CDH1* belongs, supporting the finding that synthetic lethality occurs within biological pathways (Boone *et al.*, 2007; Kelley and Ideker, 2005).

While translational and immune pathways detected by SLIPT were not supported by primary siRNA screening (Telford *et al.*, 2015), these were replicated across various analyses. Due to the differences between an experimental cell line model (Chen *et al.*, 2014; Fece de la Cruz *et al.*, 2015) and patient molecular profiles (Bass *et al.*, 2014; Koboldt *et al.*, 2012), these would not be expected to be completely concordant. Furthermore, many pathways are difficult to test in an isolated experimental system. Nevertheless, many of the genes and pathways detected by SLIPT are suitable to inform further investigations and triage of therapeutic targets against *CDH1* deficient tumours in combination with experimental screening.

Thus synthetic lethal pathways have been identified using TCGA patient molecular profiles and experimental screening results. Some of these were robustly replicated across these datasets and against *CDH1* mutation or expression analysis. However, there remains the need to identify actionable genes within these pathways, relationships with experimental candidates, and how these pathways may affect viability when lost. While the genes identified between these analyses were less concordant the results of the TCGA breast cancer analysis will be used to test pathway structure relationships and further examine the synthetic lethal genes detected in the following Chapter.

# Bibliography

- Aarts, M., Bajrami, I., Herrera-Abreu, M.T., Elliott, R., Brough, R., Ashworth, A., Lord, C.J., and Turner, N.C. (2015) Functional genetic screen identifies increased sensitivity to weel inhibition in cells with defects in fanconi anemia and hr pathways. *Mol Cancer Ther*, **14**(4): 865–76.
- Abeshouse, A., Ahn, J., Akbani, R., Ally, A., Amin, S., Andry, C.D., Annala, M., Aprikian, A., Armenia, J., Arora, A., *et al.* (2015) The Molecular Taxonomy of Primary Prostate Cancer. *Cell*, **163**(4): 1011–1025.
- Adler, D. (2005) *vioplot: Violin plot*. R package version 0.2.
- Akbani, R., Akdemir, K.C., Aksoy, B.A., Albert, M., Ally, A., Amin, S.B., Arachchi, H., Arora, A., Auman, J.T., Ayala, B., *et al.* (2015) Genomic Classification of Cutaneous Melanoma. *Cell*, **161**(7): 1681–1696.
- Akobeng, A.K. (2007) Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Pdiatrica*, **96**(5): 644–647.
- American Cancer Society (2017) Genetics and cancer. <https://www.cancer.org/cancer/cancer-causes/genetics.html>. Accessed: 22/03/2017.
- Anjomshoaa, A., Lin, Y.H., Black, M.A., McCall, J.L., Humar, B., Song, S., Fukuzawa, R., Yoon, H.S., Holzmann, B., Friederichs, J., *et al.* (2008) Reduced expression of a gene proliferation signature is associated with enhanced malignancy in colon cancer. *Br J Cancer*, **99**(6): 966–973.
- Araki, H., Knapp, C., Tsai, P., and Print, C. (2012) GeneSetDB: A comprehensive meta-database, statistical and visualisation framework for gene set analysis. *FEBS Open Bio*, **2**: 76–82.

- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, **25**(1): 25–29.
- Ashworth, A. (2008) A synthetic lethal therapeutic approach: poly(adp) ribose polymerase inhibitors for the treatment of cancers deficient in dna double-strand break repair. *J Clin Oncol*, **26**(22): 3785–90.
- Ashworth, A., Lord, C.J., and Reis-Filho, J.S. (2011) Genetic interactions in cancer progression and treatment. *Cell*, **145**(1): 30–38.
- Audeh, M.W., Carmichael, J., Penson, R.T., Friedlander, M., Powell, B., Bell-McGuinn, K.M., Scott, C., Weitzel, J.N., Oaknin, A., Loman, N., *et al.* (2010) Oral poly(adp-ribose) polymerase inhibitor olaparib in patients with *BRCA1* or *BRCA2* mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet*, **376**(9737): 245–51.
- Babyak, M.A. (2004) What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med*, **66**(3): 411–21.
- Bamford, S., Dawson, E., Forbes, S., Clements, J., Pettett, R., Dogan, A., Flanagan, A., Teague, J., Futreal, P.A., Stratton, M.R., *et al.* (2004) The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer*, **91**(2): 355–358.
- Barabási, A.L. and Albert, R. (1999) Emergence of scaling in random networks. *Science*, **286**(5439): 509–12.
- Barabási, A.L., Gulbahce, N., and Loscalzo, J. (2011) Network medicine: a network-based approach to human disease. *Nat Rev Genet*, **12**(1): 56–68.
- Barabási, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell’s functional organization. *Nat Rev Genet*, **5**(2): 101–13.
- Barrat, A. and Weigt, M. (2000) On the properties of small-world network models. *The European Physical Journal B - Condensed Matter and Complex Systems*, **13**(3): 547–560.

- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehar, J., Kryukov, G.V., Sonkin, D., *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**(7391): 603–607.
- Barry, W.T. (2016) *safe: Significance Analysis of Function and Expression*. R package version 3.14.0.
- Baryshnikova, A., Costanzo, M., Dixon, S., Vizeacoumar, F.J., Myers, C.L., Andrews, B., and Boone, C. (2010a) Synthetic genetic array (sga) analysis in *saccharomyces cerevisiae* and *schizosaccharomyces pombe*. *Methods Enzymol*, **470**: 145–79.
- Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.Y., Ou, J., San Luis, B.J., Bandyopadhyay, S., *et al.* (2010b) Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat Meth*, **7**(12): 1017–1024.
- Bass, A.J., Thorsson, V., Shmulevich, I., Reynolds, S.M., Miller, M., Bernard, B., Hinoue, T., Laird, P.W., Curtis, C., Shen, H., *et al.* (2014) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*, **513**(7517): 202–209.
- Bates, D. and Maechler, M. (2016) *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.2-7.1.
- Bateson, W. and Mendel, G. (1909) *Mendel's principles of heredity, by W. Bateson*. University Press, Cambridge [Eng.].
- Becker, K.F., Atkinson, M.J., Reich, U., Becker, I., Nekarda, H., Siewert, J.R., and Hfler, H. (1994) E-cadherin gene mutations provide clues to diffuse type gastric carcinomas. *Cancer Research*, **54**(14): 3845–3852.
- Bell, D., Berchuck, A., Birrer, M., Chien, J., Cramer, D., Dao, F., Dhir, R., DiSaia, P., Gabra, H., Glenn, P., *et al.* (2011) Integrated genomic analyses of ovarian carcinoma. *Nature*, **474**(7353): 609–615.
- Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, **57**(1): 289–300.



- Berx, G., Cleton-Jansen, A.M., Nollet, F., de Leeuw, W.J., van de Vijver, M., Cornelisse, C., and van Roy, F. (1995) E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers. *EMBO J*, **14**(24): 6107–15.
- Berx, G., Cleton-Jansen, A.M., Strumane, K., de Leeuw, W.J., Nollet, F., van Roy, F., and Cornelisse, C. (1996) E-cadherin is inactivated in a majority of invasive human lobular breast cancers by truncation mutations throughout its extracellular domain. *Oncogene*, **13**(9): 1919–25.
- Berx, G. and van Roy, F. (2009) Involvement of members of the cadherin superfamily in cancer. *Cold Spring Harb Perspect Biol*, **1**: a003129.
- Bitler, B.G., Aird, K.M., Garipov, A., Li, H., Amatangelo, M., Kossenkova, A.V., Schultz, D.C., Liu, Q., Shih, Ie, M., Conejo-Garcia, J.R., *et al.* (2015) Synthetic lethality by targeting ezh2 methyltransferase activity in arid1a-mutated cancers. *Nat Med*, **21**(3): 231–8.
- Blake, J.A., Christie, K.R., Dolan, M.E., Drabkin, H.J., Hill, D.P., Ni, L., Sitnikov, D., Burgess, S., Buza, T., Gresham, C., *et al.* (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res*, **43**(Database issue): D1049–1056.
- Boettcher, M., Lawson, A., Ladenburger, V., Fredebohm, J., Wolf, J., Hoheisel, J.D., Frezza, C., and Shlomi, T. (2014) High throughput synthetic lethality screen reveals a tumorigenic role of adenylate cyclase in fumarate hydratase-deficient cancer cells. *BMC Genomics*, **15**: 158.
- Boone, C., Bussey, H., and Andrews, B.J. (2007) Exploring genetic interactions and networks with yeast. *Nat Rev Genet*, **8**(6): 437–49.
- Borgatti, S.P. (2005) Centrality and network flow. *Social Networks*, **27**(1): 55 – 71.
- Boucher, B. and Jenna, S. (2013) Genetic interaction networks: better understand to better predict. *Front Genet*, **4**: 290.
- Bozovic-Spasojevic, I., Azambuja, E., McCaskill-Stevens, W., Dinh, P., and Cardoso, F. (2012) Chemoprevention for breast cancer. *Cancer treatment reviews*, **38**(5): 329–339.
- Breiman, L. (2001) Random forests. *Machine Learning*, **45**(1): 5–32.

- Brin, S. and Page, L. (1998) The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, **30**(1): 107 – 117.
- Brückner, A., Polge, C., Lentze, N., Auerbach, D., and Schlattner, U. (2009) Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci*, **10**(6): 2763–2788.
- Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., and Helleday, T. (2005) Specific killing of *BRCA2*-deficient tumours with inhibitors of polyadprbose polymerase. *Nature*, **434**(7035): 913–7.
- Bussey, H., Andrews, B., and Boone, C. (2006) From worm genetic networks to complex human diseases. *Nat Genet*, **38**(8): 862–3.
- Butland, G., Babu, M., Diaz-Mejia, J.J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A.G., Pogoutse, O., *et al.* (2008) esga: E. coli synthetic genetic array analysis. *Nat Methods*, **5**(9): 789–95.
- Cardiff, R.D., Couto, S., and Bolon, B. (2011) Three interrelated themes in current breast cancer research: gene addiction, phenotypic plasticity, and cancer stem cells. *Breast Cancer Res*, **13**(5): 216.
- cBioPortal for Cancer Genomics (cBioPortal) (2017) cBioPortal for Cancer Genomics. <http://www.cbioportal.org/>. Accessed: 26/03/2017.
- Cerami, E.G., Gross, B.E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N., Schultz, N., Bader, G.D., and Sander, C. (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res*, **39**(Database issue): D685–690.
- Chen, A., Beetham, H., Black, M.A., Priya, R., Telford, B.J., Guest, J., Wiggins, G.A.R., Godwin, T.D., Yap, A.S., and Guilford, P.J. (2014) E-cadherin loss alters cytoskeletal organization and adhesion in non-malignant breast cells but is insufficient to induce an epithelial-mesenchymal transition. *BMC Cancer*, **14**(1): 552.
- Chen, S. and Parmigiani, G. (2007) Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol*, **25**(11): 1329–1333.
- Chipman, K. and Singh, A. (2009) Predicting genetic interactions with random walks on biological networks. *BMC Bioinformatics*, **10**(1): 17.

- Christofori, G. and Semb, H. (1999) The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene. *Trends in Biochemical Sciences*, **24**(2): 73 – 76.
- Ciriello, G., Gatza, M.L., Beck, A.H., Wilkerson, M.D., Rhie, S.K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., *et al.* (2015) Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*, **163**(2): 506–519.
- Clark, M.J. (2004) Endogenous Regulator of G Protein Signaling Proteins Suppress G  $\alpha$ -Dependent  $\mu$ -Opioid Agonist-Mediated Adenylyl Cyclase Supersensitization. *Journal of Pharmacology and Experimental Therapeutics*, **310**(1): 215–222.
- Collingridge, D.S. (2013) A primer on quantitized data analysis and permutation testing. *Journal of Mixed Methods Research*, **7**(1): 81–97.
- Collins, F.S. and Barker, A.D. (2007) Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci Am*, **296**(3): 50–57.
- Collisson, E., Campbell, J., Brooks, A., Berger, A., Lee, W., Chmielecki, J., Beer, D., Cope, L., Creighton, C., Danilova, L., *et al.* (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, **511**(7511): 543–550.
- Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S., *et al.* (2010) The genetic landscape of a cell. *Science*, **327**(5964): 425–31.
- Costanzo, M., Baryshnikova, A., Myers, C.L., Andrews, B., and Boone, C. (2011) Charting the genetic interaction map of a cell. *Curr Opin Biotechnol*, **22**(1): 66–74.
- Courtney, K.D., Corcoran, R.B., and Engelman, J.A. (2010) The PI3K pathway as drug target in human cancer. *J Clin Oncol*, **28**(6): 1075–1083.
- Creighton, C.J., Morgan, M., Gunaratne, P.H., Wheeler, D.A., Gibbs, R.A., Robertson, A., Chu, A., Beroukhim, R., Cibulskis, K., Signoretti, S., *et al.* (2013) Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*, **499**(7456): 43–49.
- Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M.R., *et al.* (2014) The Reactome pathway knowledge-base. *Nucleic Acids Res*, **42**(database issue): D472D477.

- Crunkhorn, S. (2014) Cancer: Predicting synthetic lethal interactions. *Nat Rev Drug Discov*, **13**(11): 812.
- Csardi, G. and Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal, Complex Systems*: 1695.
- Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., and Shi, B. (2015) Breast cancer intrinsic subtype classification, clinical use and future trends. *Am J Cancer Res*, **5**(10): 2929–2943.
- Davierwala, A.P., Haynes, J., Li, Z., Brost, R.L., Robinson, M.D., Yu, L., Mnaimneh, S., Ding, H., Zhu, H., Chen, Y., *et al.* (2005) The synthetic genetic interaction spectrum of essential genes. *Nat Genet*, **37**(10): 1147–1152.
- De Leeuw, W.J., Berx, G., Vos, C.B., Peterse, J.L., Van de Vijver, M.J., Litvinov, S., Van Roy, F., Cornelisse, C.J., and Cleton-Jansen, A.M. (1997) Simultaneous loss of E-cadherin and catenins in invasive lobular breast cancer and lobular carcinoma in situ. *J Pathol*, **183**(4): 404–11.
- De Santis, G., Miotti, S., Mazzi, M., Canevari, S., and Tomassetti, A. (2009) E-cadherin directly contributes to PI3K/AKT activation by engaging the PI3K-p85 regulatory subunit to adherens junctions of ovarian carcinoma cells. *Oncogene*, **28**(9): 1206–1217.
- Demir, E., Babur, O., Rodchenkov, I., Aksoy, B.A., Fukuda, K.I., Gross, B., Sumer, O.S., Bader, G.D., and Sander, C. (2013) Using biological pathway data with Pax-tools. *PLoS Comput Biol*, **9**(9): e1003194.
- Deshpande, R., Asiedu, M.K., Klebig, M., Sutor, S., Kuzmin, E., Nelson, J., Piotrowski, J., Shin, S.H., Yoshida, M., Costanzo, M., *et al.* (2013) A comparative genomic approach for identifying synthetic lethal interactions in human cancer. *Cancer Res*, **73**(20): 6128–36.
- Dickson, D. (1999) Wellcome funds cancer database. *Nature*, **401**(6755): 729.
- Dijkstra, E.W. (1959) A note on two problems in connexion with graphs. *Numerische Mathematik*, **1**(1): 269–271.
- Dixon, S.J., Andrews, B.J., and Boone, C. (2009) Exploring the conservation of synthetic lethal genetic interaction networks. *Commun Integr Biol*, **2**(2): 78–81.

- Dixon, S.J., Fedyshyn, Y., Koh, J.L., Prasad, T.S., Chahwan, C., Chua, G., Toufighi, K., Baryshnikova, A., Hayles, J., Hoe, K.L., *et al.* (2008) Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. *Proc Natl Acad Sci U S A*, **105**(43): 16653–8.
- Dorsam, R.T. and Gutkind, J.S. (2007) G-protein-coupled receptors and cancer. *Nat Rev Cancer*, **7**(2): 79–94.
- Erdős, P. and Rényi, A. (1959) On random graphs I. *Publ Math Debrecen*, **6**: 290–297.
- Erdős, P. and Rényi, A. (1960) On the evolution of random graphs. *Publ Math Inst Hung Acad Sci*, **5**(1): 17–61.
- Eroles, P., Bosch, A., Perez-Fidalgo, J.A., and Lluch, A. (2012) Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. *Cancer Treat Rev*, **38**(6): 698–707.
- Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., *et al.* (2005) Targeting the dna repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*, **434**(7035): 917–21.
- Fawcett, T. (2006) An introduction to ROC analysis. *Pattern Recognition Letters*, **27**(8): 861 – 874. {ROC} Analysis in Pattern Recognition.
- Fece de la Cruz, F., Gapp, B.V., and Nijman, S.M. (2015) Synthetic lethal vulnerabilities of cancer. *Annu Rev Pharmacol Toxicol*, **55**: 513–531.
- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*, **136**(5): E359–386.
- Fisher, R.A. (1919) Xv.the correlation between relatives on the supposition of mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, **52**(02): 399–433.
- Fong, P.C., Boss, D.S., Yap, T.A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O’Connor, M.J., *et al.* (2009) Inhibition of poly(adp-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med*, **361**(2): 123–34.

- Fong, P.C., Yap, T.A., Boss, D.S., Carden, C.P., Mergui-Roelvink, M., Gourley, C., De Greve, J., Lubinski, J., Shanley, S., Messiou, C., *et al.* (2010) Poly(adp)-ribose polymerase inhibition: frequent durable responses in BRCA carrier ovarian cancer correlating with platinum-free interval. *J Clin Oncol*, **28**(15): 2512–9.
- Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., *et al.* (2015) COSMIC: exploring the world’s knowledge of somatic mutations in human cancer. *Nucleic Acids Res*, **43**(Database issue): D805–811.
- Fraser, A. (2004) Towards full employment: using RNAi to find roles for the redundant. *Oncogene*, **23**(51): 8346–52.
- Friedman, N., Linial, M., Nachman, I., and Pe’er, D. (2000) Using Bayesian networks to analyze expression data. *J Comput Biol*, **7**(3-4): 601–620.
- Fromental-Ramain, C., Warot, X., Lakkaraju, S., Favier, B., Haack, H., Birling, C., Dierich, A., Doll e, P., and Chambon, P. (1996) Specific and redundant functions of the paralogous Hoxa-9 and Hoxd-9 genes in forelimb and axial skeleton patterning. *Development*, **122**(2): 461–472.
- Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004) A census of human cancer genes. *Nat Rev Cancer*, **4**(3): 177–183.
- Futreal, P.A., Kasprzyk, A., Birney, E., Mullikin, J.C., Wooster, R., and Stratton, M.R. (2001) Cancer and genomics. *Nature*, **409**(6822): 850–852.
- Gao, B. and Roux, P.P. (2015) Translational control by oncogenic signaling pathways. *Biochimica et Biophysica Acta*, **1849**(7): 753–65.
- Gatza, M.L., Kung, H.N., Blackwell, K.L., Dewhirst, M.W., Marks, J.R., and Chi, J.T. (2011) Analysis of tumor environmental response and oncogenic pathway activation identifies distinct basal and luminal features in HER2-related breast tumor subtypes. *Breast Cancer Res*, **13**(3): R62.
- Gatza, M.L., Lucas, J.E., Barry, W.T., Kim, J.W., Wang, Q., Crawford, M.D., Datto, M.B., Kelley, M., Mathey-Prevot, B., Potti, A., *et al.* (2010) A pathway-based classification of human breast cancer. *Proc Natl Acad Sci USA*, **107**(15): 6994–6999.

- Gatza, M.L., Silva, G.O., Parker, J.S., Fan, C., and Perou, C.M. (2014) An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer. *Nat Genet*, **46**(10): 1051–1059.
- Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, **5**(10): R80.
- Genz, A. and Bretz, F. (2009) Computation of multivariate normal and t probabilities. In *Lecture Notes in Statistics*, volume 195. Springer-Verlag, Heidelberg.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2016) *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-5. URL.
- Glaire, M.A., Brown, M., Church, D.N., and Tomlinson, I. (2017) Cancer predisposition syndromes: lessons for truly precision medicine. *J Pathol*, **241**(2): 226–235.
- Globus (Globus) (2017) Research data management simplified. <https://www.globus.org/>. Accessed: 25/03/2017.
- Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*, **17**(6): 333–351.
- Grady, W.M., Willis, J., Guilford, P.J., Dunbier, A.K., Toro, T.T., Lynch, H., Wiesner, G., Ferguson, K., Eng, C., Park, J.G., *et al.* (2000) Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer. *Nat Genet*, **26**(1): 16–17.
- Graziano, F., Humar, B., and Guilford, P. (2003) The role of the E-cadherin gene (*CDH1*) in diffuse gastric cancer susceptibility: from the laboratory to clinical practice. *Annals of Oncology*, **14**(12): 1705–1713.
- Guaragnella, N., Palermo, V., Galli, A., Moro, L., Mazzoni, C., and Giannattasio, S. (2014) The expanding role of yeast in cancer research and diagnosis: insights into the function of the oncosuppressors p53 and BRCA1/2. *FEMS Yeast Res*, **14**(1): 2–16.
- Güell, O., Sagus, F., and Serrano, M. (2014) Essential plasticity and redundancy of metabolism unveiled by synthetic lethality analysis. *PLoS Comput Biol*, **10**(5): e1003637.

- Guilford, P., Hopkins, J., Harraway, J., McLeod, M., McLeod, N., Harawira, P., Taite, H., Scoular, R., Miller, A., and Reeve, A.E. (1998) E-cadherin germline mutations in familial gastric cancer. *Nature*, **392**(6674): 402–5.
- Guilford, P., Humar, B., and Blair, V. (2010) Hereditary diffuse gastric cancer: translation of *CDH1* germline mutations into clinical practice. *Gastric Cancer*, **13**(1): 1–10.
- Guilford, P.J., Hopkins, J.B., Grady, W.M., Markowitz, S.D., Willis, J., Lynch, H., Rajput, A., Wiesner, G.L., Lindor, N.M., Burgart, L.J., *et al.* (1999) E-cadherin germline mutations define an inherited cancer syndrome dominated by diffuse gastric cancer. *Hum Mutat*, **14**(3): 249–55.
- Guo, J., Liu, H., and Zheng, J. (2016) SynLethDB: synthetic lethality database toward discovery of selective and sensitive anticancer drug targets. *Nucleic Acids Res*, **44**(D1): D1011–1017.
- Hajian-Tilaki, K. (2013) Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian J Intern Med*, **4**(2): 627–635.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I.H. (2009) The weka data mining software: an update. *SIGKDD Explor Newsl*, **11**(1): 10–18.
- Hammerman, P.S., Lawrence, M.S., Voet, D., Jing, R., Cibulskis, K., Sivachenko, A., Stojanov, P., McKenna, A., Lander, E.S., Gabriel, S., *et al.* (2012) Comprehensive genomic characterization of squamous cell lung cancers. *Nature*, **489**(7417): 519–525.
- Hanahan, D. and Weinberg, R.A. (2000) The hallmarks of cancer. *Cell*, **100**(1): 57–70.
- Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**(5): 646–674.
- Hanna, S. (2003) Cancer incidence in new zealand (2003-2007). In D. Forman, D. Bray F Brewster, C. Gombe Mbalawa, B. Kohler, M. Piñeros, E. Steliarova-Foucher, R. Swaminathan, and J. Ferlay (editors), *Cancer Incidence in Five Continents*, volume X, 902–907. International Agency for Research on Cancer, Lyon, France. Electronic version <http://ci5.iarc.fr> Accessed 22/03/2017.



- Hansford, S., Kaurah, P., Li-Chang, H., Woo, M., Senz, J., Pinheiro, H., Schrader, K.A., Schaeffer, D.F., Shumansky, K., Zogopoulos, G., *et al.* (2015) Hereditary Diffuse Gastric Cancer Syndrome: CDH1 Mutations and Beyond. *JAMA Oncol*, **1**(1): 23–32.
- Heiskanen, M.A. and Aittokallio, T. (2012) Mining high-throughput screens for cancer drug targets-lessons from yeast chemical-genomic profiling and synthetic lethality. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **2**(3): 263–272.
- Hell, P. (1976) Graphs with given neighbourhoods i. problèmes combinatoires at theorie des graphes. *Proc Coil Int CNRS, Orsay*, **260**: 219–223.
- Higgins, M.E., Claremont, M., Major, J.E., Sander, C., and Lash, A.E. (2007) CancerGenes: a gene selection resource for cancer genome projects. *Nucleic Acids Res*, **35**(Database issue): D721–726.
- Hillenmeyer, M.E. (2008) The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science*, **320**: 362–365.
- Hoadley, K.A., Yau, C., Wolf, D.M., Cherniack, A.D., Tamborero, D., Ng, S., Leiserson, M.D., Niu, B., McLellan, M.D., Uzunangelov, V., *et al.* (2014) Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*, **158**(4): 929–944.
- Hoehndorf, R., Hardy, N.W., Osumi-Sutherland, D., Tweedie, S., Schofield, P.N., and Gkoutos, G.V. (2013) Systematic analysis of experimental phenotype data reveals gene functions. *PLoS ONE*, **8**(4): e60847.
- Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**(2): 65–70.
- Hopkins, A.L. (2008) Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol*, **4**(11): 682–690.
- Hu, Z., Fan, C., Oh, D.S., Marron, J.S., He, X., Qaqish, B.F., Livasy, C., Carey, L.A., Reynolds, E., Dressler, L., *et al.* (2006) The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics*, **7**: 96.

- Huang, E., Cheng, S., Dressman, H., Pittman, J., Tsou, M., Horng, C., Bild, A., Iversen, E., Liao, M., Chen, C., *et al.* (2003) Gene expression predictors of breast cancer outcomes. *Lancet*, **361**: 1590–1596.
- Hutchison, C.A., Chuang, R.Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L., *et al.* (2016) Design and synthesis of a minimal bacterial genome. *Science*, **351**(6280): aad6253.
- Imoto, S., Higuchi, T., Goto, T., Tashiro, K., Kuhara, S., and Miyano, S. (2004) Combining microarrays and biological knowledge for estimating gene networks via bayesian networks. *J Bioinform Comput Biol*, **2**(1): 77–98.
- International HapMap 3 Consortium (HapMap) (2003) The International HapMap Project. *Nature*, **426**(6968): 789–796.
- Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N.J., Chung, S., Emili, A., Snyder, M., Greenblatt, J.F., and Gerstein, M. (2003) A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science*, **302**(5644): 449–453.
- Jeanes, A., Gottardi, C.J., and Yap, A.S. (2008) Cadherins and cancer: how does cadherin dysfunction promote tumor progression? *Oncogene*, **27**(55): 6920–6929.
- Jerby-Arnon, L., Pfetzer, N., Waldman, Y., McGarry, L., James, D., Shanks, E., Seashore-Ludlow, B., Weinstock, A., Geiger, T., Clemons, P., *et al.* (2014) Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. *Cell*, **158**(5): 1199–1209.
- Joachims, T. (1999) Making large-scale support vector machine learning practical. In S. Bernhard, I. Kropf, J.C.B. Christopher, and J.S. Alexander (editors), *Advances in kernel methods*, 169–184. MIT Press.
- Kaelin, Jr, W. (2005) The concept of synthetic lethality in the context of anticancer therapy. *Nat Rev Cancer*, **5**(9): 689–98.
- Kaelin, Jr, W. (2009) Synthetic lethality: a framework for the development of wiser cancer therapeutics. *Genome Med*, **1**: 99.
- Kamada, T. and Kawai, S. (1989) An algorithm for drawing general undirected graphs. *Information Processing Letters*, **31**(1): 7–15.

- Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., *et al.* (2001) Functional annotation of a full-length mouse cDNA collection. *Nature*, **409**(6821): 685–690.
- Kelley, R. and Ideker, T. (2005) Systematic interpretation of genetic interactions using protein networks. *Nat Biotech*, **23**(5): 561–566.
- Kelly, S.T. (2013) *Statistical Predictions of Synthetic Lethal Interactions in Cancer*. Dissertation, University of Otago.
- Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., *et al.* (2009) Human Protein Reference Database–2009 update. *Nucleic Acids Res*, **37**(Database issue): D767–772.
- Kim, N.G., Koh, E., Chen, X., and Gumbiner, B.M. (2011) E-cadherin mediates contact inhibition of proliferation through Hippo signaling-pathway components. *Proc Natl Acad Sci USA*, **108**(29): 11930–11935.
- Koboldt, D.C., Fulton, R.S., McLellan, M.D., Schmidt, H., Kalicki-Veizer, J., McMichael, J.F., Fulton, L.L., Dooling, D.J., Ding, L., Mardis, E.R., *et al.* (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**(7418): 61–70.
- Kockel, L., Zeitlinger, J., Staszewski, L.M., Mlodzik, M., and Bohmann, D. (1997) Jun in drosophila development: redundant and nonredundant functions and regulation by two mapk signal transduction pathways. *Genes & Development*, **11**(13): 1748–1758.
- Kozlov, K.N., Gursky, V.V., Kulakovskiy, I.V., and Samsonova, M.G. (2015) Sequence-based model of gap gene regulation network. *BMC Genomics*, **15**(Suppl 12): S6.
- Kranthi, S., Rao, S., and Manimaran, P. (2013) Identification of synthetic lethal pairs in biological systems through network information centrality. *Mol BioSyst*, **9**(8): 2163–2167.
- Kroepil, F., Fluegen, G., Totikov, Z., Baldus, S.E., Vay, C., Schauer, M., Topp, S.A., Esch, J.S., Knoefel, W.T., and Stoecklein, N.H. (2012) Down-regulation of CDH1 is associated with expression of SNAIL in colorectal adenomas. *PLoS ONE*, **7**(9): e46665.

- Lander, E.S. (2011) Initial impact of the sequencing of the human genome. *Nature*, **470**(7333): 187–197.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**(6822): 860–921.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, **10**(3): R25.
- Latora, V. and Marchiori, M. (2001) Efficient behavior of small-world networks. *Phys Rev Lett*, **87**: 198701.
- Laufer, C., Fischer, B., Billmann, M., Huber, W., and Boutros, M. (2013) Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. *Nat Methods*, **10**(5): 427–31.
- Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014) voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*, **15**(2): R29.
- Le Meur, N. and Gentleman, R. (2008) Modeling synthetic lethality. *Genome Biol*, **9**(9): R135.
- Le Meur, N., Jiang, Z., Liu, T., Mar, J., and Gentleman, R.C. (2014) Slgi: Synthetic lethal genetic interaction. r package version 1.26.0.
- Lee, A.Y., Perreault, R., Harel, S., Boulier, E.L., Suderman, M., Hallett, M., and Jenna, S. (2010a) Searching for signaling balance through the identification of genetic interactors of the rab guanine-nucleotide dissociation inhibitor gdi-1. *PLoS ONE*, **5**(5): e10624.
- Lee, I., Lehner, B., Vavouri, T., Shin, J., Fraser, A.G., and Marcotte, E.M. (2010b) Predicting genetic modifier loci using functional gene networks. *Genome Research*, **20**(8): 1143–1153.
- Lee, I. and Marcotte, E.M. (2009) Effects of functional bias on supervised learning of a gene network model. *Methods Mol Biol*, **541**: 463–75.

- Lee, M.J., Ye, A.S., Gardino, A.K., Heijink, A.M., Sorger, P.K., MacBeath, G., and Yaffe, M.B. (2012) Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell*, **149**(4): 780–94.
- Lehner, B., Crombie, C., Tischler, J., Fortunato, A., and Fraser, A.G. (2006) Systematic mapping of genetic interactions in *caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nat Genet*, **38**(8): 896–903.
- Li, B., Ruotti, V., Stewart, R.M., Thomson, J.A., and Dewey, C.N. (2010) RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics*, **26**(4): 493–500.
- Li, X.J., Mishra, S.K., Wu, M., Zhang, F., and Zheng, J. (2014) Syn-lethality: An integrative knowledge base of synthetic lethality towards discovery of selective anti-cancer therapies. *Biomed Res Int*, **2014**: 196034.
- Linehan, W.M., Spellman, P.T., Ricketts, C.J., Creighton, C.J., Fei, S.S., Davis, C., Wheeler, D.A., Murray, B.A., Schmidt, L., Vocke, C.D., *et al.* (2016) Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N Engl J Med*, **374**(2): 135–145.
- Lokody, I. (2014) Computational modelling: A computational crystal ball. *Nature Reviews Cancer*, **14**(10): 649–649.
- Lord, C.J., Tutt, A.N., and Ashworth, A. (2015) Synthetic lethality and cancer therapy: lessons learned from the development of PARP inhibitors. *Annu Rev Med*, **66**: 455–470.
- Lu, X., Kensche, P.R., Huynen, M.A., and Notebaart, R.A. (2013) Genome evolution predicts genetic interactions in protein complexes and reveals cancer drug targets. *Nat Commun*, **4**: 2124.
- Lu, X., Megchelenbrink, W., Notebaart, R.A., and Huynen, M.A. (2015) Predicting human genetic interactions from cancer genome evolution. *PLoS One*, **10**(5): e0125795.
- Lum, P.Y., Armour, C.D., Stepaniants, S.B., Cavet, G., Wolf, M.K., Butler, J.S., Hinshaw, J.C., Garnier, P., Prestwich, G.D., Leonardson, A., *et al.* (2004) Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. *Cell*, **116**(1): 121–137.

- Luo, J., Solimini, N.L., and Elledge, S.J. (2009) Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction. *Cell*, **136**(5): 823–837.
- Machado, J., Olivera, C., Carvalh, R., Soares, P., Berx, G., Caldas, C., Sercuca, R., Carneiro, F., and Sorbrinho-Simoes, M. (2001) E-cadherin gene (*CDH1*) promoter methylation as the second hit in sporadic diffuse gastric carcinoma. *Oncogene*, **20**: 1525–1528.
- Markowetz, F. (2017) All biology is computational biology. *PLoS Biol*, **15**(3): e2002050.
- Masciari, S., Larsson, N., Senz, J., Boyd, N., Kaurah, P., Kandel, M.J., Harris, L.N., Pinheiro, H.C., Troussard, A., Miron, P., *et al.* (2007) Germline E-cadherin mutations in familial lobular breast cancer. *J Med Genet*, **44**(11): 726–31.
- Mattison, J., van der Weyden, L., Hubbard, T., and Adams, D.J. (2009) Cancer gene discovery in mouse and man. *Biochim Biophys Acta*, **1796**(2): 140–161.
- McLachlan, J., George, A., and Banerjee, S. (2016) The current status of parp inhibitors in ovarian cancer. *Tumori*, **102**(5): 433–440.
- McLendon, R., Friedman, A., Bigner, D., Van Meir, E.G., Brat, D.J., Mastrogianakis, G.M., Olson, J.J., Mikkelsen, T., Lehman, N., Aldape, K., *et al.* (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **455**(7216): 1061–1068.
- Miles, D.W. (2001) Update on HER-2 as a target for cancer therapy: herceptin in the clinical setting. *Breast Cancer Res*, **3**(6): 380–384.
- Muzny, D.M., Bainbridge, M.N., Chang, K., Dinh, H.H., Drummond, J.A., Fowler, G., Kovar, C.L., Lewis, L.R., Morgan, M.B., Newsham, I.F., *et al.* (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, **487**(7407): 330–337.
- Nagalla, S., Chou, J.W., Willingham, M.C., Ruiz, J., Vaughn, J.P., Dubey, P., Lash, T.L., Hamilton-Dutoit, S.J., Bergh, J., Sotiriou, C., *et al.* (2013) Interactions between immunity, proliferation and molecular subtype in breast cancer prognosis. *Genome Biol*, **14**(4): R34.
- Novomestky, F. (2012) *matrixcalc: Collection of functions for matrix calculations*. R package version 1.0-3.

- Nowak, M.A., Boerlijst, M.C., Cooke, J., and Smith, J.M. (1997) Evolution of genetic redundancy. *Nature*, **388**(6638): 167–171.
- Oliveira, C., Senz, J., Kaurah, P., Pinheiro, H., Sanges, R., Haegert, A., Corso, G., Schouten, J., Fitzgerald, R., Vogelsang, H., *et al.* (2009) Germline *CDH1* deletions in hereditary diffuse gastric cancer families. *Human Molecular Genetics*, **18**(9): 1545–1555.
- Oliveira, C., Seruca, R., Hoogerbrugge, N., Ligtenberg, M., and Carneiro, F. (2013) Clinical utility gene card for: Hereditary diffuse gastric cancer (HDGC). *Eur J Hum Genet*, **21**(8).
- Pandey, G., Zhang, B., Chang, A.N., Myers, C.L., Zhu, J., Kumar, V., and Schadt, E.E. (2010) An integrative multi-network and multi-classifier approach to predict genetic interactions. *PLoS Comput Biol*, **6**(9).
- Parker, J., Mullins, M., Cheung, M., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., *et al.* (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*, **27**(8): 1160–1167.
- Pereira, B., Chin, S.F., Rueda, O.M., Vollan, H.K., Provenzano, E., Bardwell, H.A., Pugh, M., Jones, L., Russell, R., Sammut, S.J., *et al.* (2016) Erratum: The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nat Commun*, **7**: 11908.
- Perou, C.M., Sørlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., *et al.* (2000) Molecular portraits of human breast tumours. *Nature*, **406**(6797): 747–752.
- R Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. R version 3.3.2.
- 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): e47.
- Roguev, A., Bandyopadhyay, S., Zofall, M., Zhang, K., Fischer, T., Collins, S.R., Qu, H., Shales, M., Park, H.O., Hayles, J., *et al.* (2008) Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science*, **322**(5900): 405–10.

- Roychowdhury, S. and Chinnaiyan, A.M. (2016) Translating cancer genomes and transcriptomes for precision oncology. *CA Cancer J Clin*, **66**(1): 75–88.
- Rung, J. and Brazma, A. (2013) Reuse of public genome-wide gene expression data. *Nat Rev Genet*, **14**(2): 89–99.
- Ryan, C., Lord, C., and Ashworth, A. (2014) Daisy: Picking synthetic lethals from cancer genomes. *Cancer Cell*, **26**(3): 306–308.
- Schena, M. (1996) Genome analysis with gene expression microarrays. *Bioessays*, **18**(5): 427–431.
- Scheuer, L., Kauff, N., Robson, M., Kelly, B., Barakat, R., Satagopan, J., Ellis, N., Hensley, M., Boyd, J., Borgen, P., *et al.* (2002) Outcome of preventive surgery and screening for breast and ovarian cancer in BRCA mutation carriers. *J Clin Oncol*, **20**(5): 1260–1268.
- Semb, H. and Christofori, G. (1998) The tumor-suppressor function of E-cadherin. *Am J Hum Genet*, **63**(6): 1588–93.
- Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005) Rocr: visualizing classifier performance in r. *Bioinformatics*, **21**(20): 7881.
- Slurm development team (Slurm) (2017) Slurm workload manager. <https://slurm.schedmd.com/>. Accessed: 25/03/2017.
- Sørbye, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., *et al.* (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA*, **98**(19): 10869–10874.
- Srihari, S., Singla, J., Wong, L., and Ragan, M.A. (2015) Inferring synthetic lethal interactions from mutual exclusivity of genetic events in cancer. *Biology Direct*, **10**(1): 57.
- Stajich, J.E. and Lapp, H. (2006) Open source tools and toolkits for bioinformatics: significance, and where are we? *Brief Bioinformatics*, **7**(3): 287–296.
- Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009) The cancer genome. *Nature*, **458**(7239): 719–724.



- Ström, C. and Helleday, T. (2012) Strategies for the use of poly(adenosine diphosphate ribose) polymerase (parp) inhibitors in cancer therapy. *Biomolecules*, **2**(4): 635–649.
- Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011) Differential expression in RNA-seq: a matter of depth. *Genome Res*, **21**(12): 2213–2223.
- Telford, B.J., Chen, A., Beetham, H., Frick, J., Brew, T.P., Gould, C.M., Single, A., Godwin, T., Simpson, K.J., and Guilford, P. (2015) Synthetic lethal screens identify vulnerabilities in gpcr signalling and cytoskeletal organization in E-cadherin-deficient cells. *Mol Cancer Ther*, **14**(5): 1213–1223.
- The 1000 Genomes Project Consortium (1000 Genomes) (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**(7319): 1061–1073.
- The Cancer Genome Atlas Research Network (TCGA) (2017) The Cancer Genome Atlas Project. <https://cancergenome.nih.gov/>. Accessed: 26/03/2017.
- The Catalogue Of Somatic Mutations In Cancer (COSMIC) (2016) Cosmic: The catalogue of somatic mutations in cancer. <http://cancer.sanger.ac.uk/cosmic>. Release 79 (23/08/2016), Accessed: 05/02/2017.
- The Comprehensive R Archive Network (CRAN) (2017) Cran. <https://cran.r-project.org/>. Accessed: 24/03/2017.
- The ENCODE Project Consortium (ENCODE) (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, **306**(5696): 636–640.
- The National Cancer Institute (NCI) (2015) The genetics of cancer. <https://www.cancer.gov/about-cancer/causes-prevention/genetics>. Published: 22/04/2015, Accessed: 22/03/2017.
- The New Zealand eScience Infrastructure (NeSI) (2017) NeSI. <https://www.nesi.org.nz/>. Accessed: 25/03/2017.
- Tierney, L., Rossini, A.J., Li, N., and Sevcikova, H. (2015) *snow: Simple Network of Workstations*. R package version 0.4-2.
- Tiong, K.L., Chang, K.C., Yeh, K.T., Liu, T.Y., Wu, J.H., Hsieh, P.H., Lin, S.H., Lai, W.Y., Hsu, Y.C., Chen, J.Y., *et al.* (2014) Csnk1e/ctnnb1 are synthetic lethal to tp53 in colorectal cancer and are markers for prognosis. *Neoplasia*, **16**(5): 441–50.

- Tischler, J., Lehner, B., and Fraser, A.G. (2008) Evolutionary plasticity of genetic interaction networks. *Nat Genet*, **40**(4): 390–391.
- Tomasetti, C. and Vogelstein, B. (2015) Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*, **347**(6217): 78–81.
- Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Page, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., *et al.* (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, **294**(5550): 2364–8.
- Tong, A.H., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., *et al.* (2004) Global mapping of the yeast genetic interaction network. *Science*, **303**(5659): 808–13.
- Tran, B., Dancey, J.E., Kamel-Reid, S., McPherson, J.D., Bedard, P.L., Brown, A.M., Zhang, T., Shaw, P., Onetto, N., Stein, L., *et al.* (2012) Cancer genomics: technology, discovery, and translation. *J Clin Oncol*, **30**(6): 647–660.
- Travers, J. and Milgram, S. (1969) An experimental study of the small world problem. *Sociometry*, **32**(4): 425–443.
- Tunggal, J.A., Helfrich, I., Schmitz, A., Schwarz, H., Gunzel, D., Fromm, M., Kemler, R., Krieg, T., and Niessen, C.M. (2005) E-cadherin is essential for in vivo epidermal barrier function by regulating tight junctions. *EMBO J*, **24**(6): 1146–1156.
- Tutt, A., Robson, M., Garber, J.E., Domchek, S.M., Audeh, M.W., Weitzel, J.N., Friedlander, M., Arun, B., Loman, N., Schmutzler, R.K., *et al.* (2010) Oral poly(adp-ribose) polymerase inhibitor olaparib in patients with *BRCA1* or *BRCA2* mutations and advanced breast cancer: a proof-of-concept trial. *Lancet*, **376**(9737): 235–44.
- University of California, Santa Cruz (UCSC) (2012) Uscs cancer browser. Accessed 29/03/2012.
- van der Meer, R., Song, H.Y., Park, S.H., Abdulkadir, S.A., and Roh, M. (2014) RNAi screen identifies a synthetic lethal interaction between PIM1 overexpression and PLK1 inhibition. *Clinical Cancer Research*, **20**(12): 3211–3221.
- van der Post, R.S., Vogelaar, I.P., Carneiro, F., Guilford, P., Huntsman, D., Hoogerbrugge, N., Caldas, C., Schreiber, K.E., Hardwick, R.H., Ausems, M.G., *et al.* (2015)

- Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline CDH1 mutation carriers. *J Med Genet*, **52**(6): 361–374.
- van Steen, K. (2012) Travelling the world of genegene interactions. *Briefings in Bioinformatics*, **13**(1): 1–19.
- van Steen, M. (2010) *Graph Theory and Complex Networks: An Introduction*. Maarten van Steen, VU Amsterdam.
- Vapnik, V.N. (1995) *The nature of statistical learning theory*. Springer-Verlag New York, Inc.
- Vizeacoumar, F.J., Arnold, R., Vizeacoumar, F.S., Chandrashekhar, M., Buzina, A., Young, J.T., Kwan, J.H., Sayad, A., Mero, P., Lawo, S., *et al.* (2013) A negative genetic interaction map in isogenic cancer cell lines reveals cancer cell vulnerabilities. *Mol Syst Biol*, **9**: 696.
- Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013) Cancer genome landscapes. *Science*, **339**(6127): 1546–1558.
- Vos, C.B., Cleton-Jansen, A.M., Berx, G., de Leeuw, W.J., ter Haar, N.T., van Roy, F., Cornelisse, C.J., Peterse, J.L., and van de Vijver, M.J. (1997) E-cadherin inactivation in lobular carcinoma in situ of the breast: an early event in tumorigenesis. *Br J Cancer*, **76**(9): 1131–3.
- Waldron, D. (2016) Cancer genomics: A multi-layer omics approach to cancer. *Nat Rev Genet*, **17**(8): 436–437.
- Wang, K., Singh, D., Zeng, Z., Coleman, S.J., Huang, Y., Savich, G.L., He, X., Mieczkowski, P., Grimm, S.A., Perou, C.M., *et al.* (2010) MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res*, **38**(18): e178.
- Wang, X. and Simon, R. (2013) Identification of potential synthetic lethal genes to p53 using a computational biology approach. *BMC Medical Genomics*, **6**(1): 30.
- Wappett, M. (2014) Bisep: Toolkit to identify candidate synthetic lethality. r package version 2.0.

- Wappett, M., Dulak, A., Yang, Z.R., Al-Watban, A., Bradford, J.R., and Dry, J.R. (2016) Multi-omic measurement of mutually exclusive loss-of-function enriches for candidate synthetic lethal gene pairs. *BMC Genomics*, **17**: 65.
- Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., *et al.* (2015) *gplots: Various R Programming Tools for Plotting Data*. R package version 2.17.0.
- Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks. *Nature*, **393**(6684): 440–2.
- Weinstein, I.B. (2000) Disorders in cell circuitry during multistage carcinogenesis: the role of homeostasis. *Carcinogenesis*, **21**(5): 857–864.
- Weinstein, J.N., Akbani, R., Broom, B.M., Wang, W., Verhaak, R.G., McConkey, D., Lerner, S., Morgan, M., Creighton, C.J., Smith, C., *et al.* (2014) Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*, **507**(7492): 315–322.
- Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M., Chang, K., *et al.* (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*, **45**(10): 1113–1120.
- Wickham, H. and Chang, W. (2016) *devtools: Tools to Make Developing R Packages Easier*. R package version 1.12.0.
- Wickham, H., Danenberg, P., and Eugster, M. (2017) *roxygen2: In-Line Documentation for R*. R package version 6.0.1.
- Wojtukiewicz, M.Z., Hempel, D., Sierko, E., Tucker, S.C., and Honn, K.V. (2016) Thrombin-unique coagulation system protein with multifaceted impacts on cancer and metastasis. *Cancer Metastasis Rev*, **35**(2): 213–233.
- Wong, S.L., Zhang, L.V., Tong, A.H.Y., Li, Z., Goldberg, D.S., King, O.D., Lesage, G., Vidal, M., Andrews, B., Bussey, H., *et al.* (2004) Combining biological networks to predict genetic interactions. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(44): 15682–15687.
- World Health Organization (WHO) (2017) Fact sheet: Cancer. <http://www.who.int/mediacentre/factsheets/fs297/en/>. Updated February 2017, Accessed: 22/03/2017.

- Wu, M., Li, X., Zhang, F., Li, X., Kwoh, C.K., and Zheng, J. (2014) In silico prediction of synthetic lethality by meta-analysis of genetic interactions, functions, and pathways in yeast and human cancer. *Cancer Inform*, **13**(Suppl 3): 71–80.
- Yu, H. (2002) Rmpi: Parallel statistical computing in r. *R News*, **2**(2): 10–14.
- Zhang, F., Wu, M., Li, X.J., Li, X.L., Kwoh, C.K., and Zheng, J. (2015) Predicting essential genes and synthetic lethality via influence propagation in signaling pathways of cancer cell fates. *J Bioinform Comput Biol*, **13**(3): 1541002.
- Zhang, J., Baran, J., Cros, A., Guberman, J.M., Haider, S., Hsu, J., Liang, Y., Rivkin, E., Wang, J., Whitty, B., *et al.* (2011) International cancer genome consortium data portala one-stop shop for cancer genomics data. *Database: The Journal of Biological Databases and Curation*, **2011**: bar026.
- Zhong, W. and Sternberg, P.W. (2006) Genome-wide prediction of c. elegans genetic interactions. *Science*, **311**(5766): 1481–1484.
- Zweig, M.H. and Campbell, G. (1993) Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, **39**(4): 561–577.